UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos



Implementation of artificial intelligence methodologies in climate change adaptation studies in port and coastal engineering. Implementación de metodologías de inteligencia artificial en estudios de adaptación al cambio climático en ingeniería portuaria y costera

Doctoral Thesis

Submitted for the degree of Doctor by:

Nerea Fátima Portillo Juan
Civil Engineer

Madrid, 2025



UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería de Caminos, Canales y Puertos

Doctoral Degree in Civil Engineering Systems

Implementation of artificial intelligence methodologies in climate change adaptation studies in port and coastal engineering. Implementación de metodologías de inteligencia artificial en estudios de adaptación al cambio climático en ingeniería portuaria y costera

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Nerea Fátima Portillo Juan
Civil Engineer

Under the supervision of:
Dr. Vicente Negro Valdecantos

Madrid, 2025

Title: Implementation of artificial intelligence methodologies in climate change adaptation studies in port and coastal engineering. Implementación de metodologías de inteligencia artificial en estudios de adaptación al cambio climático en ingeniería portuaria y costera.

Author: Nerea Fátima Portillo Juan

Doctoral Programme: Civil Engineering Systems

Thesis Supervision:

Dr. Vicente Negro Valdecantos, Catedrático, Universidad Politécnica de Madrid (Director)

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially funded by the UPM's Internal Program and later by the FPU scholarships from the Government of Spain (Reference FPU21/00812). Additionally, the international research stays carried out during the thesis (UCC, UGhent, and UCLA) were funded by UPM (UCC and UCLA) and by the FPU short-stay program (UGhent – Reference EST23/00054).

A mis padres, por haberme dado todo

A mi hermano, por ser mi faro

A Olga, por convertirse en familia

A María, por estar siempre a mi lado

A Julia, por ser luz

A Lu, por ser tan especial

A Ana, por su amor y fortaleza

Os quiero

Acknowledgment

Siempre que me preguntan cómo acabé haciendo el doctorado digo que fue a raíz de una crisis existencial, en la que Vicente me tendió la mano. Así que, lo primero, gracias Vicen por haber estado siempre ahí, guiándome y apoyándome, gracias por transmitirme tu amor y pasión por este mundo, pero sobre todo gracias por ser.

También me encanta inventarme porcentajes (solo en mi vida personal) y mi porcentaje favorito es el 80%. Creo que el 80% de la persona que soy hoy se la debo a mis padres, así que gracias, mamá y papá, por absolutamente todo.

Gracias Adri por ser el mejor regalo que la vida ha podido darme.

Gracias Olga por ser mi compañera de vida.

Gracias María por ser y estar desde el día uno en el que te conocí.

Gracias a mis Finas, a Pablo, a Tade, mis amigos de la infancia, siempre apoyando y estando en cada etapa de mi vida. Elena, Dioli, Ceci, María os he querido, os quiero y os querré siempre.

Gracias a Caminos+1, mis ingenieras favoritas, lo más bonito que me ha dado la carrera. Sara, siempre siendo mi referente. María, eres pura inspiración. Las Andres, Mery, Elena, gracias siempre.

Gracias a mis Gaviolas, mi día a día. Porque los días son más bonitos cuando se que os voy a ver. Gracias por las horas de deporte, por los viajes, por los Pepes, por las tonterías, porque no se qué haría sin vosotros. Somos una familia muy bonita. Gracias Julia, Lu, Ana, Cris, Rakel, Sara, Jason, Nere, Carmensita, Gon, Sebas.

Gracias a todos los que me alegráis los días en cada entrenamiento.

Gracias María por devolverme el espíritu de ser un niña.

Gracias a toda la gente que he conocido en el camino del doctorado, porque siempre digo que los sitios y las experiencias y, la vida, las hacen las personas.

Carlos, Mire, Antonio, Ana; Cove 36 Street siempre estará en mi corazón.

Lau y Ali, gracias por ser casa. Rafa y Joe, para mi Gante sois vosotros. Gracias siempre al equipo de Gante por hacerme sentir una más.

Gracias porque siempre allá donde he ido me he encontrado a gente que me ha hecho sentir querida.

GRACIAS A LA VIDA Y GRACIAS A VOSOTROS.

Abstract

Climate change is probably the greatest challenge facing society in the 21st century. One of the most affected sectors is our ports and coastal systems and Spain and the Mediterranean region are especially vulnerable. More than 15% of Spain's GDP relies on tourism, much of which is connected to coastal areas, making it a priority to address this issue if we want to avoid losses like those experienced during recent storms, such as Gloria in 2020, Filomena in 2021, or Nelson in 2024.

However, climate change is not a new phenomenon, scientists have been studying it for decades. Thanks to their efforts, we now have highly accurate numerical models capable of predicting maritime climate under various scenarios, as well as specific tools for studying the impact on our structures and ports.

Nevertheless, the high computational cost of these tools has limited adaptation studies to ports of greater economic and social importance. Smaller ports, which are likely to be more severely affected by climate change, have been left behind. Marinas and recreational terminals have shallower drafts, making them more vulnerable to the effects of rising sea levels. The increase in water levels alters diffraction patterns and operational functionality, putting their operability at risk. However, the lack of efficient computational tools has been an obstacle to analysing these smaller terminals. This is where this work focuses.

This project proposes a methodology based on Artificial Intelligence (AI) for port adaptation studies to climate change. This methodology addresses everything from climate projection to port agitation. The goal is to optimize the entire climate modelling process. This proposal offers an alternative to existing methodologies, but with lower computational cost, allowing for small-scale local studies in smaller ports.

Among the AI techniques, neural networks (ANNs) have been used with genetic algorithms (GAs). ANNs and GAs have been combined with Level III probabilistic methodologies to develop AI models, leveraging the data generation capabilities of probabilistic methodologies alongside the data exploitation capabilities of ANNs.

As a result, three main models have been developed for each phase of the study. The first model is for the propagation of maritime climate, with two other AI models to project future maritime climate and analyse its trends. The second model addresses structural adaptation to new climatic conditions using a hybrid model of ANNs and GAs. Finally, the third model consists of a hybrid Deep Learning

approach to analyse port agitation, combining conventional networks that handle wave data and convolutional networks that process port images. This approach addresses the adaptation problem, covering all phases of adaptation studies: propagation of wave climate, structural assessments and operability analysis.

Based on these models, a set of rules is proposed that can be applied to any port adaptation study in response to climate change. The ultimate goal is to facilitate the development of AI models and provide a starting point.

When comparing AI models with conventional models, it was found that AI models slightly improve the prediction of wave periods and are more efficient. In the case of climate propagation, simulation time has been reduced by up to 20 times compared to numerical models, and by up to 10 times in the case of port agitation. Additionally, hybrid models are much more efficient than those incorporating a single AI technique.

AI is the major agent of change in this century, and climate change is the major challenge. We must start implementing these high-potential methodologies. AI models are powerful tools, and their effectiveness is growing exponentially, making them one of the most promising techniques. Furthermore, we must strive to bring these models closer to technical professionals and not limit them solely to the scientific community. This is where this thesis fits in.

Resumen

El cambio climático es probablemente el mayor desafío al que se enfrenta la sociedad en el siglo XXI, y uno de los sectores más afectados es nuestro sistema portuario y costero. En particular, España es especialmente vulnerable a sus impactos. Más del 15% del PIB español depende del turismo, gran parte del cual está vinculado a las zonas costeras.

No obstante, el cambio climático no es nada nuevo, científicos llevan décadas estudiándolo. De hecho, gracias a todos sus esfuerzos, hoy contamos con modelos numéricos muy precisos, así como herramientas específicas para estudiar la afección de nuestras estructuras y puertos.

Sin embargo, su alto coste computacional ha limitado los estudios de adaptación a los puertos de mayor importancia económica y social, pero son los puertos más pequeños los que más afectados se van a ver por el cambio climático. Las terminales deportivas y recreativas tienen un calado más pequeño, por lo que, al tener una profundidad menor, el impacto del cambio climático será mayor. El incremento de la lámina de agua altera los patrones de difracción y funcionamiento, poniendo en riesgo su operatividad. Sin embargo, la falta de herramientas computacionales más eficientes ha sido un obstáculo para analizar estas terminales más pequeñas. Es aquí donde se centra este trabajo.

Este proyecto propone una metodología basada en Inteligencia Artificial (IA) para estudios de adaptación portuaria al cambio climático. Esta metodología integral aborda desde la proyección del clima hasta la agitación portuaria. El objetivo es optimizar todo el proceso de modelado climático. Esta propuesta ofrece una alternativa a las metodologías existentes, pero con un coste computacional mucho menor.

De entre las técnicas de IA se han utilizado redes neuronales (ANNs) junto con algoritmos genéticos (GAs). Se han combinado las ANNs y GAs con metodologías probabilísticas de nivel III para desarrollar los modelos de IA, aprovechando la capacidad de generación de datos de las metodologías probabilísticas junto con la capacidad de explotación de las ANNs.

Como resultado, se han desarrollado tres modelos principales para cada fase de estudio. El primero para la propagación del clima marítimo, utilizando a su vez otros dos modelos para proyectar el clima marítimo futuro y analizar sus tendencias. El segundo abarca la adaptación estructural ante las nuevas

condiciones climáticas utilizando un modelo híbrido de ANNs y GAs. Finalmente, el tercer modelo, consiste en un enfoque híbrido de *Deep Learning* para analizar la agitación portuaria, combinando redes convencionales para el oleaje y redes convolucionales que procesan las imágenes del puerto. De esta manera se aborda el problema de adaptación con un enfoque integral.

A partir de estos modelos, se proponen una serie de reglas que pueden aplicarse en todo estudio de adaptación portuaria al cambio climático. El objetivo final es facilitar el desarrollo de modelos de IA y proporcionar un punto de partida.

Al comparar los modelos de IA con los convencionales, se ha comprobado que los de IA mejoran ligeramente la predicción del periodo de oleaje y son más eficientes. En el caso de la propagación del clima, se ha logrado reducir el tiempo de simulación hasta 20 veces en comparación con modelos numéricos, y hasta 10 veces en el caso de la agitación portuaria. Además, los modelos híbridos son mucho más eficientes que los modelos que incorporan una única técnica de IA.

La IA es el gran agente del cambio de este siglo y el cambio climático el gran problema. Tenemos que estar a la altura y empezar a implementar este tipo de metodologías. Los modelos de IA son modelos muy potentes y su eficacia está creciendo de manera exponencial, convirtiéndolos en una de las herramientas más prometedoras. Además, tenemos que luchar por acercar estos modelos a los técnicos y no solo limitarlos a la comunidad científica. Es aquí donde se enmarca esta tesis.

Table of contents

1. Introduction 1

2. Methodology 9

2.1. Introduction to Artificial Neural Networks (ANNs) ... 9

2.1.1. Algorithms ... 11

2.1.2. Classification of ANNs ... 13

2.2. Wave downscaling ... 15

2.2.1. Statistical study of maritime climate ... 16

2.2.2. Synthetic data generation ... 19

2.2.3. Propagation model ... 21

2.3. Structural assessment (ULS) ... 25

2.3.1. Probabilistic characterization of coastal structures ... 26

2.3.2. Implementation of the AI-based model ... 31

2.4. Port agitation (SLS) ... 33

2.4.1. MILDwave model ... 34

2.4.2. Implementation of the AI model ... 36

2.5. Application to the Western Mediterranean and to the Port of Valencia ... 43

2.5.1. Wave downscaling ... 43

2.5.2. Adaptation of coastal protection structures ... 54

2.5.3. Port agitation ... 57

3. Scientific production 65

3.1. Index of papers ... 65

3.2. A New Climate Change Analysis Parameter: A Global or a National Approach Dilemma 68

Abstract and Keywords ... 68

1. Introduction ... 68

2. Literature review ... 71

3. Materials and Methods ... 74

4. Results ... 78

5. Discussion of results ... 93

6. Conclusions ... 97

References ... 99

3.3. Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain ... 104

Abstract and Keywords ... 104

1. Introduction ... 104

2. Materials and Methods ... 106

3. Results and Discussion ... 108

4. Conclusions ... 122

References ... 123

3.4. Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast ... 132

Abstract and Keywords 132
1. Introduction 132
2. Methodology 134
3. Results 137
4. Discussion 155
5. Conclusions 158
References 159
3.5. Review of the application of Artificial Neural Networks in ocean engineering 162
Abstract and Keywords 162
1. Introduction 162
2. ANNs in ocean engineering 169
3. Analysis of the main issues in the application of ANNs in ocean engineering 186
4. Conclusions 196
References 197
3.6. Uncertainties in the application of Artificial Neural Networks in ocean engineering 205
Abstract and Keywords 205
1. Introduction 205
2. Uncertainties 208
3. Methodology 212
4. Analysis of ANNs uncertainties in ocean engineering 213
5. Discussion 238
6. Conclusions 242
References 244
3.7. Comparison of SIMAR-WANA, ERA-5 and Waverys databases for maritime climate estimations and implications for coastal protection structures 252
Abstract and Keywords 252
1. Introduction 252
2. Area of study and datasets 254
3. Methodology 259
4. Results 262
5. Discussion 278
6. Conclusions 283
References 284
3.8. Data-driven and physics-based approach for wave downscaling: A comparative study 288
Abstract and Keywords 288
1. Introduction 288
2. Materials and methods 293
3. Results 299
4. Discussion 304
5. Conclusions 306
Appendix A 307
References 316
3.9. A new methodology for the study of future maritime climate using Artificial Neural Networks and Monte Carlo simulations under the context of climate change 321
Abstract and Keywords 321
1. Introduction 321
2. Methodology 328

3. Results ... 349
4. Discussion ... 356
5. Conclusions ... 360
References ... 361

4. Discussion **371**

4.1. Wave downscaling ... 371
4.1.1. Results ... 371
4.1.2. Comparison and guide for AI model development ... 376
4.2. Structural assessment ... 381
4.2.1. Results ... 381
4.2.2. Comparison and guide for AI model development ... 385
4.3. Study of port agitation ... 388
4.3.1. Results ... 388
4.3.2. Comparison and guide for AI model development ... 390
4.4. Contributions of the Publications ... 391

5. Conclusions **397**

References **399**

Appendices **407**

Appendix A ... 407
Wave propagation model codes ... 407
Structural assessment model codes ... 409
Port operability assessment model codes ... 428
Appendix B ... 439
Appendix C ... 440

List of Figures

Figure 1-1. Indexed studies on Web of Science by area of ocean and maritime engineering. 2

Figure 1-2. Distribution of the amount of climate change studies along the Spanish coast (Portillo Juan et al., 2022a) (Research paper 3.4.). 3

Figure 1-3. Distribution of studies that apply AI in ocean and maritime engineering classified by topic. 5

Figure 1-4. Evolution of the number of research studies that apply AI techniques in ocean and maritime engineering (Portillo Juan et al., 2024). 6

Figure 2-1. Scheme of the methodology proposed. 9

Figure 2-2. Approach of the methodology proposed. 17

Figure 2-3. ANN model for wave downscaling. 24

Figure 2-4. Methodology sum-up for phase 2. 26

Figure 2-5. Simulation of a rubble-mound breakwater. 27

Figure 2-6. Pressure diagram on a vertical breakwater taken from Juan Nerea et al. (2024). 29

Figure 2-7. Scheme of a convolutional layer. 37

Figure 2-8. Scheme of a max pooling layer. 38

Figure 2-9. Guidelines to develop an ANN model for harbour agitation. 42

Figure 2-10. AI implementation in ports adaptation studies. 42

Figure 2-11. Location of buoys. 44

Figure 2-12. Data partition of the ANN model. 45

Figure 2-13. Optimization of the ANN model architecture in Valencia. 45

Figure 2-14. Optimization of the ANN architecture in Cabo de Palos. 45

Figure 2-15. Training error boxplot for each algorithm (Valencia case study). 46

Figure 2-16. Validation error for each algorithm (Valencia case study). 46

Figure 2-17. Test error for each algorithm (Valencia case study). 46

Figure 2-18. Training error boxplot for each algorithm (Cabo de Palos case study). 47

Figure 2-19. Validation error for each algorithm (Cabo de Palos case study). 47

Figure 2-20. Test error for each algorithm (Cabo de Palos case study). 47

Figure 2-21. ANN models for wave downscaling. 48

Figure 2-22. ANN errors for selected arch. in Valencia (up) and Cabo de Palos (down). 49

Figure 2-23. Statistical adjustment of parameters A, B and C for long-term distribution. 50

Figure 2-24. Statistical adjustment of parameters α, ß and γ for extreme distribution. 51

Figure 2-25. Performance of the model in Valencia (up) and Cabo de Palos (down). 53

Figure 2-26. Trunk section. Taken from (Albert et al., 2012). 54

Figure 2-27. Head section of the North breakwater. Taken from (Albert et al., 2012). 54

Figure 2-28. Optimization of the vertical breakwater neural network architecture. 55

Figure 2-29. Optimization of the compound breakwater neural network architecture. 56

Figure 2-30. Swell rose in Valencia for long-term and extreme distribution. 58

Figure 2-31. Bitmap file for the port layout (wave direction 22.5º). 60

Figure 2-32. Optimization of the number of epochs and batch size. 60

Figure 2-33. Optimization of the learning rate. 61

Figure 2-34. Optimization of the FFNN architecture. 62

Figure 2-35. CNN architecture optimization. 63

Figure 2-36. Deep Learning model for harbour agitation. 64

Figure 4-1. Hs (up) and Tp (down) evolution in Valencia under climate change. 373

Figure 4-2. Hs (up) and Tp (down) evolution in Cabo de Palos under climate change. 373

Figure 4-3. Extreme Hs (up) and Tp (down) evolution in Valencia under climate change. 375

Figure 4-4. Extreme Hs (up) and Tp (down) evolution in Cabo de Palos under climate change. 376

Figure 4-5. Regression plots in Valencia (left) and Cabo de Palos (right). 378

Figure 4-6. Error histogram in Valencia (left) and Cabo de Palos (right). 379

Figure 4-7. Regression plot for SWAN model. Hs up and Tp down. 379

Figure 4-8. Increment of height needed for the vertical breakwater under different wave scenarios for 2035 and 2050. 382

Figure 4-9. Increment of height needed for the compound breakwater under different wave scenarios for 2035 and 2050. 383

Figure 4-10. Increment of weight needed for the compound breakwater under different wave scenarios for 2035 and 2050. 384

Figure 4-11. Increment of computational time with the increment of inputs (100 epochs). 386

Figure 4-12. Distribution of the errors of the ANN-GA model. 387

Figure 4-13. Port agitation in Valencia under different wave climate scenarios. 388

Figure 4-14. Worst cases of port agitation in Valencia (Tp= 8.9 s). 389

Figure 4-15. Comparison of the numerical model output (left) ant the AI model output (right). 390

List of Tables

Table 2-1. Transfer functions used in ANNs. 10

Table 2-2. Parameters that characterize breakwater behaviour. 27

Table 2-3. Historical records used for the development of the ANN model. 44

Table 2-4. Goodness of fit evaluation for the long-term distribution. 50

Table 2-5. Goodness of fit evaluation for the extreme distribution. 51

Table 2-6. Variability of Hs, Tp and Dir trough Monte Carlo simulations. 52

Table 2-7. ANN wave propagation models applied to climate change. 53

Table 2-8. Training and validation error of the model. 53

Table 2-9. Model characteristics. 57

Table 2-10. Errors of the ANN-GA model for structural adaptation. 57

Table 2-11. Study cases. 58

Table 2-12. MILDwave coefficients calibration. 59

Table 2-13. Characteristics of the model. 63

Table 3-1. Research Papers. 65

Table 4-1. Evolution of wave climate in Valencia under climate change. 371

Table 4-2. Evolution of wave climate in Cabo de Palos under climate change. 371

Table 4-3. Evolution of extreme wave climate in Valencia under climate change. 374

Table 4-4. Evolution of extreme wave climate in Cabo de Palos under climate change. 374

Table 4-5. Wave climate under climate change scenario in the Western Mediterranean. 376

Table 4-6. Statistical analysis of the models for wave downscaling. 380

Table 4-7. Organization of scientific output 391

Abbreviations and Acronyms

AI	Artificial Intelligence
AGI	Artificial General Intelligence
ANI	Artificial Narrow Intelligence
ANN	Artificial Neural Network
aP	Air Pressure
ASI	Artificial Super Intelligence
AST	Air – Sea Temperature Interaction
aT	Air temperature
BFGS	Broyden, Fletcher, GolFarb and Shanno
BP	Backpropagation
BR	Bayesian Regularization
CG	Conjugate Gradient
CGB	Powell-Beale Restarts
CGF	Fletcher-Reeves
CGP	Polak-Ribiére
CNN	Convolutional Neural Network
DIA	Discrete Interaction Approximation
Dir	Wave direction
DTs	Decision Trees
DTM	Digital Terrain Model
ECCLIPSE	assEssment of CLImate change in Ports of Southwest Europe
FFNN	Feed Forward multilayer perceptron Neural Network
GAs	Genetic Algorithms
GANs	Generative Adversarial Networks
GCM	Global Circulation Model
GDM	Gradient Descent with Momentum

GDP	Gross Domestic Product
GHG	Greenhouse Gases
Hs	Significant wave height
Hs^0	Significant wave height in deep waters
Hs^{dw}	Significant wave height downscaled
LM	Levenberg-Marquardt
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MSE	Mean Square Error
MSL	Mean Sea Level
MT	Meteorological Tide
NN	Neural Network
OSS	One Step Secant algorithm
P	Wave power
R	Correlation Coefficient
RBF	Radial Basis Function
RBP	Resilient Backpropagation
RCM	Regional Circulation Model
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RSL	Rolling Sea Level
RSST	Rolling Deviation SST
SCG	Scaled Conjugate Gradient Descent
SL	Sea Level
SLC	Sea Level Change

SLR	Sea Level Rise
SLS	Serviceability Limit State
SSC	Sliding Safety Coefficient
SST	Sea Surface Temperature
Tp	Peak wave period
Tp^0	Peak wave period in deep waters
Tp^{dw}	Peak wave period downscaled
ULS	Ultimate Limit State
UNFCC	United Nations Framework Convention on Climate Change
UPM	Universidad Politécnica de Madrid
VAEs	Variational Auto-Encoders
Wd	Wind direction
Ws	Wind speed
WT	Wind – Temperature interaction
WW	Wind – Wave interaction
α	Learning rate
θp	Peak wave direction
θp^0	Peak wave direction in deep waters
θp^{dw}	Peak wave direction downscaled

1. Introduction

The current society is a society of constant change. If we compare life, customs, tools, science, ways of interacting, etc., with those of the late 20th century, little or nothing is the same; both, we and our environment, have changed. The foundation of all progress is science. Science, engineering, technology, and research drive a society to progress, they are the main drivers of improvements, they prevent stagnation and lead us to achieve well-being. Civil engineers have played a significant role in achieving this well-being; we made possible the development of the entire infrastructure system in the second half of the 20th century. This involved building roads, designing large dams and sanitary systems, creating ports and beaches, and generating spaces of great social and natural value. Thanks to all these efforts and technical rigor, the society of the 2000s reached a level of well-being much higher than that of the previous generation. However, we continue to experience constant change. From 1990 to the present day, the needs of society have completely changed, but what about us? Have we evolved with society, or have we fallen behind?

In 1992, the United Nations Framework Convention on Climate Change (UNFCCC) recognized climate change for the first time as a global issue (UNFCC, 1992). Since then, 28 climate summits have taken place, along with two major global agreements: the Kyoto Protocol and the Paris Agreement. These agreements primarily aim to limit worldwide emissions of Greenhouse Gases (GHGs). (Boehringer, 2003; Schleussner et al., 2016). All efforts have been focused on limiting the temperature increase to 1.5ºC, something that, as of today, seems unlikely to achieve. In fact, not only have GHG emissions not decreased compared to the base year (1990), but they have increased by 68%, and per capita emissions have risen by 16% (Portillo Juan et al., 2022b) (Research paper 3.2.).

One of the systems most affected by climate change is the port and coastal system. The increase in GHG emissions is causing sea level rise, erosion of coastal systems, polar ice melting, and more frequent extreme weather events, among other phenomena (Bindoff et al., 2019; Collins et al., 2019; Hock et al., 2019; Meredith et al., 2019; Oppenheimer et al., 2019).

- Approximately 40% of the world's population lives within 100 km of the coast, and it is expected that by 2050, this number will increase by 75% (Losada et al., 2011; Masselink & Hughes, 2003).

- If necessary adaptation measures are not taken, it is projected that between 0.2% and 4.6% of the world's population will experience annual flooding by the year 2100, with corresponding losses of 0.3% to 9.3% of the global Gross Domestic Product (GDP) (Hinkel et al., 2014).

- Between 1984 and 2015, there was a net loss of 14,000 km^2 of land due to erosion, and approximately 24% of beaches are eroding at rates exceeding 0.5 m/year (Luijendijk et al., 2018; Mentaschi et al., 2018).

All of these are just some of the facts that demonstrate the need to take measures to protect our coastal systems against climate change.

Referring to our port system; between 80% and 90% of the world's goods are transported by sea, and it is expected to increase by 4% within five years (Garcia-Alonso et al., 2020; IMO, 2012; UNCTAD, 2019). This means that a significant portion of the global economy depends on the proper functioning of navigation routes and ports. A clear example of this was the recent grounding of the Ever Given in the Suez Canal in 2021, which resulted in substantial financial losses. However, unlike coastal systems, there are not as many studies analysing the potential impacts of climate change on our ports (Figure 1-1).

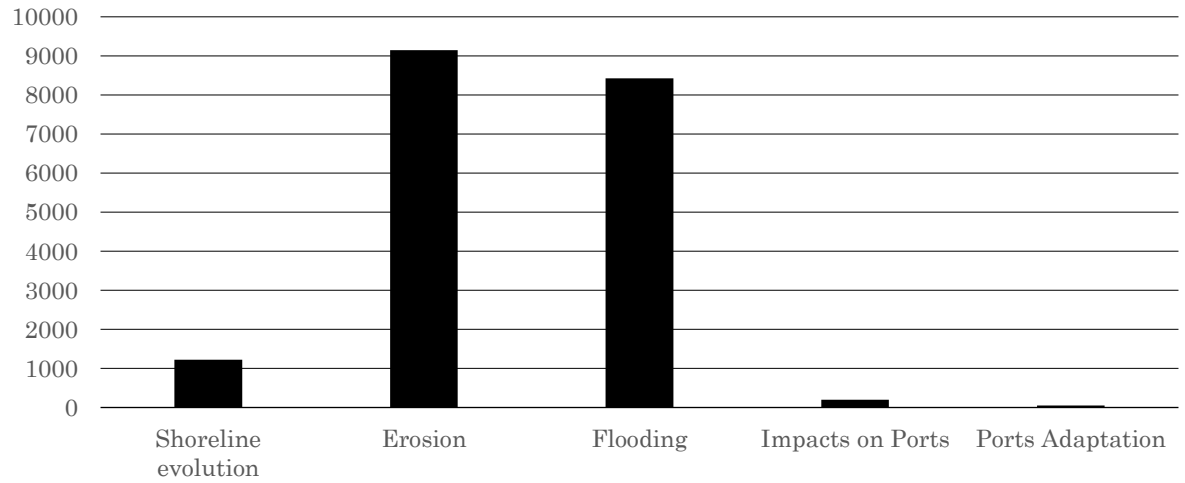


Figure 1-1. Indexed studies on Web of Science by area of ocean and maritime engineering.

In comparison to the scientific effort dedicated to studying coastal systems and beaches, the number of studies involving port areas is much lower (Portillo Juan et al., 2022c) (Research paper 3.3.). The need for conducting climate studies becomes even more significant if we focus on the ports located in vulnerable areas.

The Mediterranean Sea is considered one of the main hotspots of climate change (Giorgi, 2006). The Mediterranean coast lies between North Africa, whose climate is arid and dry, and Central Europe, which is characterized by a rainy and humid climate. This makes it affected by atmospheric and oceanic interactions that make it highly vulnerable to climate change. (Giorgi & Lionello, 2008). The impact of rising temperatures in the Mediterranean Sea is expected to be 20% greater than in the rest of the world. (Cramer et al., 2018).

For all these reasons, it is essential that measures are taken to face the new challenge of the 21st century. The port areas of our coast are one of the most vulnerable areas. However, as it has been pointed out, the number of scientific studies that analyse the impact of climate change on our ports is not very abundant. Furthermore, within our coast there are areas that have awakened greater scientific interest and others that have been left in the shadows (Figure 1-2).



Figure 1-2. Distribution of the amount of climate change studies along the Spanish coast (Portillo Juan et al., 2022a) (Research paper 3.4.).

Within the Mediterranean Sea, the Maresme, given its complexity and unique climatic, maritime and geotechnical patterns, has aroused greater interest among

scientists, making the area of Valencia a less studied area and one in which more specific studies are urgently needed.

Climate change studies should be conducted at regional and local levels, as the effects of climate change largely depend on the specific area, its topography, atmospheric patterns, circulation patterns, etc. Regarding ports, which is the topic at hand, the effects of climate change on different ports vary significantly, and local studies are necessary to understand the specific implications of climate change in each case.

Small ports and marinas are of particular importance, as they are disproportionately affected by rising sea levels. Even small increases in sea level can lead to significant relative changes in the depth of shallow ports, altering diffraction patterns and their overall behaviour. Despite their vulnerability, very few studies analyse the impact of climate change on recreational and sports marinas, as their economic and social value is lower than that of large port systems. Consequently, research efforts tend to focus on major shipping terminals, which are less vulnerable as they are better protected. The numerical models used in the study of port agitation and in the analysis of port structures are highly complex and require a significant amount of computational resources, making it challenging to conduct studies on small ports and areas.

However, in recent years, a new tool has emerged: Artificial Intelligence (AI). Its advantage over traditional models lies in its low computational cost. These methodologies can greatly reduce the time and computational resources currently needed for climate change studies in port and coastal engineering. By integrating AI models, we can overcome computational barriers.

AI is a contemporary term encompassing a diverse range of technologies and that it could be defined as the capability of machines and computer systems to perform complex tasks that historically only humans were able to do. Although the concept of AI first appeared in 1956 at the Dartmouth Conference, it is from the 1990s onwards that the first breakthroughs began to emerge (McCarthy et al., 1956). Since the development of the first AI models until today, AI techniques have completely revolutionized the scientific world and our society. These two new agents, climate change and AI, seem to be the protagonists of the 21st-century.

In our field, the application of AI techniques is a step behind that in other disciplines such as medicine or telecommunications, but even if we compare with similar fields such as hydrological resources or atmospheric sciences, the popularity and development of Machine Learning (ML) techniques in ocean engineering is much lower (Portillo Juan & Negro Valdecantos, 2022) (Research paper 3.5.). Within maritime and coastal engineering, the popularity of AI techniques varies greatly depending on the sector, the naval sector is the most advanced in this topic, with the development of autonomous vehicles and optimization systems with AI. Within coastal engineering, maritime climate prediction is where the vast majority of AI models have been applied, but in studies of coastal structures, the application of ML models is very uncommon (Figure 1-3) (Portillo Juan et al., 2024).

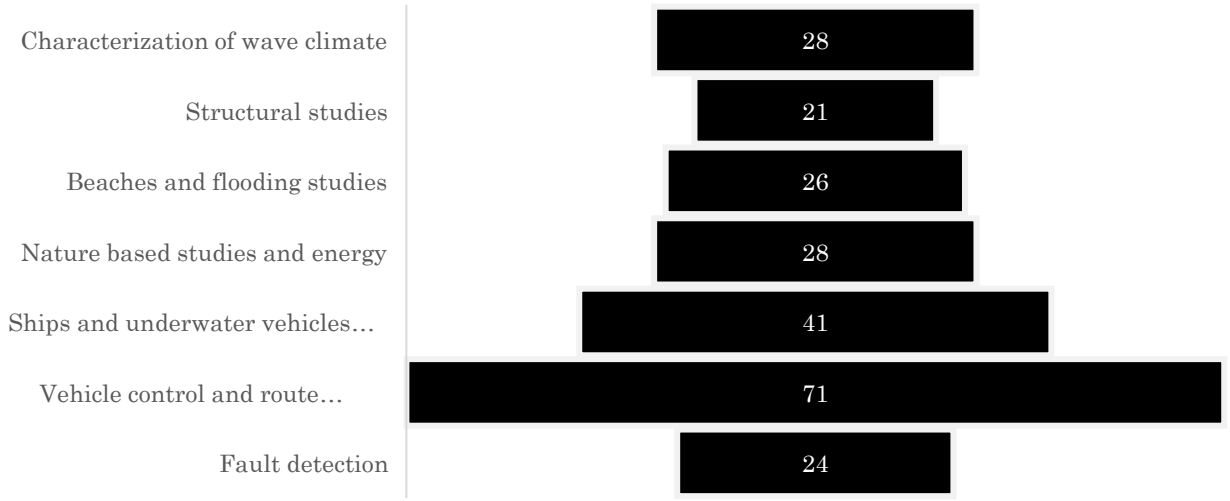


Figure 1-3. Distribution of studies that apply AI in ocean and maritime engineering classified by topic.

Within maritime climate and coastal structure characterization studies, the AI technique par excellence is Artificial Neural Networks (ANNs) (Portillo Juan et al., 2024). Neural networks (NNs) were introduced for the first time by McCulloch and Pitts (1943), but it not was until Rumelhart et al. (1986) developed the backpropagation algorithm, when the use of ANNs became popular (Møller, 1993). ANNs try to mimic the human brain with computational elements that emulate the neurons (by elements called nodes) and the unions between them (by transfer functions). ANNs are made of different layers: the input layer in which the known data is introduced into de model; the intermediate layers, known as hidden layers; and the output layer in which the final value searched is obtained. Each layer consists of different units (neurons or nodes) and each one is connected to the

following layer by a transfer function. (Dixit & Londhe, 2016; Medina et al., 2004; Portillo Juan & Negro Valdecantos, 2022) (Research paper 3.5.).

One of the main issues associated with ANNs is the inherent uncertainty in the models. Neural networks are known as "black-box models" because there is no clear explanation of how they work. Input data is provided, and results are obtained without any physical explanation, with very little traceability of the outcomes. Moreover, the development of the model itself involves significant associated uncertainty. This uncertainty and lack of understanding of how the model works may be the main reasons why numerical and traditional models are often preferred in the study of coastal engineering problems (Portillo Juan, Matutano, et al., 2023) (Research paper 3.6.). However, AI models are developing at unprecedented rates and the number of scientific studies in engineering applying AI has grown exponentially in recent years in our field (Figure 1-4).

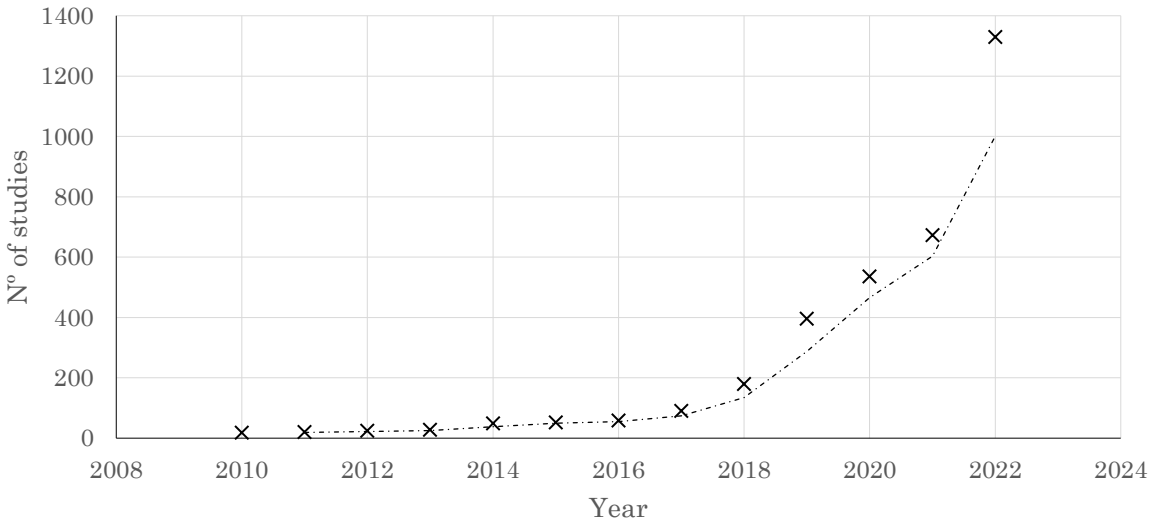


Figure 1-4. Evolution of the number of research studies that apply AI techniques in ocean and maritime engineering (Portillo Juan et al., 2024).

These models are showing very promising results, leading to very accurate models with much lower computational costs than traditional models. (Portillo Juan, Olalde Rodríguez, et al., 2023) (Research paper 3.8.). It is important that in our field we do not lag behind other disciplines and begin to implement these new models. We need to take advantage of the full potential of AI and address challenges that, so far, have limited numerical models.

The present thesis merges the two key agents of change in our current society: artificial intelligence and climate change. Its main objective is to integrate AI into

climate change research, specifically within the realm of adapting coastal structures and ports, thereby laying the groundwork for future research and providing guidance for studies in port and coastal engineering. The thesis delves into the application of AI models in the three key dimensions of port engineering in the context of climate change.

Firstly, it explores wave propagation from deep waters to coastal protection structures, proposing an alternative model that can complement traditional numerical models. Secondly, it examines the impacts of climate change on these structures, focusing on their redesign and adaptation to new climatic conditions, offering a new model that combines AI with probabilistic tools. Lastly, the thesis investigates port agitation, proposing a surrogate model to Boussinesq or mild-slope models. The ultimate goal is to provide a complementary tool to existing models, facilitating the development of preliminary studies and enabling climate change adaptation studies in small ports and marinas.

The validation of this new approach has been carried out in one of the most vulnerable and least scientifically studied regions, the Levante area of Spain, with a specific focus on the adaptation of the Port of Valencia at both the structural and operational levels. The Port of Valencia is one of the most important in Spain and is actively involved in the ECCLIPSE project. The ECCLIPSE project employs AI techniques to create early prediction models that facilitate the adaptation of ports to climate change. Therefore, this project is perfectly aligned with and complements the research direction of this doctoral thesis. To the best of the authors' knowledge, this doctoral thesis represents the first comprehensive port analysis using AI techniques in both structural and operational domains.

The aim of this thesis is to create a guide for future applications of AI in maritime and coastal engineering. The thesis has been developed with an educational approach, striving to explain the key concepts of AI within our field and to bring all the knowledge of machine learning, computational engineering, and computer science into our sector, thereby creating an interdisciplinary work that bridges pure maritime and coastal engineering with emerging data engineering disciplines.

After an extensive review of the scientific literature, it was observed that existing papers in our field that apply AI often do so without detailed or educational explanations, making it difficult for researchers and practitioners who wish to

start applying these new methodologies. This thesis aims to address this issue by applying AI methodologies in three phases, with a didactic approach, so that future technicians and readers can gradually understand — from the most basic concepts of AI model design to more complex ideas — using both simpler tools and more advanced software with great potential but greater complexity.

Furthermore, the content of this thesis can be extrapolated to other fields, serving as a guide for anyone interested in beginning to develop AI models and implementing these technologies in their area. The authors' aim is to guide the reader along the path of AI, facilitating its learning and implementation, and making AI accessible to anyone who wishes to apply it. As scientists and researchers, it is our responsibility to bring innovation and the most advanced models to society. There is little value in developing sophisticated models if they are not applied in the real world.

2. Methodology

Port adaptation studies to climate change can be divided into three main phases. First, the maritime climate and climate change projections are studied to model the propagation of present and future waves from deep waters to coastal waters. In the next phase, the structural behaviour of the port against the climatic event of interest is evaluated, analysing the structural and hydraulic behaviour of coastal protection structures. Finally, the functional and operational behaviour of the port is studied, including its agitation state and how this affects its operations.

The methodology proposed in this thesis aims to implement AI techniques in these three phases of climate change assessments in ports (Figure 2-1).

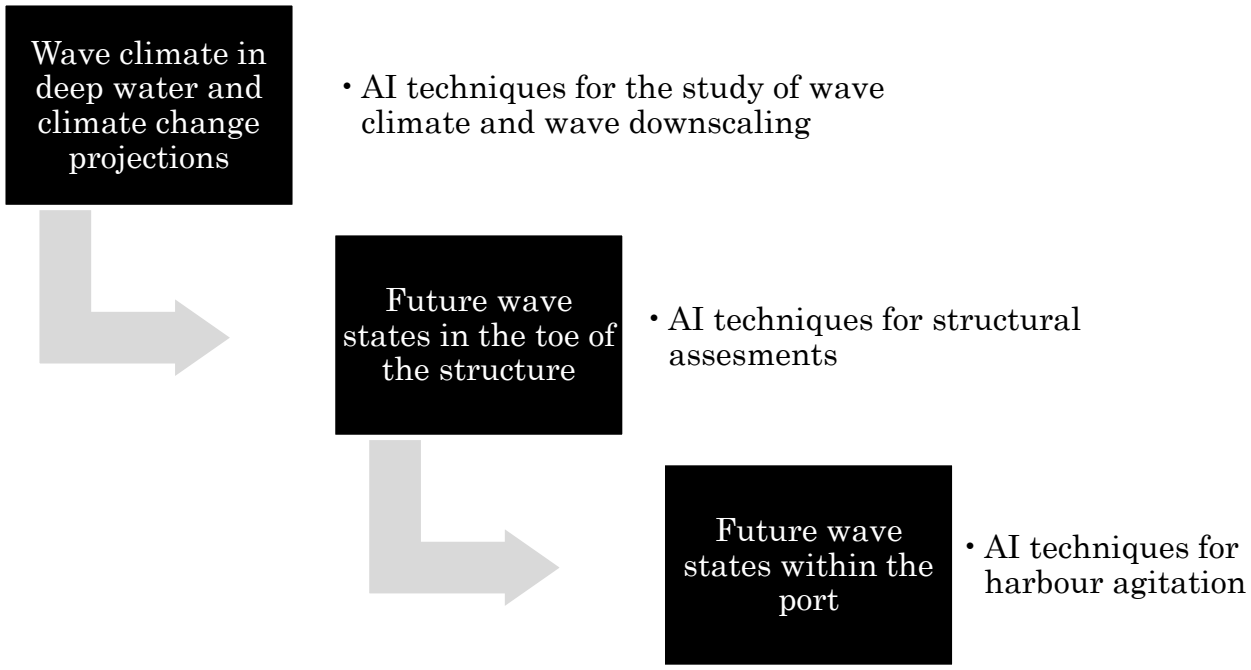


Figure 2-1. Scheme of the methodology proposed.

2.1. Introduction to Artificial Neural Networks (ANNs)

*The content of this section primarily corresponds to Papers 3.5 and 3.*6.

As pointed out in Section 1., there are many different AI techniques and, from all of them, the most popular among ocean and coastal engineers are ANNs. ANNs are a supervised learning technique classified as a machine learning methodology within artificial intelligence. Supervised learning methodologies are applied in

labelled datasets, which means that there are clear pairs of inputs-outputs and ANNs are suggested when the relationship between input-output pairs is intricate and the process to model is complex and involves different variables (Portillo Juan & Negro Valdecantos, 2022).

ANNs consist of multiple layers comprised of interconnected nodes, often referred to as "neurons" due to their resemblance to the functioning of the human brain. ANNs transmit information from one layer to another, ultimately producing the output in the final layer based on the input introduced into the initial layer. To pass the information from one layer to another ANNs use different transfer functions (Equations 1 to 4) that are presented in Table 2-1 (Yogitha & Mathivanan, 2018).

Table 2-1. Transfer functions used in ANNs.

Function	Expression		Application
Linear	$F(x) = x$	[1]	Output layers
Logistic sigmoid	$F(x) = \frac{1}{1 + e^{-x}}$	[2]	Output and hidden layers
Hyperbolic tangent sigmoid	$F(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$	[3]	Hidden layers
Radial basis	$F(x) = g(\|\|x - c\|\|)$	[4]	Output and hidden layers

ANNs are data-driven models that have the capacity to learn from extensive datasets consisting of input-output pairs. They are fed by huge datasets and, with these data, they are trained, validated and tested. Normally 70-80% of the data is used to train the model and 30-20% to validate it and test it (Mandal & Prabaharan, 2006). To train, validate and test the model different algorithms are used. These algorithms can be divided into four main groups: Gradient descent algorithms, Conjugate gradient algorithms (CG), Quasi-Newton's algorithms and Levenberg-Marquardt algorithm (LM). These algorithms are explained in the next subsection.

2.1.1. Algorithms

Gradient descent algorithms

The most common gradient descent algorithm is the Backpropagation (BP), it is also one of the most widely used in ANNs in general because it is the easiest and simplest to implement. It is called backpropagation algorithm because the error is propagated back to update the weights and the biases using an optimization technique (gradient descent), which minimizes the error (Dixit & Londhe, 2016). The backpropagation propagates the input forward through the network and the sensitivities backward, it updates the network weights and biases in the direction in which the performance function decreases most rapidly, which is the negative of the gradient. Its main drawbacks is that it is a local algorithm, which means that it can reach a local optimum instead of the global optimum, and that it is slow in converging (Hagan et al., 1996).

It also exists the variant gradient descent with momentum (GDM), which objective is to faster the convergence of the algorithm. The momentum smooths the oscillations of the algorithm and allows the network to respond not only to the local gradient, but also to recent trends in the error surface, avoiding getting stuck in a local optima (Demuth & Beale, 1992).

One of the main problems of the conventional backpropagation algorithm is that it is too slow. Therefore, there are other algorithms known as "faster training" that avoid these problems such as Resilient Backpropagation (RBP). ANNs usually use sigmoid functions in the hidden layers. These functions transform values from an infinite range to a range of $0 - 1$ and of -1 to 1. This can cause the gradient to have a very small magnitude, which can cause small changes in the weights and biases, generating problems when using the gradient descent technique. The objective of this algorithm is to avoid these effects by determining the weights and bias by a separate update value (Demuth & Beale, 1992).

Conjugate gradient (CG) algorithms

Conjugate gradient descent algorithms are also classified as "faster training". Common gradient descent updates the network weights and biases in the direction in which the performance function decreases most rapidly, which is the negative of the gradient. However, this does not necessarily produce the fastest

convergence. Conjugate gradient algorithms also search along conjugate directions. There are different conjugate gradient functions.

- Fletcher-Reeves (CGF): It computes the previous search direction by multiplying it by a factor that is the ratio of the norm squared of the current gradient to the norm squared of the previous gradient (Fletcher & Reeves, 1964).

- Polak-Ribiére (CGP): In this case the factor is the inner product of the previous change in the gradient with the current gradient divided by the norm squared of the previous gradient (Fletcher & Reeves, 1964).

- Powell-Beale Restarts (CGB): It restarts the direction of searching to the negative of the gradient if there is very little orthogonality left between the current gradient and the previous gradient (Powell, 1977).

- Scaled conjugate gradient descent (SCG): It uses the model-trust region approach to scale the step size (Møller, 1993).

Quasi-Newton's algorithms

Newton's algorithms usually converge faster than gradient methods, but they are much more computationally expensive because they have to compute the Hessian matrix. Quasi-Newton's algorithms are based on Newton's method, but do not require calculation of second derivatives. There are two main algorithms.

- Broyden, Fletcher, GolFarb and Shanno (BFGS): It is based on conjugate directions combined with line search. It requires more computation in each iteration and more storage, but it generally converges in fewer iterations. It is especially appropriate for small networks (Møller, 1993).

- One step secant algorithm (OSS): It deals with the problem of the storage needed for the BFGS algorithm. It assumes that at each iteration, the previous Hessian matrix was the identity matrix (Demuth & Beale, 1992).

Levenberg-Marquardt algorithm (LM)

It is a variation of Newton's method that was designed for minimizing functions that are sums of squares of other nonlinear functions. When the performance function has the form of a sum of squares the Hessian matrix can be approximated

as the product of the Jacobian matrix, that is much less complex than computing the Hessian matrix (Hagan et al., 1996). In the last years the LM algorithm has been widely used, showing excellence performance.

ANNs can perform very well in the training data set, but fail to perform in the test data set, which means that they do not perform well when adding new data. This happens because the ANN model memorizes the patterns within the pairs of inputs-outputs but does not learn the real relationship between them. When the ANN is not able to generalize the problem is known as "overfitting". This problem can be solved using regularization or early stopping. Regularization is a technique in which the different features are penalized by multiplying it by a factor λ, smoothing the function and making the network less complex. One of the most common regularization techniques used in ocean engineering is the Bayesian regularization (BR).

Apart from different algorithms and transfer functions, it exists different types of ANNs, that are used depending on the purpose of the study. The most common type of ANNs is the Feed Forward Multilayer Perceptron Neural Network (FFNN), but there also other network types that are explained below.

2.1.2. Classification of ANNs

FFNNs are characterized because they only process information in one direction, which means that there are not back connections (Teslyuk et al., 2021). They are the simplest and easiest ANNs and the most popular in ocean and coastal engineering. Their main drawback is that they are very sensitive to initial weights, and they may fail when processing too complex relationships such as those that exist when working with images.

Radial Basis Function Neural Networks (RBF NN) compute the output as a linear combination of basis functions generated by hidden neurons. They are especially recommended when non-linearities must be modelled (Kalra et al., 2005).

Convolutional Neural Networks (CNNs) are ANNs that use convolutional layers to extract the characteristics of the pairs of inputs-outputs by applying kernels or filters. They are the most common tool used in image processing because they can process huge amounts of data with very high levels of accuracy.

Recurrent Neural Networks (RNNs) are ANNs with internal loops (Marhon et al., 2013). The main difference with FFNNs is that in FFNNs the signal is passed only in a forward direction, whereas in RNNs the output obtained is fed back into the network through the input layer (Gopinath & Dwarakish, 2015). Their main drawback is that they can have difficulties in capturing long-term dependencies because the influence of a given input decays exponentially as it cycles around the network's recurrent connections (vanishing gradient problem). They are the best option when the data to train the model is noisy and usually the best option to forecast time-series (Portillo Juan et al., 2024).

Long short-term memory (LSTM) NN are developed based on the previously explained RNNs. Instead of the neurons of a conventional NN, they have long and short-term memory cells and forget and outputs gates to write, read and reset operations for the cells (Gao et al., 2021; Graves, 2012). These gates allow the LSTM NN to store information for a long period of time if it is required, which ends with the vanishing gradient problem. The main advantage with respect to RNNs is that they can be used to classify, process, or predict time series when there are time delays of unknown duration between important events and that they can save information in any length of time (Marhon et al., 2013). In the last years, they have become very popular to predict wave climate.

Once the ANNs are trained, validated and tested they can discern relationships within the data and can predict output values provided with new inputs. Various types of ANNs have been applied throughout different phases of port adaptation studies to climate change.

However, a primary drawback of ANNs is the lack of a clear process for optimizing various hyperparameters, such as model architecture, transfer functions, and algorithms. Typically, this optimization process relies on a trial-and-error approach, which can be tedious and time-consuming. Nonetheless, alternative techniques can expedite this process, such as optimization algorithms (Portillo Juan, Matutano, et al., 2023). Among these, Genetic Algorithms (GAs) stand out as one of the most common and effective optimization methods, and thus, it is the one that was also employed in this thesis.

The concept of GAs was firstly introduced by Holland (1992). GA is an algorithm that tries to mimic the genetic evolution process. It is based on the nature theory of Darwin in which the strongest and best adapted to nature are the ones that

survive. This translated into GA theory means that the characteristics that better fit to the fitness functions are the ones that prevail. In this algorithm there are three main elements; the parameters of the model (chromosomes), the fitness function, and the biological-inspired operators (crossover, mutation and selection), which are the tools used to reach the optimum solution (Katoch et al., 2021). The fitness function is the optimization function and the function that evaluates the quality of the parameters. For the process of selecting the best parameters of the model three main operators are used; crossover, which combines different sets of data creating a new generation of data; mutation, which produces a random change in the data, and selection that evaluates the best dataset and select the ones with the highest scores to create a new generation by applying crossover and mutation (Wang et al., 2020; Yedjour & Benyettou, 2018).

In the next subsections the specific methodology applied to each of the phases of the studies of adaptation of ports to climate change is explained in detail.

2.2. Wave downscaling

The content of this section primarily corresponds to Papers 3.8 and 3.9.

The first step to study the behaviour and adaptation of a port for a given climate event is to obtain the maritime climate within the port. Conventionally, maritime climate is derived from offshore buoy records and then it is propagated to the area of study. This process is known as wave downscaling and two phases can be distinguished. In the first phase, the propagation from deep water to coastal waters is calculated with models based on the energy balance equation, which allow large areas to be covered, but with relatively high error levels (phase-averaged models). The second phase of climate propagation is done at a local level, from coastal waters to the point of interest, with more sophisticated models (phase-resolving models), but also with a much higher computational cost (Portillo Juan, Olalde Rodríguez, et al., 2023). This subsection is focused on this first phase (phase-averaged models), while subsection 2.4. is focused on the second phase (phase-resolving models).

Phase-averaged models are based on the energy balance equation, which defines the wave spectrum evolution by the advection and refraction phenomena, and by a source function "S" that aggregates other physical processes that affect the wave spectrum (Janssen, 2008). Some of these most popular models are WaveWatch III (Tolman, 1991), WAM (Komen et al., 1996) or SWAN (Booij et al., 1999). The term

"S" usually includes nonlinear wind input, wave interactions and wave dissipation by bottom friction, white-capping and breaking (Umesh & Behera, 2020). This approach, by incorporating all these phenomena into the S component, has a lower computational cost compared to phase-resolving models, which allows its application over larger areas. However, it is also less accurate and carries much more uncertainty, which limits its application, as it cannot cover all spatial and temporal scales (Wang et al., 2022).

Since phase-averaged models with an energy balance approach have the drawbacks of numerical models (high computational cost) but are still limited in their accuracy, what it is proposed in this work is to develop an AI-based model using ANNs that can act as a surrogate model that solves the computational cost problem. This section is divided into three subsections, where each part of the study and propagation of maritime climate, which could potentially apply neural networks, is explained. These sections are mainly the study and statistical analysis of historical records and distributions, the creation of a synthetic data series that simulates future climate, and the design of the propagation model.

2.2.1. Statistical study of maritime climate

Traditionally, the study of maritime climate in the context of climate change has been conducted using different emission scenarios defined by the IPCC. These scenarios serve as the basis for generating a series of global and regional atmospheric models (GCMs and RCMs), which lead to various climate projections that translate into different future sea states (Lara et al., 2019; Lucio et al., 2019). These wave projections are then propagated to the point of interest using numerical models. Although this process is highly sophisticated and precise, it is also notably computationally expensive. To address this complexity and reduce computation time, an alternative and complementary approach is proposed (Figure 2-2).



Figure 2-2. Approach of the methodology proposed.

It is proposed to revisit the trend analysis and statistical study of climate, which were the basis of climate studies before the introduction of emission scenarios. However, trend analysis was abandoned due to its lack of reliability for making precise climate extrapolations (Casas-Prat & Sierra, 2010, 2012). This abandonment was mainly due to the lack of adequate tools to perform these extrapolations. However, the advent of AI, and particularly of ANNs — data models specialized in identifying complex relationships and optimizing the use of databases — has enabled a renewed application of this approach.

The first step in studying future maritime climate and propagating it is to download the historical records from Puertos del Estado and preprocess them. It is recommended to use buoy records whenever possible, but if not, other sources such as satellites or numerical models may be used (Juan Nerea et al., 2024) (Research paper 3.7). Once the data is downloaded, it must be preprocessed so that the ANNs can utilize it. The data feeding any AI model must meet three characteristics:

1. **Homogeneous**
   They must be consistent and uniform in type and structure, taken from the same source, and with a constant sampling period.
2. **Without gaps in time**
   If data is missing due to recording failures, it should be interpolated using linear regression techniques or neural networks.
3. **Balanced**
   There should be significant samples of all data. For example, in a set of positive and negative cases, it is not valid to have 80% positive cases and only 20% negative one, they need to be balanced (as discussed in Section 2.3).

If data is missing, which is very common in maritime climate instrumental records, techniques like interpolation or trend analysis should be applied. This interpolation can be done using traditional techniques like linear regression methodologies, or more advanced techniques like neural networks (Equation 5).

$$H\,{}_s^t, T_p^t, \theta_p^t = f^{FFNN}\left(H_s^{t-1}, T_p^{t-1}, \theta_p^{t-1}, H_s^{t+1}, T_p^{t+1}, \theta_p^{t+1}\right) \quad [5]$$

Where Hs is the significant wave height, Tp is the peak period, and θp is the direction.

Recent studies (Ahmed et al., 2023; Alemam et al., 2023) have demonstrated that neural networks are much more effective, prompting the authors of this thesis to encourage the use of ANNs in the preprocessing of data series. For the development of interpolation ANNs, the objective is to determine a function that, given two wave states at adjacent time instants (t-1 and t+1), predicts the wave state at instant t. The input data for the model are the wave states at instants t-1 and t+1, and the output is the wave state at instant t.

In other words, the network consists of six inputs and three outputs, with parameters requiring optimization. Since this is a relatively simple process to model, it is recommended to use the simplest network, specifically Feedforward Neural Networks (FFNNs) with a single hidden layer. Additionally, it is suggested to use the hyperbolic tangent sigmoid function and the linear function as activation functions.

In this incipient phase of the thesis and to start using and developing neural network models, the Matlab Deep Learning Toolbox (MathWorks, 2021) is suggested. This toolbox provides sufficient functionality to achieve the objectives of this phase of the study and offers a user-friendly interface compared to other available tools. It significantly simplifies the process of initiating work with ANNs compared to alternative software options. Additionally, it provides preset values for the more complex parameters. Thus, this initial phase of ANN development will focus on optimizing the number of nodes in the hidden layer. This optimization is carried out through a loop in MATLAB that tests different numbers of nodes (it is recommended to try between 1 and 100). The training algorithm also needs to be optimized. Using Portillo Juan and Negro Valdecantos (2022); Portillo Juan, Olalde Rodríguez, et al. (2023) as a reference, it is suggested to test the following algorithms: Backpropagation (BP), Resilient Backpropagation (RBP), Scaled Conjugate Gradient (SCG), Broyden-Fletcher-Goldfarb-Shanno (BFGS), Bayesian Regularization (BR), and Levenberg-Marquardt (LM).

2.2.2. Synthetic data generation

Once the data is preprocessed, a temporal and trend analysis must be conducted, obtaining statistical parameters and creating the corresponding graphs to analyse existing trends. Subsequently, distribution fitting and extreme value analysis are performed. After fitting the distributions, their future values are extrapolated. This can be done using simple techniques like linear regression or more advanced methods such as neural networks.

As mentioned in the previous section, climate models that study future maritime climate generally do so based on different emission scenarios predefined by the Intergovernmental Panel on Climate Change (IPCC, 2000). This approach entails a high computational cost due to the need to use and combine two types of climate models: global circulation models (GCMs) and regional circulation models (RCMs). The use of different GCMs/RCMs leads to various projections of future maritime climate, necessitating the consideration of different combinations of these models. This significantly increases the computational cost.

The alternative approach that is proposed in this thesis involves studying historical data series and trends and extrapolating these maritime climate trends into the future using AI tools. With the climate distributions defined and extrapolated, a synthetic data series is generated to simulate maritime climate from the present to the study's time horizon. This phase is carried out using level III probabilistic methodologies, specifically Monte Carlo simulations, which provides a simple and quick approach— the primary objective.

The generation of synthetic data and temporal extrapolation can also be addressed using neural networks, specifically through LSTMs (Equation 6). This type of architecture is particularly well-suited for modelling time series due to its ability to capture long-term dependencies between successive observations.

$$X^{t+1}, Y^{t+1}, Z^{t+1} = f^{LSTM}(X^t, Y^t, Z^t) \quad [6]$$

Where X, Y, Z are the parameters that define the climate distribution.

A detailed description of the LSTM network, along with other AI techniques for the development of time-sequence models, is provided in Annex C. Therefore, this section will focus exclusively on the development and analysis of the Monte Carlo simulation method, which was initially used in this study. The aim is to facilitate the reader's understanding, leaving more complex models such as LSTMs as supplementary material.

Monte Carlo simulations are a computational algorithm based on random sampling. This method is used to model the probability of various outcomes in processes that are complicated to define due to the involvement of many random variables. As described by Díaz et al. (2022), this methodology is developed through six key steps:

1. Problem Definition: Articulate the problem in terms of random variables, identify the main variables that define the process, and assign random values to them.
2. Probabilistic Characterization: Quantify the probabilistic characteristics of the identified random variables. This step involves characterizing the statistical distribution of these variables.
3. Variable Generation: Transform the random variables that define the problem into real variables according to the previously defined distribution.
4. Deterministic Evaluation: Evaluate the problem deterministically for each set of variables and compile the resulting dataset.
5. Extraction of Probabilistic Information: Extract probabilistic information from the simulations and characterize the new virtual dataset.
6. Evaluation of Process Accuracy: Finally, assess the accuracy of the entire process through validation procedures.

By using Monte Carlo simulations, a virtual database that encompasses wave climate data from the present moment to the target year can be constructed. The primary variables that define the process are Hs, θp and Mean Sea Level (MSL) with the latter included to account for sea level rise. Tp is modelled as a function of Hs, following the relationship proposed by Puertos del Estado (Equation 7).

$$T_p = a \times (H_s)^b \quad [7]$$

Random values between 0 and 1 are assigned to Hs and θp, for each year from the present to the target year. These values between 0 and 1 represent the probability of occurrence. These probabilities are then transformed into the variables Hs and θp by applying the corresponding distribution function. In the case of Valencia, the Weibull distribution is used for Hs and the uniform distribution adjusted to the wave rose is used for θp.

Given the context of climate change, it is also necessary to account for Mean Sea Level (MSL) in order to incorporate the effects of sea level rise. For historical

records, the MSL can be taken from the corresponding tide gauge. However, for the synthetic data series, the estimated sea level rise needs to be implemented. In the Valencia area, an approximate rise of 0.25 m is expected by 2050, so this rise is progressively implemented throughout the simulations (Portillo Juan et al., 2022a; Portillo Juan et al., 2022c) (Research papers 3.3. and 3.4.).

As many simulations as the quantity of records from the present year to the target year must be conducted. Once the simulations are completed, the historical and synthetic data series must be unified so that a database is available, covering from the first year of available records to the target year of study. It is important to note that these two steps—the statistical study and the synthetic data generation by Monte Carlo—must be performed for both deep and coastal waters, as the final model aims to propagate waves from deep waters to coastal areas.

2.2.3. Propagation model

Once the necessary data to feed the model has been obtained, the final propagation model, governed by Equation 8, is designed. In this equation, the superscript *dw* denotes "propagated" (downscaled), while the superscript 0 refers to deep waters.

$$H_s^{dw}, T_p^{dw} = f^{ANN}\left(H_s^0, T_p^0, \theta_p^0, MSL\right) \quad [8]$$

For wave climate studies, it is advisable to opt for a shallow ANN with a single hidden layer, as the complexity of the model being processed is not excessively high. Introducing additional hidden layers and adopting deeper architectures can lead to increased computational time and costs, along with a heightened risk of overfitting, especially when the underlying process does not demand deep architectures. Hence, FFNNs with one hidden layer emerge as the preferred choice for characterizing and propagating wave climate. Regarding transfer functions, the hyperbolic tangent sigmoid function and linear function have demonstrated superior performance in this context (Portillo Juan & Negro Valdecantos, 2022) (Equations 1 and 3).

The number of neurons in the input and output layers is already determined by the physics of the problem. The number of input neurons of the ANN is 4: wave height, peak period, and direction in deep water and MSL ($H_s^0, T_p^0, \theta_p^0, MSL$). The number of output neurons is 2: wave height and peak period in shallow waters [H_s^{dw}, T_p^{dw}].

In this case, we do not recommend using wind data, as the majority of the effort in data-driven and AI-based models lies in data pre-processing. While incorporating wind data could potentially improve the model's accuracy, it would also significantly increase computational costs, something this approach aims to avoid.

For wave downscaling, a distinct ANN must be developed for each site. ANNs are data-driven models, and when the physics or characteristics of the modelled process vary, it becomes necessary to construct a new ANN. This is because ANNs learn patterns that correlate inputs and outputs, and discrepancies in these patterns can lead to model errors. Hence, different ANNs need to be developed for each local site. Just as the bathymetry is modified in numerical models, in AI models, the NN must be adjusted to reflect the unique characteristics of each specific area, as each NN learns patterns valid only for those processes that can be characterized with the same parameters.

Once the datasets are preprocessed, the optimization of the hyperparameters of the ANNs must be conducted. In the Matlab module used, the more complex parameters of FFNN networks, such as epochs, batch size, etc., which require a deeper understanding of neural networks (Section 2.4), are given by default. Therefore, this phase will not focus on these parameters. This first phase of implementing AI in ocean models will simply focus on optimizing the number of neurons in the hidden layer and in optimizing the algorithm of the ANN, following an approach very similar to that of subsection 2.2.1 – statistical study of the maritime climate.

For the optimization of the ANN architecture, the authors of this thesis propose to develop a loop in Matlab that evaluates the error of the model (training and validation error) along the different number of neurons (Appendix A). It is proposed to test from 1 to 100 hidden neurons and select that ANN with the least training and validation error. It must be noted that the training error typically decreases as complexity is introduced into the model, whereas the validation error tends to initially decrease, reach a minimum, and then rise again. If this pattern emerges, it could indicate overfitting of the model. Therefore, it is crucial to pay attention to this phenomenon and ensure that the validation error does not surpass a certain threshold.

For the evaluation of the model performance, the Matlab Deep Learning Toolbox uses the Mean Square Error (MSE) by default (Equation 9), but the authors also suggest using the Root Mean Square Error (RMSE), correlation coefficient (R),

Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Bias (Equations 10 to 14) to have a more complete understanding of the performance of the model.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (\mathrm{o}_i - \mathrm{u}_i)^2 \quad [9]$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\mathrm{o}_i - \mathrm{u}_i)^2} \quad [10]$$

$$R = \frac{\sum_{i=1}^{n}(o_i-\bar{o})(u_i-\bar{u})}{\sqrt{\sum_{i=1}^{n}(o_i-\bar{o})^2}\sqrt{\sum_{i=1}^{n}(u_i-\bar{u})^2}} \quad [11]$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|o_i - u_i| \quad [12]$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|o_i-u_i|}{u_i} \quad [13]$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(o_i - u_i) \quad [14]$$

Where o_i represents the output of the model and u_i denotes the real value.

Once the architecture of the ANN is optimized, the algorithm that operates best must be selected (Appendix A). The recommendation is to test different algorithms: RBP, SCG, BFGS, BR and LM are the algorithms to be tested, which are the algorithms that have proven to perform better in wave characterization studies (Portillo Juan & Negro Valdecantos, 2022).

With all the hyperparameters selected, the next step is the training of the ANN. By default, Matlab allocates 70% of the data for training the model and the remaining 30% for validation and testing purposes. This data separation is carried out randomly; however, the division can be modified into contiguous blocks, which is considered a more suitable approach for wave climate characterization.

Finally, it must be said that for the training of the ANN, Matlab uses different initialization weights and biases that may affect the performance of the model, to reduce the influence of these parameters, there are two options: To use optimization algorithms or to repeat the training process enough number of times and use the average values of the models. This second approach is the most rudimentary, but also the easiest to implement and to understand. In addition, it

is a repetitive process that helps to get familiar and assimilate some of the basic concepts of ANNs. Therefore, trial and error was the methodology selected in this first phase of AI implementation. In Section 2.3, the implementation of optimization algorithms will proceed.

The methodology for developing an ANN model for wave downscaling is outlined in Figure 2-3. This approach stands out as the simplest among the three methodologies employed in the different study's phases, making it an ideal starting point for users embarking on their journey into AI-based models. Beginning with straightforward models like this one allows users to become acquainted with ANNs and AI before delving into more complex and detailed model training. This gradual progression enables a deep understanding of each model parameter over time. Therefore, the AI implementation methodology proposed in this thesis is recommended for all users seeking to initiate themselves in this type of models. The implementation of AI models has been structured into three phases, with increasing model complexity at each stage. This phased approach is designed to be extrapolated and applied beyond the scope of maritime and coastal engineering, offering a framework for other fields seeking to incorporate AI into their practices.

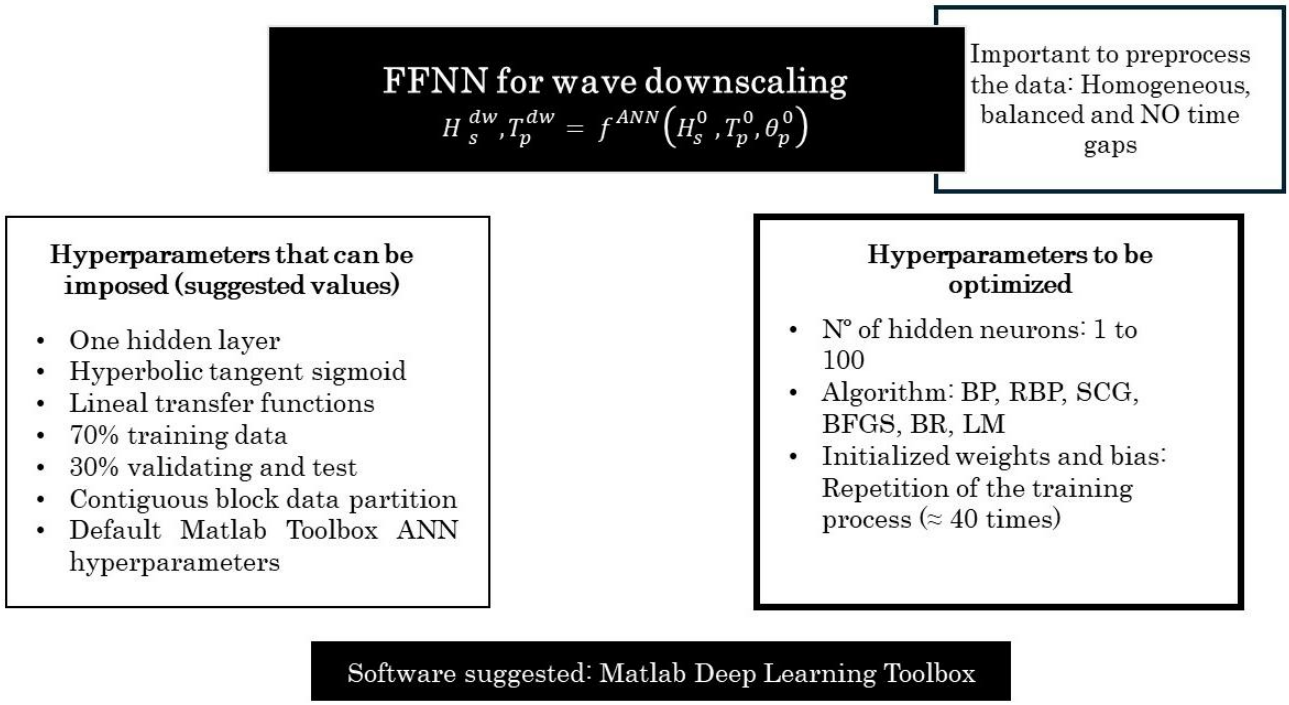


Figure 2-3. ANN model for wave downscaling.

2.3. Structural assessment (ULS)

The content of this section primarily corresponds to Paper 2 in Appendix B.

The following level, and more complex, when implementing ANNs in coastal engineering would involve the structural and hydraulic study of coastal protection structures. Once the maritime climate has been propagated to the toe of the structure through FFNNs, the focus shifts to evaluating the impact of these sea states on breakwaters and determining the necessary adaptation measures in response to climate change. As highlighted throughout this thesis, this second phase in the implementation of AI in coastal and maritime studies introduces greater sophistication to the neural network models by incorporating genetic algorithms (GAs). Rather than relying on multiple training iterations and averaging the resulting outputs to reduce the influence of initial weights and biases, the proposed approach integrates GAs directly into the network design. This enhancement introduces a higher degree of complexity to the model, enabling direct optimization of weights and biases, and thereby eliminating the need for repeated training. This methodological advancement moves beyond the traditional trial-and-error strategies typically used in earlier, educational applications of AI, offering a more efficient and robust alternative for real-world engineering problems.

The model proposed is already a much more complex model than the previous one as it incorporates the interaction of two completely different systems, the coastal protection structure and the maritime climate itself. Numerical models that study the interaction between structures and maritime climate often entail substantial computational expenses, restricting their application to highly specific studies. The integration of AI-based models offers a solution to this issue, enabling the development of adaptation strategies for all our coastal protection structures in response to climate change, a matter of particular significance given the present situation.

The sum-up of the methodology of this second phase of study is presented in Figure 2-4. Block 1 corresponds to the methodology already explained in Section 2.2. and block 2 to this new phase.

This section will be divided in two subsections. In Section 2.3.1 it will be explained how to characterize the coastal structures of a given area with a probabilistic

approach and in Section 2.3.2, the integration of the breakwater characterization (Block 2) and the wave climate characterization (Block 1) into the AI-based model.

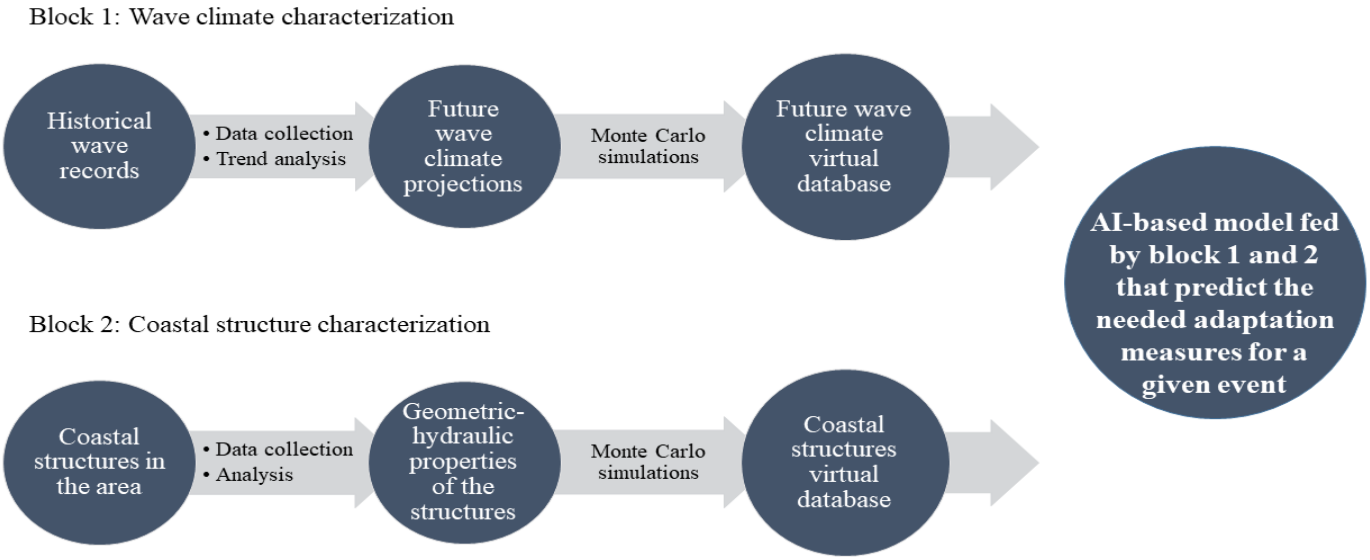


Figure 2-4. Methodology sum-up for phase 2.

2.3.1. Probabilistic characterization of coastal structures

For probabilistic characterization, information must be collected from all coastal protection structures in the area of interest. Neural networks learn patterns that relate data, and when the parameters that model a process change, so do the relationships between them. Thus, when these parameters and relationships change, a new ANN is needed to model the process of interest. In this case, our goal is to adapt the breakwaters of a specific port, and to characterize the coastal structures in the port area. Therefore, the authors recommend collecting information from all protection structures along the coastal strip where the port is located. Breakwaters in a coastal strip are usually designed according to similar criteria to withstand the characteristic climatic phenomena of that area, allowing them to be assimilated by a single neural network. Collecting data from the entire coastal strip encompassing the port and the relevant breakwater enables to generate a representative statistical sample of the area. This dataset allows for the probabilistic characterization of the coastal structures, which are influenced by similar phenomena. As a result, all the data gathered are valid for training the neural network model.

However, it is important to distinguish between the different types of coastal protection structures. Although data is collected along the entire coastline, a

separate ANN model should be developed for each specific type of structure during the model development phase. Once all relevant data has been gathered, the primary geometric and hydraulic variables that characterize breakwaters must be identified. Table 2-2 summarizes the key parameters selected to define each type of breakwater.

Table 2-2. Parameters that characterize breakwater behaviour.

Type	**Variable**
Vertical	h, d, h', R_c, A, W_v, q_{max}
Rubble Mound	h, d, h', R_c, W, q_{max}, γ
Compound	h, d, h', R_c, A, W_m, W_{50}, q_{max}, γ

Where h is the design water depth; d, the distance from the berm to the Mean Sea Level (MSL); h', the distance from the bedding layer to the MSL; R_c, the freeboard; A, the width of the monolith; W_v, the weight of the monolith of the vertical breakwater; q_{max}, the maximum allowable overtopping; γ, the specific weight of the blocks; W, the weight of the armour of the rouble mound breakwater; W_m, the weight of the monolith of the compound breakwater; and W_{50}, the weight of the armour of the compound breakwater (Figure 2-5).

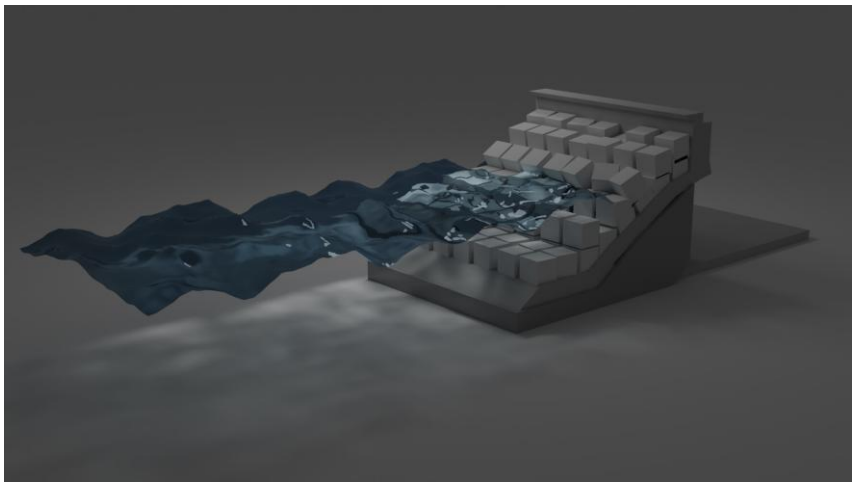


Figure 2-5. Simulation of a rubble-mound breakwater.

To conduct the Monte Carlo simulations, an in-depth statistical analysis of the existing correlations among the key variables was undertaken. All the identified variables were found to be interdependent, except for γ and the armour weight which were directly correlated. As a result, only the armour unit weight was selected for characterization and incorporation into the Monte Carlo simulations. Following this, an analysis of their distributions and correlations was conducted.

Once the independent variables are defined in the Monte Carlo simulations, all the intermediate variables necessary to calculate the required adaptation measures of the coastal structure under study must be calculated. The ultimate goal is to obtain the safety factor against sliding (SSC), the required increase in breakwater height (Δh), and the necessary increase in armour weight (Δw) to adapt coastal structures to the new situation of climate change – the SSC in the case of vertical and compound breakwaters and the increase of armour weight in the case of rubble mound and compound breakwaters.

To effectively carry out this task, it is first necessary to filter the various breakwater configurations generated through Monte Carlo simulations, in order to eliminate those that lack physical plausibility. Monte Carlo simulations, while efficient for data generation, can sometimes produce configurations that are not feasible in practice. Therefore, it is crucial to apply filters to ensure the validity of the generated data.

Once the breakwater configurations have been filtered, the next step involves combining each valid configuration with every possible wave scenario presented in Section 2.2. This entails developing a code capable of systematically generating these combinations, considering all permutations of breakwater configurations and wave cases.

This approach serves two primary objectives. First, it enables the generation of a substantial volume of data—potentially exceeding one million cases. Second, it ensures that the analysis is based solely on physically meaningful data, thereby providing reliable insights into the interactions between various breakwater configurations and wave scenarios. Following this, the relevant equations for calculating $SSC, \Delta h$ and Δw. can be applied. The proposed equations for $SSC, \Delta h$ and Δw are the following.

<u>Vertical breakwaters</u>

The design wave height can be defined following Goda formulae (Equations 15 to 17) (Goda, 1970, 1975; Takagi et al., 2011).

$$H_D = mín(H_b, H_{max}) \quad [15]$$

$$H_{max} = H_{1/250} \cong 1.8H_s \quad [16]$$

$$H_b = 0.17L_0\left\{1 - exp\left[-1.5\frac{\pi h}{L_0}\left(1 + 15\tan\theta^{4/3}\right)\right]\right\} \quad [17]$$

Where L_0 is the wavelength in deep water and $\tan\theta$ the slope of the ground.

The Goda-Takashi wave load formula (Goda, 2010) is also recommended to compute the loads on the vertical breakwater (Equations 18 to 25).

To calculate the pressure, first the auxiliar parameters α_1, α_2 and α_3 must be defined (Equations 18 to 20).

$$\alpha_1 = 0.60 + \frac{1}{2}\left[\frac{\frac{4\pi h}{L}}{\sin\frac{4\pi h}{L}}\right]^2 \quad [18]$$

$$\alpha_2 = \min\left(\frac{h_b - d}{3h_b}\left(\frac{H_D}{d}\right)^2, \frac{2d}{H_D}\right) \quad [19]$$

$$\alpha_3 = 1 - \frac{h'}{h}\left(1 - \frac{1}{\cosh\frac{2\pi h}{L}}\right) \quad [20]$$

Once α_1, α_2 and α_3 are defined, the pressure on the breakwater can be calculated (Equations 21 to 25). Where β is the incident wave direction, ρ_w is the water density, g is the gravity and η is the wave run-up (Equation 21).

$$\eta = 0.75\,(1 + \cos\beta)H_D \quad [21]$$

$$P_1 = \frac{1}{2}(1 + \cos\beta)(\alpha_1 + \alpha_2 \cos\beta^2)\rho_w g H_D \quad [22]$$

$$P_2 = \frac{P_1}{ch\ 2\pi\frac{h}{L}} \quad [23]$$

$$P_3 = \alpha_1 \alpha_3 \rho_w g H_D \quad [24]$$

$$P_4 = \frac{\eta - h_c}{\eta} P_1 \quad [25]$$

Figure 2-6 shows the diagram of the pressures, P_1, P_2, P_3 and P_4. For the forces, the pressure area is integrated.

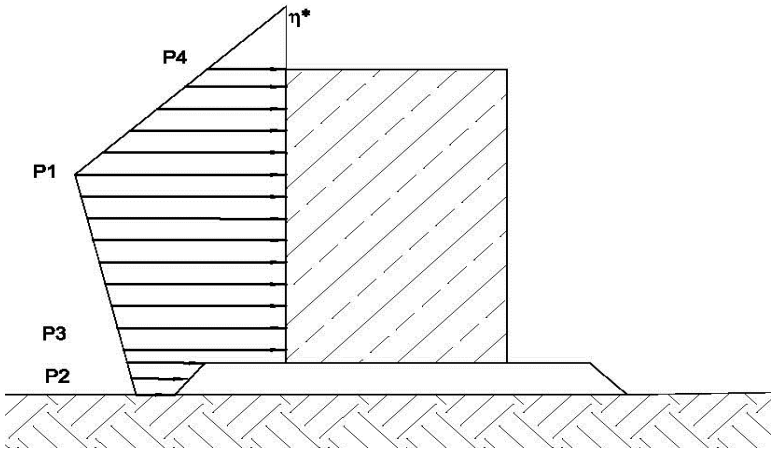


Figure 2-6. Pressure diagram on a vertical breakwater taken from Juan Nerea et al. (2024).

To calculate the safety coefficient for sliding (SSC), Goda equation can also be used (Goda, 2001) (Equation 26).

$$C_{SD} = \frac{\sum \delta F_v}{\sum F_H} = \frac{0.6(W_{50} - p_u)}{\sum F_H} \quad [26]$$

Where p_u is the uplifting pressure and equals to p_3 and C_{SD} is the sliding safety coefficient (SSC).

To calculate the overtopping, and the maximum admissible overtopping, it is suggested to apply the EurOtop manual (van der Meer et al., 2016) and Franco et al. (2010) formula (Equation 27).

$$\frac{q}{\sqrt{gHs^3}} = 0.082 \exp\left(-3 \frac{R_c}{H_s}\right) \quad [27]$$

<u>Rubble mound breakwaters</u>

The design wave height can be defined following Goda formulae (Equation 28 and 29) (Goda, 1970, 1975; Takagi et al., 2011).

$$H_D = mín\left(H_b, H_{1/10}\right) \quad [28]$$

$$H_{1/10} = 1.27 H_s \quad [29]$$

For the calculation of the block damage N_{od} the Van der Meer formula is recommended (Equations 30 to 32) (Jcntsje & Meer, 1988).

$$\frac{H_s}{\Delta D_{n50}} = \left(6.7 \frac{N_{od}^{0.4}}{N^{0.3}} + 1\right) s_m^{-0.1} \quad [30]$$

$$s_m = \frac{2\pi H_s}{g T_{m,0}^2} \quad [31]$$

$$\Delta = \left(\frac{\gamma}{\gamma_w} - 1\right) \quad [32]$$

Where γ_w is the specific weight of water and N 1000 waves for the Mediterranean.

Finally, for the overtopping, the EurOtop manual (van der Meer et al., 2016) and Franco et al. (2010) formula are suggested (Equation 33).

$$\frac{q}{\sqrt{gHs^3}} = 8 \times 10^{-5} \exp\left(3.10 \left(\frac{Ru_{ʒ0}^{2\%} - Rc}{Hs}\right)\right) \quad [33]$$

Where $Ru_{ʒ0}^{2\%}$ is defined as a function of the number of Iribarren ʒ (Allsop & Hettiarachchi, 1989) (Equations 34 and 35).

$$Ru_{ʒ0}^{2\%} = aʒ \; if \; ʒ < 1.5 \quad [34]$$

$$Ru_{ʒ0}^{2\%} = bʒ^c \; if \; ʒ > 1.5 \quad [35]$$

Compound breakwaters

The design wave height can be defined using Goda formulae (Equations 36 and 37) (Goda, 1970, 1975; Takagi et al., 2011).

$$H_D = mín(H_b, H_{max}) \quad [36]$$

$$H_{max} = H_{1/250} \cong 1.8H_s \quad [37]$$

For the calculation of the block damage N_{od} and overtopping q, the same formula as with rubble mound breakwater can be applied.

To calculate the sliding safety coefficient of the monolith, the same Goda-Takashi wave load formulas as in the vertical breakwater are suggested (Goda, 2010). The wave height used to calculate the sliding safety coefficient is the transmitted wave height (H_t) trough the armour (Equations 38 and 39) (van der Meer et al., 2001).

$$c_t = 0.46 - 0.3\frac{R_c}{H_s} \quad [38]$$

$$H_t = c_t H_s \quad [39]$$

Finally, to define the adaptation measures for the vertical, rubble mound, and compound breakwaters, it is necessary to calculate the required increases in height and weight to achieve the prescribed safety coefficients and limit overtopping within the predefined maximum allowable limits. This involves applying the presented equations to determine the necessary height and weight for the breakwater to meet the safety standards. These calculated values are then compared to the actual height and weight of the breakwater to determine the required increases.

After completing these calculations, all the requisite sets of inputs and outputs to construct the AI-based model are obtained. The wave climate conditions (Hs, Tp and θ_p) at the toe of structure along with the breakwater configurations serve as inputs, while the desired outputs of the AI model are the SSC, Δh, and Δw needed to adapt the breakwater to climate change.

2.3.2. Implementation of the AI-based model

Now, with all the databases generated (Block 1 and Block 2 from Figure 2-4). The first step always to develop an AI-based model is to preprocess the datasets that are going to feed the model. As said before, these datasets must have three main characteristics: homogenous, balanced and with no gaps. In this case, the datasets are homogeneous and do not have gaps, as this issue was already addressed in the

time series in Section 2.2.1, and the new virtual database generated in Section 2.2.2 has been made consistent with the dataset. However, the dataset is likely to be imbalanced; in this study, depending on the breakwater configuration, between 75% and 90% of the cases were negative—indicating no need for adaptation measures—while only 10% to 25% were positive cases, requiring such measures.

This imbalance between the number of cases requiring adaptation measures and those that do not, can lead to the ANN predicting false negatives. As it was explained, neural networks learn patterns that relate data in the most efficient way possible. Therefore, if most data indicate that no adaptation measures are needed, the network will deduce that this is almost always the case and will predict that no adaptation is necessary even when it is.

To address this issue and prevent biased datasets, two primary techniques are commonly employed: undersampling and oversampling. Undersampling involves reducing the number of samples in the majority class, which are the negative cases in this context. Conversely, oversampling entails increasing the number of samples in the minority class using various data mining techniques. The choice between these methods depends on the quantity of data samples available. Oversampling often demands more effort, as it involves generating new data from the minority class. In contrast, undersampling simply entails randomly selecting different data cases from the available dataset. Therefore, if enough data is available, the authors particularly recommend undersampling due to its simpler implementation.

Once the entire dataset has been preprocessed, the next step is to design the AI model. Both, TensorFlow (Abadi et al., 2015) and Keras DL (Francois, 2015) in the environment Python 3.12 (Booij et al., 1999) with Matlab (MathWorks, 2021) are proposed for implementing FFNNs with one hidden layer, employing hyperbolic tangent sigmoid and linear transfer functions. The design phase includes optimizing the number of neurons in the hidden layer and selecting the most suitable training algorithm (Appendix A).

Following the methodology outlined in Section 2.2, the optimization of weights and biases is further enhanced by GAs, moving beyond traditional trial-and-error approaches towards more robust optimization techniques. The number of inputs in this second phase of the study increased from four (wave parameters; Hs, Tp, θp and MSL) to 12-13, depending on the breakwater type. Consequently, the computational cost increases significantly, approximately 100 times more than in the wave downscaling ANN model. Therefore, repeating the training process 50

times, as done previously, would consume considerable time and computational resources. To address this issue, it is proposed to integrate GAs directly into the code, enabling the optimization of the model's initial weights and biases. This strategy ensures a more efficient optimization process while substantially reducing the computational load associated with repeated training cycles (See Appendix A).

GAs were implemented with 100 iterations, a population size of 200, 80% crossover, 20% mutation, and the Roulette Wheel selection function. Equation 40 defines the new hybrid neural network model.

$$CS_D, \Delta h, \Delta W = f^{FFNN}\left(\delta^{dique}, H_s^{dw}, T_p^{dw}, \theta_p, NMM\right) \quad [40]$$

2.4. Port agitation (SLS)

The content of this section primarily corresponds to Paper 3 in Appendix B.

The third and final phase of the study on adapting ports to climate change focuses on assessing their operational capacity and anticipating the impact of changing climate conditions on their functionality. After examining the port in ULS, where the structural and hydraulic performance of the protective structures under climate change was analysed, the focus now shifts to the port in SLS. In this phase, the impact of climate shifts on its overall operability and serviceability will be assessed.

As anticipated in Section 2.2., the established methodology to study harbour agitation is with numerical models. The existing numerical models are phase-resolving models that are very computationally intensive. They solve diffraction and other non-linearities of wave climate which incurs in a very high computational cost.

Among the various types of models considered, it was decided to apply MILDwave model developed by Troch (1998). The choice of MILDwave stems from its consideration of diverse transformation phenomena, including absorption, refraction, shoaling, reflection, transmission, diffraction, and breaking, while maintaining efficiency in computational resources and user-friendliness. As such, it offers an ideal solution for simulating port agitation, avoiding excessive computational costs and mitigating significant numerical complexities often associated with other models. Nonetheless, the methodology outlined in this section is adaptable and can be applied to alternative numerical models as well.

Subsection 2.4.1 explains the MILDwave model and subsection 2.4.2 explains the AI model. Thus, everything related to the development of the AI model, which is subsection 2.4.2, is valid for any numerical model that the reader decides to use.

2.4.1. MILDwave model

MILDwave is a mild-slope equation model based on the mild-slope equations of Radder and Dingemans (1985) (Equations 41 to 43).

$$\begin{cases} \frac{\partial \eta}{\partial t} = B\varphi - \nabla \cdot A\nabla\varphi \\ \frac{\partial \varphi}{\partial t} = -g\eta \end{cases} \quad [41]$$

$$A = \frac{C \cdot C_g}{g} \quad [42]$$

$$B = \frac{\omega^2 - k^2 C \cdot C_g}{g} \quad [43]$$

Where η is the surface elevation; φ, the velocity potential at the still water level; C, the phase velocity; C_g, the group velocity; ω, the angular frequency; and k, the wave number.

MILDwave uses finite differences to solve these equations, using derivative central differences for both, space, and time. The velocity potential is calculated at every time step $(n+1)\Delta t$ and the surface elevation at every two-time step $\left(n+\frac{1}{2}\right)\Delta t$ following equations 44 and 45.

$$\varphi_{i,j}^{n+1} \cong \varphi_{i,j}^{n} - g\eta_{i,j}^{n+\frac{1}{2}}\Delta t \quad [44]$$

$$\eta_{i,j}^{n+\frac{1}{2}} \cong \eta_{i,j}^{n-\frac{1}{2}} + B_{i,j}\varphi_{i,j}^{n}\Delta t - \frac{A_{i+1,j}-A_{i-1,j}}{2\Delta x}\frac{\varphi_{i+1,j}^{n}-\varphi_{i-1,j}^{n}}{2\Delta x}\Delta t - A_{i,j}\frac{\varphi_{i-1,j}^{n}-2\varphi_{i,j}^{n}+\varphi_{i+1,j}^{n}}{(\Delta x)^2} - \frac{A_{i,j+1}-A_{i,j-1}}{2\Delta y}\frac{\varphi_{i,j+1}^{n}-\varphi_{i,j-1}^{n}}{2\Delta y}\Delta t - A_{i,j}\frac{\varphi_{i,j-1}^{n}-2\varphi_{i,j}^{n}+\varphi_{i,j+1}^{n}}{(\Delta y)^2}\Delta t \quad [45]$$

Where Δx and Δy are the spatial dimensions; Δt, the time step; and the superscripts n, i and j denote the time and spatial (x and y) step respectively.

For the wave generation MILDwave uses the source term addition method, generating waves from the offshore boundary and propagating the source term with the energy velocity. It can generate both, regular and irregular waves.

Regular waves are calculated following equation 46 and irregular waves following equation 47.

$$\eta_i = a\sin(kx - \omega t + \phi) = a\sin(k\cos\theta\, x - k\sin\theta\, x - \omega t + \phi) \quad [46]$$

Where a is the wave amplitude; ϕ, the phase shift; and θ, the angle at what wave propagates from the x-axis (Troch, 2018).

$$\eta_i = \sum_{n=1}^{N} \sqrt{2S(f_n)\Delta f}\,\sin(k_n x\cos\theta + k_n y\sin\theta - \omega_n t + \phi_n) \quad [47]$$

Where $S(f_n)$ is the input frequency spectrum – in this case the JONSWAP spectrum was used (Equation 48); Δf, the frequency interval (Equation 49); ω_n, the angular frequency (Equation 50); and ϕ_n, the random phase.

$$S(f) = \frac{0.0624}{0.230+0.0336\gamma-\frac{0.185}{1.9+\gamma}} \cdot H_s^2 \cdot f_p^4 \cdot f^{-5} \cdot \gamma^{exp-\frac{(f-f_p)^2}{2\sigma^2 f_p^2}} exp\left(-\frac{5}{4}\left(\frac{f_p}{f}\right)^4\right) \quad [48]$$

Where f_p is the peak wave frequency; γ, the peak enhancement factor; and σ, the spectral width parameter.

$$\Delta f = \frac{f_{max}-f_{min}}{N-1} \quad [49]$$

$$\omega_n = 2\prod f_n = 2\prod(n-1)\Delta f + 2\prod f_{min} \quad [50]$$

MILDwave uses absorbing boundaries to generate a homogeneous wave field, allowing waves travelling through the domain without numerical disturbances. For that purpose, absorbing sponge layers are used in the edges of the model. Numerical absorption is achieved by applying the calculated surface elevation at each time step to an absorption function. This function smoothly transitions from a value of 1 to 0 across the span of the sponge layer, gradually attenuating the wave energy. Furthermore, periodic lateral boundaries are implemented to ensure homogeneity of the wave field, particularly for oblique waves. This mechanism facilitates the transfer of numerical information from one side of the domain to the opposite side, enabling waves to propagate seamlessly across the computational domain.

For the application of MILDwave, the different wave cases to study must be defined. MILDwave operates by introducing a unitary wave ($Hs = 1m$) to determine the amplification factor, Kd, within the port. Therefore, study cases focus on varying wave periods and directions. The wave periods derived from the ANN downscaling model are analysed. The objective is to select the most representative peak periods and directions, aided by the swell rose, to subsequently study them.

Once the wave study cases have been defined, a bitmap file of the port is created and distinct colours are assigned to different types of structures (vertical breakwaters, beaches, armour layer breakwaters, etc.). Each colour corresponds to

a different diffraction coefficient that must be obtained through model calibration. With these preparations complete, we have all the necessary data to execute the numerical simulations that will feed the AI model.

2.4.2. Implementation of the AI model

The difficulty and complexity of the process to be modelled have increased with each phase of the present thesis. First, an AI model for wave downscaling was developed, followed by a hybrid ANN-GA model to adapt coastal protection structures to climate change. In this final phase of the study, the aim is to develop an AI model to study port agitation. The study of port agitation is a complex process that involves numerous systems and variables, such as port design, bathymetry, coastal protection structures, maritime climate, etc. Consequently, incorporating all this data requires a much more developed and advanced AI model than those implemented so far.

The first step in developing the model is to incorporate all the information related to the port. The most effective way to integrate all the port-related information—design, coastal protection structures, bathymetry, etc.—was through images. These images are then combined with wave data to obtain the final agitation model. The port photos were integrated into the AI model using a specific type of network, convolutional networks. While traditional FFNNs can be applied to incorporate wave data, they fall short when working with images due to the large volume of data involved. This is why Convolutional Neural Networks (CNNs) were chosen for image processing, given their proven efficiency in handling such tasks.

CNNs, unlike conventional ANNs, offer a distinct architecture tailored for tasks like image processing. While FFNNs rely on fully connected layers where neurons are interconnected between adjacent layers, CNNs employ a sequence of specialized layers: convolutional and pooling layers, designed for feature extraction and dimensional reduction (Ruscio et al., 2023). In contrast to FFNNs, where the output is determined by connections between individual neurons, convolutional layers in CNNs are defined by three dimensions: width, height, and depth. Width and height correspond to the dimensions of the input images, while depth denotes the number of filters applied to extract the features of these images.

Deeper layers can capture more intricate features, but this depth comes with increased computational demands. To manage this, pooling layers are often employed after convolutional layers to reduce computational overhead. Typically,

as CNNs progress through layers, depth increases, with initial layers capturing fundamental image features and later layers refining the details.

Crucially, convolutional layers operate on local regions of input, with shared weights and biases across the layer. This differs from FFNNs, where each neuron has its own set of weights and biases (Isikdogan et al., 2017). In the convolutional layer each filter is convolved across the width and height of the input volume following Equation 51 (Bento et al., 2021).

$$y_k^{(m)} = f_j\left(\sum_{c=1}^{C} W_k^{(c,m)}\, x_{k-1}^{(c)} + b_k^{(c)}\right) \qquad [51]$$

Where $y_k^{(m)}$ is the output of the convolutional layer k; $x_{k-1}^{(c)}$, the input from layer $k-1$; $W_k^{(c,m)}$,the m^{th} filter of channel c; and $b_k^{(c)}$, the bias term.

In the convolutional layer, the weighted summation of the pixels is done following the filter – convolutional kernel – and producing a new output (Figure 2-7) (Yao et al., 2019). Once the filter – convolutional kernel – is applied to the image, the output of the convolutional layer is passed to the pooling layer.

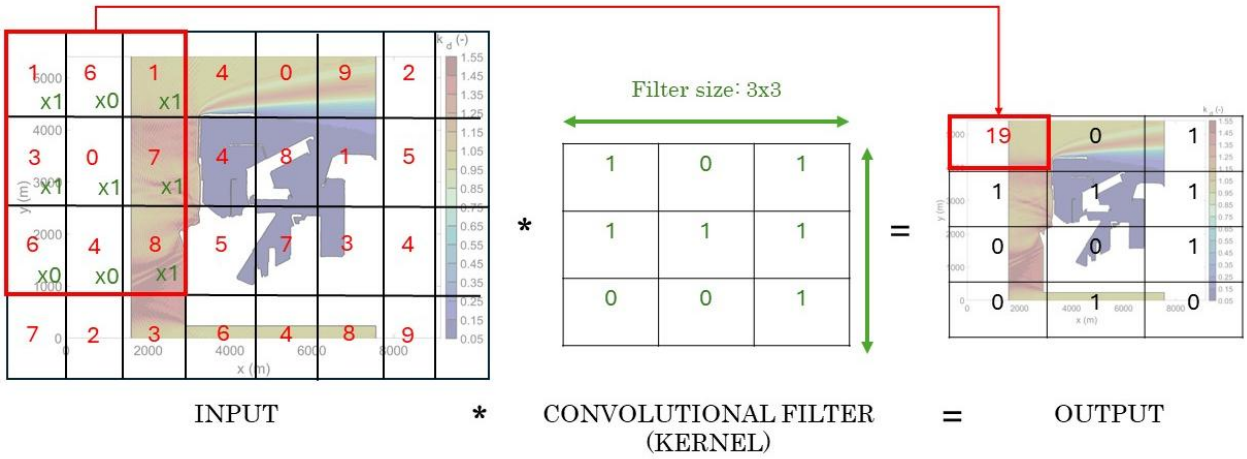


Figure 2-7. Scheme of a convolutional layer.

On the other hand, the pooling layer, as said before, downsamples the output produced by the convolutional layers to avoid prohibitive models in terms of computational costs. Pooling layers divide the output of the convolutional layer into non overlapping regions that depends on the filter size and take commonly the average value (average pooling layers) or the maximum value (max pooling layers) (Figure 2-8).

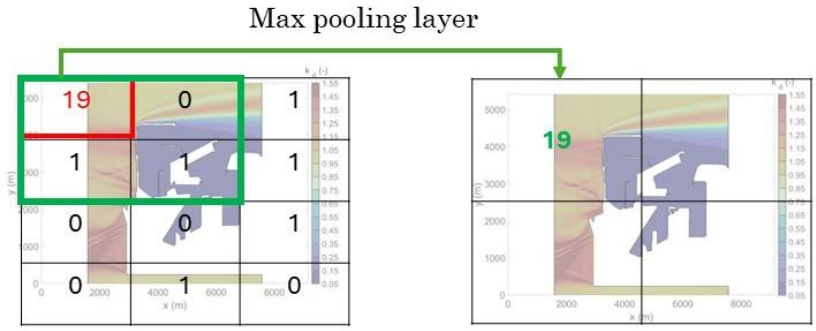


Figure 2-8. Scheme of a max pooling layer.

To highlight the need to use specific types of neural networks when dealing with images due to the significant computational cost they entail, it is important to note that computers treat images as pixels, and within each pixel, there are three colour channels (green, red, and blue). Thus, each pixel implies three digital components, a number from 0 to 255 representing the red, green, and blue colours (RGB code), whose combination gives the pixel value. Hence, handling a single image of 1000x1000 pixels would be the same as working with 3 million data.

In this final phase of the study, instead of following the traditional approach commonly employed in maritime engineering, where the most common type of neural network is the FFNN with a single hidden layer, a more complex architecture is suggested, in line with the difficulty of the process to be modelled. Specifically, a deep FFNN composed of two hidden layers. These layers should then be concatenated with the CNN that processes the images of the ports. This approach allows to harness the benefits of both deep FFNNs and CNNs, creating a hybrid model capable of capturing the intricacies of the process to be modelled while mitigating the risk of overfitting.

TensorFlow (Abadi et al., 2015) and Keras DL libraries (Francois, 2015) in Python 3.12 environment (Booij et al., 1999), are recommended, as they offer great flexibility even though it is less user-friendly. In this study, a CORE i7 8th Gen Intel® CPU with UHD Graphics 620 and 16 GB of RAM was used to process all the models. This information will be relevant when addressing computational costs in Section 4. This software and environment enable to adjust all the ANN hyperparameters much more easily. In previous phases, the default hyperparameters provided by the Matlab Deep Learning Toolbox were applied. The hyperparameters to optimize in this section are highly complex, requiring prior experience with ANNs. Now that two ANN models were developed and it is

the concluding phase of the thesis, these parameters can be optimized, understanding well what they mean.

As pointed out in all the Methodology subsections, the first step in AI models development is always to preprocess the data that it is going to feed the model. In the previous subsections, it was explained how the data feeding the FFNN branch should be preprocessed, but it also necessary to preprocess the data feeding the CNN, the images. For CNNs all input images (port situation and geometric arrangement) as well as output images (port agitation) must have the same dimensions. Therefore, the first step is to standardize all images to the same dimensions. Once this is done, all the images must be normalized dividing them by 255 to reduce computational costs. It is also recommended normalizing wave data.

With all the datasets preprocessed, the model can be developed. To evaluate the performance of the model it is proposed to use the MSE as the loss function (Equation 9). This loss function was preferred over other commonly used metrics in image evaluation, such as the structural similarity index, because it offers simplicity and a stronger focus on individual pixels. This choice aligns closely with the primary objective of this study, which is to identify areas within a port that are susceptible to agitation, prioritizing detailed pixel-level analysis against general trend analysis.

The hyperparameters that are optimized and explained in this phase are the number of epochs, the batch size, the learning rate, the architecture of the FFNN, the architecture of the convolutional layers (height, width and height), the size of the kernels, the stride, the architecture of the pooling layers, the optimizer and the activation functions.

The number of epochs determines how many times the neural network passes through the training dataset to discern patterns. Conversely, the batch size represents the quantity of data samples utilized by the ANN during each iteration of the epoch. For instance, if the number of data samples in the training set is 100 and the batch size is set to 10, each epoch will consist of 10 iterations (100 data samples / 10 batch size = 10 iterations) and this process will be repeated the number of epochs established. A low number of epochs may result in underfitting, where the neural network fails to capture the underlying patterns of the problem. On the other hand, a great number of epochs incurs higher computational costs and may lead to overfitting, where the model merely memorizes patterns without

true understanding. Similar considerations apply to batch size: larger sizes accelerate training but demand greater memory capacity, while smaller batch sizes often yield better generalization but can introduce noise due to increased iteration frequency within each epoch. Consequently, both parameters must be fine-tuned for each model. For their optimization, it is proposed to evaluate epoch values ranging from 10 to 500 and batch sizes varying from 1 to encompassing the entire training dataset.

The learning rate α determines the size of each step in the ANN training. Large learning rates adjust the ANN more aggressively and, consequently, are more computational efficient but can lead to instabilities or divergence of the model. On the other hand, low learning rates are usually more stable and improve generalization capabilities of the model but are also slower and require more computational resources. It is recommended to test learning rates from 0.0001 to 0.01.

Once these hyperparameters are optimized, the architecture of the FFNN and CNN branch must also be tested.

A FFNN architecture aligned with a CNN is proposed. As CNNs dimensions typically adhere to a power-of-2 pattern that increases with layer depth, it is suggested employing a similar architecture pattern for the FFNN. As pointed out, the authors contend that a deep learning approach is better suited to effectively capture harbour agitation, aligning it with the capabilities of CNNs. Therefore, it is recommended a two-hidden-layer ANN and optimizing the number of hidden neurons following a power of 2 pattern, such as [2-4], [4-8], [8-16], and so forth.

With the FFNN branch optimized, the CNN architecture must also be optimized. First, the number of convolutional and pooling layers must be chosen. 3 convolutional layers followed by 2 pooling layers are enough to model port agitation. Architectures more complex than this one usually do not increase significantly the performance of the neural network but do increase the computational costs. With the general structure of the CNN defined, the height, width and depth of the convolutional layers are tested following the power of 2 pattern mentioned before – [2-4-8], [4-8-16], [8-16-32], etc.

Apart from the height, width and depth of the convolutional layers, they use filters of different sizes that determine the area over which the kernel operates. Typical filter sizes go from 2x2 to 8x8. Larger filter sizes can potentially capture more information, but they also escalate computational time, demand more memory,

and may yield less detailed information. Consequently, small filter sizes – 3x3 are suggested. The study focuses on discerning harbour agitation from images, prioritizing detailed image information over global patterns, hence the choice. This choice aligns with the goal of detecting areas of heightened agitation and distinguishing them from areas with less activity. However, for other types of study larger filter size may be more convenient.

Finally, there is another parameter involved in this type of ANNs that is the stride. The stride refers to the number of pixels by which the filter moves across the input image. It is recommended configurating this value to 1, which is the standard setting in CNNs. A stride of 1 implies that the filter is applied to each pixel individually, without skipping any pixel. This approach enables the network to capture patterns more accurately.

With the architecture of the convolutional layers now fully defined, the selection between average and max pooling layers must be done. Max pooling layers, as previously explained, extract the maximum value from a specific region, making them ideal for preserving the most important features of an image while disregarding less relevant ones. Conversely, average pooling layers compute the average value within a specific region, making them suitable for capturing general characteristics and trends of the images. Given that our study aims to discern harbour agitation from images, prioritizing detailed image information over global patterns, the optimal choice is max pooling layers.

With max pooling layers selected as the preferred option, their size must also be chosen. Typically, pooling layers, like filters or kernels, are square-shaped, and their size dictates the region of the image that will be reduced to a single maximum value. Essentially, the larger the size of the pooling layer, the less detailed information it retains, but also, the faster it is. Here, the same approach as with the kernel size is followed, using a detailed pooling layer with a size of 2x2.

From all the optimizers and activation functions Adam solver and RelU functions are suggested.

Figure 2-9. sums up all the methodology explained in this subsection.

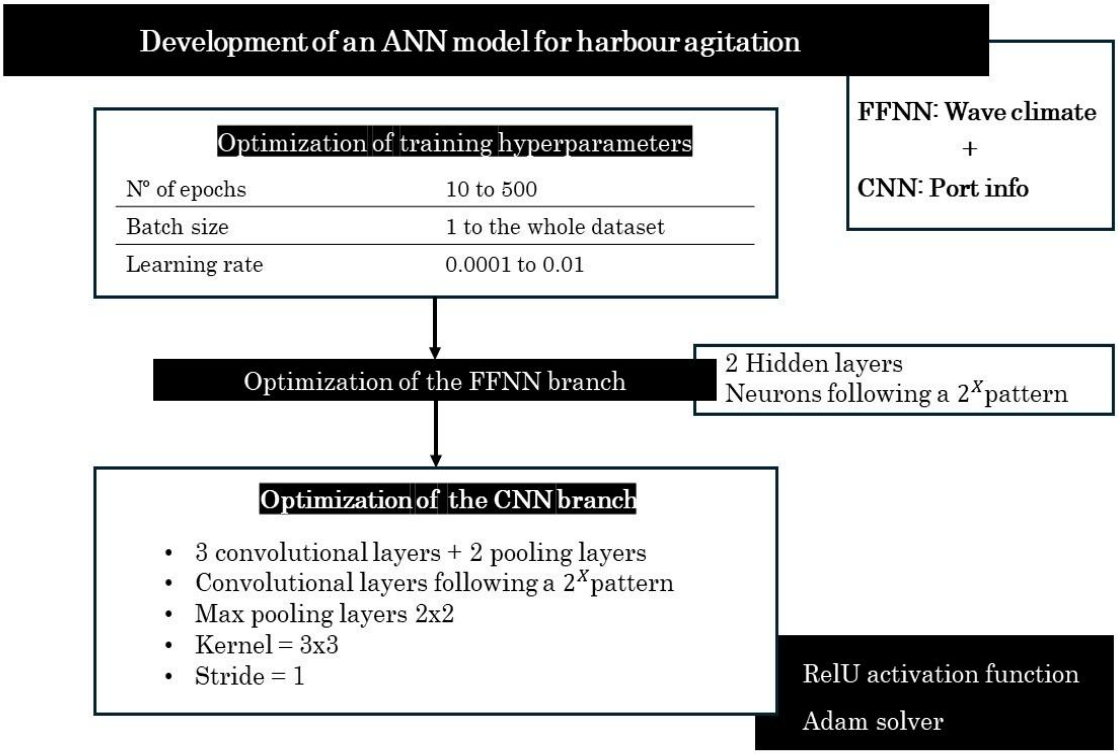


Figure 2-9. Guidelines to develop an ANN model for harbour agitation.

Figure 2-10 sums up the global methodology proposed in this thesis.

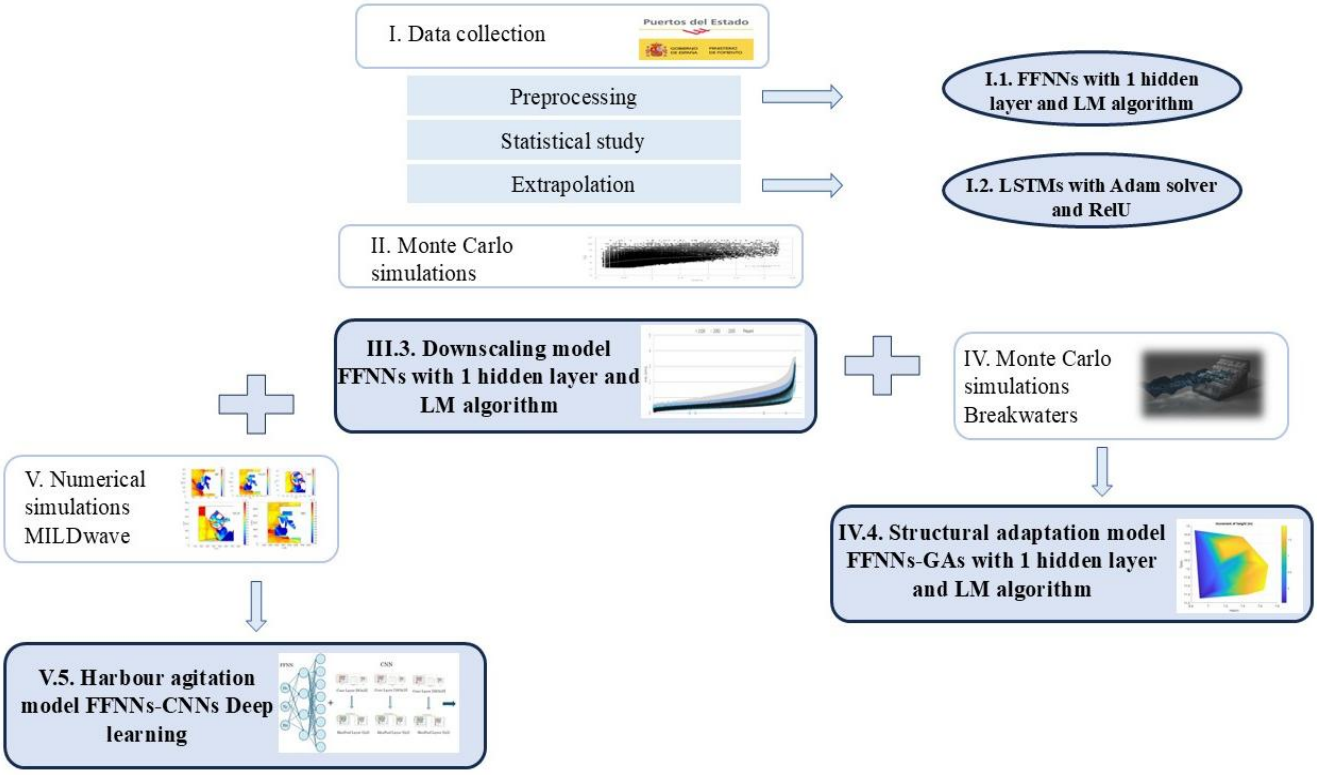


Figure 2-10. AI implementation in ports adaptation studies.

2.5. Application to the Western Mediterranean and to the Port of Valencia

Once the general methodology proposed for developing AI models in the study of port adaptation to climate change has been outlined, this section will detail the specific models developed for the Mediterranean Sea and the Port of Valencia. In Section 4, a comparison between the proposed AI models and traditional approaches will be conducted, and a set of guidelines along with a practical guide for implementing AI methodologies will be presented. While these guidelines are applicable to any technical field, they will focus on port and coastal engineering.

Although the thesis centres on the coastal and port sectors, the principles and processes described for developing AI models can be applied to other disciplines. One of the main objectives of this work is to facilitate the introduction to artificial intelligence for any professional, regardless of their area of expertise.

2.5.1. Wave downscaling

This initial study phase was divided into two parts. First, a wave propagation model was developed without considering climate change (Subsection 2.5.1.1. corresponding to paper 3.8.) (Portillo Juan, Olalde Rodríguez, et al., 2023)

After verifying the validity and application of ANNs, a second propagation model was developed, specifically addressing climate change (Subsection 2.5.1.2., corresponding to paper 3.9.) (Portillo Juan & Negro Valdecantos, 2024).

2.5.1.1. Wave propagation model

In this first phase of the study, an ANNs model in the Spanish Mediterranean Sea was developed and validated. To conduct the study the deep-water and coastal buoys of Valencia and Cabo de Palos were used (Figure 2-11). Historical hourly wave height (Hs), peak period (Tp) and Direction (θp or Dir) records were downloaded from Spanish authority web page (Puertos del Estado. Oceanografía. Registros históricos de oleaje).

The ANN is fed by deep water historical records as inputs and the coastal buoys data as outputs, in such a way that it will learn to predict Hs and Tp in coastal depth knowing the Hs, Tp and Dir of the waves in deep waters. Table 2-3 shows the characteristics of the historical records used.

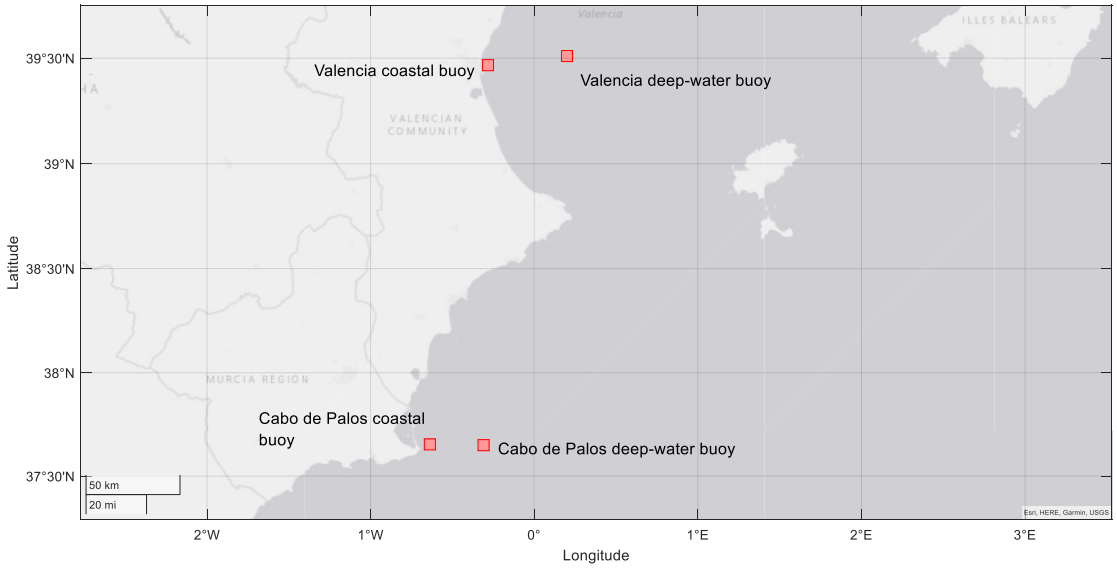


Figure 2-11. Location of buoys.

Table 2-3. Historical records used for the development of the ANN model.

Buoy	Depth	Hs	Tp	Dir	Series availability
Valencia	260 m	X	X	X	15/09/2005 to 14/07/2022
Coastal – Valencia	20 m	X	X	X	08/06/2005 to 30/10/2013
Cabo de Palos	230 m	X	X	X	18/07/2006 to 14/07/2022
Coastal – Cabo de Palos	67 m	X	X	-	14/11/1985 to 18/06/2012

As stated in Section 2, Methodology, all data were preprocessed prior to the development of the ANN model. A code was developed using interpolation techniques in the Python environment to eliminate time gaps in the buoy records that could affect the ANN's performance. Subsequently, the data were filtered to remove false values and errors; specifically, wave height values below 0.1 m and periods under 1 s were excluded. Once preprocessing was completed, the optimization process of the ANN was initiated.

An ANN was developed for each location—one for Valencia and another for Cabo de Palos. The data were divided into 70% for training, 15% for validation, and 15% for testing, using contiguous blocks. The initial records were used for training the ANN, while the remaining 30% were allocated for validation and testing (Figure 2-12).

First, the architecture was optimized (Figure 2-13 and Figure 2-14).

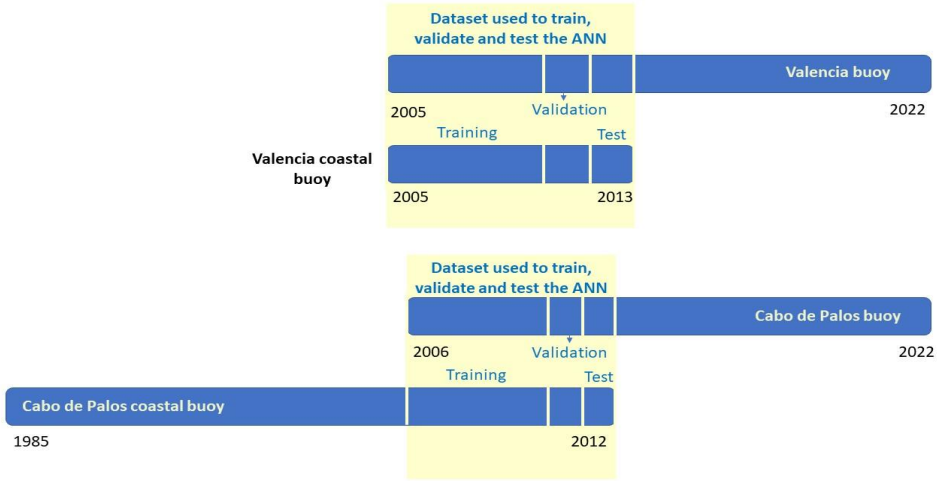


Figure 2-12. Data partition of the ANN model.

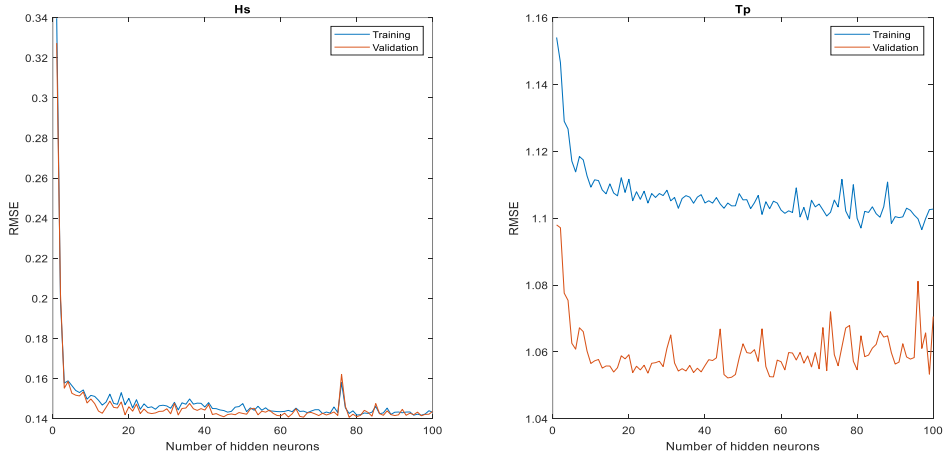


Figure 2-13. Optimization of the ANN model architecture in Valencia.

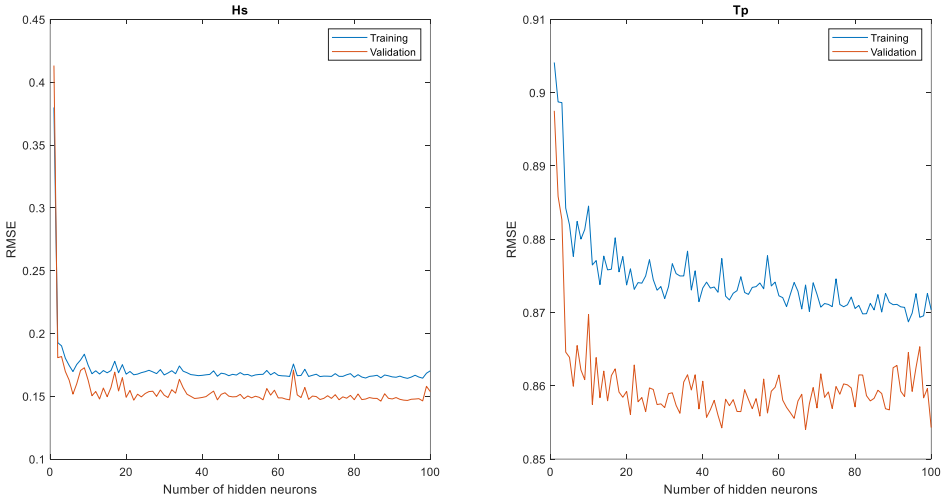


Figure 2-14. Optimization of the ANN architecture in Cabo de Palos.

With the architecture optimized, different algorithms were tested by evaluating the training, validation, and testing errors (Figures 2-15 to 2-20).

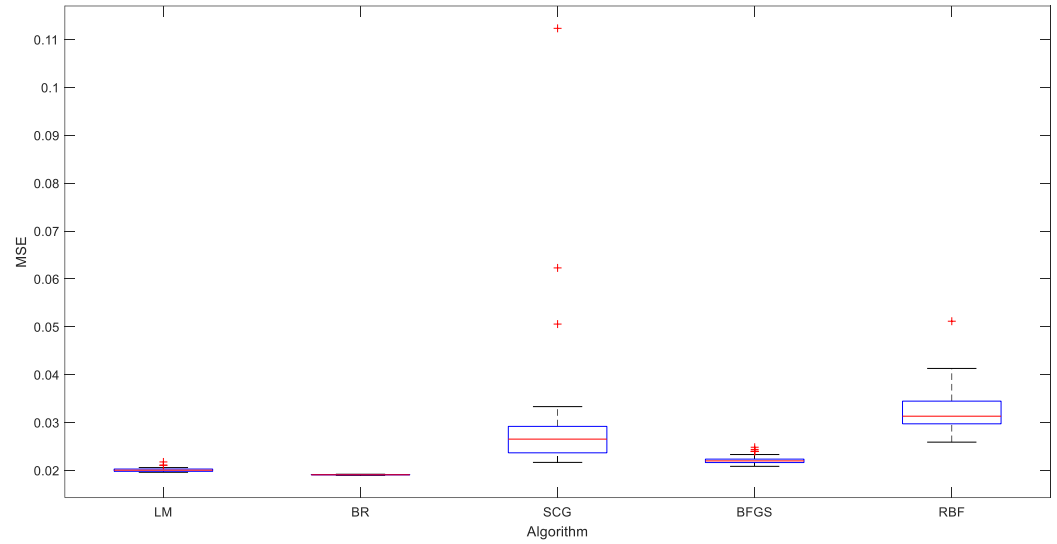


Figure 2-15. Training error boxplot for each algorithm (Valencia case study).

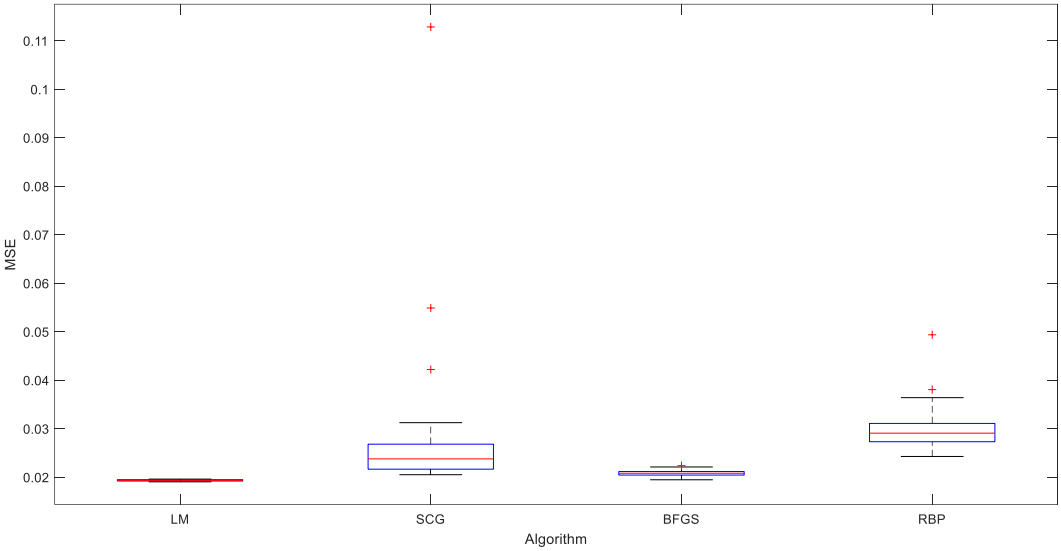


Figure 2-16. Validation error for each algorithm (Valencia case study).

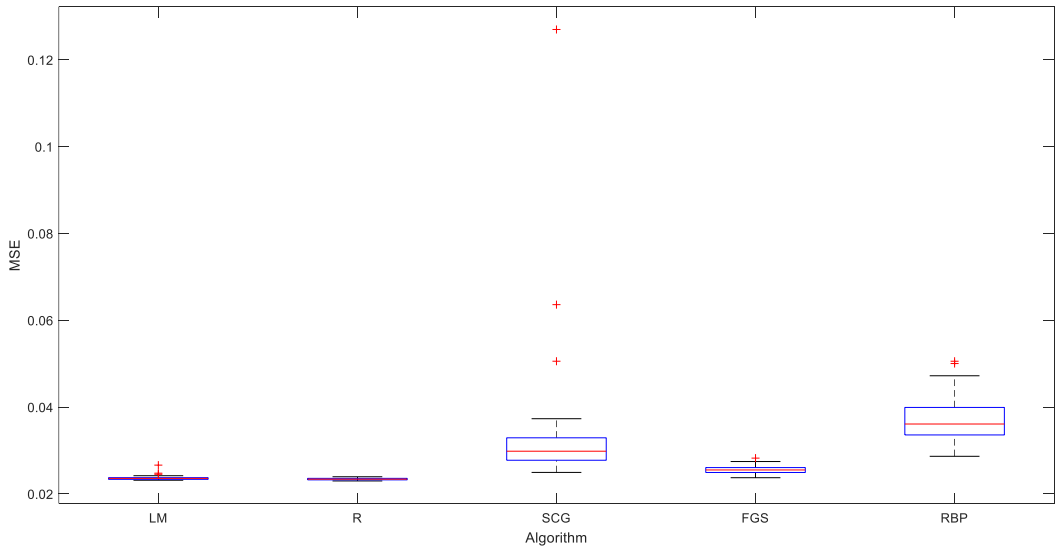


Figure 2-17. Test error for each algorithm (Valencia case study).

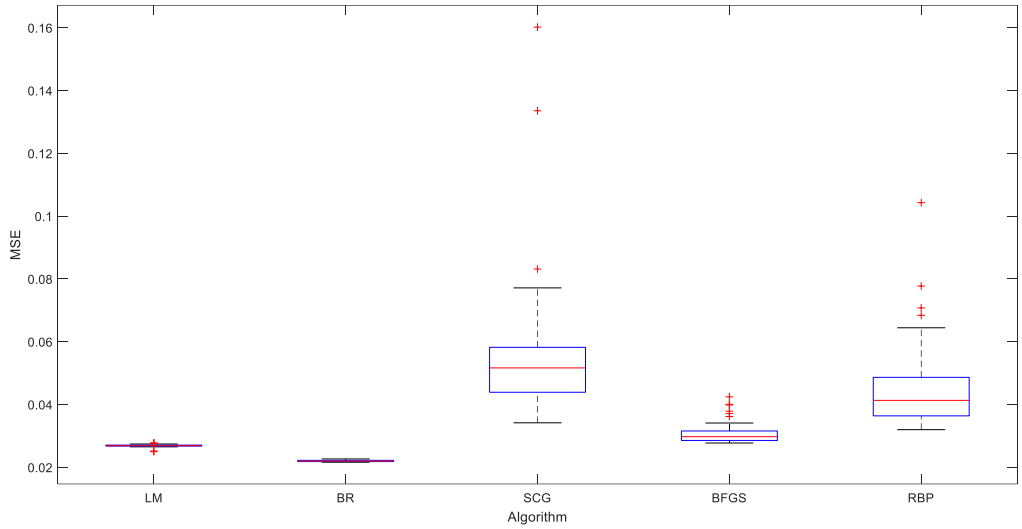


Figure 2-18. Training error boxplot for each algorithm (Cabo de Palos case study).

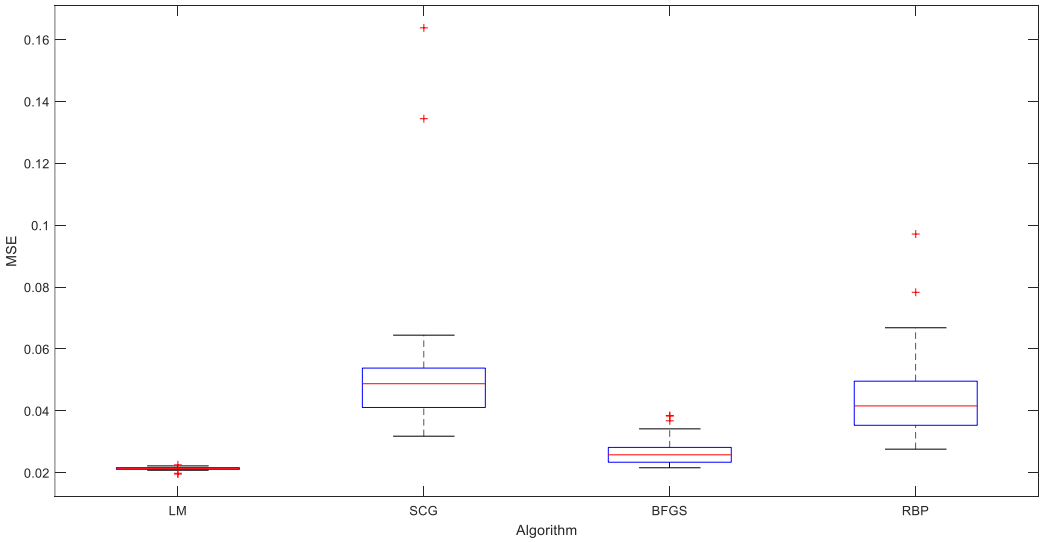


Figure 2-19. Validation error for each algorithm (Cabo de Palos case study).

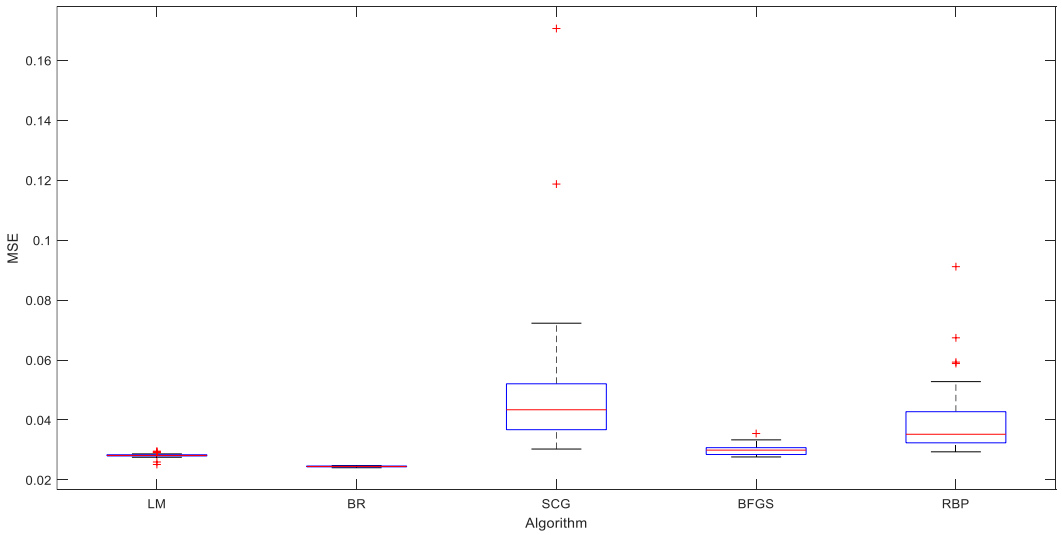


Figure 2-20. Test error for each algorithm (Cabo de Palos case study).

Finally, it resulted that the optimum models were an ANN with LM algorithm and a [3-33-2] architecture for Valencia buoy and an ANN with LM algorithm and a [3-22-2] architecture for Cabo de Palos (Figure 2-21).

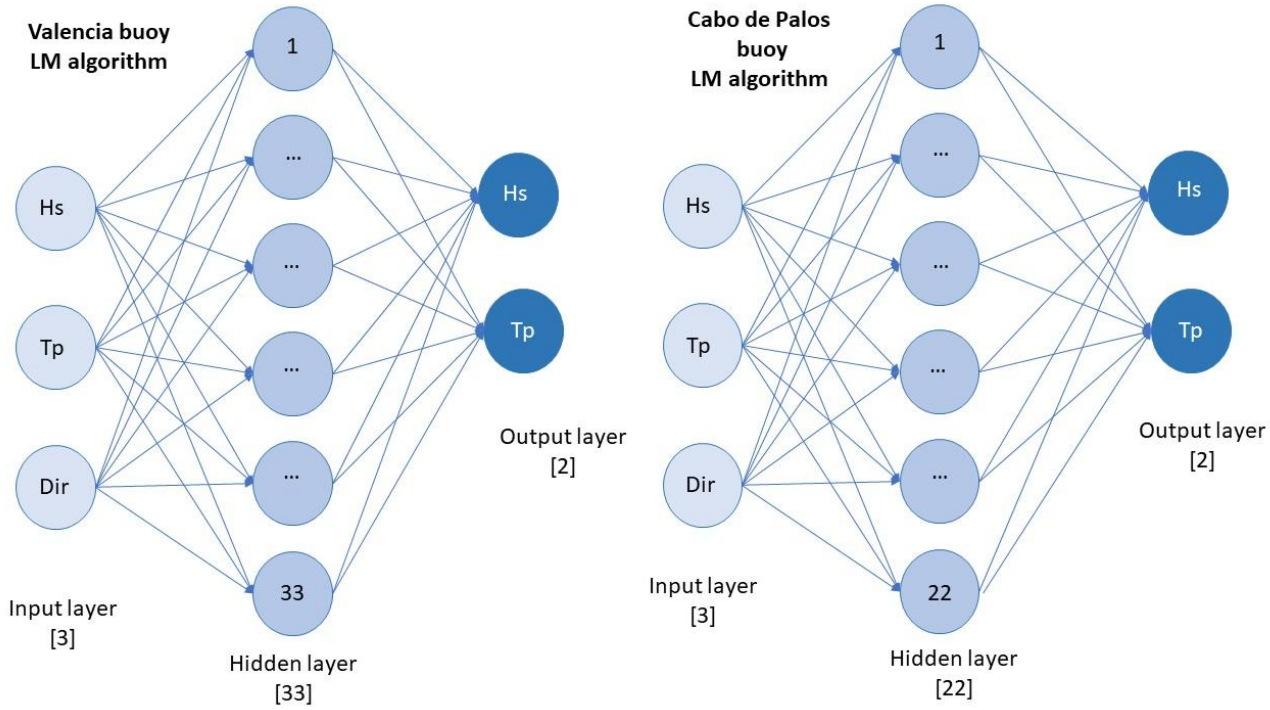


Figure 2-21. ANN models for wave downscaling.

Once the neural network models were designed and optimized, the errors of the model were evaluated (Figure 2-22). After analysing the errors, it was confirmed that the model's errors were on par with those of current numerical models. In some cases, the AI model even produced more accurate results than the numerical models (Section 4). Consequently, the model was deemed valid for subsequent application in predicting future wave behaviour (Section 2.5.1.2).

Figure 4-12 shows the model errors for selected network architectures. The maximum error (upper whisker), minimum error (lower whisker), and average error (central line of the boxplot) all fall within highly acceptable ranges.

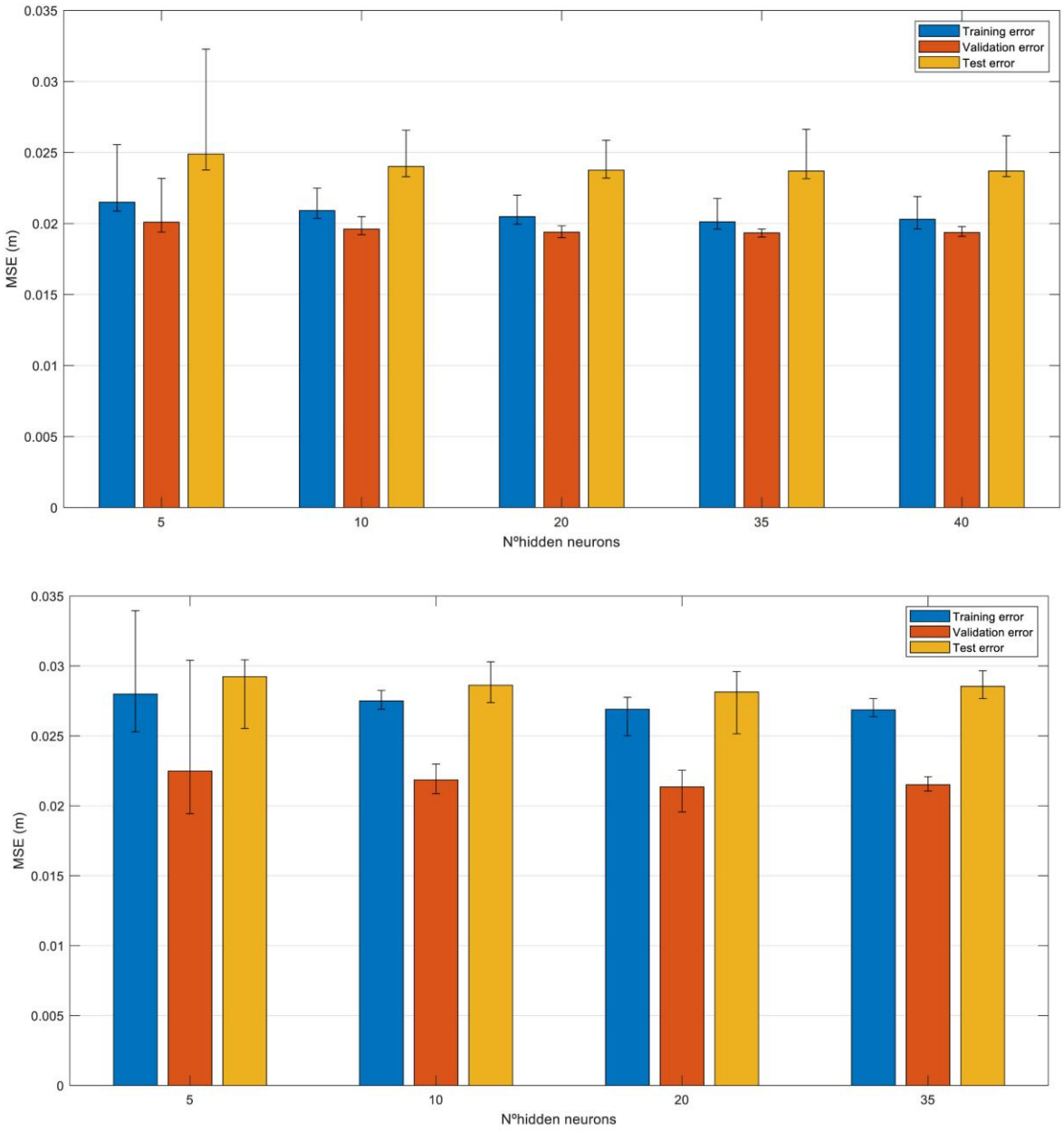


Figure 2-22. ANN errors for selected arch. in Valencia (up) and Cabo de Palos (down).

2.5.1.2. Wave propagation model applied to climate change

Once the application of ANNs for wave downscaling models was proved (Portillo Juan, Olalde Rodríguez, et al., 2023) (Paper 3.8.), ANNs were applied for climate change forecasting (Portillo Juan & Negro Valdecantos, 2024) (Paper 3.9.). The same data as in section 2.5.1.1. was used to conduct the statistical analysis.

First, the data records were analysed, and future values of the wave distributions were extrapolated for both long-term and extreme regimes. Secondly, the ANN

model was constructed and optimized. Finally, the optimized ANN model was applied to estimate the future wave climate in the Spanish Mediterranean Sea.

Long-term wave regimes in the Spanish Mediterranean Sea are typically described using Weibull distribution (Portillo Juan & Negro Valdecantos, 2024). Weibull distribution is characterized by 3 parameters (Equation 51): the scale parameter (A), the centring parameter(B) and the shape parameter (C) .

$$F_e(x) = 1 - exp(-((x - B)/A)^C) \quad [51]$$

Where the input x is the wave height and the output $F_e(x)$ is the probability of no exceeding this value.

The values of A, B and C for the historical records were given by Puertos del Estado official reports (*Informes anuales Oceanografía Puertos del Estado*) and to obtain the values of A, B and C, from 2024 to 2050, polynomial regression was applied (Figure 2-23). The goodness of fit evaluation is presented in Table 2-4.

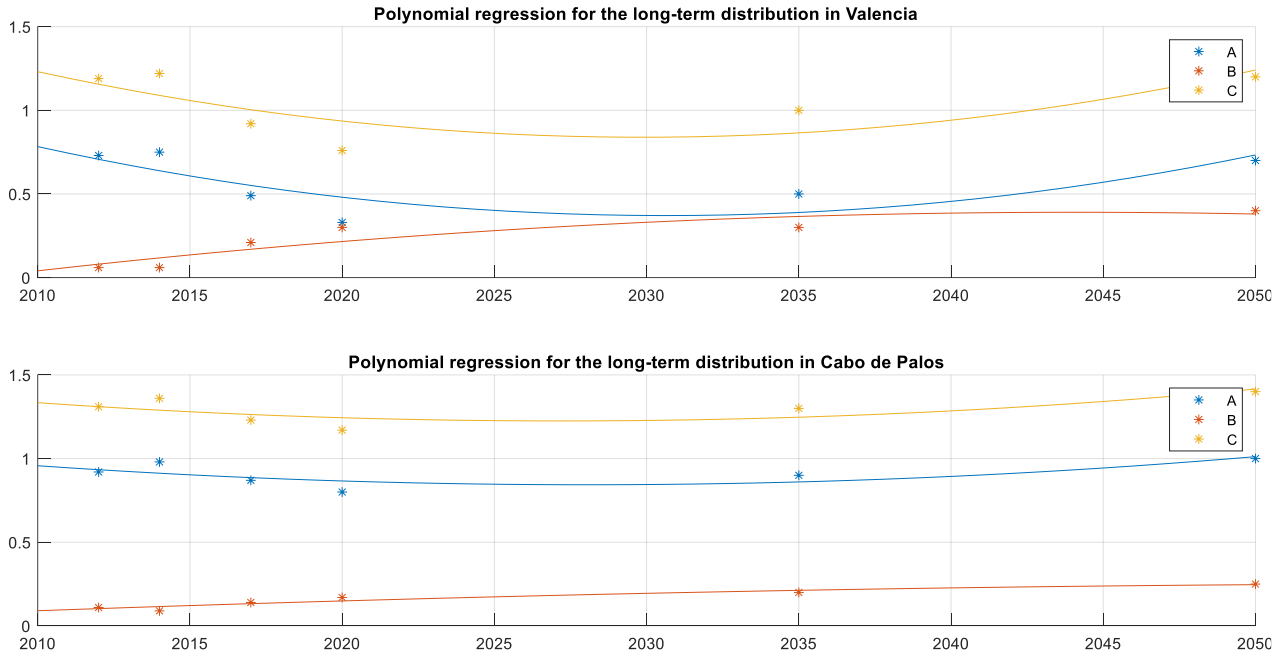


Figure 2-23. Statistical adjustment of parameters A, B and C for long-term distribution.

Table 2-4. Goodness of fit evaluation for the long-term distribution.

Parameters	R in Valencia	R in Cabo de Palos
A	0.8	0.77
B	0.91	0.96
C	0.75	0.77

For the extreme distribution a Weibull distribution was also chosen to model sea states (Equation 52). Extreme Weibull distribution is characterized by 3 parameters: the threshold (α), the scale parameter (β), and the shape parameter(γ).

$$F_e(x) = 1 - exp(-((H_\alpha - \alpha)/\beta)^\gamma) \quad [52]$$

To study and make projections for the extreme long-term distribution, the 95th percentile of the buoy records were taken as reference, that corresponds to a return period of 20 years according to the DNV standards (Hopstad et al., 2018) and Puertos del Estado reports (*Informes anuales Oceanografía Puertos del Estado*). This return period is of particular importance because is the return period that the consortium establishes to define extreme events and above which it covers the costs of the damage caused in maritime structures. To project values for future years, polynomial regression was also employed. The results for the extreme distribution are presented in Figure 2-24 and Table 2-5.

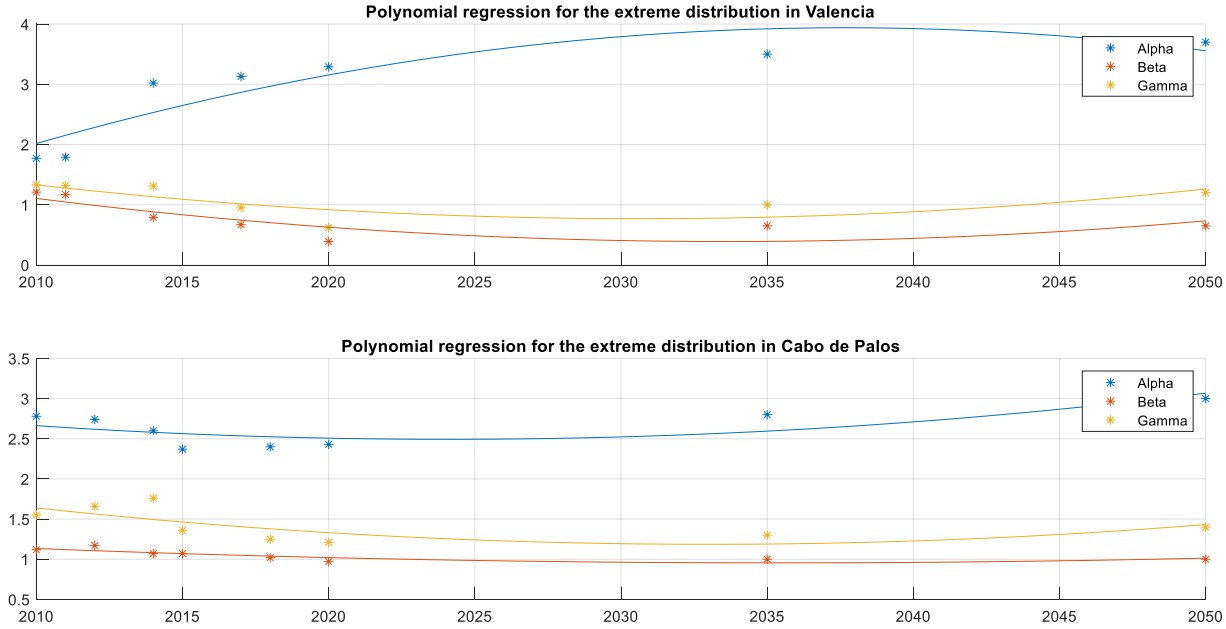


Figure 2-24. Statistical adjustment of parameters α, ß and γ for extreme distribution.

Table 2-5. Goodness of fit evaluation for the extreme distribution.

Parameters	R in Valencia	R in Cabo de Palos
$\boldsymbol{\alpha}$	0.90	0.79
$\boldsymbol{\beta}$	0.83	0.84
$\boldsymbol{\gamma}$	0.77	0.70

After analysing future wave climate and estimating the three parameters that characterize long-term and extreme wave distributions, Monte Carlo simulations were employed to generate the synthetic data series representing the future wave climate.

As introduced in the methodology, various techniques are available to generate synthetic data series, with some of the most widely used in ocean and coastal engineering including Bootstrapping sampling, Generative Adversarial Networks (GANs), Markov Chain Methods, Variational Auto-Encoders (VAEs), and Monte Carlo simulations. For this study, however, Monte Carlo simulations were chosen to maintain simplicity and make the process more user-friendly.

As many simulations as those corresponding to the time span between the last year of available records (2022) and the year 2050 were run, resulting in 227,760 simulations that correspond to the hourly records of the virtually generated maritime climate database. This criterion was adopted to ensure consistency with the historical records, providing both a real database collected by buoys for the present maritime climate and a dataset generated through Monte Carlo simulations for the future maritime climate. Both datasets were subsequently used to feed the AI model.

The simulations were conducted a total of 20 times to validate and assess result accuracy. To confirm the reliability of the simulations, a sensitivity analysis was executed. The findings revealed minimal variability in the key parameters defining the wave climate (Hs, Tp, and Dir). As a result, Monte Carlo simulations were deemed sufficiently valid and accurate for further analysis (Table 2-6). SLR was included following Portillo Juan et al. (2022c) (Paper 3.3.).

Table 2-6. Variability of Hs, Tp and Dir trough Monte Carlo simulations.

Variability	Hs (m)	Tp (s)	Dir (º)
Mean	± 0.005 m	± 0.005 s	± 0.1º
Maximum	± 0.6 m	± 0.2 s	± 15º s
Minimum	± 0.0002 m	± 0.001 s	± 0.03º s
Standard deviation	± 0.002 m	± 0.002 s	± 0.08º s
Median	± 0.002 m	± 0.004 s	± 0.08º s

The ANN model was built following the same procedure as in Section 2.5.1.1. Since the procedure is the same, it will not be repeated it here, this section will be limited to present the results of the architecture and algorithm optimization (Table 2-7).

Table 2-7. ANN wave propagation models applied to climate change.

Location	Architecture	Algorithm	Transfer function
Valencia	4-33-2	Levenberg-Marquardt (LM)	Hyperbolic tangent sigmoid. Linear
Cabo de Palos	4-22-2	Levenberg-Marquardt (LM)	Hyperbolic tangent sigmoid. Linear

The overall performance of the model is summed up in Table 2-8 and Figure 2-25.

Table 2-8. Training and validation error of the model.

Variable	Error	Valencia	Cabo de Palos
Hs	RMSE Training	0.144	0.167
	RMSE Validation	0.142	0.147
Tp	RMSE Training	1.10	0.873
	RMSE Validation	1.05	0.863

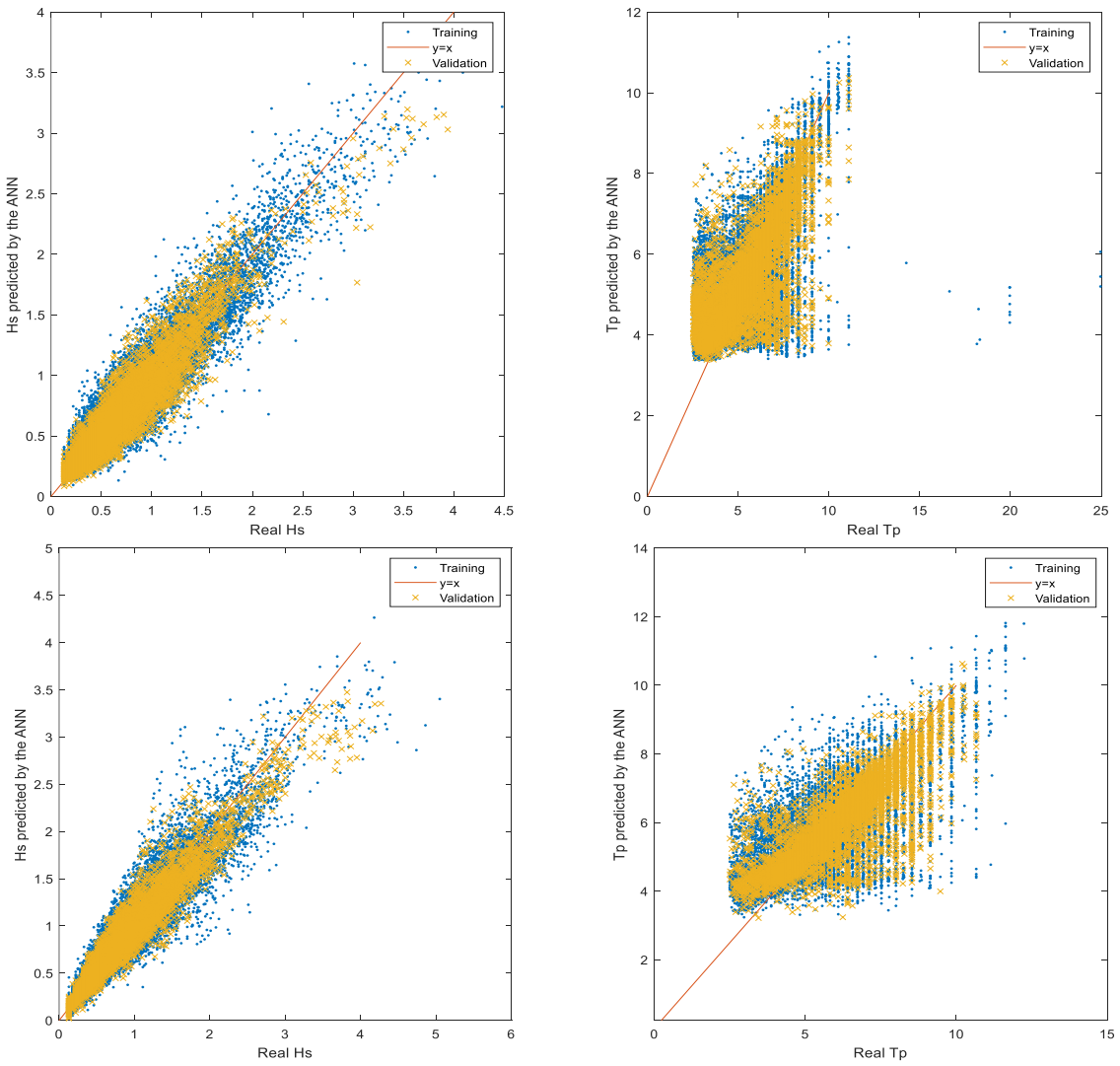


Figure 2-25. Performance of the model in Valencia (up) and Cabo de Palos (down).

The performance of the model is very accurate at both, Hs and Tp, with a maximum MSE of 0.021 m and 1.21 s. The dispersion associated with the variable Tp is higher due to the nature of this variable, but not due to the quality of the model.

2.5.2. Adaptation of coastal protection structures

Once the wave states were propagated to the point of interest, the analysis was focused on the Valencia area, specifically on the two most critical sections of the North Breakwater: the head and the final section of the trunk.

The design conditions of the North breakwater for both sections were Hs=6.84 m and Tp=7.44 s for a return period of 475 years (Albert et al., 2012). A range of wave states corresponding to a return period of 475 years within the timeframes of 2035 and 2050 were systematically analysed and applied to the ANN-GA model to obtain the adaptation measures needed.

The end of the trunk of the North is presented in Figure 2-26.

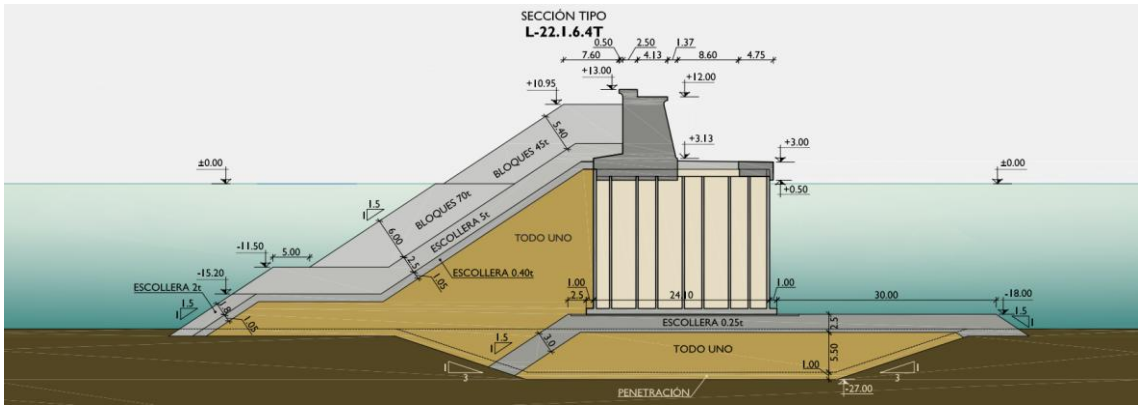


Figure 2-26. Trunk section. Taken from (Albert et al., 2012).

The head of the North breakwater is presented in Figure 2-27.

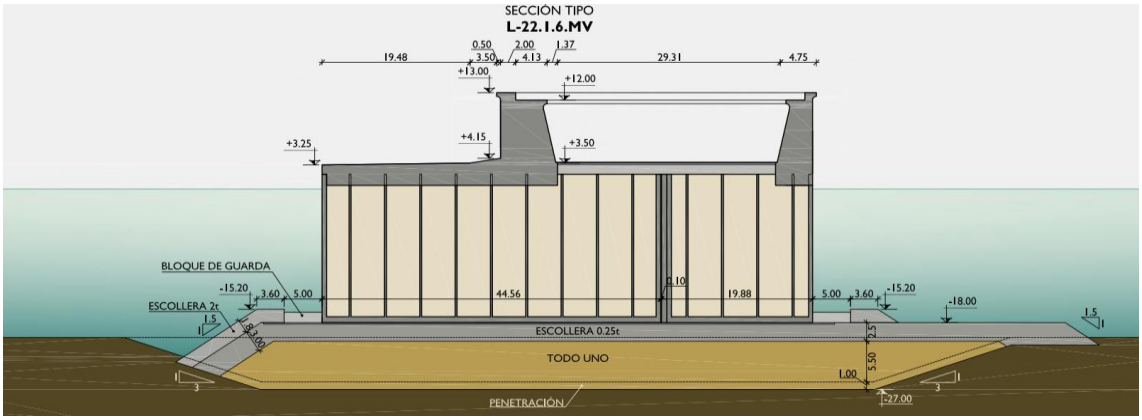


Figure 2-27. Head section of the North breakwater. Taken from (Albert et al., 2012).

For the ANN model design, the same approach was followed as in Section 2.5.1. FFNNs with one hidden layer and hyperbolic tangent sigmoid and linear transfer functions were used. Here, two different ANN models were distinguished, one for the head of the breakwater, which is a compound breakwater and one for the final section of the trunk, which is a vertical breakwater.

The number of hidden neurons were optimized following a similar approach as in the wave downscaling study. However, due to the increased complexity and computational expense of the ANN model, the methodology. proposed was slightly modified. Instead of testing a range from 1 to 100 hidden neurons, a loop termination criterion was implemented. This criterion was set to halt when the error increased for more than three consecutive iterations, surpassing the maximum local error of previously tested networks (Figures 2-28 and 2-29). For the optimization of the architecture, a maximum RMSE of 0.6 was required for the SSC, 0.3 m for the height increase and 0.6 t for the weight increase.

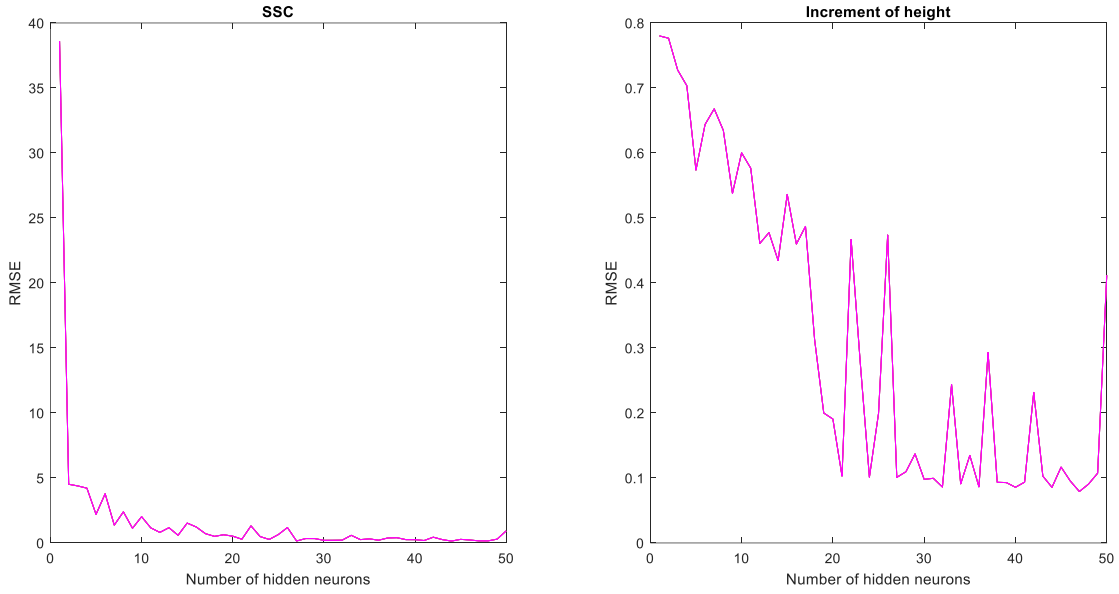


Figure 2-28. Optimization of the vertical breakwater neural network architecture.

It must be noted that the great fluctuations that can be seen on the right graph (increment of height) are mainly due to the scale of the y-axis. This y-axis scale was chosen to better appreciate why the iterations were cut in 50 hidden neurons, but it can give a false image of instability when it is not the case as the maximum difference of errors between peaks is 0.3 m.

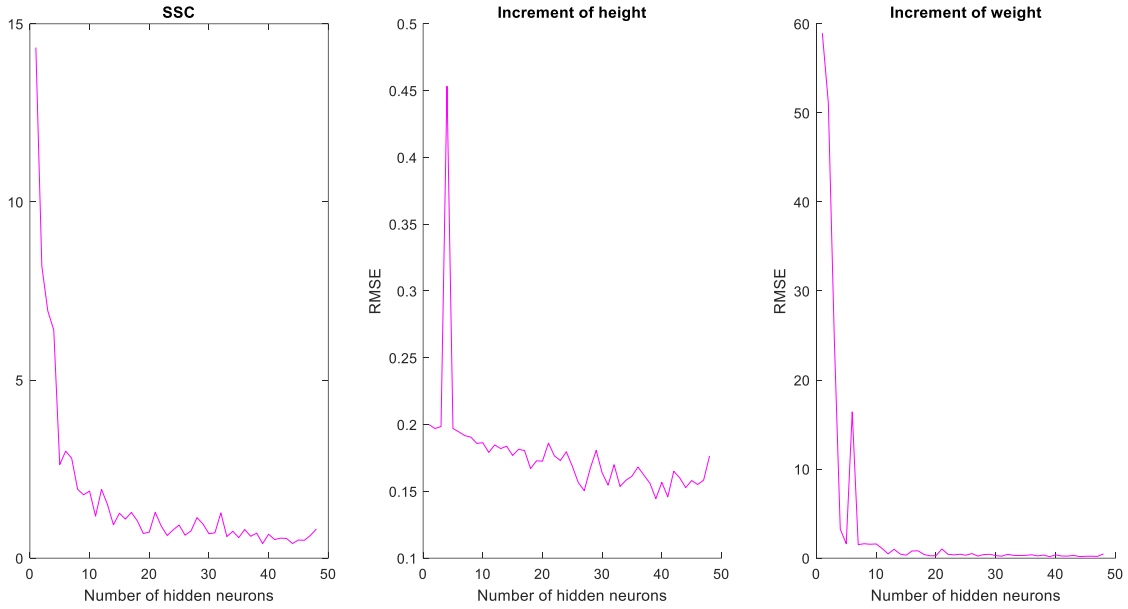


Figure 2-29. Optimization of the compound breakwater neural network architecture.

The number of epochs was set to 100. The data was divided randomly, and the data partitioning was 70-15-15. 70% of the data samples were used to train the network, 15% to validate it and 15% to test it. With the NN optimized the optimization of the algorithm was done as in Section 2.5.1. and the best algorithm was also the LM.

Once the architecture and the algorithm of the model was defined, Genetic Algorithms (GAs) were implemented to optimize the weights and biases of the two ANN models. For the GA, a population size of 200 was chosen, with a maximum number of iterations of 100, a crossover percentage of 80% and a mutation percentage of 20% and the selection function employed was the roulette wheel selection (Equation 53)

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \qquad [53]$$

Where p_i is the probability of choosing a chromosome i for the next generation, f_i is the fitness of i and N is the size of the current generation.

The characteristics of the model are in Tables 2-9 and 2-10.

Table 2-9. Model characteristics.

Model characteristics	Vertical Breakwater	Compound breakwater
	ANN	
Architecture	[12-47-2]	[13-39-3]
Algorithm	Levenberg–Marquardt	Levenberg–Marquardt
Transfer functions	Sigmoidal and linear	Sigmoidal and linear
Nº epochs	100	100
Data partitioning	Random 70-15-15	Random 70-15-15
Loss function	RMSE and MSE	RMSE and MSE
Output	SSC, Increment of height	SSC, Increment of height, Increment of weight
	GA	
Nº of iterations	100	100
Population size	200	200
Crossover percentage	80%	80%
Mutation percentage	20%	20%
Selection function	Roulette wheel	Roulette wheel
Normalization of the data	Yes	Yes

Table 2-10. Errors of the ANN-GA model for structural adaptation.

	Vertical breakwater model		Compound breakwater model		
	SSC	Δ Height	SSC	Δ Height	Δ Weight
r	0.986		0.874		
RMSE training	0.45	0.19 m	0.53	0.15 m	0.52 t
RMSE validation	0.44	0.19 m	0.53	0.15 m	0.55 t

2.5.3. Port agitation

After completing phases I and II of the study, the investigation of port agitation proceeded. In the previous section, the port's behaviour under ULS was analysed, along with the potential impact of the new climate conditions on its structural

performance. The focus now shifts to studying the port under SLS and analysing how the same new climate conditions will affect the port's operability and serviceability.

The aim of this final phase of study is to develop a surrogate AI model. For that purpose, the first step is to run the numerical simulations.

MILDwave, which is the numerical model used in this study, operates by introducing a unitary wave ($Hs = 1m$) to determine the amplification factor, Kd, within the port. Therefore, the testing cases focus on varying wave periods and directions. The Tp cases were selected based on the data collected from *Puertos del Estado. Oceanografía. Registros históricos de oleaje* and on the projections made for future time scenario. For the wave direction, the same approach was followed as with the wave period, studying the swell rose (Figure 2-30), and dividing the main directions in sectors of 22.5º.

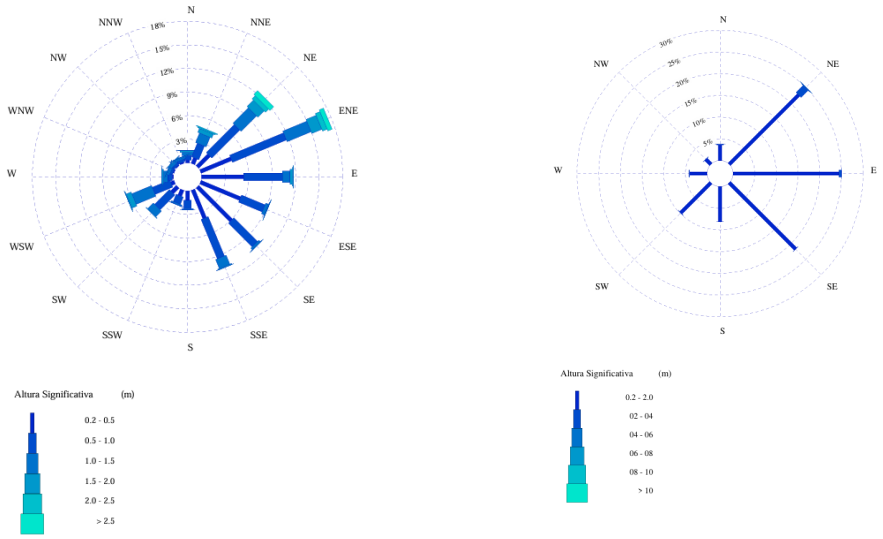


Figure 2-30. Swell rose in Valencia for long-term and extreme distribution.

6 different wave periods and 8 different wave directions according to the data collected were defined, giving a total of 48 different numerical models. For future references these model cases will be denoted as A.B cases, being A, the number of cases corresponding to the wave period and B to the wave direction (Table 2-11).

Table 2-11. Study cases.

Case	Wave period (Tp-A)	Wave direction (Dir-B)
1.	8.9 s	22.5º
2.	7.4 s	45º

3.	6.1 s	67.5º
4.	5.6 s	90º
5.	5.0 s	112.5º
6.	4.6 s	135º
7.	-	157.5º
8.	-	180º

For example, case 3.6. would be the wave case with a wave period of 6.1 s (3.) and a direction of 135º (.6).

Referring to the port layout information, bmp images based on Google Earth information were used (*Google Earth 7.3.6.9796 64-bit*, 2016)

Regarding the model settings, 1000 waves per case were produced, with a precautionary filtering of initial waves to mitigate numerical noise. Sponge layers of 10 wavelengths were applied and 4 different absorption coefficients were used to model the different structures and objects within the port of Valencia. The different absorption coefficients were tuned and validated before running the model to obtain the reflection coefficients that we were looking for, that were 0.3 for armour layer breakwaters – represented by colour blue and green – and 0.9 for vertical walls – represented by colour red (Figure 2-31 and Table 2-12).

Table 2-12. MILDwave coefficients calibration.

Case	**Tp**	**Absorption coefficient**				
		Blue	Green	Red	White	Black
1	4.6	0.97	0.7	0.45	1	0
2	5	0.97	0.7	0.25	1	0
3	5.6	0.96	0.68	0.2	1	0
4	6.1	0.95	0.7	0.15	1	0
5	7.4	0.88	0.2	0.12	1	0
6	8.9	0.88	0.2	0.11	1	0



Figure 2-31. Bitmap file for the port layout (wave direction 22.5º).

After defining all the wave data scenarios and port layout, calibrating absorption coefficients, and updating model settings, all simulations were run. These numerical simulations provide the data to train the AI model.

For the data preprocessing, all images were standardized to the same dimensions, which were 1728x1152 pixels. Subsequently, all the images were normalized dividing them by 255 to reduce computational costs. Wave data was also normalized.

With all datasets pre-processed, the model was designed following a trial-and-error approach, starting from an initial configuration and proceeding with the optimization of various hyperparameters. The process began by defining the number of epochs and the batch size (Figure 2-32).

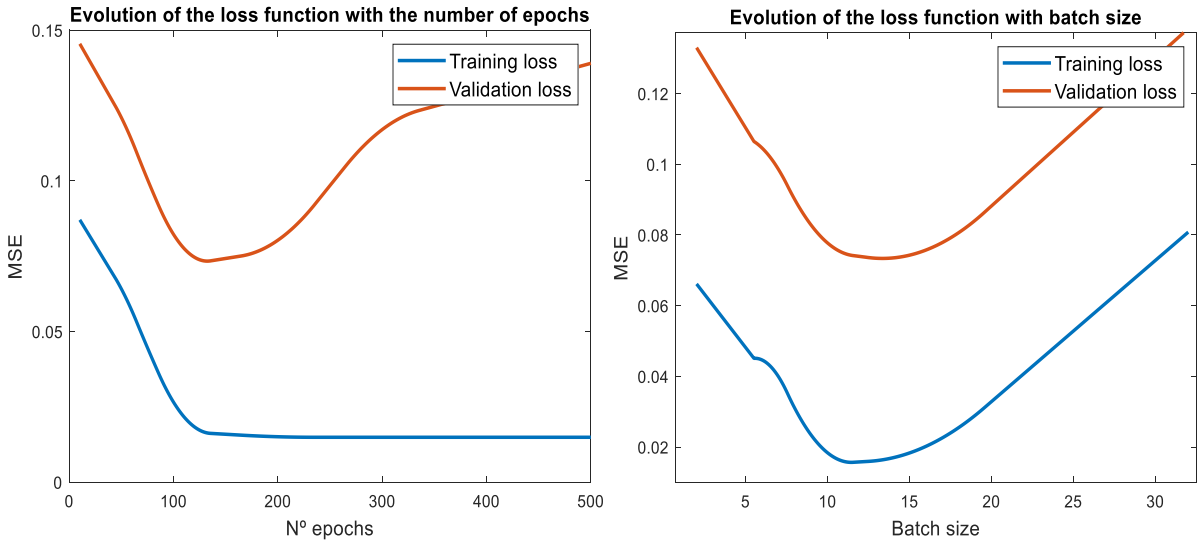


Figure 2-32. Optimization of the number of epochs and batch size.

As it can be observed, the training error steadily decreases with the number of epochs, yet beyond approximately 150 epochs, the reduction in error becomes marginal. Concurrently, the validation error begins to rise from around 100 epochs onwards, signalling overfitting. Consequently, the number of epochs was capped at 100 to mitigate overfitting.

Regarding batch size, both the training and validation errors reached their minimum at 12. Hence, the batch size was fixed at 12.

After defining the number of epochs and batch size the learning rate (α) was optimized, following the methodology explained in Section 2.4. (Figure 2-33).

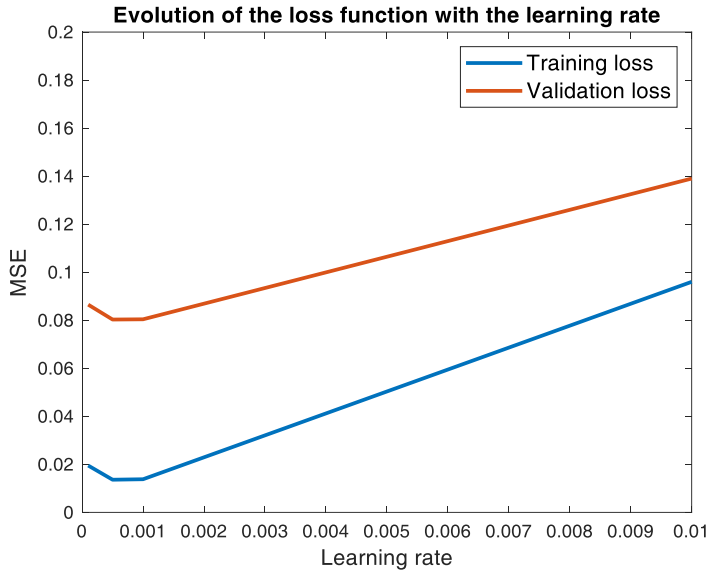


Figure 2-33. Optimization of the learning rate.

Subsequently, the model architecture was optimized. The model consists of two distinct branches: an FFNN to incorporate wave climate data into the neural network, and a CNN to process images. For the FFNN optimization, two hidden layers were selected, and various architectures were tested in accordance with typical CNN architecture schemes, ranging from 4-8 neurons to 32-64 neurons (Figure 2-34).

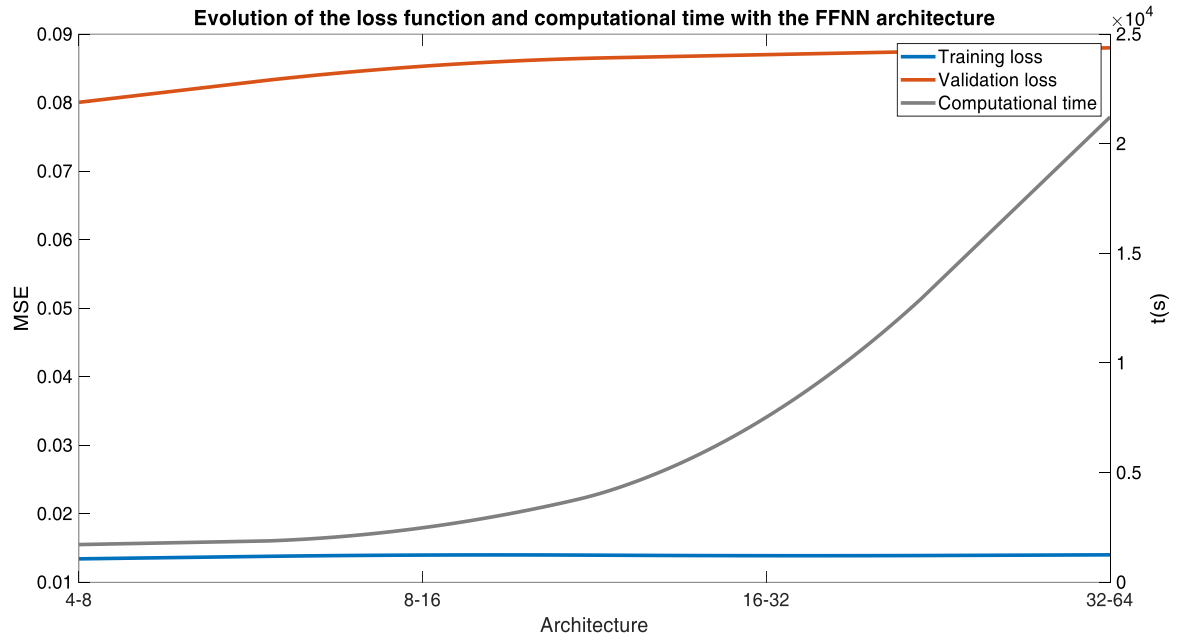


Figure 2-34. Optimization of the FFNN architecture.

As can be seen from the figure, the training error remains relatively consistent across all FFNN architectures, while the validation error slightly rises with an increase in the number of neurons, suggesting potential overfitting. Furthermore, computational time escalates exponentially from the 8-16 architecture onwards. Consequently, it was opted for the simplest architecture for the FFNN, 4-8.

For the CNN architecture optimization, the same approach as with the FFNN was followed, testing different CNN architectures. The number of layers was set to three convolutional layers and two pooling layers and proceeded to fine-tune the architecture. As illustrated in Figure 2-35, significant increase in computational time is evident beyond the 8-16-32 architecture, going from 5,951 s for the 8-16-32 architecture to 86,400 s for 16-32-64. However, the decrease of errors with a more complex CNN architecture is almost undetectable. Therefore, the architecture of the convolutional neural network was determined to be 8-16-32.

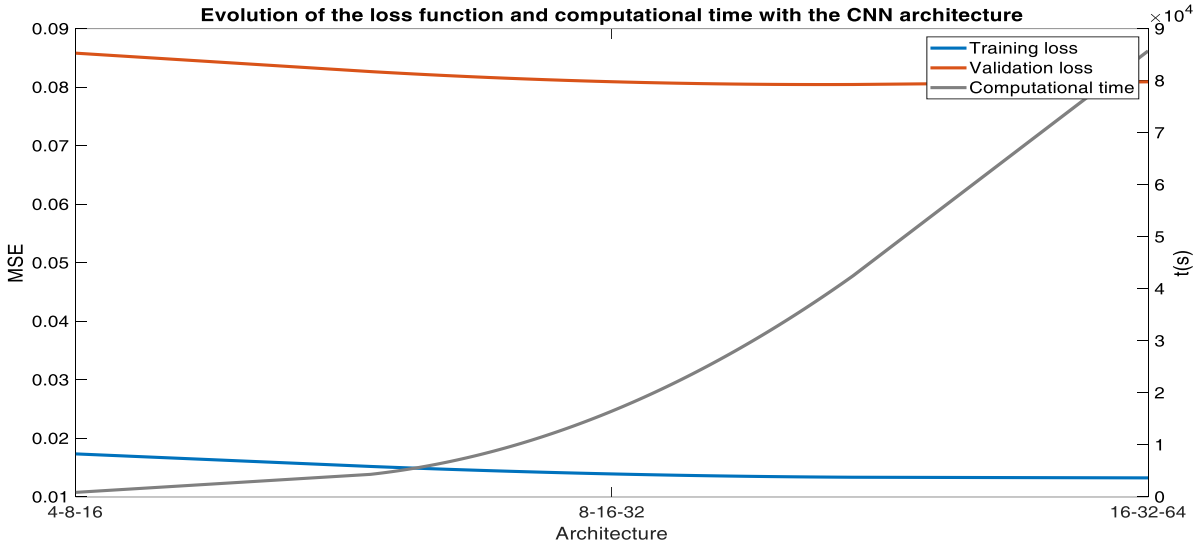


Figure 2-35. CNN architecture optimization.

The CNN has filters with a size of 3x3 and strides of 1 and the pooling layers have a size of 2x2. Among the different optimizers and activation functions available, Adam optimizer and RelU gave the best results. The final characteristics and representation of the model is presented in Table 2-13 and Figure 2-36.

Table 2-13. Characteristics of the model.

Hyperparameter	**Value**
Nº epochs	100
Batch size	12
Learning rate	0.0005
Architecture	FNN [4-8] – CNN [Conv (8) – MaxPool – Conv (16) – MaxPool– Conv (32)]
Convolutional filter size	3x3
MaxPooling size	2x2
Stride	1
Optimizer	Adam
Activation function	RelU
Data partition	70% training 30% validation
Errors	**Value**
Training	0.0134
Validation	0.0800

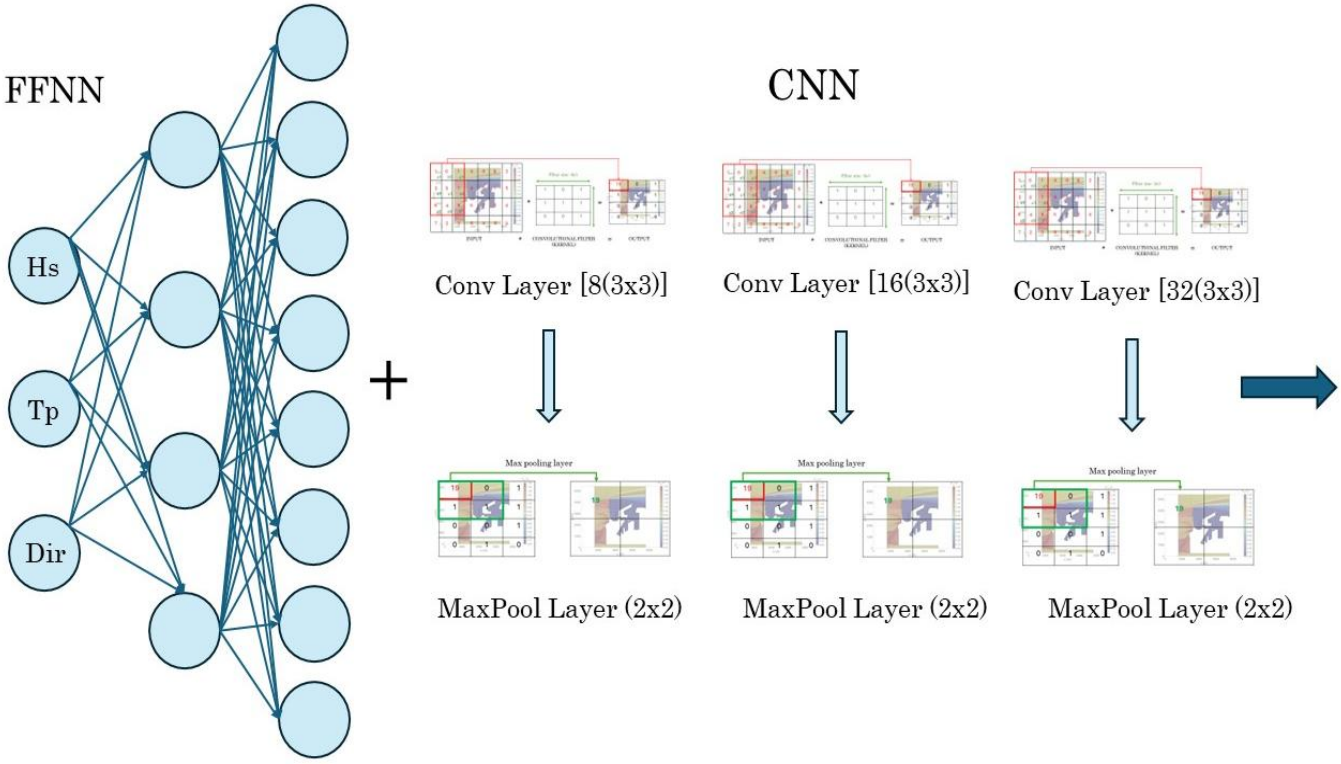


Figure 2-36. Deep Learning model for harbour agitation.

3. Scientific production

3.1. Index of papers

Table 3-1. Research Papers.

	Paper 1
Title:	**A New Climate Change Analysis Parameter: A Global or a National Approach Dilemma**
Authors:	*Nerea Portillo Juan*, Vicente Negro Valdecantos, José María del Campo Yagüe
Journal:	Energies 15 (4), 1522, 2022. MDPI
ISSN:	1996-1073
Quartile:	SCI: 102/427, Quartile Q1; SCOPUS: 27/151, T1.
DOI:	doi.org/10.3390/en15041522
	Paper 2
Title:	**Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain**
Authors:	*Nerea Portillo Juan*, Vicente Negro Valdecantos, José María del Campo Yagüe
Journal:	Sustainability 14 (12), 7507, 2022. MDPI
ISSN:	2071-1050
Quartile:	SCI: 127/366, Quartile Q2; SCOPUS: 99/747, T1.
DOI:	doi.org/10.3390/su14127507
	Paper 3
Title:	**Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast**
Authors:	*Nerea Portillo Juan*, Vicente Negro Valdecantos, José María del Campo Yagüe
Journal:	Water 14 (21), 3453, 2022. MDPI

ISSN:	2073-4441
Quartile:	SCI: 67 /256, Quartile Q2; SCOPUS: 111/779, T1
DOI:	doi.org/10.3390/w14213453
	Paper 4
Title:	**Review of the application of Artificial Neural Networks in ocean engineering**
Authors:	*Nerea Portillo Juan*, Vicente Negro Valdecantos,
Journal:	Ocean Engineering 259, 111947, 2022. Elsevier
ISSN:	0029-8018
Quartile:	SCI: 6/106, Quartile Q1; SCOPUS:16/100, T1.
DOI:	doi.org/10.1016/j.oceaneng.2022.111947
	Paper 5
Title:	***Uncertainties in the application of artificial neural networks in ocean engineering***
Authors:	*Nerea Portillo Juan*, Clara Matutano, Vicente Negro Valdecantos,
Journal:	Ocean Engineering 284, 115193, 2023. Elsevier
ISSN:	0029-8018
Quartile:	SCI: 6/106, Quartile Q1; SCOPUS:16/100, T1.
DOI:	doi.org/10.1016/j.oceaneng.2023.115193
	Paper 6
Title:	***Comparison of SIMAR-WANA, ERA-5 and Waverys databases for maritime climate estimations and implications for coastal protection structures***
Authors:	*Nerea Portillo Juan*, Javier Olalde Rodríguez, Vicente Negro Valdecantos
Journal:	Journal of Waterway, Port, Coastal and Ocean Engineering 150 (1), 04023021, 2024. ASCE.
ISSN:	1943-5460

Quartile:	SCI: 28/106, Quartile Q2; SCOPUS: 36/100 T2
DOI:	doi.org/10.1061/JWPED5.WWENG-2017
	Paper 7
Title:	***Data-driven and physics-based approach for wave downscaling: A comparative study.***
Authors:	*Nerea Portillo Juan*, Javier Olalde Rodríguez, Vicente Negro Valdecantos, Gregorio Iglesias
Journal:	Ocean Engineering 285, 115380, 2023. Elsevier.
ISSN:	0029-8018
Quartile:	SCI: 6/106, Quartile Q1; SCOPUS:16/100, T1.
DOI:	doi.org/10.1016/j.oceaneng.2023.115380
	Paper 8
Title:	***A new methodology for the study of future maritime climate using Artificial Neural Networks and Monte Carlo simulations under the context of climate change.***
Authors:	*Nerea Portillo Juan*, Vicente Negro Valdecantos
Journal:	Ocean Modelling, 102384, 2024. Elsevier
ISSN:	1463-5011
Quartile:	SCI: 49/344, Quartile Q1; SCOPUS: 27/133, T1
DOI:	doi.org/10.1016/j.ocemod.2024.102384

3.2. A New Climate Change Analysis Parameter: A Global or a National Approach Dilemma

Abstract and Keywords

Climate change is an issue nowadays present in almost all of the media daily, but information can be manipulated very easily. It is a fact that, in the last decades, greenhouse gas emissions have multiplied, and to tackle climate change efficiently, it is necessary to analyze their origin and their relationship with regards to countries, population, production, etc. When analyzing a country's emissions, not only the total emissions, but also the emissions in relation to its population, production, etc., should be considered. In this paper, a new parameter (CE_2N) that merges total emissions, and emissions per capita and per GDP is proposed and applied, obtaining, for the first time, a unified and universal parameter that considers the emission efficiency and total emissions at the same time and can be used in all countries. We validated this new parameter with its implementation in previous environmental models, and the results obtained showed that CE_2N would help to increase the transparency and objectivity of these models, giving more weight to emission efficiency, rather than other, more subjective criteria previously used. In addition, CE_2N could be implemented in future international agreements, being beneficial not only for the scientific community, but also for policymakers.

Keywords: climate change; international agreements; countries emissions; greenhouse gases; emissions per capita; emissions per gross domestic product; new environmental parameter; sustainable policies; climate emergency; COP results

1. Introduction

Defining climate and climate change is a complex issue that requires study and meditation (Todorov, 1986 [1] and Werndl, 2016 [2]). The correct definition of climate change is crucial for both, climate policies and scientific studies. In fact, there are whole papers that deal with the problem of climate and climate change definition. Werndl, 2016 [2] stated that the definition of climate has to meet five conditions: be empirically and mathematically valid, classify different climates, not depend on human knowledge, and be applicable to the past, present, and future. He gave five definitions of climate and discussed their advantages and inconveniences. Lorenz already defined climate change mathematically in 1970 [3]. Werndl, 2016 [2] and Dymnikov and Gritsoun, 2001 [4] stated that there is climate

change when there are different distributions of climate variables for two successive time periods. More recently, the United Nations Framework Convention on Climate Change (UNFCC) defined climate change as "a change of climate which is attributed directly or indirectly to human activity that alters the composition of the global atmosphere and which is in addition to natural climate variability observed over comparable time periods" [5], and the IPCC defined climate change as "a change in the state of the climate that can be identified by changes in the mean and/or the variability of its properties and that persists for an extended period, typically decades or longer" [6].

Climate change has always existed, since the origin of the Earth, 4500 M years ago. However, almost all literature about climate change and global warming only focuses on the last few decades. Although it is true that climate change has been influenced for a long time since the agricultural revolution [7,8], the main factors responsible for it are greenhouse gas (GHG) emissions, and it is has been after the industrial revolution that the amount of GHGs emitted to the atmosphere increased considerably. Between 1750 and 2011, cumulative anthropogenic CO_2 emissions to the atmosphere were 2040 ± 310 $GtCO_2$, and about half of these emissions have occurred in the last 40 years. The total anthropogenic GHG emissions continued to increase between 1970 and 2010, with the largest absolute increases occurring in the last decades. In fact, each of the last three decades has been successively warmer at the Earth's surface than any previous decade since 1850. Combined and globally averaged land and ocean surface temperature data, calculated from a linear trend, show a warming of 0.85 (0.65 to 1.06) °C over the period of 1880–2012. This increase in emissions has not only produced the warming of the planet, but it has also altered ecosystems, provoked the acidification of the ocean, the death of animal species, etc. [9].

Emissions vary per country and understanding the contribution of each country to climate change is essential to face it. However, there is still no agreement on the criteria that should be used to judge countries' environmental policies, whether by total emissions, emissions per capita, or emissions per gross domestic product (GDP).

Reducing GHG emissions is directly related to Sustainable Development Goals (SDG). SDG were adopted in 2015 in the Agenda 2030 by the General Assembly of the United Nations (UN) [10,11]. Reducing GHG emissions is considered in SDG 13: Climate Action, which is positively correlated with SDG 7: Affordable and Clean Energy [11]. The literature about emissions, sustainable development goals,

and energy is very rich. Among others, Bilan et al. [12], Zoundi [13], and Liu et al. [14] studied it and concluded that the use of renewable energy allowed decreasing the emissions of GHG. Ziolo et al. [15] and Akaev et al. [16] supported the need to increase energy efficiency and to develop a good energy transition if sustainable development is to be reached. Peña-Ramos et al. [17] studied the energy transition in Spain. Drozdz et al. [18] studied the determinants of decarbonization and of the development of a sustainable energy system in Poland, and Zimon et al. [19] studied the relationship between emissions, SDG, business models, and supply chains.

There are also many scientific articles about GHG emissions, their impact on the economy, energy, and population, their spatial distribution, and the relationship between all of these variables. In fact, the study of the linkage between economic growth and emissions goes back to the 1990s, with Grossman and Krueger [20,21] and Douglas [22].

Some of the most relevant studies about the distribution of GHG emissions, population, and GDP are the following: Zhang et al. [23] studied the advantages and drawbacks of using each variable (total emissions, emissions per capita, and emissions per GDP) as an indicator. Moghaddam et al. [24] proposed a new GHG intensity indicator that integrated the concepts of GDP and per capita in a new one that they called activities. Liang et al. [25] carried out an analysis of the emissions of each nation. Arango Miranda et al. [26] analyzed the relationship between carbon dioxide (CO_2) emissions, energy consumption, and economic growth. Sun et al. [27] developed an index that incorporated the concepts of ecology, economy, and equity to analyze the spatial distribution of emissions in China. Leitão et al. [28] studied the linkage between emissions, economic growth, tourism, and renewable energy in the European Union. Bădîrcea et al. [29] studied the relationship between climate change, emissions, and the blue economy, and Karman et al. [30] presented a climate model to evaluate regional competitiveness.

All these previous studies either conducted complex environmental models on specific areas/countries, or analyzed only some of the three main parameters to be considered when talking about emissions (population, GDP, and total emissions). However, none of them proposed neither a universal formula that could be applied to every country of the world, nor a formula that integrated GDP, population, and total emissions in only one parameter. This paper proposes a new universal formula to quantify emissions that integrates these three main variables in a

unique parameter: CE_2N (where C means carbon dioxide, E means equivalent, and N means normalized).

The main objective of this study is to define a universal criterion that puts an end to disputes about what should be the environmental variable (total emissions, or emissions per capita or per GDP) that underlies international agreements, unifying the variables used until now into a single CE_2N parameter that can also be applied in environmental models developed by the scientific community. The results of environmental models vary a lot depending on the criteria used and, as CE_2N unifies all these criteria, it will end with this problem.

For the first time in the literature, a global parameter that unifies the three main environmental variables and that can be applied in every country of the world, not only in some countries or specific areas, is proposed.

The uniqueness of the study lies in the fact that it proposes a unified and simple criterion that can benefit both politicians and the scientific community, helping to narrow the gap between science and society and making the understanding of emissions distribution accessible to everyone. CE_2N is very simple and has an advantage over other studies—its clarity. There is no need to use sophisticated and unclear equations to obtain CE_2N, although it can also be implemented in more complex models (Section 4.3.), also making environmental models more objective.

The limitations of the technique applied in this study are that the allocation of the total emissions, emissions per capita, and per GDP weights depends on user criteria (Section 3.2 Weighting of total emissions, emissions per capita, and per GDP), and that, to obtain CE_2N, only the emissions of the present year are considered; consequently, CE_2N gives only an image of the environmental policies of a country in the present. However, historical emissions could be considered when applying CE_2N in environmental models (Section 4.3.)

The structure of the paper is as follows: Section 2 presents the literature review; Section 3 describes the methodology; Section 4 reveals the results obtained; Section 5 presents the discussion of results and policy recommendations; and Section 6 states the conclusions reached and suggests new research lines.

2. Literature review

The literature on climate change, gas emissions of nations, and equity is voluminous. Some of the most significant contributions are those made by Sagar

[31], who studied the allocation scheme of emissions; Vaillancourt and Waaub [32], who developed a model to allocate emissions based on the equity principle; Markandya [33], who conducted an analysis of the distribution of climate change; Mattoo and Subramanian [34], who conducted a review about equity and climate change; Remuzgo et al. [35], who studied the evolution of GHG emissions during 1990–2011; Alcaraz et al., who developed a model of climate justice per capita [36] and applied it to the Mediterranean [37]; and Jena et al. [38], who designed a nonlinear model to forecast the CO_2 emissions of the 17 key emitting countries.

There are three main approaches in the study of the distribution of gas emissions and equity [33,34]:

Emissions per capita: Based on the principle that every person has an equal right to emit GHG. There are some authors that support this approach, such as Dubash, N.K. [39], Saran, S. [40], and Alcaraz et al. [36]. However, judging environmental responsibility focusing on emissions per capita is far from equity, because the capacity to produce GHG and also to reduce emissions highly depends on the level of development of a country [41,42,43]. In addition, the main percentage of GHG emissions is not caused by individuals.

Emissions per GDP: Supported by Cao, J. [44], among others. Based on the idea that countries with higher levels of production have the right to emit more GHG, but, as it happens with the emissions per capita scope, this theory helps to increase the gap between high and low-income countries.

Historic responsibility: Dubash, N.K. [39], Cooper, R. N. [45], and Stern, N. [46] are some of the authors that agree with this scope. It considers that the allocation of future emissions should be inversely related to past emissions. However, it is not clear from when emissions should be accounted from. Although many authors consider that it should be since the industrial revolution (Cao, J. [44], Kanitkar et al. [47]), humankind has been altering the climate since the origin of the human species and not only over the last 100 years, so it would not be rigorous to only consider a few years. However, as stated by Shue [48], emissions from developing countries should be treated differently to those from developed countries; the latter should support developing countries with their adaptation costs, mainly through financial and technological transfers (Falkner [49]).

The fact that the emissions per capita and per GDP approaches are not fair is supported by the Environmental Kuznets Curve Theory [50], which states that the emissions of CO_2 follow an inverted U-shape with economic growth. Therefore, the

level of emissions of a country will depend on its level of development. The literature on this topic is very abundant. Some of the most recent and relevant studies were carried out by Dogan et al. [51], Mania et al. [52], Jóźwik et al. [53], and Knight and Schor [54].

The relationship between globalization and GHG emissions has also been studied, but there are no solid conclusions. Some studies, such as the ones carried out by Rahman [55], Villanthenkodath et al. [56], and Baydoun et al. [57], concluded that globalization helps to reduce GHG emissions, while others, such as the ones by Pata, U.K. [58] or Kihombo et al. [59], concluded the opposite.

Similar to globalization, there are no strong conclusions linking financial development and emissions. Some studies defend a negative relationship (Xu et al. [60] and Salahuddin et al. [61]), while others support a positive relationship (Chebbi et al. [62], Jiang and Ma [63], and Ameyaw and Lao. [64]).

International Agreements, Kyoto, and Paris

In order to tackle the problem of climate change, it is essential to study the international agreements that have been reached throughout history and the political and social situation. This is the main rationale for this section.

In 1992, climate change was recognized as a problem for the first time, when the UNFCC was adopted [65]. Since then, 26 conferences have taken place and two big international agreements have been adopted: the Kyoto Protocol and the Paris Agreement [66,67,68].

The Kyoto Protocol was adopted in 1997; it was the first international agreement in which nations committed to reducing their emissions by 5% (referring to 1990 emissions) for the period of 2008–2012. However, the Kyoto Protocol distinguished between developed and developing countries. Only developed countries who were annexed in UNFCC had quantified commitments, while developing countries only had to try to improve their policies. In addition, the Kyoto Protocol defined different tools to help nations to reach their targets [66]:

International Emissions Trading: Based on the right of emitting gases and its commercialization. When a country reduces its emissions in a percentage higher than its target, it can sell the surplus of CO_2 on the carbon market.

Clean Development Mechanism (CDM): It allows an annexed country to implement an emission-reduction project on developing countries. Such projects can earn

saleable certified emission reductions, which can be counted towards meeting reduction targets [69].

Joint Implementation (JI): It allows an annexed country to earn emission reduction units from an emission-reduction or emission removal project in another annexed country [70].

In 2012, the Kyoto Protocol was extended until 2020 with the Doha Amendment. Countries that ratified it committed to reducing their emissions by 18% for 2020. However, the Doha Amendment never entered into force because the annexed countries only represented 15% of global emissions [66,67].

The other big international agreement is the Paris Agreement that came into effect in 2016. It leaves the distinction between developed and developing countries and also the common objective of reducing the emissions by a certain percentage. With this new agreement, each country has to commit to making its best effort to develop environmental policies through nationally determined contributions [68].

3. Materials and Methods

The new parameter proposed, CE_2N, is obtained by applying the following equation:

$$CE_2N[units\ of\ normalized\ CO_2] = \alpha CO_2[Mt/year] + \beta CO_2\ per\ capita\ [t/year] + \gamma \frac{CO_2}{kUSD}\ [t/year] \quad (1)$$

Where α, ß, and γ are constants that are obtained by normalizing and weighting each variable.

3.1. Normalization of Total Emissions, Emissions Per Capita, and Per GDP

As the total emissions, and emissions per capita and per GDP have different orders of magnitude, the first step is to normalize them. To this end, the maximum and minimum values of each variable are studied.

The emissions per capita is the variable that has medium values (total emissions are usually higher than the emissions per capita and the emissions per GDP are usually lower). Therefore, the constant of the emissions per capita ß is taken as the base parameter. As ß is taken as the reference value, ß is equal to 1.

Normalization of Total Emissions (α):

- Total emissions are, on average, 41 times the emissions per capita.

- About 90% of the total emissions are between 0.2 and 82 times the emissions per capita.
- Total emissions' standard deviation is 154 units.

Therefore, to make the total emissions and emissions per capita have the same order of magnitude, the total emissions' constant α has to be divided by 0.2 and 82. This gives a range of values α of 0.012 to 5. From this acceptable range of values, the value of 0.07 was chosen as the recommended value for α, because it is the medium value plus 15% of deviation units, which was considered the best statistical criteria. Below, the whole process of normalization is explained step by step.

$$\text{Final objective: } \alpha * Total\ emissions = \beta * Emissions\ per\ capita \quad (2)$$

(3)

$$\begin{cases} \beta = 1 \\ Max.Total\ emissions = 82 * Emissions\ per\ capita \quad (4) \\ Min.Total\ emissions = 0.2 * Emissions\ per\ capita \end{cases}$$

(5)

Substituting (3) and (4) in (2) gives

$$\alpha * 82 * \ emissions\ per\ capita = 1 * emissions\ per\ capita$$

$$\alpha = \frac{1}{82} = 0.012$$

Substituting (3) and (5) in (2) gives

$$\alpha * 0.2 * \ emissions\ per\ capita = 1 * emissions\ per\ capita$$

$$\alpha = \frac{1}{0.2} = 5$$

$$Range\ of\ values\ of\ \alpha = 0.012\ to\ 5$$

$$\text{Average value} = 41; \text{Deviation units} = 154; 0.15 * \text{Deviation units} = 23.1$$

$$\alpha = \frac{1}{41} + \frac{1}{23.1} = 0.07$$

Normalization of the Emissions per kUSD (γ):

- Emissions per kUSD are, on average, 1/21 times the emissions per capita.

- About 90% of the emissions per kUSD are around 1/3 and 1/50 times the emissions per capita.
- The standard deviation of emissions per kUSD is 22 units.

Taking the parameter of emissions per capita ß as the base parameter, γ range of values should be between 3 and 50. From this range, the value of 23 was chosen as the recommended value for γ because it is the medium value of the range proposed plus 10% of the deviation units, which was considered the best statistical criteria. (Same criteria as those applied to total emissions, but with 5% fewer deviation units because, for the total emissions, the standard deviation is more relevant).

Final objective: $\gamma * emissions\ per\ GDP = \beta * emissions\ per\ capita$ (6)

$$\begin{cases} \beta = 1 & (7) \\ Max.Emissions\ per\ GDP = \frac{1}{3} * Emissions\ per\ capita & (8) \\ Min.Emissions\ per\ GDP = \frac{1}{50} * Emissions\ per\ capita & (9) \end{cases}$$

Substituting (7) and (8) in (6) gives

$$\gamma * \frac{1}{3} * \ emissions\ per\ capita = 1 * emissions\ per\ capita$$

$$\gamma = 3$$

Substituting (3) and (5) in (2) gives

$$\gamma * \frac{1}{50} * \ emissions\ per\ capita = 1 * emissions\ per\ capita$$

$$\gamma = 50$$

$$Range\ of\ values\ of\ \gamma = 3\ to\ 50$$

$$\text{Average value} = \frac{1}{21}; \text{Deviation units} = 22; 0.10 * \text{Deviation units} = 2.2$$

$$\gamma = 21 + 2.2 = 23$$

In the table below (Table 1), the values obtained for α, ß, and γ after normalizing the variables are presented.

Table 1. Values of α, ß, and γ after normalization.

Constant	Range Without Weight Allocation	Recommended Value of the Range

α	0.012 to 5	0.07
β	Base parameter (1)	Base parameter (1)
γ	3 to 50	23

3.2. Weighting of Total Emissions, Emissions Per Capita and Per GDP

The second step is to assign the corresponding weight to each of the variables. It was considered that the best relative weight for the total emissions (α) was 70%, and 15% for the emissions per capita (β) and per kUSD (γ) (Table 2).

Table 2. Parameter values proposed.

Parameter	Range Without Weight Allocation	Recommended Value of the Range	Relative Weight	Recommended Value
α	0.012 to 5	0.07	0.7	**0.05**
β	Base parameter (1)	Base parameter (1)	0.15	**0.15**
γ	3 to 50	23	0.15	**3.50**

This distribution is based on the fact that the priority right now is to cut CO_2 emissions. This is why the total emissions relative weight is much higher (0.7). It is based on the authors' criteria and knowledge, but if potential users of CE_2N do not agree with this distribution of weights, it can be changed. If a user considers that a better distribution would be 60% for total emissions and 20% for emissions per capita and per GDP, the new values of α, β, and γ can be obtained by multiplying the recommended value of the range (column 3 of Tables 1 and 2) by the new relative weight:

α: 0.07*0.6 = 0.04

β: 1*0.2 = 0.2

γ: 23*0.2 = 4.6

$$CO_2N = 0.05CO_2 + 0.15CO_2\ per\ cápita + 3.5\ CO_2/kUSD \quad (10)$$

Once the formula was defined, it was validated by introducing it into previous environmental models. Finally, the results obtained were analyzed and compared to the results obtained from the original models.

4. Results

To analyze how the implementation of CE_2N will change international agreements and environmental models, a previous analysis of emissions statistics and of compliance with agreements needs to be presented.

Section 4.1 presents the analysis of the total emissions, and emissions per capita and per kUSD; Section 4.2 presents the analysis of the compliance of the Kyoto Protocol; Section 4.3 focuses on the new parameter, CE_2N; and, finally, Section 4.4 presents different graphs to illustrate the application of CE_2N.

4.1. Total Emissions, and Emissions Per Capita and Per kUSD

The distribution of the global total emissions in 1990 and 2019 is presented below (Figure 1).

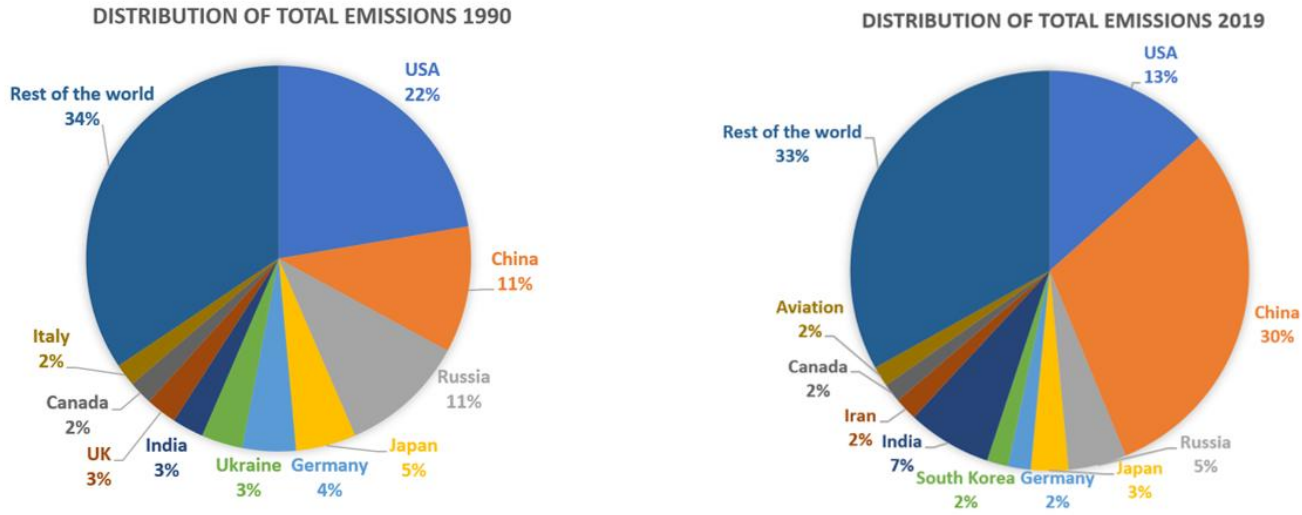


Figure 1. Distribution of global emissions 1990 vs. 2019.

Since 1990 (base year), global emissions have increased by 68% and the world population has increased by 45%, which means that, today, humanity is not only emitting more CO_2, but also more CO_2 per person (16% greater than in 1990) [71]. In addition, as it can be seen from the figure above, the three most polluting countries (China, the USA, and India) account for 50% of global emissions. These three countries emit as much as the remaining 191 countries, which means that, without their collaboration, the climate change problem cannot be addressed.

The tables below show the countries that reduced and increased their emissions the most in 2019 compared to 1990 (Table 3 and Table 4, where Mt is Megaton and t is ton).

Table 3. Ten first countries with the greatest reduction in emissions.

Max. Reduction of Emissions (Mt/year)	
Country	Reduction 1990
EU27+UK	1104
Russia	601
Ukraine	586
Germany	315
UK	223
Romania	108
Italy	98
North Korea	89
France	71
Czechia	57

Table 4. Ten countries with the greatest increment in emissions.

Max. Increment of Emissions (Mt/Year)	
Country	Increment 1990
China	9130
India	1997
Iran	497
Indonesia	461
Saudi Arabia	441
South Korea	381
Vietnam	284
Turkey	265
Brazil	250
Mexico	194

If, instead of the global total emissions, the emissions per capita are studied, the results are very different. China or the USA are no longer the most polluting countries, and, in the ranking of countries with the worst values of emissions per capita, new countries appear [71] (Table 5).

Table 5. Ten countries with the worst values of emissions per capita in 1990 and 2019.

1990		2019	
Country	**t CO_2/Capita/Year**	**Country**	**t CO_2/Capita/Year**
Palau	142.826	Palau	58.879
Curaçao	37.321	New Caledonia	55.25
Qatar	35.695	Qatar	38.823
Luxembourg	30.777	Curaçao	36.382
United Arab Emirates	30.601	Trinidad & Tobago	23.806
Estonia	24.576	Kuwait	23.289
Bahrain	24.173	United Arab Emirates	22.992
St. Pierre & Miquelon	23.556	Bahrain	21.637
USA	20.057	Gibraltar	19.884
Canada	16.373	Oman	18.549

When the emissions per kUSD are analyzed, the results are the same as in the previous section. There are new countries that appear in the list of countries with the worst values of emissions per kUSD [71] (Table 6).

Table 6. Ten countries with the worst values of emissions per kUSD in 1990 and 2019.

1990		2019	
Country	**t CO_2/kUSD**	**Country**	**t CO_2/kUSD**
Palau	10.273	Palau	4.089
Bosnia & Herzegovina	3.024	New Caledonia	1.666
Curaçao	1.882	Curaçao	1.514
Uzbekistan	1.757	Syria	1.198
Turkmenistan	1.718	Armenia	1.129
China	1.488	Turkmenistan	0.98
Estonia	1.475	Mongolia	0.905
Mongolia	1.208	Trinidad & Tobago	0.897
Belarus	1.203	Barbados	0.853
Kazakhstan	1.141	Iran	0.685

From Tables 3 and 4, it can be seen that the countries that reduced most their emissions with respect to 1990 are European, and the countries that incremented

their emissions the most are China and India, two of the three most polluting countries in 2019. Tables 5 and 6 show that the countries with worst the emission efficiencies are developing countries. These two facts are related to the Environmental Kuznets Curve Theory [50] that states that the emissions of CO_2 follow an inverted U-shape with economic growth. In India and China, economic growth is based on industrial development, which means increasing GHG emissions. On the other hand, European countries were already developed countries in 1990, and their economy and technology allow them to continue growing without emitting more CO_2 into the atmosphere.

In the table below (Table 7), the countries with the largest values of emissions and their relative position according to other variables (emissions per capita and per kUSD) are presented. It can be seen that the statistics differ a lot depending on the variable analyzed.

Table 7. Position of the countries with the highest values of emissions in the emissions per capita and per kUSD in 2019.

Countries with the Highest Values of CO_2 Emissions in 2019	Relative Position Among Countries with Worst Values t CO_2/Capita/Year 2019	Relative Position Among Countries with Worst Values t CO_2/kUSD/Year 2019
China	42	18
USA	16	60
EU27+UK	53	115
India	124	32
Russia	22	24
Japan	33	68
Germany	37	108
Iran	39	12
South Korea	21	45
Indonesia	110	83

4.2. Kyoto annexed countries emissions

With the aim of analyzing the effectiveness of the Kyoto Protocol and its compliance, the emissions of each annexed country in 1990 and 2019 are presented in Table 8 (the emissions are given in Mt/year).

Table 8. Kyoto Protocol emissions analysis.

Country	% Emissions Base Year Commitment	Base Year Emissions	2019 Emissions	*%2019-base*
Germany	80	1018.221	702.6	**69.00**
Australia	99.5	354.23	433.379	**122.34**
Austria	80	62.927	72.363	**115.00**
Belgium	80	115.996	104.415	**90.02**
Bulgary*	80	82.346	43.314	**52.60**
Belaruse*	88	109.02	66.335	**60.85**
Cyprus	80	4.54	7.413	**163.24**
Croatia*	80	25.158	19.119	**76.00**
Denmark	80	53.59	31.119	**58.07**
Slovakia*	80	60.545	35.985	**59.44**
Slovenia*	80	16.624	15.365	**92.43**
Spain	80	230.354	259.31	**112.57**
Estonia*	80	38.468	18.503	**48.10**
Finland	80	57.254	43.415	**75.83**
France Monaco	80	386.367	314.736	**81.46**
Greece	80	79.191	65.568	**82.80**
Hungary*	80	72.075	53.183	**73.79**
Ireland	80	32.904	36.548	**111.07**
Iceland	80	2.357	3.925	**166.53**
Italy	80	430.061	331.563	**77.10**
Kazakhstan	95	251.29	277.365	**110.38**
Latvia*	80	20.132	8.379	**41.62**
Lithuania*	80	35.305	13.772	**39.01**
Luxembourg	80	11.75	9.74	**82.89**
Norway	80	37.324	47.991	**128.58**
Holland	80	161.195	156.415	**97.03**
Poland*	80	371.381	317.654	**85.53**
Portugal	80	43.692	48.472	**110.94**

United Kingdom and North Ireland	80	588.068	364.906	**62.05**
Czech Republic*	80	162.835	78.631	**48.29**
Romania*	80	187.291	78.631	**41.98**
Sweden	80	58.104	44.749	**77.02**
Switzerland and Liechtenstein	84.2	44.955	39.371	**87.58**
Ukraine*	80	783.21	196.401	**25.08**
Total	**82**			**72**

As it can be seen, the Kyoto Protocol was, by far, fulfilled. The annexed countries reduced their emissions by 72% with respect to the emissions in 1990. Twenty of 34 countries reached their objectives, which means 60% of the countries. The remaining 40% did not fulfill their promises, which can be considered a high percentage. However, 85% of the countries that did achieve their targets reduced their GHG emissions by more than 20%, and some of the countries reduced their emissions by more than 50%: Ukraine (75%), Lithuania (61%), and Romania (60%) are some of these countries. This is the main reason why, despite the fact that many countries did not deliver as promised, the Kyoto Protocol did deliver, because the countries that did reduce their emissions did so at high rates. These results and their implications will be further discussed.

The USA, Russia, Japan, New Zealand, and Canada are not included in the table because they abandoned the Kyoto Protocol before its end.

Countries with "*" are countries that were considered by Kyoto Protocol as developing countries, so they did not have real commitments, they only had to try to develop sustainable policies. Countries with the %2019-base in green are countries that fulfilled their commitment, and countries with this number in red are countries that did not.

4.3. CE_2N

As the results differ a lot depending on the variable analyzed, the formula presented in the section "Methodology" has been proposed in order to integrate all

variables in one parameter: CE_2N. This new parameter CE_2N aims to define a unique classification of the most polluting countries and also proposes a unified criterion that could be applied in environmental models. CE_2N considers all the variables (GDP, population, and total emissions), and not just one of them. It is important to consider the total emissions, GDP, and population in one parameter to account for different factors and to be more impartial when assessing countries' environmental policies.

Configuring CE_2N with $\alpha = 0.05$, $\beta=0.15$ and $\gamma=3.5$, the obtained classification of the 20 most-polluting countries for the year 2019 would be as follows (Table 9):

Table 9. CE_2N Classification.

Country	CE_2N
China	579.77045
USA	258.5589
EU27+UK	166.67385
India	131.4827
Russia	93.05085
Japan	59.8236
Germany	38.7682
Iran	36.95795
South Korea	35.53405
Saudi Arabia	34.76765
Canada	32.705
Indonesia	32.3173
South Africa	28.9105
Mexico	25.47545
Australia	25,46055
Brazil	24.7873
Palau	23.20985
Turkey	22.16045
United Kingdom	19.4723
Kazakhstan	18.09745

If the CE_2N classification is compared with the total emissions classification per country, the results are as follows (in red are those countries that change their position in the ranking) (Table 10).

Table 10. Twenty top emitters countries and differences with the CE_2N classification.

Country	Mt/Year
China	11535
USA	5107.2
EU27+UK	3303.9
India	2597.3
Russia	1792.0
Japan	1153.7
Germany	702.6
Iran	701.98
South Korea	651.87
Indonesia	625.66
Saudi Arabia	614.60
Canada	584.84
South Africa	494.86
Mexico	485.00
Brazil	478.14
Australia	433.37
Turkey	415.78
United Kingdom	364.90
Italy, Vatican City and S.Marino	331.56
Poland	317.65

As per the table, CE_2N gives a new methodology to assess countries' emissions; considering the emissions per capita and per GDP changes the usual classification.

Comparing Tables 9 and 10, it can be seen that using CE_2N narrows the emissions gap between countries. While the difference in the total emissions between China and USA is 6428 Mt, it is 321 in equivalent carbon units. In fact, from the eighth position, the CE_2N differences between countries are less than 1 equivalent carbon unit (Table 9), while those in Table 10 are more than 100 Mt. This implies that countries that change their environmental policies and make an effort to improve

their emissions statistics can easily change their position in the CE_2N Classification (Table 9), while it would be much more difficult in the usual classification (total emissions, Table 10). Therefore, using CE_2N as the new reference environmental parameter could promote the improvement of environmental policies, as it would be much easier to see direct results than with the previous environmental criteria. However, the position of the top eight emitting countries is very difficult to change due to their high level of total emissions.

Sensitivity analysis

The CE_2N formula is a formula that integrates the concept of total emissions, emissions per capita, and emissions per GDP with the aim of knowing how polluting a country is. The sensitivity of this formula depends on the values of the parameters chosen. A sensitivity analysis with the recommended values of the parameters for two scenarios was conducted:

- Scenario 1: Europe continues with its policy of reducing its total emissions, reaching an 8% of reduction, and Asia and America continue increasing their emissions with an increment of 5%.
- Scenario 2: Most developed countries reduce their emissions by 5% and developing countries increase their emissions by 5%.

Total emissions sensitivity

For both scenarios, the formula was "highly sensitive" to the total emissions. However, the position of the 10 countries with the highest values of emissions hardly changed due to their high values of emissions. The difference in the amount of gases emitted by the 10 most emitting countries was large. For example, China emits 6428 Mt of CO_2 more than the USA [71], which is a great difference that makes it very difficult to change the relative position of these countries. As emissions become lower, the difference in emissions between countries also decreases, which makes the formula proposed much more sensitive. The positions of the countries began to change from the ninth position.

Emissions per capita and per GDP sensitivity

With the values of the parameters proposed, the formula of CE_2N is less sensitive for the emissions per capita and per GDP than for total emissions. This is due to the reason that it was considered that the most appropriate relative weights were 70% for the total emissions and 15% for the emissions per capita and per GDP.

However, if the user does not agree with this and considers that the emissions per capita and per GDP should be more important in the formula, it would be enough to use a lower value of α and higher values of ß and γ.

Validation of the New Parameter CE_2N

In this section, CE_2N is introduced into the environmental models developed by Sagar [31] and Vaillancourt and Waaub [32].

The results obtained from the models are consistent with the changes that the introduction of the CE_2N into the models imply. Therefore, the new formula proposed for CE_2N is considered validated.

- Sagar, AD. (2000)

Sagar, AD. [31] developed a model to allocate future GHG emissions based on equal per capita emissions. Three factors were considered in this environmental model:

- Population
- Per capita income levels ($pcGNP_{PPP}$)
- Per capita accumulative responsibility (pcCR)

The basic allocation formula proposed by Sagar is [31]:

$$F_i = \frac{Pop_i * f\left(pcGNP_{PPP_i}\right)/g(pcCR_i)}{\sum_{i=1}^{n}\left[Pop_i * f\left(pcGNP_{PPP_i}\right)/g(pcCR_i)\right]} \quad (11)$$

(Sagar, A.D., 2000 [31])

where F_i is the fraction of the global emissions allocated to country i, and f and g are the functions of *pcGNPPPP* and *pcCR*, respectively.

In this research paper, Equation (11) was redefined, replacing function g with another function g' that depends on the parameter CE_2N. GDP was also used instead of $pcGNP_{PPP}$. The formula obtained after applying all those changes is as follows:

$$F_{iN} = \frac{Pop_i * f(GDP)/g'(CE_2N)}{\sum_{i=1}^{n}[Pop_i * f(GDP)/g'(CE_2N)]} \quad (12)$$

The obtained results are presented below (Table 11).

Table 11. Sagar results vs. results with CE_2N.

Year 1995		
Country	**F_i (%)**	**F_{iN} (%)**

USA	2.18	4.79
Japan	2.34	3.91
Norway	0.05	0.07
France	0.92	2.55
Canada	0.26	0.44
Germany	0.69	2.44
Netherlands	0.20	0.31
Italy	1.07	2.34
UK	0.53	1.49
Australia	0.17	0.24
Czech Republic	0.04	0.09
Poland	0.12	0.34
Russia	0.33	1.35
Ukraine	0.07	0.29
Singapore	0.04	0.04
Qatar	0.003	0.001
South Korea	0.99	0.64
Chile	0.43	0.23
Saudi Arabia	0.18	0.42
Malaysia	0.65	0.32
South Africa	0.24	0.33
Mexico	1.52	2.83
Brazil	5.36	8.39
Trinidad & Tobago	0.004	0.001
Ecuador	0.28	0.13
Philippines	2.78	1.43
Kazakhstan	0.03	0.05
Indonesia	5.78	7.25
China	15.30	9.64
Vanuatu	0.008	0.000
Egypt	1.01	1.50
Solomon Islands	0.019	0.000

Senegal	0.32	0.03
India	16.71	21.26
Kenya	0.89	0.39
Nepal	0.89	0.15
Bangladesh	4.54	3.11
Uganda	0.72	0.10
Nigeria	1.04	1.99
Burundi	0.16	0.01
Mozambique	0.32	0.02
Ethiopia	1.04	0.43

- Vaillancourt and Waaub (2003)

Vaillancourt and Waaub [32] also developed a model to allocate GHG emissions that was based on equity. They defined a total of 11 criteria to allocate these emissions and gave different weights to them that depended on the level of development of the country. (South and North criteria, Table 12).

Table 12. Vaillancourt results vs. CE_2N results.

Region	**North criteria (%)**	**South criteria (%)**	**CE_2N criteria (%)**
Africa	2.7	8.5	4.8
Latin America	2.6	4.2	5.4
Asia	1.7	5.3	8.5
Australia-NZ	0.5	0.7	1.9
Canada	1.1	1.1	2.5
FSU	6.5	7.3	8.3
China	10.2	24.9	16.3
South Korea	0.4	0.7	2.8
USA	49.7	23	13.3
Eastern Europe	1.1	1.7	3.2
Western Europe	16.8	9.4	8.3
India	1.4	5.8	7.3
Japan	2.1	1.9	4.9
Mexico	1	1.8	3.6

Middle East	2.1	3.7	8.9

In this research paper, it is proposed to substitute those criteria by the parameter CE_2N and apply the same model. The model can be summed up in two main steps:

1. Definition of priority functions

$$a(t) = \frac{1}{ave - min} \quad (13) \qquad ; \; b = 1 - max(a) \quad (14)$$

(Vaillancourt and Waaub, 2003, [32])

$$n_i\,(k,t) = v_i(k,t) * a(t) + b(t) \; if \; v_i(k,t) > average$$
$$n_i\,(k,t) = 0 \; if \; v_i(k,t) < average \quad (15)$$

(Vaillancourt and Waaub, 2003, [32])

where $v_i(k,t)$ is the variable; $n_i\,(k,t)$, the normalized variable; i, the country; k, the criteria and t, the year.

The priority functions are defined as:

$$g_{ij}(k,t) = \Big(n_i(k,t) - n_j(k,t)\Big) * a'(t) + b'(t) \; if \; n_i(k,t) > \; n_j(k,t) \quad (16)$$
$$g_{ij}(k,t) = 0 \; if \; n_i(k,t) < \; n_j(k,t)$$

(Vaillancourt and Waaub, 2003, [32])

$$a'(t) = \frac{1}{p - q} \quad (17)$$

(Vaillancourt and Waaub, 2003, [32])

$$b'(t) = a - p\big(a'(t)\big) \quad (18)$$

where p and q are the maximum and minimum differences between countries respectively.

Adding all the criteria considered (k):

$$g_{ij}(t) = \sum_k w_k g_{ij}(k,t) \; (19) \quad (19)$$

(Vaillancourt and Waaub, 2003, [32])

where w_k is the weight of criteria k

$$f_{ij} = (1 - \alpha_t)\Big(f_{ij}(t-1)\Big) + \alpha_t g_{ij}(t) \quad (20)$$

(Vaillancourt and Waaub, 2003, [32])

where α_t varies from 1 to 0 from 2000 to 2050

Finally,

$$x_{ij} = f_{ji}(t) - f_{ij}(t)\ (21) \tag{21}$$

(Vaillancourt and Waaub, 2003, [32])

2. Allocation of emissions

$$a_i(t) = e_i(t-1)\frac{E(t)}{E(t-1)}\ (22) \tag{22}$$

(Vaillancourt and Waaub, 2003, [32])

where $e_i(t-1)$ is the previous emissions of the region and $E(t)$ is the global emissions.

$$e_i(t) = a_i(t)\left[1 + \sum_j \frac{a_j(t)}{E(t)} x_{ij}(t)\right] \tag{23}$$

(Vaillancourt and Waaub, 2003, [32])

If, instead of using the 11 criteria that Vaillancourt and Waaub [32] defined, the parameter CE_2N is used as the conceptual base of the model, the allocation of CO_2 emissions for the year 2020 is as follows:

4.4. Graphs

To have a visual knowledge of the worst countries in terms of total emissions, and emissions per capita and per GDP, "traffic lights diagrams" are presented below (Figures 2–4). Countries that have unacceptable values of any of the variables are in the red zone and those that have medium or acceptable values are in the green area. The further from the origin, the worse the position of the country.

* All the data presented in the graphs are in emissions per year

Total emissions vs. emissions per capita

With the aim of scaling properly the graphs and making them easier to visualize, the three countries with the greatest values of total emissions and the two countries with the highest emissions per capita were removed and are not shown in the graph below.

These countries were China, the USA, EU27 + UK, Palau, and New Caledonia [71]. If they were represented, the five of them would be in the red zone.

Countries such as Australia, Canada, or South Korea that do not have high rates of total emissions are in the red zone because their emissions compared to their population are high, and they will be the countries less benefited if CE_2N is implemented in environmental models.

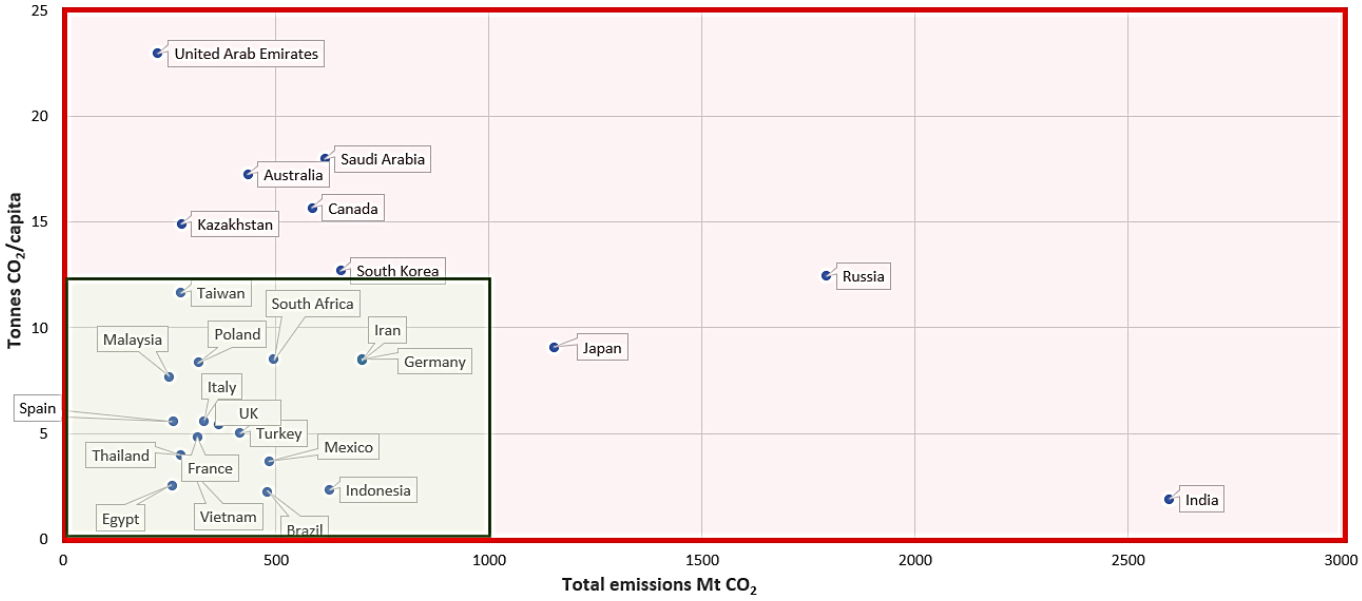


Figure 2. Traffic light diagram of the total emissions vs. emissions per capita.

Emissions per capita vs. emissions per kUSD

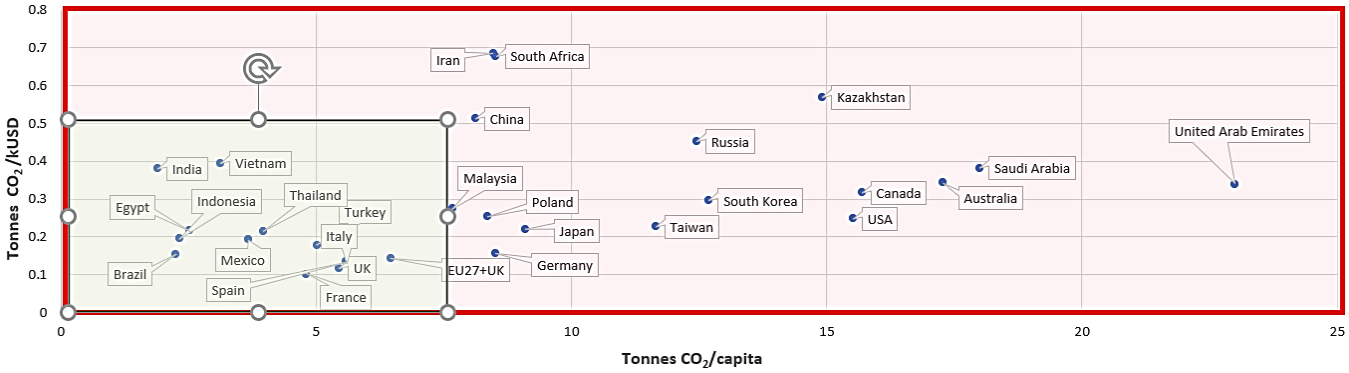


Figure 3. Traffic light diagram of the emissions per capita vs. emissions per kUSD.

Total emissions vs. emissions per kUSD

The same procedure as before was followed. China, the USA, EU27+UK, Palau, and New Caledonia were removed in order to obtain a better visualization of the graph.

As explained in Figure 2, countries with high values of emissions per GDP, such as South Africa or Iran, would be the countries most penalized.

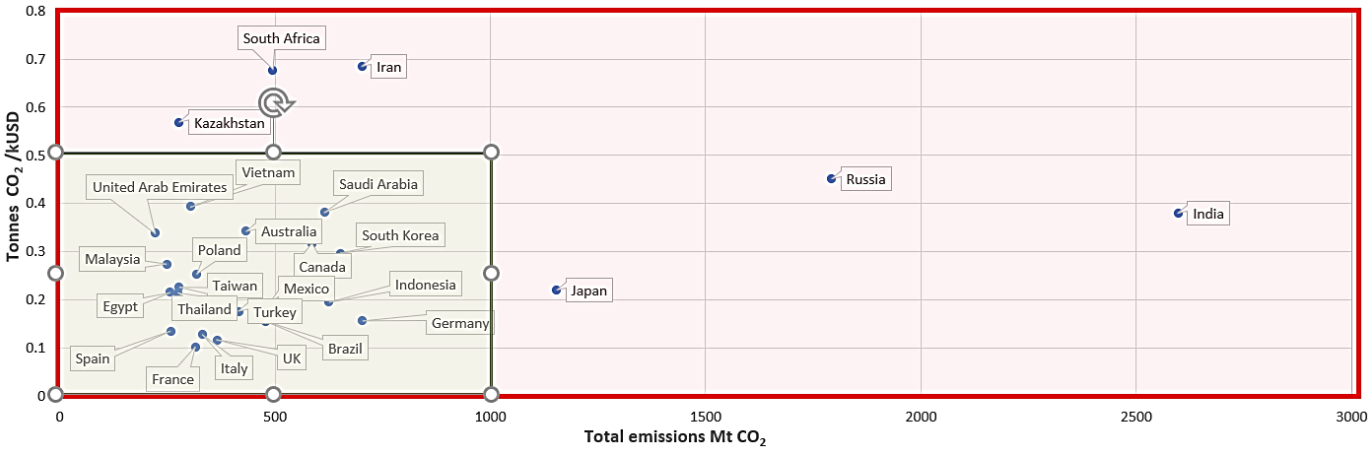


Figure 4. Traffic light diagram of total emissions vs. emissions per kUSD.

5. Discussion of results

From 1990 to 2019, global CO_2 emissions have increased by 15,333 Mt, which is 68% more compared to those in 1990. However, the Kyoto Protocol was satisfied. This is due to the fact that the countries that had reduction targets in the Kyoto Protocol and fulfilled their promises reduced their emissions by a high percentage, but the rest of the world did not.

As stated by Zhang et al. [23] and Sagar [31], the level of emissions of a country is closely related to its industrialized level and to its level of development. Economic growth is positively related to CO_2 emissions (Kim et al. [72]). Countries such as China or India that were considered as developing countries in 1990 and, therefore, did not have to fulfill any emission commitment in the Kyoto Protocol have become powerful countries and, consequently, have increased enormously their emissions. China has risen them by 480%, going from being responsible for 10% of the total emissions in 1990 to 30% in 2019, India from 2% to 7%, and Iran from 0.9% to 1.8%. On the other hand, annexed countries that ratified the Kyoto Protocol were responsible for 68% of emissions in 1990, but only accounted for 11% of the emissions in 2019. Countries such as Germany, Italy, or Russia considerably reduced their emissions, but their efforts were not enough. Countries who were annexed to the Kyoto Protocol reduced their emissions to 72% of the base year (the initial objective was only to reduce it to 82%), but global emissions increased by 68% with respect to those in 1990. This is clear proof that the Kyoto Protocol became outdated.

The effectiveness of the Kyoto Protocol is a very controversial topic and it is difficult to measure (Almer and Winkler [73]). Therefore, the results obtained in

this research contradict some of those of previous studies, such as Grunewald and Martínez-Zarzoso [74,75], who stated that the Kyoto Protocol had a reducing effect on CO_2 emissions in both developing and developed countries, and Maamoun [76], who considered the Kyoto Protocol a success. However, the results of this study are in line with the results obtained in other research: Kim et al. [72] limited the environmental benefits of the Kyoto Protocol to annexed countries, and Almer and Winkler [73] and Thakur [77] even questioned these positive effects.

With respect to the emissions per capita and emissions per kUSD, countries such as Curaçao or Palau are those with the worst values. There are two main types of countries that have high values of emissions per capita or emissions per kUSD:

- Countries with high levels of total emissions, such as China (high numerator in emissions/capita or kUSD).
- Developing countries with very low levels of production or who are poorly populated, which makes the emissions per capita or per kUSD high, even with a low level of total emissions (Palau or Curaçao). (Low denominator in emissions/capita or kUSD).

This diversity among countries with regards to the emissions per capita and per kUSD ranking could also be explained by the Environmental Kuznets Curve [50] mentioned before. This theory states that the emissions of CO_2 follow an inverted U-shape with economic growth. Therefore, countries that are right now developing and that have pre-industrial economies will have high rates of emissions per kUSD and usually also per capita. This is the reason why countries such as Kazakhstan, St. Pierre, and Miquelon or Bahrain appear in Table 5 and Table 6. In general, the countries with the worst values of emissions efficiency are developing countries, because their economy is not efficient and implies high rates of emissions.

Focusing on the new parameter proposed, CE_2N:

The classification of the 10 most polluting countries obtained by applying the new formula (CE_2N) is very similar to the classification obtained if only total emissions are analyzed. However, from the ninth position onwards, the new estimation proposed of CE_2N differs from the classification by total emissions. The use of CE_2N allows considering the emissions per capita and emissions per kUSD at the same time as the total emissions, giving a better idea of which countries should improve their environmental policies, and showing which countries should improve either their total emissions, emissions per capita, or emissions per kUSD.

In addition, the variable CE_2N is much more elastic than the total emissions; differences in the carbon equivalent units between countries are much less than the differences in Mt of CO_2. This means that, if CE_2N is implemented in international agreements, it could promote the development of environmental measures, as improving CE_2N values is simpler and it can be seen as more achievable than improving total emissions statistics.

The low elasticity of the 10 most emitting countries is because, as explained in Section 4.3., they emit huge amounts of Mt of GHG to the atmosphere and, consequently, the difference in the amount of gases emitted by them is large, which makes it very difficult to change their relative position in the new classification.

If this is translated to reality and policy measures, it means that, until countries such as China, the USA, or India, commit to climate change and develop real political measures to reduce their emissions, the efforts made by the rest of the countries will be useless. The obtained results are in line with those obtained in earlier studies: Paltsev et al. [78] pointed out that the participation or non-participation of China in global climate architecture can result in a 1.1 °C to 1.3 °C change by the end of the century. Brenton [79] and Liu et al. [80] also pointed out that the cooperation of the USA and China, as the two largest producers of GHG, is key to reaching sustainable development. Without the largest GHG emitters' involvement, ambitious global climate goals are vastly more difficult, if not impossible, to achieve. Only 10 countries out of 194 are responsible for more than 60% of the total CO_2 emissions. China is the biggest pollutant; it is responsible for 30% of global emissions, followed by the USA, with 13%, and India, with 7%. Without additional policy intervention, China's CO_2 emissions will continue to grow until 2040 or 2050 and will approximately double from their 2010 level (Liu et al., 2017 [80]).

CE_2N can also be used in other environmental models.

In the case of the Sagar model [31], the use of CE_2N and the removal of the historic component implies assigning more weight to the population, GDP, and emissions of a country. When allocating CO_2 emissions, the model gives greater percentages to countries that have high levels of production and that are densely populated, such as Japan or Germany, and penalizes countries with high levels of emissions, such as China (See Table 11). It benefits countries with good values of emissions efficiency; the more efficient they are, the more they can emit. In addition, the definition of the historic responsibility function is not entirely clear, so substituting

this function with CE_2N improves the transparency, clarity, and objectivity of the model.

In the case of the Vaillancourt model [32], the introduction of the parameter CE_2N helps to improve the clarity and objectivity of the model. Vaillancourt defined different weights and criteria that depended on the level of development of the countries and their results varied a lot depending on the criteria and weights applied. The introduction of CE_2N to the model ends the problem of the variability of the results with different criteria, unifies all these criteria, simplifies the model, and makes it more objective (see Table 12). The consequences of basing the Vaillancourt model on CE_2N are similar to the implementation of this parameter in the Sagar model; countries that have a good balance of total emissions, emissions per capita, and per GDP are benefited. Countries that have unacceptable values in any of these variables (countries in the red zone in Figure 2, Figure 3 and Figure 4) will be penalized, and the allocation of emissions will be much less for them.

Policy Recommendations

To address climate change, every country has to cooperate and leave nationalist approaches behind. There are many articles that study the relationship between globalization and GHG emissions [56,57,58,59]. However, none of them analyzed in depth the concept of globalization; previous articles analyzed the impacts that a global economy can have on the environment, but they did not mention that only the economy is global. Politics are not global, and here is the main problem. Thanks to globalization, nowadays, society has access to many services and goods that it would not have if the economy was only local. Therefore, the current global economic model is hard to change and nor is it the solution. What should be promoted is global politics. Emissions and climate change do not have borders, so to cut emissions and tackle climate change, global politics are needed; the final objective is that every state in the world reduces its emissions and that they cooperate to face global warming. Every country should develop a realistic and affordable plan adapted to its history and situation without looking at other countries or the disadvantages with respect to other states. Moreover, international cooperation is needed to push the most polluting countries to cut their emission, even if this has a negative impact on their economy. Without the commitment of the big economies (China, the USA, India, etc.), very little can be done, as they are responsible for more than 40% of total emissions.

Climate change is a global problem and, as a global problem, it needs a global answer. Emissions from China affect the inhabitants of Australia, and global and without borders politics and cooperation are the only solution. Another difficulty is that there are many interests involved, depending on the climate variable that is analyzed, and the statistics differ a lot. The capacity of manipulating information through the media added to the economic and political interests give very different speeches that may raise questions about the reality of climate change.

On the other hand, developed countries should help developing countries to reach their peak of emissions as soon as possible with the aim of being able to reduce their emissions as soon as possible (Environmental Kuznets Curve Theory [13,14,50,51,52,53]). They should promote technology and knowledge transfer, and help to implement the projects needed to cut emissions. Until developing countries implement sustainable policies, climate change would be hard to face. Countries such as Germany or Denmark (developed countries) have already reduced their emissions and, nowadays, developing countries have to improve.

In addition, the use of renewable energy, the enhancement of energy efficiency, and the promotion of the blue economy have been proven to help sustainable development [12,13,14,15,16,28,29,55,56,57,58]. Consequently, countries should focus on developing their economy from this base, avoiding policies that promote economic expansion at the price of environmental deterioration. There is a need to create policies that strike a balance between economic expansion, environmental measures, energy transition, and technology and knowledge equity.

In conclusion, climate change is a problem that can only be addressed with international cooperation, a global economy, and global politics. Effective measures can only be achieved if nationalities and ethnicities are forgotten and if nationalist approaches are left. The answer to this problem mostly relies on the most powerful countries and, until they cooperate, very little can be achieved in the climate change fight.

6. Conclusions

Climate change is a reality and the countries that are most affected are small island states or mountain communities that are only responsible for 0.1% of global emissions or less. All these societies could disappear in 100 years and they can do very little to stop climate change. Developed countries need to take action and

mobilize all of their resources, scientific knowledge, and technology to address this problem. The first step to facing the climate change problem is to analyze global emissions with the final aim of defining a universal criterion that could be applied to all countries and is seen as fair by all countries. Until this is achieved, effective international agreements will never be reached. In addition, it is the scientific community that should be responsible for the definition of this criterion, because it is the one with the necessary knowledge.

This article has helped to tackle this problem. It has studied the distribution of GHG emissions, the effectiveness of international agreements, and defined a new universal environmental criterion, the new parameter proposed, CE_2N.

The main contribution of this paper is the definition of the new environmental parameter, CE_2N. There are many environmental criteria, which leads to problems in both the scientific community and society and politics. As it has been recently seen, during COP26, world leaders do not agree on which environmental criteria should rule international agreements, which prevents the understanding between countries and progress on environmental policies. On the other hand, scientific environmental model results vary a lot depending on the criteria used. In this paper, we have addressed these two problems taking the three most important environmental variables: total emissions, and emissions per capita and per GDP, and unified them in a unique and universal parameter: CE_2N. The results obtained show a unique classification of the countries that can be used to implement environmental policies. This classification considers all of the environmental criteria used until now and proposes a new way to judge environmental policies in which emissions efficiency is considered. This new parameter benefits countries with average values of total emissions, and emissions per capita and per GDP (countries in the green area in the traffic diagrams presented, such as Poland, Italy, or the UK), while penalizes countries with unacceptable values of one of these variables (countries in the red area in the traffic diagrams, such as Kazakhstan or Australia). In addition, it is much more sensitive than total emissions, so improvements in environmental policies implemented by countries will have a direct impact on their CE_2N values, which can promote real environmental measures. Finally, the introduction of CE_2N into previous environmental models helps to increase the objectivity and clarity of the model. This paper also calls for a reflection; it points out, as a solution to climate change, a real global economy and politics, real interdisciplinary knowledge, and cooperation between social and technical disciplines.

Future research should focus on climate change equity, globalization, and the unification of environmental criteria. The results of environmental models vary a lot depending on the criteria used; this is the main reason why this research paper focuses on the development of a parameter that unifies all of these criteria in only one value. Future articles should keep studying how to define a unique criterion to quantify emissions all across the world.

References

1. Todorov, A.V. Reply. *J. Appl. Clim. Meteorol.* **1986**, *25*, 258–259.
2. Werndl, C. On Defining Climate and Climate Change. *Br. J. Philos. Sci.* **2016**, *67*, 337–364.
3. Lorenz, E. Climatic Change as a Mathematical Problem, *J. Appl. Meteorol.* **1970**, *9*, 325–329.
4. Dymnikov, V. and Gritsoun, A. Atmospheric Model Attractors: Chaoticity, Quasi-Regularity and Sensitivity, *Nonlinear Process. Geophys.* **2001**, *8*, 201–208.
5. UNFCCC. *United Nations Framework Convention on Climate Change 1992,* Article 1. Available online: Conveng.pdf (unfccc.int). (accessed on 5 February 2022).
6. IPCC. 2018: Annex I: Glossary In: Global Warming Of 1.5°C. An IPCC Special Report on The Impacts of Global Warming of 1.5°C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, In the Context of Strengthening The Global Response To The Threat of Climate Change, Sustainable Development, And Efforts To Eradicate Poverty; IPCC: Geneva, Switzerland, 2018. In Press.
7. Parry, M.L. The problem of climate in history—An editorial. *Clim. Chang.* **1982**, *4*, 107–110, doi:10.1007/BF00140584.
8. Zizzamia, D. Climate Change in Human History: Prehistory to the Present. *Environ. Hist.* **2018**, *23*, 912–914, doi:10.1093/envhis/emy082.
9. IPCC. *Climate Change 2014: Synthesis Report;* Cambridge University Press, Cambridge, UK, 2014, 151p.
10. United Nations General Assembly. Transforming Our World: The 2020 Agenda for Sustainable Development; United Nations General Assembly: New York, NY, USA, 2015.
11. Fonseca, L.M.; Domingues, J.P.; Dima, A.M. Mapping the Sustainable Development Goals Relationships. *Sustainability* **2020**, *12*, 3359, doi:10.3390/su12083359.
12. Bilan, Y.; Streimikiene, D.; Vasylieva, T.; Lyulyov, O.; Pimonenko, T.; Pavlyk, A. Linking between Renewable Energy, CO2 Emissions, and Economic Growth: Challenges for Candidates and Potential Candidates for the EU Membership. *Sustainability* **2019**, *11*, 1528, doi:10.3390/su11061528.
13. Zoundi, Z. CO2 Emissions, Renewable Energy and the Environmental Kuznets Curve, A Panel Cointegration Approach. Renew. *Sustain. Energy Rev.* **2017**, *72*, 1067–1075, doi:10.1016/j.rser.2016.10.018.
14. Liu, X.; Zhang, S.; Bae, J. The Impact of Renewable Energy and Agriculture on Carbon Dioxide Emissions: Investigating the Environmental Kuznets Curve in Four Selected ASEAN Countries. J. *Clean. Prod.* **2017**, *164*, 1239–1247, doi:10.1016/j.jclepro.2017.07.086.
15. Ziolo, M.; Jednak, S.; Savić, G.; Kragulj, D. Link between Energy Efficiency and Sustainable Economic and Financial Development in OECD Countries. *Energies* **2020**, *13*, 5898, doi:10.3390/en13225898.
16. Akaev, A.A.; Davydova, O.I. A Mathematical Description of Selected Energy Transition Scenarios in the 21st Century, Intended to Realize the Main Goals of the Paris Climate Agreement. *Energies* **2021**, *14*, 2558, doi:10.3390/en14092558.

17. Peña-Ramos, J.A.; del Pino-García, M.; Sánchez-Bayón, A. The Spanish Energy Transition into the EU Green Deal: Alignments and Paradoxes. *Energies* **2021**, *14*, 2535, doi:10.3390/en14092535.
18. Drożdż, W.; Kinelski, G.; Czarnecka, M.; Wójcik-Jurkiewicz, M.; Maroušková, A.; Zych, G. Determinants of Decarbonization—How to Realize Sustainable and Low Carbon Cities? *Energies* **2021**, *14*, 2640, doi:10.3390/en14092640.
19. Zimon, D.; Tyan, J.; Sroufe, R. Drivers of Sustainable Supply Chain Management: Practices for Alignment with UN Sustainable Development Goals. *Int. J. Qual. Res.* **2020**, *14*, 219–236, doi:10.24874/IJQR14.01-14.
20. Grossman, G.M.; Krueger, A.B. Environmental Impacts of a North American Free Trade Agreement. *Natl. Bur. Econ. Res.* **1991**, 3914, doi:10.3386/w3914.
21. Grossman, G.M.; Krueger, A. Economic Growth and the Environment. *Q. J. Econ.* **1995**, *110*, 353–377, doi:10.2307/2118443.
22. Douglas, H.E.; Selden, T.; Stoking the fires? CO2 Emissions and Economic Growth. *J. Public Econ.* **1995**, *1*, 85–101, doi:10.1016/0047-272701449-X.
23. Zhang, Z.Q.; Qu, J.S.; Zeng, JJ.A.; Quantitative comparison and analysis on the assessment indicators of greenhouse gases emission. *J. Geogr. Sci.* **2008**, *18*, 387–399, doi:10.1007/s11442-008-0387-8.
24. Moghaddam, R.F.; Moghaddam, F.F.; Cheriet, M. A modified GHG intensity indicator: Toward a sustainable global economy based on a carbon border tax and emissions trading. *Energy Policy*. **2013**,57:363-80, doi:10.1016/j.enpol.2013.02.012.
25. Liang, S.; Qu, S.; Zhu, Z.Q.; Guan, D.B.; Xu, M.; Income-Based Greenhouse Gas Emissions of Nations. *Environ. Sci. Technol.* **2017**, *51*, 346–355, doi:10.1021/acs.est.6b02510.
26. Arango-Miranda, R.; Hausler, R.; Romero-Lopez, R.; Glaus, M.; Ibarra-Zavaleta, S.P. Carbon Dioxide Emissions, Energy Consumption and Economic Growth: A Comparative Empirical Study of Selected Developed and Developing Countries. "The Role of Exergy". *Energies* **2018**, *11*, 2668, doi:10.3390/en11102668.
27. Sun, L.; Cui, H.; Ge, Q.; Adenutsi, C.D.; Hao, X. Spatial Pattern of a Comprehensive fE Index for Provincial Carbon Emissions in China. *Energies* **2020**, *13*, 2604, doi:10.3390/en13102604.
28. Leitão, N.C.; Lorente, D.B. The Linkage between Economic Growth, Renewable Energy, Tourism, CO2 Emissions, and International Trade: The Evidence for the European Union. *Energies* **2020**, *13*, 4838, doi:10.3390/en13184838.
29. Bădîrcea, R.M.; Manta, A.G.; Florea, N.M.; Puiu, S.; Manta, L.F.; Doran, M.D. Connecting Blue Economy and Economic Growth to Climate Change: Evidence from European Union Countries. *Energies* **2021**, *14*, 4600, doi:10.3390/en14154600.
30. Karman, A.; Miszczuk, A.; Bronisz, U. Regional Climate Change Competitiveness—Modelling Approach. *Energies* **2021**, *14*, 3704, doi:10.3390/en14123704.
31. Sagar, AD. Wealth, responsibility, and equity: Exploring an allocation framework for global GHG emissions. *Clim. Change.* **2000**, *45*, 511–527, doi:10.1023/a:1005616413549.
32. Vaillancourt, K.; Waaub, J.P. Equity in international greenhouse gases abatement scenarios: A multicriteria approach. *European Journal of Operational Research*. **2004**, *153*, 489–505, doi:10.1016/s0377-221700170-x.
33. Markandya, A. Equity and Distributional Implications of Climate Change. *World Dev.* **2011**, 39, 1051–1060, doi:10.1016/j.worlddev.2010.01.005.
34. Mattoo, A.; Subramanian, A. Equity in Climate Change: An Analytical Review. *World Dev.* **2012**, *40*, 1083–1097, doi:10.1016/j.worlddev.2011.11.007.
35. Remuzgo, L.; Trueba, C.; Sarabia, J.M. Evolution of the global inequality in greenhouse gases emissions using multidimensional generalized entropy measures. *Phys. A-Stat. Mech. Its Appl.* **2016**, *444*, 146–157, doi:10.1016/j.physa.2015.10.017.

36. Alcaraz, O.; Buenestado, P.; Escribano, B.; Sureda, B.; Turon, A.; Xercavins, J. Distributing the Global Carbon Budget with climate justice criteria. *Clim. Change.* **2018**, *149*, 131–145, doi:10.1007/s10584-018-2224-0.
37. Alcaraz, O.; Sureda, B.; Turon, A.; Ramírez, C.; Gebellí, M. Equitable mitigation to achieve the 1.5 °C goal in the Mediterranean Basin. *Clim. Chang.* **2021**, *165*, 62, doi:10.1007/s10584-021-03070-8.
38. Jena, P.R.; Managi, S.; Majhi, B. Forecasting the CO2 Emissions at the Global Level: A Multilayer Artificial Neural Network Modelling. *Energies* **2021**, *14*, 6336, doi:10.3390/en14196336.
39. Dubash, N.K. Toward a progressive indian and global climate politics. New Delhi: Centre for Policy Research Climate Initiative 2009, Working paper, Available online: Toward a progressive Indian and global climate politics | Eldis (accessed on 22 March 2021).
40. Saran, S. Global governance and climate change. *Glob. Gov. Rev. Multilater. Int. Inst.* **2009**, *15*, 457–460, doi:10.1163/19426720-01504006.
41. Campbell-Lendrum, D.; Corvalan, C. Climate Change and Developing-Country Cities: Implications for Environmental Health and Equity. *J. Urban Health-Bull. New York Acad. Med.* **2007**,84, I109-I17, doi:10.1007/s11524-007-9170-x.
42. Baležentis, T.; Liobikienė, G.; Štreimikienė, D.; Sun, K. The impact of income inequality on consumption-based greenhouse gas emissions at the global level: A partially linear approach. *J. Environ. Manag.* **2020**, 267, 110635, doi:10.1016/j.jenvman.2020.110635.
43. Formetta, G.; Feyen, L. Empirical evidence of declining global vulnerability to climate-related hazards. *Glob. Environ. Change.* **2019**, *57*, 101920, doi:10.1016/j.gloenvcha.2019.05.004.
44. Cao, J. Reconciling Human Development and Climate Protection: Perspectives from Developing Countries On Post-2012 International Climate Change Policy. Harvard Kennedy School of Government Discussion Paper 08-25. Harvard University: Cambridge, MA, USA. Available online: http://belfercenter.ksg.harvard.edu/publication/18685/ (accessed on 24 March 2021).
45. Cooper, R.N. The case for charges on greenhouse gas emissions. Massachusetts: Cambridge, Massachusetts, USA: Harvard Project on Climate Agreements, 2008. Available online: The Case for Charges on Greenhouse Gas Emissions | Harvard Environmental Economics Program (accessed on 24 March 2021).
46. Stern, N. The Global Deal. Climate Change and The Creation of a New Era Of Progress And Prosperity; Public Affairs: New York, NY, USA, 2009, ISBN 978-1433265365.
47. Kanitkar, T. et al; Carbon budgets for climate change mitigation—A GAMS-based emissions model. *Curr. Sci.* **2013**, *104*, 1200–1206. Available online: https://www.researchgate.net/publication/281229901_Carbon_budgets_for_climate_change_mitigation_-_a_GAMS-based_emissions_model (Accessed on 26/03/21)
48. Shue, H. The International Politics of the Environment: Actors, Interests, and Institutions; Hurrell, A.; Kingsbury, B, Eds; Clarendon Press: Oxford, UK, 1992. pp. 373–397, ISBN: 9780198273653.
49. Falkner, R. The unavoidability of justice—And order—In international climate politics: From Kyoto to Paris and beyond. *Br. J. Politics Int. Relat.* **2019**, *21*, 270–278, doi:10.1177/1369148118819069.
50. Stern, D.I. *The Environmental Kuznets Curve*; International Society for Ecological Economics Internet Encyclopaedia of Ecological Economics: Boston, MA, USA, 2003.
51. Dogan, E.; Inglesi-Lotz, R. The impact of Economic Structure to Environmental Kuznets Curve (EKC) Hypothesis: Evidence from European Countries. *Environ. Sci. Pollut. Res.* **2020**, *27*, 12717–12724, doi:10.1007/s11356-020-07878-2.
52. Mania, E. Export Diversification and CO2 Emissions: An Augmented Environmental Kuznets Curve. *J. Int. Dev.* **2020**, *32*, 168–185, doi:10.1002/jid.3441.

53. Jóźwik, B.; Gavryshkiv, A.-V.; Kyophilavong, P.; Gruszecki, L.E. Revisiting the Environmental Kuznets Curve Hypothesis: A Case of Central Europe. *Energies* **2021**, *14*, 3415, doi:10.3390/en14123415.
54. Knight, K.W.; Schor, J.B. Economic Growth and Climate Change: A Cross-National Analysis of Territorial and Consumption-Based Carbon Emissions in High-Income Countries. *Sustainability* **2014**, *6*, 3722–3731. doi:10.3390/su6063722.
55. Rahman, M.M. Environmental degradation: The role of electricity consumption, economic growth, and globalization. *J. Environ. Manag.* **2020**, *253*, 109742, doi:10.1016/j.jenvman.2019.109742.
56. Villanthenkodath, M.A.; Ansari, M.A.; Shahbaz, M.; Vo, X.V. Do tourism development and structural change promote environmental quality? Evidence from India. *Environ. Dev. Sustain.* **2021**, *4*, 1–32, doi:10.1007/s10668-021-01654-z.
57. Baydoun, H.; Aga, M. The Effect of Energy Consumption and Economic Growth on Environmental Sustainability in the GCC Countries: Does Financial Development Matter? *Energies* **2021**, *14*, 5897, doi:10.3390/en14185897.
58. Pata, U.K. Linking renewable energy, globalization, agriculture, CO2 emissions and ecological footprint in BRIC countries: A sustainability perspective. *Renew. Energy* **2021**, *173*, 197–208, doi:10.1016/j.renene.2021.03.125.
59. Kihombo, S.; Ahmed, Z.; Chen, S.; Adebayo, T.S.; Kirikkaleli, D. Linking financial development, economic growth, and ecological footprint: What is the role of technological innovation? *Environ. Sci. Pollut. Res.* **2021**, 1–11, doi:10.1007/s11356-021-14993-1.
60. Xu, Z.; Baloch, M.A.; Danish Meng, F.; Zhang, J.; Mahmood, Z. Nexus between financial development and CO2 emissions in Saudi Arabia: Analyzing the role of globalization. Environ. *Sci. Pollut Res.* **2018**, *25*, 28378–28390, doi:10.1007/s11356-018-2876-3.
61. Salahuddin, M.; Alam, K.; Ozturk, I.; Sohag, K. The effects of electricity consumption, economic growth, financial development and foreign direct investment on CO2 emissions in Kuwait. *Renew. Sustain. Energy Rev.* **2018**, *81*, 2002–2010, doi:10.1016/j.rser.2017.06.009.
62. Chebbi, H.E. Long and Short–Run Linkages between Economic Growth, Energy Consumption and CO2 Emissions in Tunisia. *Middle East Dev. J.* **2010**, *2*, 139–158, doi:10.1142/S1793812010000186.
63. Jiang, C.; Ma, X. The Impact of Financial Development on Carbon Emissions: A Global Perspective. *Sustainability* **2019**, *11*, 5241, doi:10.3390/su11195241.
64. Ameyaw, B.; Yao, L. Analyzing the Impact of GDP on CO2 Emissions and Forecasting Africa's Total CO2 Emissions with Non-Assumption Driven Bidirectional Long Short-Term Memory. *Sustainability* **2018**, *10*, 3110, doi:10.3390/su10093110.
65. Naciones Unidas. Convención Marco de las Naciones Unidas sobre el Cambio Climático; New York, NY, USA, 1992, 25p.
66. Ministerio para la Transición ecológica y el reto demográfico. Protocolo de Kioto. Available online: Protocolo de Kioto (miteco.gob.es). (accessed on 20 January 2021).
67. UNFCC. Doha Climate Change Conference—November 2012. Available online: Doha Climate Change Conference - November 2012 | UNFCCC(accessed on 24 January 2021).
68. Ministerio para la Transición ecológica y el reto demográfico. Acuerdo de París. Available online: Principales elementos del Acuerdo de París (miteco.gob.es) (accessed on 31 January 2021).
69. UNFCC. The Clean Development Mechanism. Available online: The Clean Development Mechanism | UNFCCC (Accessed on 22/01/21).
70. UNFCC. Joint Implementation. Available online: Joint implementation | UNFCCC. Available online: Joint implementation | UNFCCC (accessed on 22 January 2021).
71. Crippa, M.; Guizzardi, D.; Muntean, M.; Schaaf, E.; Solazzo, E.; Monforti-Ferrario, F.; Olivier, J.G.J.; Vignati, E. Fossil CO2 Emissions of All World Countries—2020 Report,

EUR 30358 EN. Publications Office of the European Union; Luxembourg, 2020, ISBN 978-92-76-21515-8, doi:10.2760/143674.
72. Kim, Y.; Tanaka, K.; Matsuoka, S. Environmental and economic effectiveness of the Kyoto Protocol. *PLoS ONE* **2020**, 15, doi:10.1371/journal.pone.0236299.
73. Almer, C.; Winkler, R. Analyzing the effectiveness of international environmental policies: The case of the Kyoto Protocol. *J. Environ. Econ. Manag.* **2017**, 82,125–151, doi:10.1016/j.jeem.2016.11.003.
74. Grunewald, N.; Martínez-Zarzoso, I. Carbon dioxide emissions, economic growth and the impact of the Kyoto Protocol. *Working Paper for Spanish Ministry of Education and Science*; Spanish Ministry of Education and Science: Madrid, Spain, 2011; pp: 1–26.
75. Grunewald, N.; Martinez-Zarzoso, I. Did the Kyoto Protocol fail? An evaluation of the effect of the Kyoto Protocol on CO2 emissions. *Environ. Dev. Econ.* **2016**, *21*, 1–22, doi:10.1017/S1355770X15000091.
76. Maamoun, N. The Kyoto protocol: Empirical evidence of a hidden success. *J. Environ. Econ. Manag.* **2019**,95:227-56, doi:10.1016/j.jeem.2019.04.001.
77. Thakur, S. From Kyoto to Paris and Beyond: The Emerging Politics of Climate Change. India *Q. -A J. Int. Affairs.* **2021**; 77, 366–383, doi:10.1177/09749284211027252.
78. Paltsev, S.; Morris, J.; Cai, Y.; Karplus, V.; Jacoby, H. The role of China in mitigating climate change. *Energy Econ.* **2012**, *34*, S444-S50, doi:10.1016/j.eneco.2012.04.007.
79. Brenton, S.A. 'Great Powers' in climate politics. *Clim Policy*, **2013**, *13*, 541–546, doi:10.1080/14693062.2013.774632.
80. Liu, Q.; Gu, A.; Teng, F.; Song, R.; Chen, Y. Peaking China's CO2 Emissions: Trends to 2030 and Mitigation Potential. *Energies.* **2017**, *10*, 209, doi:10.3390/en10020209.

3.3. Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain

Abstract and Keywords

Climate change is one of the issues of greatest concern to today's society. The increase in temperatures has affected sea levels, polar masses and extreme events, among others. There are many scientific studies that analyze the impacts of climate change on coastal communities, but most of them focus on beach erosion and coastal recession. Scientific literature on the effects of climate change on ports and harbors and their adaptation is much less abundant. Ports are essential for the economy and society of their cities, so studying the impact of climate change on them is an urgent need. The Mediterranean and the Spanish Mediterranean coast is one of the areas that will be most affected by climate change in the future. In addition, the Spanish economy depends a lot on its tourism and, thus, on its coastal cities. Therefore, the study of the impact of climate change on Spanish ports and coastal communities is essential. This article presents a review of the studies carried out until now on the effects of climate change on Spanish ports, and it identifies research gaps and weaknesses and suggests new research lines.

Keywords: climate change adaptation; Mediterranean Coast; sea-level rise; ports adaptation; Spanish coast adaptation

1. Introduction

Climate change has always existed, but since the industrial revolution the amount of greenhouse gases (GHG) emitted to the atmosphere has increased considerably. The greenhouse effect is a natural effect and without it, Earth would have freezing temperatures (around 33 °C less than today). When humans emit greenhouse gases, what they are producing is an "overeffect" that causes global warming. There are four main greenhouse gases: carbon dioxide (CO_2) (accounting for 80% of the emissions caused by humans), methane (CH_4), oxide of nitrogen (N_2O) and fluorinated gases [1].

The increase of emissions of these GHG is causing global warming, which is affecting the ocean and cryosphere. This increase of temperatures is provoking sea level rise (SLR) due to the thermal expansion of the ocean and the thaw of polar regions and high mountain areas. Another consequence of the rising temperatures

is the increment of both the magnitude and frequency of extreme events. This is causing, at the same time, negative impacts in water quality, ecosystems and cultures of different ethnicities, great damages in human infrastructures and social and economic losses [2–6].

To avoid climate change effects there are two main strategies: mitigation and adaptation. Mitigation strategies focus on the reduction of emissions, while adaptation strategies focus on the adjustment of the existing systems to the expected effects caused by these emissions [7]. Even if the world stopped emitting GHG today, the sea will continue rising and extreme events will continue increasing. Therefore, defining proper and adequate adaptation plans is key for the future.

One of the systems most threatened by global warming is coastal systems. A significant percentage of world's population and infrastructures are located next to the coast; around 40% of global population lives within 100 km (km) of the coast. In addition, in 2050 this value is expected to increase to 75% [8,9]. On the other hand, the im-pact of climate change will not be the same in all countries. There are areas that will be most affected by the effects of climate change and one of these areas is the Mediterranean basin. It has been calculated that temperatures in the Mediterranean will increase by 1.4 °C with respect to preindustrial levels, which is 20% more than in the rest of the world [10]. The Mediterranean coast is located between the North of Africa, whose climate is arid and dry and central Europe, which has a rainy and humid climate. Therefore, it is affected by atmospheric and oceanic interactions, which makes it very vulnerable to climate change [11]. In fact, it has been defined as one of most worrying hotspots when referring to the climate change problem [12]. Therefore, scientific literature on climate change and the Mediterranean is extensive.

However, most of the literature about climate change and coastal management focuses on the effect of climate change on shoreline evolution. Some examples are Mentaschi et al., 2018 [13] who concluded that a total of 28,000 km^2 of land has been lost due to coastal erosion or Luijendijk et al., 2018 [14] that found that one out of four beaches in the world is eroding.

Although many scientists studied the effect of climate change in beaches and coastal ecosystems, there is much less research on the effects that climate change will have on ports and their operability. Seaports are one of the infrastructures most threatened by climate change, and given their key role in global supply

chains, adapting and protecting ports against climate change is crucial for the future of our society [15]. Navigation is one of the main modes of transport, especially in the transport of goods; 80–90% of goods worldwide are traded by sea [16,17], and it is forecasted to annually increase by 4% over the next five years [18]. Therefore, to avoid economic and social losses, it is crucial to study the impact that climate change will have on ports and to propose adaptation plans.

The aim of this article is to collect up-to-date knowledge about how climate change will affect the port infrastructure of the Spanish Mediterranean. For the first time, a regional, temporal, methodological and thematic analysis has been carried out. It has been identified in which areas of the Spanish Mediterranean there is a lack of specific studies. The main methodologies used in these studies have also been analyzed and compared. In addition, a thematical analysis of these studies has been conducted, identifying which topics have been the most dealt with in scientific literature and which are the most important for the future and need more attention and resources, as well as what future lines of research should be focused on.

2. Materials and Methods

For the present research, a structured literature review based on the methodology of Adebisi et al., 2019 [19] was conducted. As explained in Adebisi et al., 2019 [19], the structured literature review has three stages: planning, review and information syn-thesis. In the first stage, planning, search criteria and databases were defined. The databases used were mainly Web of Science (WOS), Scopus and PubMed. The search criteria was the topic of the articles, and the keywords used were "Levante climate change," "Mediterranean climate change" and "Spain climate change." In the second stage of the review, the articles were searched and filtered. A total of 14 articles coincided with the search of "Levante climate change," 9875 with "Spain climate change" and 30,896 with "Mediterranean climate change."

The distribution of the major concepts of all the studies obtained is presented in Figure 1. The great majority of studies were related to environmental sciences, followed by climatology, ecology environmental sciences, marine ecology and computational biology.

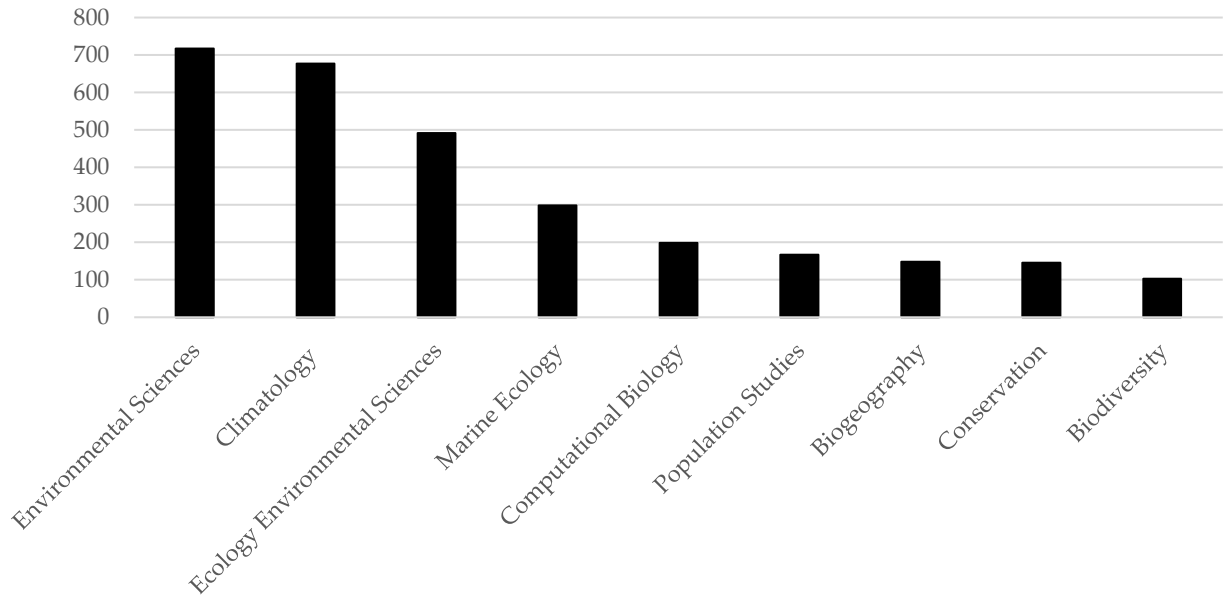


Figure 1. Major concepts of the studies from WOS.

All these articles were filtered by study area, selecting only the ones related to oceanography, maritime infrastructures and ocean and maritime engineering, which gave a total of 272 articles. Further, 21 articles were about Mediterranean climate, 63 about SLR and flooding, 41 about coastal erosion and shoreline projections, 32 about extreme events and 115 about adaptation and vulnerability (Figure 2). Finally, in the third stage, all the information of these 272 articles was analyzed and synthesized.

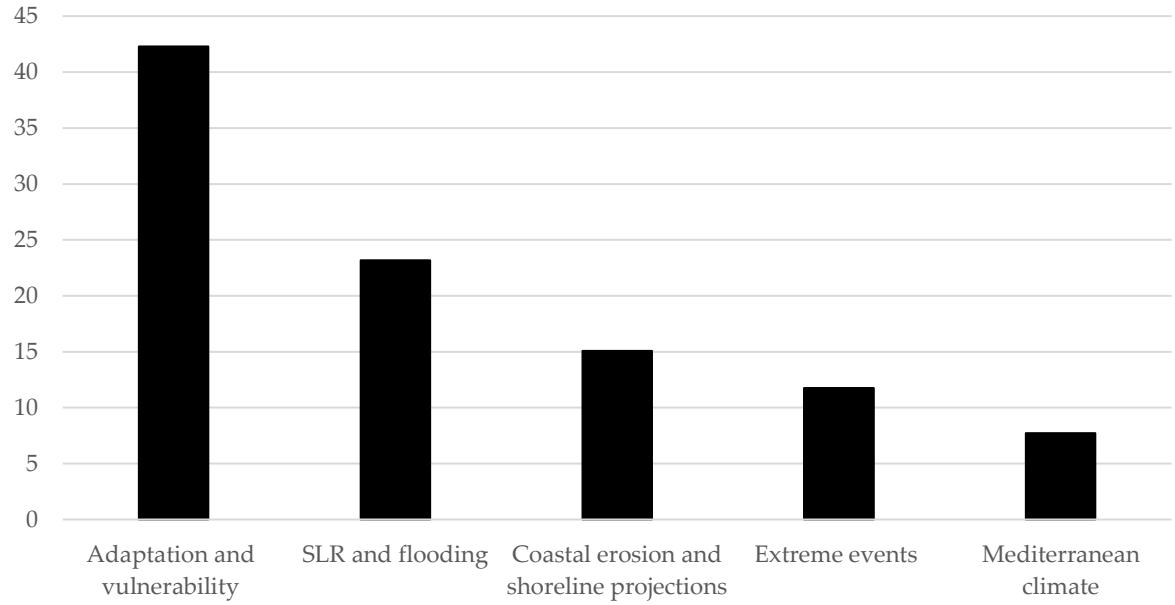


Figure 2. Classification of reviewed articles by topic.

3. Results and Discussion

3.1. Sea Level Rise (SLR) and Flooding

As was explained in Section 1, one of the direct consequences of climate change and the increment of Earth's temperature is the increase of sea level (SL). Total water level of the sea can be divided into mean sea level (MSL), wave setup and storm surge [20]. An increase of any of these three components will cause an increase in SL.

The elements that cause SL changes are usually divided into mass component (due to mass changes, such as ice thaw) and steric component (due to density changes); the steric component can also be divided into thermosteric and halosteric components [21].

All the causes of changes in SL can be summed up in: ice sheets (Greenland, West Antarctic and East Antarctic), glacier and ice cap, global mean thermal expansion, land water storage, regional ocean steric and dynamic effects, surface mass balance, glacial isostatic adjustment (GIA), sediment compaction and tectonics [22].

This increase can cause negative impacts on environmental systems, social communities, infrastructures, etc. Therefore, much scientific effort has been put into its estimation. To know the MSL of a region, there are different databases, and the most used one is the Permanent Service for Mean Sea Level (PSMSL) [23]. Apart from SL databases, scientific literature about SLR is abundant. The number of studies of SLR has increased since 1997 due to the necessity to develop climate change adaptation strategies. Figure 3 presents the evolution in time of the number of studies of SLR presented in this section (Table 1).

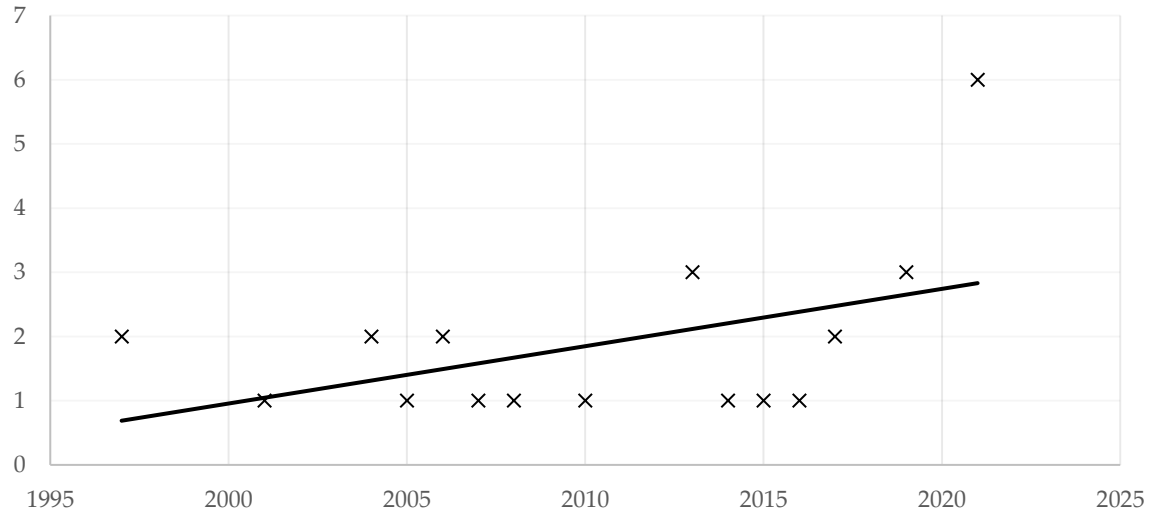


Figure 3. Evolution in time of the number of studies concerning SLR.

For the twentieth century, the value of global SLR estimated by most research is 1–2 mm/year [24–26]. For the period 1993–2010 the global SLR estimated is around 3 mm/year [27–30]. Apart from estimating SLR from altimetry measures, tide gauges, satellites, etc., it is important to project future SLR. Some of the most relevant estimations and projections of SLR from the scientific literature are presented in Table 1 [31].

In general, the global SLR projection has increased with time. Studies between 2005–2015 estimated a SLR around 75 cm for 2100 under RCP8.5, while studies between 2015–2020 estimated a SLR around 85 cm.

Although global SL estimations are important, regional variability is huge. Vacchi et al., 2021 [32] studied this variability along the central and Western Mediterranean with a Bayesian spatio-temporal statistical model with 401 sea level index points and pointed out that, to study the effects of SLR, regional studies are necessary.

Some of the most important regional studies in Spain are presented in Table 1. Apart from these studies, Marcos et al., 2021 reconstructed sea-level historic series in Alicante and Santander [33] and studied the SLR of Cadiz during the period 1880–2018 [34]; they obtained different values of SLR (0.7 mm/year) from those of Vargas-Yanez, 2021 [35] (1.28 mm/year); these differences may be due to the influence of the methodology followed to fill the measure gaps. Table 1 also presents regional studies that project SLR for the future.

Table 1. SLR review

Author	Region	Time Frame	Scenario	SLR	Methodology
Peltier, 2001 [26]	Global	20th century	Present estimations	1–2 mm/year	Tide gauges and GIA model
Miller and Douglas, 2004 [25]					Tide gauges with observations of temperature and salinity
Church and White, 2006 [24]				1.1 mm/year	Tide gauges
Church and White, 2006 [28]		1993–2010		3.1 mm/year	Satellite altimeter and island sea-level measurements
Hay et al., 2015 [29]		1993–2010		3 mm/year	Tide gauges and physics-based and model-derived signals (Probabilistic)
Tsimpilis et al., 2013 [30]		1993–2011		2.8–3 mm/year	Altimetry, temperature, salinity and gravity measurements
Cazenave et al., 2004 [27]		1993–2003		2.8–3.1 mm/year	Satellite altimeter
IPCC 2013 [36]		2100	RCP8.5	53–97 cm	Numerical models
IPCC 2019 [6]				61–110 cm	Numerical models
IPCC 2021 [37]				63–101 cm (SSP5–8.5)	Numerical models
Rahmstorff et al., 2007 [38]				50–140 cm	Semi-empirical model
Horton et al., 2014 [39]				50–150 cm	Experts survey
Kopp et al., 2016 [40]				52–131 cm	Statistical synthesis of regional sea-level reconstructions
Mengel et al., 2016 [41]				57–131 cm	Semi-empirical model
Bamber et al., 2019 [42]				21–163 cm	Structured expert judgement (Probabilistic)
Ibañez et al., 1997 [43]	Ebro Delta	20th century	Present estimations	3 mm/year	Measure rates of sedimentation, accretion,

					vertical elevation and subsidence
Jiménez et al., 1997 [44]					Tide gauges and Bruun's rule
Vázquez et al., 2021 [45]	Cádiz	1960–2020		3.5 mm/year	Tide gauges, linear regression and global records
Vargas-Yanez, 2021 [35]		1880–2018		1.28 mm/year	Tide gauges and statitical linear models
Marcos et al., 2021 [34]				0.7 mm/year	Tide gauges and near records
Chust et al., 2019 [46]	Basque Coast	2nd half of 20th century		2–2.5 mm/year	Tide gauges and LiDAR (altimetry)
Marcos et al., 2005 [47]					Tide gaguges and Empirical Orthogonal Function analysis
Leorri et al., 2008 [48]					Foraminifera-based transfer function
Vousdoukas et al., 2017 [49]	Europe	2050	RCP8.5	21–24 cm	Numerical models
Luque et al., 2021 [20]	Balearic			18–36 cm	Numerical models
Chust et al., 2010 [50]	Basque Coast	2100		28.5–48.7 cm	Temperature projections and LiDAR (altimetry)
Vitousek et al., 2017 [51]	Europe			53–77 cm	Numerical models
Luque et al., 2021 [20]	Balearic			46–103 cm	Numerical models

As it can be seen from the table above, most SL predictions are global, but to de-velop efficient climate change studies these global estimations must be translated to regional studies. Figure 4 shows the number of regional studies of SLR in Spain; the colour green means that a considerable number of regional studies have been con-ducted and the colour red that zero studies have been conducted. Most SLR studies of the Mediterranean coast are conducted in the Maresme and Delta del Ebro areas because they are the most problematic zones on this coast. However, the rest of the Mediterranean coast is also a hotspot for climate change and has a

need of specific regional studies to accurately predict SLR. In addition, as it can be seen from Table 1, all the studies obtained similar results, except those by Marcos et al. [34] and Vargas-Yanez et al. [35], whose differences could be due to the different methodologies applied.

Referring to the methodology followed to estimate SLR, a large number of the studies presented used tide gauges records with different supplementary techniques. In previous years, numerical models have been the most used methodology to predict SLR because they have improved enormously and have become one of the most accurate tools to forecast sea level rise. Figure 5 shows the main methodologies used to estimate SLR.



Figure 4. Number of SLR studies in the coast of Spain.

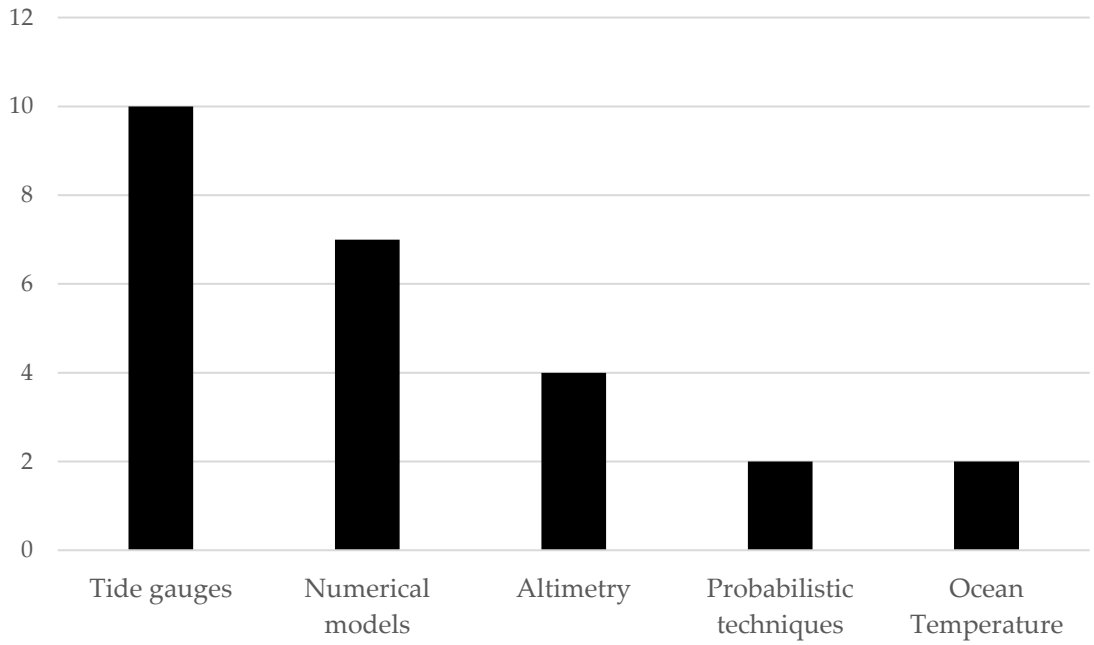


Figure 5. Frequency of methodologies used to estimate SLR.

One of the main direct impacts of SLR is the flooding of coastal cities, ports, dunes and any coastal system. Hence, many scientists have studied the flooding of different areas.

Hinkel et al., 2014 [52] studied flood damage globally and highlighted that under a SLR of 25–123 cm, 0.2 to 4.6% of the population would be flooded annually in 2100, and annual losses of Gross Domestic Product (GDP) would be of 0.3 to 9.3%. Focusing on the Mediterranean coast, Jorda et al., 2013 [21] studied the Mediterranean basin and predicted an increase of its salinity due to the increase of evaporation and decrease of precipitation. Galassi et al., 2014 [53] also studied SLR projections in the Mediterranean by 2050 and concluded that in the near future the mass component will clearly exceed the steric component.

In Spain, Guil-Guirado et al., 2021 [54] conducted a study on the Spanish Mediterranean coast and concluded that the worst seasons for flooding were autumn and winter and the worst months September to November. The cities most affected were Valencia, Barcelona and Alicante, and the worst sector was the tertiary sector. Pérez-Morales et al., 2018 [55] stated that the increase in flood exposure in Murcia and Alicante was linked to two construction peaks, one between 1978–1982 and the other between 1997–2007. Two years later, Ribas et al. [56] reached the same conclusion for Catalonia, Valencia, Alicante and Murcia, and stated that the urbanization in areas with a high risk of flooding significantly increased the exposure of these cities.

Chust et al., 2010 [50] studied the effects of SLR on the Basque coast and pointed out that 15.6% of sandy beaches will be affected by 2099.

Vinet et al., 2019 [57] mapped the mortality related to floods in Catalonia and the Balearic Islands and obtained a fatality rate of 0.407 and 0.623, respectively. This rate means that one person per million people dies every 1/F years due to floods.

Sayo et al., 2018 [58] and Grasses et al., 2020 [59] studied the effects of SLR in the Ebro Delta. López-Dóriga et al., 2020 [60] developed a pseudo-dynamic to evaluate the response capacity of active shorelines to Relative Sea Level Rise (RSLR) and applied it to low-lying coastal areas of Catalonia.

Ballesteros et al., 2018 conducted a multicomponent flood-risk assessment of the Maresme coast [61]; they developed a source-pathway-receptor-consequence model and concluded that in this area, the risk of flooding was low, except in some areas such as the Tordera delta.

Finally, Hernández-Mora et al., 2021 [62] developed a flood model using LISFLOOD and S-Beach in Tossa de Mar, Catalonia.

As happens with the prediction of SLR, most of the studies about the impact of SLR are focused on very specific points of the Spanish Mediterranean coast, more specifically in Catalonia, and they are mainly the Maresme coast and the Delta del Ebro hotspot. Figure 6 illustrates this lack of studies of the rest of the Mediterranean coast.

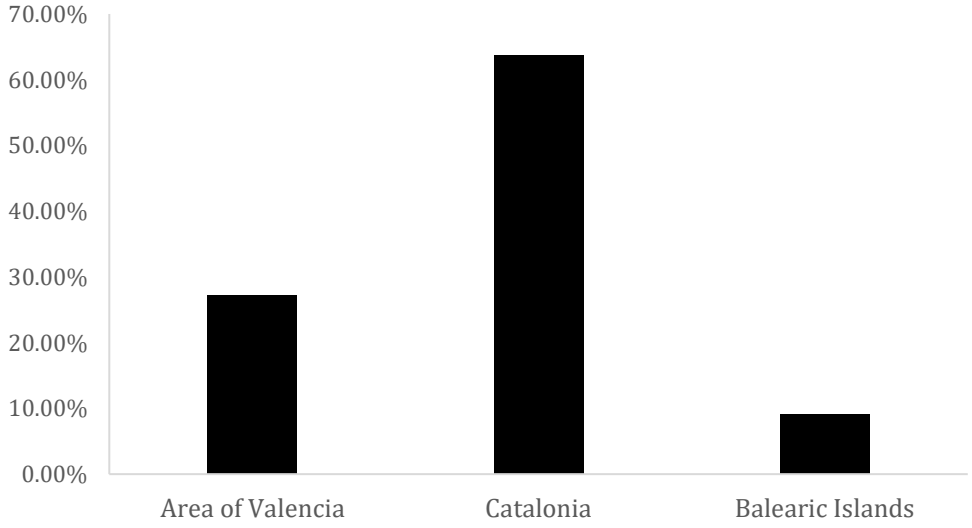


Figure 6. Distribution of studies of SLR impacts in the Spanish Mediterranean coast.

3.2. Impacts on Ports and and Harbours and the Need to Adapt

Coastal cities and ports are key for the current society and global economy. Their importance as economic hubs is increasing and they are becoming places where large parts of the population congregate and create economic and social value. In fact, in 2005, thirteen out of the twenty most populated cities in the world were port cities [63]. Ports also play a strategic role for economic growth and development in all scales, whether global, regional or local. They are also key transportation nodes for goods, they link local and national supply chains to global markets and are responsible for 80–90% of transportation of goods [64]. Therefore, if ports are affected by any problem that prevents their operation, there will be significant social and economic consequences [65]. Protecting them and developing proactive plans is essential and is a current priority.

Port cities will be one of the most affected systems under high-emissions scenarios and under climate change effects. Nursey-Bray et al., 2013 [66] affirmed that there are five main areas in ports that will be affected by climate change: Infrastructure, environment, population, safety and occupational health, and supply chains. Nicholls et al., 2008 [63,67] studied 136 port cities with more than one million inhabitants and assessed that in 2005 about 40 million people were threaten by a 100-year return-period coastal flood and that a total of 3000 billion USD was also exposed. This means that around 5% of global GDP is exposed to coastal flooding and this exposure could grow by the 2070s to 35,000 billion USD, which means that a 9% of projected global GDP will be at risk.

As it has been said in Section 1. Introduction, despite ports being one of the systems most affected by climate change, the majority of scientific studies focus on erosion and flooding events, not on the effects that climate change can have on port cities or the derived consequences. In this section, a review of the most relevant studies about the impact of climate change in ports is presented.

Global studies

Globally, Australia, USA and Japan are pioneers in studying the effects of climate change in port cities. Australian and American authors mostly focused on developing manuals to define adaptation strategies and much of their work was done between 2010–2015. On the other hand, Japanese authors focused on structure stability. Some of the most relevant studies are explained below.

In America, Haveman and Shatz, 2006 [68] compiled a standard reference manual for understanding the issue of port security in the USA and particularly in

California. Messner et al., 2013 [69] developed a methodology to evaluate vulnerability and risk in San Diego Port. In Australia, Nursey-Bray et al., 2013 [66] studied vulnerabilities and the adaption of Australian ports and Mcevoy et al., 2013 [70] wrote a document to test and refine assessment methodologies. In Japan, Takagi et al. [71–73] and Mase et al., 2013 [74] studied the stability of breakwaters under new climate change conditions (SLR and increase in wave heights and storm surge).

Becker et al. [75,76] presented the effects of climate change in ports at a global scale and conducted a survey to detect port risks. They specifically highlighted the need for more awareness about the complex implications that climate change may have on ports and related transport networks. Finally, Mutombo et al., 2017 [77] conducted a survey to analyze port risk over 29 countries.

In Europe, Prahl et al., 2018 [78] analyzed the economic loss due to coastal flood height in 600 European cities, and the adaptation costs that it will have. They did this by defining damage and protection cost curves. After this, Abadie et al., 2019 [79] studied the distribution of the damage curves using the Generalized Extreme Value (GEV) distribution model. Another relevant study at the European level is one conducted by Reckien et al., 2015 [80], who exanimated the barriers and drives of adaptation and mitigation measures. They pointed out the unemployment rate of a city and low values of GDP as key barriers to the development of urban plans. Christodoulou et al., 2019 [81] also studied the impacts of SLR in European ports and concluded that in the Black Sea and the Mediterranean, the effects of climate change will be less aggressive, but much more frequent.

The problem of these general studies is that climate change impacts have a huge regional variability and to develop an efficient adaptation pathway, each case must be studied in detail and no general rules can be applied. Therefore, all these studies must be complemented with regional and site-specific studies, and to apply all the general guidelines, they have to be adapted to each case.

Spanish studies

Focusing on Spain and in the Mediterranean, the most relevant studies are presented below.

Sánchez-Arcilla et al., 2016 [82] studied the effects of climate change on the Mediterranean. They noted that SLR will affect the water depths of ports and inevitably affect wave propagation patterns, and it may have consequences on infrastructure stability.

Sierra et al., 2015 [83] studied harbor agitation, and the results obtained showed a gentle descent in harbor agitation in the majority of the ports of the Catalan coast. In 2017 [84], they modelled the impact of climate change on Barcelona and concluded that it will remain safe and maintain operability levels under climate change. SLR represents a small percentage of the water depth of Barcelona port (around 10%). Therefore, the relative increase is small, and it will not significantly affect wave patterns and diffraction coefficients. Casas-Prat and Sierra [85,86] also conducted studies on the Catalan coast in which they analyzed the trend of wave storminess and direction. They highlighted the importance of also including directional studies in climate change assessments because most coastal infrastructures are designed with directional records of wave height, and changes in the direction of the waves could affect their stability.

León-Mateos et al., 2021 [15] criticized the lack of quantitative and validated studies in port adaptation research. To address this problem, they defined a Port Resilience Index and validated it using the port of A Coruña. López et al., 2015 [87] applied artificial neural networks (ANNs) to assess operability in the port of Ferrol.

The studies above presented are focused on the impacts that climate change can have on ports. However, another important part of research is to study possible adaptation measures to mitigate these impacts. The United Nations Conference on Trade and Development (UNCTAD) [88] stated in the Meeting on Climate Change and Adaptation that adaptation was critical, and that adaptation and mitigation should be dealt with simultaneously. In the section below, studies that are more focused on the adaptation process than on the impacts produced by climate change are presented.

Becker et al., 2012 [76] defined three main options for port adaptation: update storm defenses, elevate to compensate and relocate entirely. Foti et al., 2020 [89] performed a review on coastal defense techniques and called for an improvement of existing structures and the combination of different types of solutions. Tanik et al., 2017 [90] studied and compared the adaptation measures in Germany, France, Spain, Italy, Denmark, USA and Kenya.

Sánchez-Arcilla et al., 2016 [91] studied different pathways to adaptation on the Catalan coast, criticized the reactive management followed in Spain and called for active and sustainable interventions. López-Dóriga et al., 2020 [92] studied the investment in adaptation on the Spanish coast. The investment in coastal

adaptation between 2010–2018 was estimated at 56 M €, and they concluded that it was not enough to face climate change. Abadie et al., 2020 [93] studied 62 main coastal cities in Portugal and Spain under RCP 2.6, 4.5 and 8.5., and calculated the estimated costs if no adaption measures were taken and if adaption strategies were implemented. They concluded that the investment in adaptation will be much smaller than the economic damage caused by inaction.

Sierra et al., 2017 [94] studied the effect of seagrass meadows in two harbors on the Catalan coast. These meadows were found to be a great measure to reduce wave height. Velasco et al., 2018 [95] conducted an assessment of the effectiveness of structural and nonstructural measures in Barcelona with a 1D/2D-coupled model that used depth-damage curves, detailed flood depth maps and land-use maps.

There are also different research studies concerning adaptation and administrative and governance issues in Spain.

Salvia et al., 2021 [96] studied the mitigation policies within Mediterranean Europe and saw that France and Spain were the most active in terms of climate-mitigation planning. Losada et al., 2019 [97] also studied the adaptation strategy of Spain to climate change. In 2020, Izaguirre et al. [98] developed a methodology to assess climate change impacts in seaports. Sauer et al., 2021 [99] studied the actors that participate in the management of the port of Barcelona and its adaptation.

In addition, it has to be said that the port of Valencia is part of the ECCLIPSE European Project, of which the main objective is to develop a common framework for assessing the impacts associated with climate change and the adaptation to such impacts of ports in the southwest of Europe [100].

Table 2 presents a summary of the main conclusions and methodologies of the studies presented in this section.

Table 2. Summary of Spanish studies of climate change and ports.

Type	Authors	Region	Main Conclusions	Methodology
Impacts of climate change on Spanish ports	Sánchez-Arcilla et al., 2016 [82]	Mediterranean	Overtopping and flooding are the main direct impacts of SLR.	Review of previous studies.

	Sierra et al., 2015, 2017 [83,84]	Catalan ports	Small ports will be affected by SLR because the relative increment of depth is high.	SWAN and Boussinesq models and linear wave theory.
	Casas-Prat and Sierra, 2010, 2012 [85,86]		Directional changes will cause a mean increase of about 50% in harbor agitation.	Statistical and linear regression analysis and Boussinesq model.
	León-Mateos et al., 2021 [15]	A Coruña	Port Resilience Index to identify the areas in which improvement is necessary.	Identification of critical processes, risk scenarios and resilience indexes. Crossing of the datasets.
	López et al., 2015 [87]	Ferrol	Implementation of ANNs, more efficient and less data needed than traditional methods.	Collection of the data from buoys, training and validation of the ANN.
Adaptation of Spanish ports	López-Dóriga et al., 2020 [92]	Spain	Investment in adaptation directly related to the regional GDP. Andalusia invests much more.	Data collection and correlation and statistical analysis.
	Abadie et al., 2020 [93]	Iberian Peninsula	Using median values can underestimate coastal risks. Invest in adaptation measures is needed.	Stochastic approach. Define SL cities percentiles and damage functions.
	Sierra et a., 2017 [94]	Catalan ports	Seagrass meadows can attenuate 40% of waveheight. Higher density, higher attenuation.	SWAN and Boussinesq models.
	Sánchez-Arcilla et al., 2016 [91]		Resilience is higher for energetic coasts. Sustainable practices are long-term practices.	Comparative analysis.

	Velasco et al., 2018 [95]	Barcelona	Nonstructural strategies are better than structural. Structural are more efficient, but more expensive.	1D/2D coupled model (Infoworks ICM).
Adaptation and governance issues	Salvia et al., 2021 [96]	Mediterranean Europe	The west of Europe is more active in mitigation policies (stronger governance and regulation).	Data and plans review at regional level. Statistical analysis.
	Losada et al., 2019 [97]	Spain	Cooperation between stakeholders and policymakers, public and private sector, and at regional level is key.	Development of Spanish Strategy for Coastal Adaptation.
	Sauer et al., 2021 [99]	Barcelona	Lack of diversity, vertical coordination and communication in climate change adaptation.	Social Network Analysis, quantitative metrics and semi-structured surveys.

From all the scientific articles analyzed in this paper, most of them focused on the damage that climate change could cause on ports (Figure 7). However, only a few of them studied the relationship of the agents implied in the operation of ports, which has proven to be a weakness. The operation of ports is very complex; it involves public and private agents and for the development of an efficient strategy to adapt ports to climate change, an effort to study the relationships between these agents is needed. The number of studies that focus on the development of adaptation strategies is also scarce. The first step to deal with climate change is to improve the collaboration between the agents implied in port operations, and to do that, future research lines should focus on the study of port networks and how to improve the collaboration between the public and private sector and the academics.

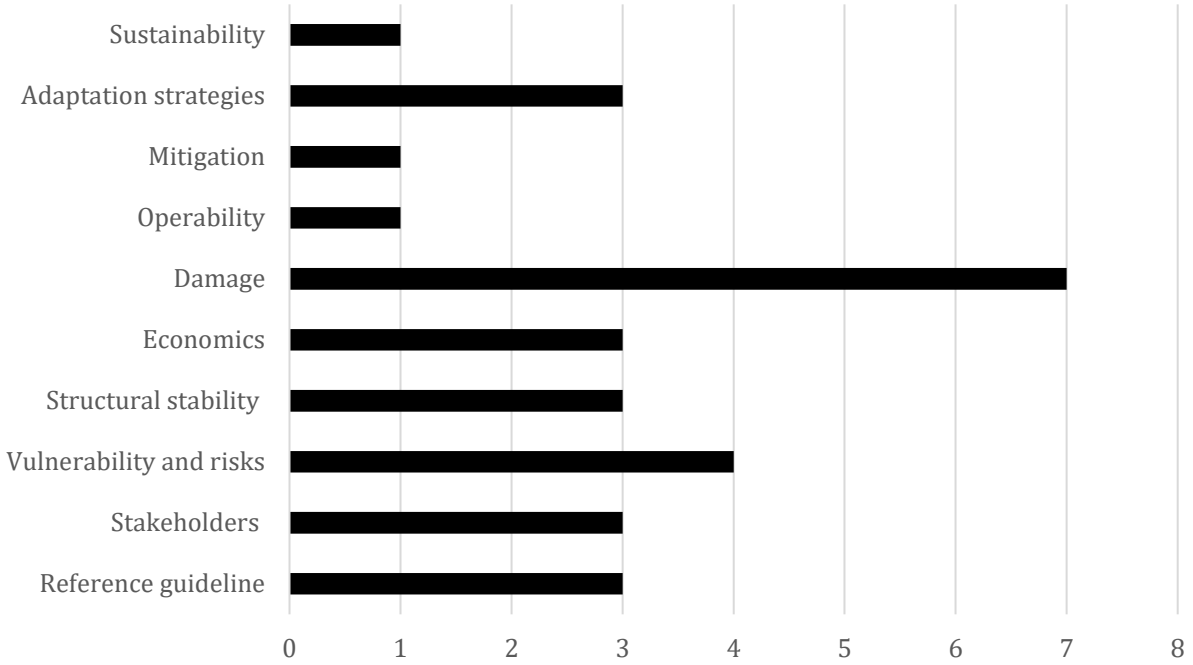


Figure 7. Topic of the main studies focused on climate change and ports.

Referring to the methodology, the majority of the studies used numerical models, more specifically, SWAN for wave propagation and Boussinesq models to study harbor agitation; SWAN models cannot solve diffraction problems and Boussinesq models are very computationally expensive, so they can only be applied in very small areas. The main problem of these numerical models is that they need a great amount of high-quality data, so when this data is not available, other techniques such as linear theory or analytical models have to be applied. Moreover, ANNs have proved to outperform traditional models and they are becoming more popular, so future studies should use this methodology to improve the accuracy of the results when numerical models cannot be applied, and even as an alternative to them.

On the other hand, all of the studies presented focus on big ports because they have a higher economic and social value. There are almost no scientific papers in the Spanish Mediterranean coast that study the impact of climate change on small ports and, as stated by Sierra et al. [83], these will be the ports most affected by climate change because the SLR will represent a high percentage of each depth. Therefore, more studies on small ports are needed.

Finally, as detailed in Section 3.1, most of the studies about port adaptation are located in very specific areas of the Mediterranean coast, though there are

thousands of km of the Spanish Mediterranean coast and ports without specific regional studies.

4. Conclusions

Ports are key spaces in coastal cities. They are not only crucial for logistic chains and the transport of goods nationally and internationally, but also, they are essential for the economic and social functions of their cities. Around ports, spaces of great social and economic value are generated for a city. Many coastal cities have taken advantage of the port area to develop street markets, bars, discotheques and even amusement parks for children. Ports concentrate many activities, both social and economic, and generate a lot of value. They create employment and contribute to the GDP of coastal cities, promoting urbanization and industrialization.

Ports are one of the systems most threatened by climate change. To face climate change impacts, mitigation and adaptation policies are key. However, mitigation policies take too long to work and there are climate change effects that do not have solutions even with the implementation of mitigation policies. Therefore, the development of adaption plans is key for the future of coastal cities. The scientific community agrees on this (Becker et al., 2013 [76]; Sánchez-Arcilla et al. [91,101]; Sierra et al. [83,84,94]; Casas-Prat et al. [85,86]; Toimil et al., 2017 [102]; Camus et al., 2019 [64]; Abadie et al., 2020 [93]) and calls on countries to develop active measures.

Spain and the Spanish Mediterranean coast is one of the hotspots of climate change. If no adaption measures are taken, huge economic and social losses will occur in future scenarios, such as the ones that happened with Gloria or Emma squall. To avoid these losses, the adaption of the coastal defense structures and adaption strategies are needed.

Although scientific literature about climate change is abundant, there are only a few articles that focus on the effect that climate change will have on ports and on their operability. Much more effort has been put into studying how coastlines are going to change with climate change, such as in the study of beach erosion and of the flooding of cities by rainfall events.

However, ports and their adaptation to climate change have received much less attention and most of the studies that exist concerning climate change and its effects on ports focus on the damage and impacts that SLR and extreme events will have. Nonetheless, it has been proven that one of the reasons that makes the study

of ports very difficult is the complexity of the network of the agents implied in the operation of ports. There are very few articles that study these relationships and how to improve coordination and cooperation between the actors involved in the operation of a port. Until this is improved, all studies and efforts will be in vain. Therefore, future research lines should focus on this topic.

Referring to the methodologies used in these studies, the most common methods used are numerical models (in both sea level prediction and impacts and adaptation assessments). However, in recent years, ANNs have been proven to provide more accurate results, be less computationally expensive and have the capacity to model processes even with less-quality data. ANNs have become a widely used tool in water resources research and should also be applied in ocean and coastal engineering.

In the case of Spain, the scientific papers that exist concerning ports and climate change focus on very specific hotspots, most of them in the Catalan coast (the Ebro Delta, Maresme, etc.), but there is still a lack of studies in most of the ports along the coast. Therefore, more studies should be carried out in these geographical gaps (Balearic Islands and the area of Valencia).

In addition, most studies focus on ports of general interest, but the ports that will be most affected by climate change are small fishing ports and marinas, because due to their dimensions, changes in SL and wave height are much more significant and their relative importance is higher. Therefore, future research lines should also focus on the study of small harbors, even if they are not as important as ports of general interest, as they are the ports most threatened by climate change.

References

1. IPCC. *Las Evaluaciones del IPCC de 1990 y 1992*; C.U. Press: Cambridge, UK, 1992; p. 196.
2. Bindoff, N.L.; Cheung, W.W.L.; Kairo, J.G.; Arístegui, J.; Guinder, V.A.; Hallberg, R.; Hilmi, N.; Jiao, N.; Karim, M.S.; Levin, L.; et al. Changing Ocean, Marine Ecosystems, and Dependent Communities. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*; Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N.M., Eds.; Intergovernmental Panel on Climate Change: Switzerland, 2019.
3. Collins, M.; Sutherland, M.; Bouwer, L.; Cheong, S.-M.; Frölicher, T.; Jacot Des Combes, H.; Koll Roxy, M.; Losada, I.; McInnes, K.; Ratter, B.; et al. Extremes, Abrupt Changes and Managing Risk. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*; Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A.,

Petzold, J., Rama, B., Weyer, N.M., Eds.; United Nations: San Francisc, CA, USA, 2019.

4. Hock, R.; Rasul, G.; Adler, C.; Cáceres, B.; Gruber, S.; Hirabayashi, Y.; Jackson, M.; Kääb, A.; Kang, S.; Kutuzov, S et al. High Mountain Areas. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*; Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N.M., Eds.; Intergovernmental Panel on Climate Change: Switzerland 2019.
5. Meredith, M.; Sommerkorn, M.; Cassotta, S.; Derksen, C.; Ekaykin, A.; Hollowed, A.; Kofinas, G.; Mackintosh, A.; Melbourne-Thomas, J.; Muelbert, M.M.C.; et al. Polar Regions. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*; Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N.M., Eds.; Intergovernmental Panel on Climate Change: Switzerland 2019.
6. Oppenheimer, M.; Glavovic, B.C.; Hinkel, J.; van de Wal, R.; Magnan, A.K.; Abd-Elgawad, A.; Cai, R.; Cifuentes-Jara, M.; DeConto, R.M.; Ghosh, T et al. Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*; Pörtner, H.-O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., Weyer, N.M., Eds.; Intergovernmental Panel on Climate Change: Switzerland 2019.
7. IPCC. *AR4 Climate Change 2007: Synthesis Report*; Pachauri, R.K., Reisinger, A., Eds.; Intergovernmental Panel on Climate Change Geneva, Switzerlnad, 2007.
8. Losada, M.A.; Baquerizo, A.; Ortega-Sanchez, M.; Avila, A. Coastal Evolution, Sea Level, and Assessment of Intrinsic Uncertainty. *J. Coast. Res.* **2011**, *59*, 218–228. https://doi.org/10.2112/si59-023.1.
9. Masselink, G.; Hughes, M. An Introduction to Coastal Processes and Geomorphology, 2003; p. 354; Routledge, London, doi: 10.4324/9780203783740 .
10. Cramer, W.; Guiot, J.; Fader, M.; Garrabou, J.; Gattuso, J.-P.; Iglesias, A.; Lange, M.A.; Lionello, P.; Llasat, M.C.; Paz, S.; et al. Climate change and interconnected risks to sustainable development in the Mediterranean. *Nat. Clim. Change* **2018**, *8*, 972–980. https://doi.org/10.1038/s41558-018-0299-2.
11. Giorgi, F.; Lionello, P. Climate change projections for the Mediterranean region. *Glob. Planet. Change* **2008**, *63*, 90–104. https://doi.org/10.1016/j.gloplacha.2007.09.005.
12. Giorgi, F. Climate change hot-spots. *Geophys. Res. Lett.* **2006**, *33*. https://doi.org/10.1029/2006GL025734.
13. Mentaschi, L.; Vousdoukas, M.I.; Pekel, J.-F.; Voukouvalas, E.; Feyen, L. Global long-term observations of coastal erosion and accretion. *Sci. Rep.* **2018**, *8*, 12876. https://doi.org/10.1038/s41598-018-30904-w.
14. Luijendijk, A.; Hagenaars, G.; Ranasinghe, R.; Baart, F.; Donchyts, G.; Aarninkhof, S. The State of the World's Beaches. *Sci. Rep.* **2018**, *8*, 6641. https://doi.org/10.1038/s41598-018-24630-6.
15. Leon-Mateos, F.; Sartal, A.; Lopez-Manuel, L.; Quintas, M.A. Adapting our sea ports to the challenges of climate change: Development and validation of a Port Resilience Index. *Mar. Policy* **2021**, *130*, 104573. https://doi.org/10.1016/j.marpol.2021.104573.

16. Garcia-Alonso, L.; Moura, T.G.Z.; Roibas, D. The effect of weather conditions on port technical efficiency. *Mar. Policy* **2020**, *113*, 103816. https://doi.org/10.1016/j.marpol.2020.103816.

17. IMO. *International Shipping Facts and Figures—Information Resources on Trade, Safety, Security, Environment, Maritime Knowledge Centre*; International Maritime Organization. London, UK: 2012.

18. UNCTAD. *Review of Maritime Transport*; United Nations Publication, Geneve, Switzerland: 2019.

19. Adebisi, N.; Balogun, A.-L.; Min, T.H.; Tella, A. Advances in estimating Sea Level Rise: A review of tide gauge, satellite altimetry and spatial data science approaches. *Ocean. Coast. Manag.* **2021**, *208*, 105632. https://doi.org/10.1016/j.ocecoaman.2021.105632.

20. Luque, P.; Gomez-Pujol, L.; Marcos, M.; Orfila, A. Coastal Flooding in the Balearic Islands During the Twenty-First Century Caused by Sea-Level Rise and Extreme Events. *Front. Mar. Sci.* **2021**, *8*, 676452. https://doi.org/10.3389/fmars.2021.676452.

21. Jorda, G.; Gomis, D. On the interpretation of the steric and mass components of sea level variability: The case of the Mediterranean basin. *J. Geophys. Res. Ocean.* **2013**, *118*, 953–963. https://doi.org/10.1002/jgrc.20060.

22. Kopp, R.E.; Horton, R.M.; Little, C.M.; Mitrovica, J.X.; Oppenheimer, M.; Rasmussen, D.J.; Strauss, B.H.; Tebaldi, C. Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites. *Earth's Future* **2014**, *2*, 383–406. https://doi.org/10.1002/2014EF000239.

23. PSMSL. Permanent Service for Mean Sea Level. Available online: https://psmsl.org/ (accessed on 2 June 2022).

24. Church, J.A.; White, N.J. A 20th century acceleration in global sea-level rise. *Geophys. Res. Lett.* **2006**, *33*. https://doi.org/10.1029/2005GL024826.

25. Miller, L.; Douglas, B.C. Mass and volume contributions to twentieth-century global sea level rise. *Nature* **2004**, *428*, 406–409. https://doi.org/10.1038/nature02309.

26. Peltier, W.R. Chapter 4 Global glacial isostatic adjustment and modern instrumental records of relative sea level history. In *International Geophysics*; BDouglas, C.; Kearney, M.S.; Leatherman, S.P., Eds.; Academic Press: Cambridge, MA, USA, 2001; Volume 75, pp. 65–95.

27. Cazenave, A.; Cabanes, C.; Dominh, K.; Gennero, M.C.; Le Provost, C. Present-Day Sea Level Change: Observations and Causes. *Space Sci. Rev.* **2003**, *108*, 131–144. https://doi.org/10.1023/A:1026238318585.

28. Church, J.A.; White, N.J. Sea-Level Rise from the Late 19th to the Early 21st Century. *Surv. Geophys.* **2011**, *32*, 585–602. https://doi.org/10.1007/s10712-011-9119-1.

29. Hay, C.C.; Morrow, E.; Kopp, R.E.; Mitrovica, J.X. Probabilistic reanalysis of twentieth-century sea-level rise. *Nature* **2015**, *517*, 481–484. https://doi.org/10.1038/nature14093.

30. Tsimplis, M.N.; Calafat, F.M.; Marcos, M.; Jorda, G.; Gomis, D.; Fenoglio-Marc, L.; Struglia, M.V.; Josey, S.A.; Chambers, D.P. The effect of the NAO on sea level and on mass changes in the Mediterranean Sea. *J. Geophys. Res. Ocean.* **2013**, *118*, 944–952. https://doi.org/10.1002/jgrc.20078.

31. Antonioli, F.; De Falco, G.; Lo Presti, V.; Moretti, L.; Scardino, G.; Anzidei, M.; Bonaldo, D.; Carniel, S.; Leoni, G.; Furlani, S.; et al. Relative Sea-Level Rise and Potential Submersion Risk for 2100 on 16 Coastal Plains of the Mediterranean Sea. *Water* **2020**, *12*, 2173. Retrieved from https://www.mdpi.com/2073-4441/12/8/2173.

32. Vacchi, M.; Joyse, K.M.; Kopp, R.E.; Marriner, N.; Kaniewski, D.; Rovere, A. Climate pacing of millennial sea-level change variability in the central and western Mediterranean. *Nat. Commun.* **2021**, *12*, 4013. https://doi.org/10.1038/s41467-021-24250-1.

33. Marcos, M.; Puyol, B.; Amores, A.; Pérez Gómez, B.; Fraile, M.Á.; Talke, S.A. Historical tide gauge sea-level observations in Alicante and Santander (Spain) since the 19th century. *Geosci. Data J.* **2021**, *8*, 144–153. https://doi.org/10.1002/gdj3.112.

34. Marcos, M.; Puyol, B.; Wöppelmann, G.; Herrero, C.; García-Fernández, M.J. The long sea level record at Cadiz (southern Spain) from 1880 to 2009. *J. Geophys. Res. Ocean.* **2011**, *116*. https://doi.org/10.1029/2011JC007558.

35. Vargas-Yáñez, M.; Tel, E.; Moya, F.; Ballesteros, E.; Garcia-Martinez, M. Long-Term Changes, Inter-Annual, and Monthly Variability of Sea Level at the Coasts of the Spanish Mediterranean and the Gulf of Cádiz. *Geosciences* **2021**, *11*, 350. https://doi.org/10.3390/geosciences11080350.

36. IPCC. *Climate Change 2013: The Physical Science Basis.Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013; p. 1535.

37. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I.; et al., Eds.; Cambridge University: Cambridge, UK; New York, NY, USA, 2021.

38. Rahmstorf, S. A Semi-Empirical Approach to Projecting Future Sea-Level Rise. *Science* **2007**, *315*, 368–370, https://doi.org/10.1126/science.1135456.

39. Horton, B.P.; Khan, N.S.; Cahill, N.; Lee, J.S.H.; Shaw, T.A.; Garner, A.J.; Kemp, A.C.; Engelhart, S.E.; Rahmstorf, S. Estimating global mean sea-level rise and its uncertainties by 2100 and 2300 from an expert survey. *Npj Clim. Atmos. Sci.* **2020**, *3*, 18. https://doi.org/10.1038/s41612-020-0121-5.

40. Kopp, R.E.; Kemp, A.C.; Bittermann, K.; Horton, B.P.; Donnelly, J.P.; Gehrels, W.R.; Hay, C.C.; Mitrovica, J.X.; Morrow, E.D.; Rahmstorf, S. Temperature-driven global sea-level variability in the Common Era. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1434–E1441. https://doi.org/10.1073/pnas.1517056113.

41. Mengel, M.; Levermann, A.; Frieler, K.; Robinson, A.; Marzeion, B.; Winkelmann, R. Future sea level rise constrained by observations and long-term commitment. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2597–2602. https://doi.org/10.1073/pnas.1500515113.

42. Bamber, J.L.; Oppenheimer, M.; Kopp, R.E.; Aspinall, W.P.; Cooke, R.M. Ice sheet contributions to future sea-level rise from structured expert judgment. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11195–11200. https://doi.org/10.1073/pnas.1817205116.

43. Ibàñez, C.; Canicio, A.; Day, J.W.; Curcó, A. Morphologic development, relative sea level rise and sustainable management of water and sediment in the Ebre Delta, Spain. *J. Coast. Conserv.* **1997**, *3*, 191. https://doi.org/10.1007/BF02905244.

44. Jiménez, J.; Sánchez-Arcilla, A.; Valdemoro, H.I.; Gracia, V.; Nieto, F. Processes reshaping the Ebro delta. *Mar. Geol.* **1997**, *144*, 59–79. https://doi.org/10.1016/S0025-3227(97)00076-5.

45. Vazquez Pinillos, F.J.; Marchena Gomez, M.J. TERRITORIAL IMPACTS OF SEA-LEVEL RISE IN MARSH ENVIRONMENTS. THE CASE OF THE BAY OF CADIZ, SPAIN. *Cuad. De Investig. Geogr.* **2021**, *47*, 523–543. https://doi.org/10.18172/cig.4531.

46. Chust, G.; Ángel, B.; Liria, P.; Galparsoro, I.; Marcos, M.; Caballero, A.; Castro, R. Human impacts overwhelm the effects of sea-level rise on Basque coastal habitats (N Spain) between 1954 and 2004. *Estuar. Coast. Shelf Sci.* **2009**, *84*, 453–462. https://doi.org/10.1016/j.ecss.2009.07.010.

47. Marcos, M.; Gomis, D.; Monserrat, S.; Álvarez-Fanjul, E.; Pérez, B.; García-Lafuente, J. Consistency of long sea-level time series in the northern coast of Spain. *J. Geophys. Res. Ocean.* **2005**, *110*. https://doi.org/10.1029/2004JC002522.

48. Leorri, E.; Horton, B.P.; Cearreta, A. Development of a foraminifera-based transfer function in the Basque marshes, N. Spain: Implications for sea-level studies in the Bay of Biscay. *Mar. Geol.* **2008**, *251*, 60–74. https://doi.org/10.1016/j.margeo.2008.02.005.

49. Vousdoukas, M.I.; Mentaschi, L.; Voukouvalas, E.; Verlaan, M.; Feyen, L. Extreme sea levels on the rise along Europe's coasts. *Earths Future* **2017**, *5*, 304–323. https://doi.org/10.1002/2016ef000505.

50. Chust, G.; Caballero, A.; Marcos, M.; Liria, P.; Hernandez, C.; Borja, A. Regional scenarios of sea level rise and impacts on Basque (Bay of Biscay) coastal habitats, throughout the 21st century. *Estuar. Coast. Shelf Sci.* **2010**, *87*, 113–124. https://doi.org/10.1016/j.ecss.2009.12.021.

51. Vitousek, S.; Barnard, P.L.; Limber, P.; Erikson, L.; Cole, B. A model integrating longshore and cross-shore processes for predicting long-term shoreline response to climate change. *J. Geophys. Res. Earth Surf.* **2017**, *122*, 782–806. https://doi.org/10.1002/2016JF004065.

52. Hinkel, J.; Lincke, D.; Vafeidis, A.T.; Perrette, M.; Nicholls, R.J.; Tol, R.S.J.; Marzeion, B.; Fettweis, X.; Ionescu, C.; Levermann, A. Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3292–3297. https://doi.org/10.1073/pnas.1222469111.

53. Galassi, G.; Spada, G. Sea-level rise in the Mediterranean Sea by 2050: Roles of terrestrial ice melt, steric effects and glacial isostatic adjustment. *Glob. Planet. Change* **2014**, *123*, 55–66. https://doi.org/10.1016/j.gloplacha.2014.10.007.

54. Gil-Guirado, S.; Perez-Morales, A.; Pino, D.; Pena, J.C.; Martinez, F.L. Flood impact on the Spanish Mediterranean coast since 1960 based on the prevailing synoptic patterns. *Sci. Total Environ.* **2021**, *807*, 150777. https://doi.org/10.1016/j.scitotenv.2021.150777.

55. Pérez-Morales, A.; Gil-Guirado, S.; Olcina-Cantos, J. Housing bubbles and the increase of flood exposure. Failures in flood risk management on the Spanish south-

eastern coast (1975–2013). *J. Flood Risk Manag.* **2018**, *11*, S302–S313. https://doi.org/10.1111/jfr3.12207.

56. Ribas, A.; Olcina, J.; Saurí, D. More exposed but also more vulnerable? Climate change, high intensity precipitation events and flooding in Mediterranean Spain. *Disaster Prev. Manag.* **2020**, *29*, 229–248.
57. Vinet, F.; Bigot, V.; Petrucci, O.; Papagiannaki, K.; Llasat, M.C.; Kotroni, V.; Boissier, L.; Aceto, L.; Grimalt, M.; Llasat-Botija, M.; et al. Mapping Flood-Related Mortality in the Mediterranean Basin. Results from the MEFF v2.0 DB. *Water* **2019**, *11*, 2196. Retrieved from https://www.mdpi.com/2073-4441/11/10/2196.
58. Sayol, J.M.; Marcos, M. Assessing Flood Risk Under Sea Level Rise and Extreme Sea Levels Scenarios: Application to the Ebro Delta (Spain). *J. Geophys. Res. Ocean.* **2018**, *123*, 794–811. https://doi.org/10.1002/2017jc013355.
59. Grases, A.; Gracia, V.; Garcia-Leon, M.; Lin-Ye, J.; Pau Sierra, J. Coastal Flooding and Erosion under a Changing Climate: Implications at a Low-Lying Coast (Ebro Delta). *Water* **2020**, *12*, 346. https://doi.org/10.3390/w12020346.
60. Lopez-Doriga, U.; Jimenez, J.A. Impact of Relative Sea-Level Rise on Low-Lying Coastal Areas of Catalonia, NW Mediterranean, Spain. *Water* **2020**, *12*, 3252. https://doi.org/10.3390/w12113252.
61. Ballesteros, C.; Jimenez, J.A.; Viavattene, C. A multi-component flood risk assessment in the Maresme coast (NW Mediterranean). *Nat. Hazards* **2018**, *90*, 265–292. https://doi.org/10.1007/s11069-017-3042-9.
62. Hernandez-Mora, M.; Meseguer-Ruiz, O.; Karas, C.; Lambert, F. Estimating coastal flood hazard of Tossa de Mar, Spain: A combined model—Data interviews approach. *Nat. Hazards* **2021**, *109*, 2153–2171. https://doi.org/10.1007/s11069-021-04914-3.
63. Nicholls, R.; Herweijer, C.; Ranger, N.; Hallegatte, S.; Corfee-Morlot, J.; Chateau, J.; Muir-Wood, R. Ranking Port Cities with High Exposure and Vulnerability to Climate Extremes: Exposure Estimates. *OECD Environ. Dir. OECD Environ. Work. Pap.* **2008**. https://doi.org/10.1787/011766488208.
64. Camus, P.; Tomás, A.; Diaz-Hernandez, G.; Rodriguez, B.; Izaguirre, C.; Losada, I.J. Probabilistic assessment of port operation downtimes under climate change. *Coast. Eng.* **2019**, *147*, 12–24.
65. Chhetri, P.K.; Corcoran, J.; Gekara, V.O.; Maddox, C.G.; McEvoy, D. Seaport resilience to climate change: Mapping vulnerability to sea-level rise. *J. Spat. Sci.* **2015**, *60*, 65–78.
66. Nursey-Bray, M.; Blackwell, B.; Brooks, B.; Campbell, M.L.; Goldsworthy, L.; Pateman, H.; Rodrigues, I.; Roome, M.; Wright, J.T.; Francis, J.; et al Vulnerabilities and adaptation of ports to climate change. *J. Environ. Plan. Manag.* **2013**, *56*, 1021–1045. https://doi.org/10.1080/09640568.2012.716363.
67. Hanson, S.; Nicholls, R.; Ranger, N.; Hallegatte, S.; Corfee-Morlot, J.; Herweijer, C.; Chateau, J. A global ranking of port cities with high exposure to climate extremes. *Clim. Change* **2011**, *104*, 89–111. https://doi.org/10.1007/s10584-010-9977-4.
68. Haveman, J.; Shatz, H. Protecting the Nation's Seaports: Balancing Security and Cost @BULLET @BULLET @BULLET. **2006**.
69. Messner, S.F.; Moran, L.; Reub, G.; Campbell, J. Climate change and sea level rise impacts at ports and a consistent methodology to evaluate vulnerability and risk. In

Proceedings of the Name of the Conference CP 2013, Uppsala, Sweden, 16–20 September 2013.

70. McEvoy, D.; Mullett, J. Enhancing the resilience of seaports to a changing climate: Research synthesis and implications for policy and practice. *Aust. Gov. Natl. Clim. Change Adapt. Res. Facil.* **2013**, *11*.
71. Takagi, H.; Esteban, M.; Shibayama, T. Proposed methodology for evaluating the potential failure risk for existing caisson-breakwaters in a storm event using a level III reliability-based approach. *Coastal Engineering* 2009, pp. 3655-3667.
72. Takagi, H.; Kashihara, H.; Esteban, M.; Shibayama, T. Assessment of future stability of breakwaters under climate change. *Coast. Eng. J.* **2011**, *53*, 21.
73. Takagi, H.; Shibayama, T.; Esteban, M. An Expansion of the Reliability Design Method for Caisson-Type Breakwaters towards Deep Water using the Fourth Order Approximation of Standing Waves. In Proceedings of the Asian and Pacific Coasts 2007, Nanjing, China, 21–24 September 2007.
74. Mase, H.; Tsujio, D.; Yasuda, T.; Mori, N. Stability analysis of composite breakwater with wave-dissipating blocks considering increase in sea levels, surges and waves due to climate change. *Ocean. Eng.* **2013**, *71*, 58–65. https://doi.org/10.1016/j.oceaneng.2012.12.037.
75. Becker, A.; Inoue, S.; Fischer, M.; Schwegler, B. Climate change impacts on international seaports: Knowledge, perceptions, and planning efforts among port administrators. *Clim. Change* **2012**, *110*, 5–29. https://doi.org/10.1007/s10584-011-0043-7.
76. Becker, A.H.; Acciaro, M.; Asariotis, R.; Cabrera, E.; Cretegny, L.; Crist, P.; Esteban, M.; Mather, A.; Messner, S.; Naruse, S.; et al. A note on climate change adaptation for seaports: A challenge for global ports, a challenge for global society. *Clim. Change* **2013**, *120*, 683–695. https://doi.org/10.1007/s10584-013-0843-z.
77. Mutombo, K.; Ölçer, A. Towards port infrastructure adaptation: A global port climate risk analysis. *WMU J. Marit. Aff.* **2017**, *16*, 161–173. https://doi.org/10.1007/s13437-016-0113-9.
78. Prahl, B.F.; Boettle, M.; Costa, L.; Kropp, J.P.; Rybski, D. Damage and protection cost curves for coastal floods within the 600 largest European cities. *Sci. Data* **2018**, *5*, 180034. https://doi.org/10.1038/sdata.2018.34.
79. Abadie, L.; Galarraga, I.; Markandya, A.; Sainz de Murieta, E. Risk measures and the distribution of damage curves for 600 European coastal cities. *Environ. Res. Lett.* **2019**, *14*, 064021. https://doi.org/10.1088/1748-9326/ab185c.
80. Reckien, D.; Flacke, J.; Olazabal, M.; Heidrich, O. The Influence of Drivers and Barriers on Urban Adaptation and Mitigation Plans—An Empirical Analysis of European Cities. *PLoS ONE* **2015**, *10*, e0135597. https://doi.org/10.1371/journal.pone.0135597.
81. Christodoulou, A.; Christidis, P.; Demirel, H. Sea-level rise in ports: A wider focus on impacts. *Marit. Econ. Logist.* **2019**, *21*, 482–496. https://doi.org/10.1057/s41278-018-0114-z.
82. Sanchez-Arcilla, A.; Pau Sierra, J.; Brown, S.; Casas-Prat, M.; Nicholls, R.J.; Lionello, P.; Conte, D. A review of potential physical impacts on harbours in the Mediterranean Sea under climate change. *Reg. Environ. Change* **2016**, *16*, 2471–2484. https://doi.org/10.1007/s10113-016-0972-9.

83. Sierra, J.P.; Casas-Prat, M.; Virgili, M.; Moesso, C.; Sanchez-Arcilla, A. Impacts on wave-driven harbour agitation due to climate change in Catalan ports. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 1695–1709. https://doi.org/10.5194/nhess-15-1695-2015.

84. Sierra, J.P.; Genius, A.; Lionello, P.; Mestres, M.; Mosso, C.; Marzo, L. Modelling the impact of climate change on harbour operability: The Barcelona port case study. *Ocean. Eng.* **2017**, *141*, 64–78. https://doi.org/10.1016/j.oceaneng.2017.06.002.

85. Casas-Prat, M.; Sierra, J.P. Trend analysis of wave direction and associated impacts on the Catalan coast. *Clim. Change* **2012**, *115*, 667–691. https://doi.org/10.1007/s10584-012-0466-9.

86. Casas-Prat, M.; Sierra, J.P. Trend analysis of wave storminess: Wave direction and its impact on harbour agitation. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 2327–2340. https://doi.org/10.5194/nhess-10-2327-2010.

87. López, M.; Iglesias, G. Artificial neural networks applied to port operability assessment. *Ocean Eng.* **2015**, *109*, 298. https://doi.org/10.1016/j.oceaneng.2015.09.016.

88. UNCTAD. Climate Change Impacts and Adaptation: A Challenge for Global Ports. Available online: https://unctad.org/system/files/official-document/dtltlb2011d2_en.pdf (accessed on 2 June 2022).

89. Foti, E.; Musumeci, R.E.; Stagnitti, M. Coastal defence techniques and climate change: A review. *Rend. Lincei. Sci. Fis. E Nat.* **2020**, *31*, 123–138.

90. Tanik, A.; Tekten, D. Climate Change Adaptation Practices in Various Countries. In Proceedings of the Name of the Conference 2nd International Conference on Green Energy Technology (ICGET), Sapienza Univ Rome, Rome, Italy, 18–20 July 2017.

91. Sánchez-Arcilla, A.; García-León, M.; Gracia, V.; Devoy, R.; Stanica, A.; Gault, J. Managing coastal environments under climate change: Pathways to adaptation. *Sci. Total Environ.* **2016**, *572*, 1336–1352. https://doi.org/10.1016/j.scitotenv.2016.01.124.

92. Lopez-Doriga, U.; Jimenez, J.A.; Bisaro, A.; Hinkel, J. Financing and implementation of adaptation measures to climate change along the Spanish coast. *Sci. Total Environ.* **2020**, *712*, 135685. https://doi.org/10.1016/j.scitotenv.2019.135685.

93. Maria Abadie, L.; Sainz de Murieta, E.; Galarraga, I. The Costs of Sea-Level Rise: Coastal Adaptation Investments vs. Inaction in Iberian Coastal Cities. *Water* **2020**, *12*, 1220. https://doi.org/10.3390/w12041220.

94. Sierra, J.P.; Garcia-Leon, M.; Gracia, V.; Sanchez-Arcilla, A. Green measures for Mediterranean harbours under a changing climate. *Proc. Inst. Civ. Eng. -Marit. Eng.* **2017**, *170*, 55–66. https://doi.org/10.1680/jmaen.2016.23.

95. Velasco, M.; Russo, B.; Cabello, A.; Termes, M.; Sunyer, D.; Malgrat, P. Assessment of the effectiveness of structural and nonstructural measures to cope with global change impacts in Barcelona. *J. Flood Risk Manag.* **2018**, *11*, S55–S68. https://doi.org/10.1111/jfr3.12247.

96. Salvia, M.; Olazabal, M.; Fokaides, P.A.; Tardieu, L.; Simoes, S.G.; Geneletti, D.; Hurtado, S.D.G.; Viguie, V.; Spyridaki, N.-A.; Pietrapertosa, F.; et al. Climate mitigation in the Mediterranean Europe: An assessment of regional and city-level plans. *J. Environ. Manag.* **2021**, *295*, 113146. https://doi.org/10.1016/j.jenvman.2021.113146.

97. Losada, I.J.; Toimil, A.; Muñoz, A.; Garcia-Fletcher, A.P.; Diaz-Simal, P. A planning strategy for the adaptation of coastal areas to climate change: The Spanish case. *Ocean. Coast. Manag.* **2019**, *182*, 104983. https://doi.org/10.1016/j.ocecoaman.2019.104983.

98. Izaguirre, C.; Losada, I.J.; Camus, P.; González-Lamuño, P.; Stenek, V. Seaport climate change impact assessment using a multi-level methodology. *Marit. Policy Manag.* **2020**, *47*, 544–557. https://doi.org/10.1080/03088839.2020.1725673.

99. Sauer, I.J.; Roca, E.; Villares, M. Integrating climate change adaptation in coastal governance of the Barcelona metropolitan area. *Mitig. Adapt. Strateg. Glob. Change* **2021**, *26*, 16. https://doi.org/10.1007/s11027-021-09953-6.

100. Union, E. Ecclipse. Available online: https://ecclipse.eu/overview/ (accessed on 2 June 2022).

101. Sanchez-Gomez, E.; Somot, S.; Josey, S.A.; Dubois, C.; Elguindi, N.; Deque, M. Evaluation of Mediterranean Sea water and heat budgets simulated by an ensemble of high resolution regional climate models. *Clim. Dyn.* **2011**, *37*, 2067–2086. https://doi.org/10.1007/s00382-011-1012-6.

102. Toimil, A.; Losada, I.J.; Díaz-Simal, P.; Izaguirre, C.; Camus, P. Multi-sectoral, high-resolution assessment of climate change consequences of coastal flooding. *Clim. Change* **2017**, *145*, 431–444. https://doi.org/10.1007/s10584-017-2104-z.

3.4. Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast

Abstract and Keywords

Due to climate change, hydroclimatic extremes are becoming more frequent and intense and their characterization and analysis is essential for climate modelling. One of the regions that will be most affected by these extremes is the Mediterranean coast of Spain. Therefore, this paper analyses the significant wave height (Hs), peak wave period (Tp) and sea level (SL) extremes and their correlation along the Spanish Mediterranean coast. After conducting this analysis, it is finally concluded that and adjustment of the extreme long-term distribution of Tp is urgently needed to create accurate models and projections, which must be considered in combination with the intense extremes happening in the Levantine basin when modelling this area and designing new projects.

Keywords: extremes; wave height; sea level; wave period; Mediterranean coast; Spain; modelling

1. Introduction

Climate change is causing temperatures to rise, which in turn is leading to rising sea levels and hydroclimatic extreme events, among others. In this context, it is crucial to analyse and properly model these events and their impacts on society, infrastructure and ecosystems.

One of the main hotspots of climate change is the Iberian Peninsula and its Mediterranean coast due to its location and orography [1]. Villeta et al., 2018 [2], proved an increase in the number of annual days with extreme minimum and maximum temperatures in the north of Spain. Quereda et al., 2018 [3] showed an increase of temperatures in Valencia. Lorenzo et al., 2021 [4] stated that in the whole Iberian Peninsula heatwaves will have a higher heat excess and heatwave days will become more intense in the near future. Kelebek et al., 2021 [5] highlighted that the aridity index over the Mediterranean has increased since the 1970s. As noted above, this increase of temperatures is provoking sea level rise (SLR) and hydroclimatic extreme events. Therefore, there are many scientific papers that study the consequences that climate change will have in Spain. In this

section, some of the studies carried out in Spain, specifically focused on hydroclimatic extreme events, are presented.

In 2008, Lionello et al. [6] simulated the changes of wave climate for future scenarios and saw a slightly downward trend in wave height. The same year, Kundzewicz et al. [7] studied the impacts of hydroclimatic extreme events on different economic sectors. In 2016, Rueda et al. [8] developed a methodology to characterize extreme wave heights by defining weather types of the synoptic circulation conditions, and Lin Ye et al. [9] developed a statistical model of extreme events that applied to the Catalan Coast. In 2016 also, Vousdoukas et al. [10] projected extreme storm surge levels along Europe and concluded a negative trend in the Bay of Biscay and no changes in the Mediterranean Sea. In 2018, Dental et al. [11] discussed the problems of using model data to evaluate extreme waves and two years later, they developed a methodology to integrate model data with buoy data. Wolff et al. [12] also developed a database for assessing the impact of extreme waves and SLR in the Mediterranean in 2018, and Sayol and Marcos [13] achieved the same for the Ebro Delta.

In 2020, Abdalazeez et al. [14] stated that gentle beaches acted like a "low-pass filter" for extreme wave heights and de Lalouvière et al. [15] pointed out the necessity of maintaining reef health to assure wave energy dissipation. In the last few years, there have also been many studies focused on SLR and extreme flooding, as it is one of the greatest challenges for today's society [16]. Eguibar et al. studied these flood hazards in Oliva, Valencia [17]; Hernandez-Mora et al. [18] in Tossa del Mar, Catalonia; Lopez-Doriga and Jimenez [19] in the Ebro Delta; Luque et al. [20] in the Balearic islands; and Carmen Llasat [21] studied the evolution of these phenomena in the Mediterranean region.

In 2021, de Alfonso et al. [22] studied the effects of the extreme storm Gloria in Spain. Amarouche, and Akpinar [23], and De Leo et al. [24] characterized extreme wave events in the Mediterranean Sea and pointed out an increasing trend. In 2022, Hsu et al. [25] studied the effect of extreme waves in vertical breakwaters stability, and Iglesias et al. [26] modelled climate change projections in estuarine regions.

Although, as has been seen above, there are a large number of climate studies along the Spanish coast, the Levante area is one of the least studied areas. The Levantine basin is an enclosed basin affected by regional wave and wind patterns that are different from the global patterns. Despite this condition, most of the

studies of the Levantine basin focus on very specific points, such as the Ebro Delta or the Catalan Maresme, leaving many areas without climatic studies [1,27,28]. Figure 1 presents a colour scale to illustrate the relative number of climate studies that have been carried out on each Spanish coast.



Figure 1. Number of studies in Spain. Taken from [1]

This article seeks to address this problem and attempts to tackle this lack of studies in the Levantine basin, focusing on the study and analysis of extreme storms along the Mediterranean coast of Spain. This analysis of extreme hydroclimatic events will help to fill the gap of studies in the Levantine basin and to characterize this part of the Spanish coast. Studying extreme storms is essential to avoid economic and human loss and structural damage. Particularly in this area, where major storms have already caused such things.

The structure of the paper is as follows: Section 2 describes the methodology. Section 3 shows the results obtained and the analysis carried out. Section 4 presents the discussion of the results, and Section 5 highlights the main conclusions and suggests new research directions.

2. Methodology

For this study, all the monthly maximums of Hs, Tp, and sea level (SL) along the Mediterranean Sea of the network of Spanish ports have been collected and analysed. The data records range from the operating start of buoys and tide-gauges

to the present. These records have also been compared to the annual reports and summaries of Spanish ports.

All the data were taken from the Spanish ports' website [29]. Data have been collected from the available buoys and tide gauges along the Spanish Mediterranean coast. Of all the available buoys and tide gauges, all those who met the following criteria were chosen for the analysis:

1. The buoys must have sufficiently long records, at least 10 years (without large data gaps). This criterion was chosen because in order to draw conclusions regarding extreme events, a data series of at least 10 years is necessary.
2. The buoys must be located in deep water, sufficiently far from the coast so that they are not affected by refraction, diffraction and shoaling phenomena that could alter the final results.

It should be noted that the Spanish port buoys are composed of scalar buoys of the Waverider (Datawell) type and directional buoys of the Triaxys™ (Axys) type.

Figure 2 shows all the buoys and tide gauges chosen for the study, and the buoys and tide gauges used are also listed in Table 1 and Table 2, respectively.

Table 1. List of buoys used in the study.

Buoy	Longitude	Latitude	Depth	Records
1. Gulf of Cadiz	6.96° O	36.49° N	466 m	1996–2022
2. Alborán	5.03° O	36.27° N	585 m	1997–2006
3. Cabo de Gata	2.34° O	36.57° N	545 m	1998–2022
4. Cabo de Palos	0.31° O	37.65° N	242 m	2006–2022
5. Valencia	0.20° E	39.51° N	349 m	2005–2022
6. Mahon	4.42° E	39.71° N	310 m	1993–2022
7. Tarragona	1.47° E	40.69° N	685 m	2004–2022
8. Begur	3.65° E	41.90° N	1355 m	2001–2022

Table 2. List of tide gauges used in the study.

Tide Gauge	Longitude	Latitude	Records
1. Tarifa	5.60° O	36.01° N	2009–2020
2. Algeciras	5.40° O	36.18° N	2009–2020
3. Almeria	2.48° O	36.83° N	2006–2020
4. Valencia	0.31° O	39.44° N	1993–2020
5. Mahon	4.27° E	39.89° N	2009–2020
6. Sagunto	0.21° O	39.63° N	2007–2020
7. Tarragona	1.21° E	41.08° N	2011–2020
8. Barcelona	2.17° E	41.34° N	1993–2020



Figure 2. Buoys and tide gauges used in the study.

The study was conducted in three different blocks: first, the values of Hs and Tp were studied; then the SL variable; and finally, the correlation between the extremes of all the variables (Hs, Tp and SL).

To analyse and construct the data of Hs and Tp, the first step was to choose the corresponding buoy of the section of oceanography of the Spanish ports' website.

Once the buoy was selected, the report was downloaded and the option "monthly maximums" was selected to download the data [30]. All the data were plotted and studied statistically. In addition, the parameters of α, ß, γ and λ; Hs and Tp for 20, 50 and 100 years and their evolution in time were studied in each buoy.

It is important to examine the evolution of α, ß, γ and λ because they are the statistical parameters that characterize the extreme long-term distribution. α is the Hs threshold, ß is the scale parameter and γ is the shape coefficient. α, ß and γ are the three parameters that characterize Weibull distribution (Equation 1)

$$F_w(H_s \geq x) = 1 - exp\left(-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}\right) \quad (1)$$

λ is the number of the average number of storms occurring in a year and it is normally used in the POT (Peaks Over the Threshold) distribution (Equation 2).

$$H_{s|Tr|} = \alpha + \beta\left[-ln\frac{1}{\lambda \mathrm{Tr}}\right]^{1/\gamma} \quad (2)$$

Tp for 20, 50 and 100 years are the Tp expected in 20, 50 and 100 years, respectively. These values are obtained applying the POT distribution (Equation 2) for the 20, 50 and 100 years return period. Once the Hs is obtained, Equation 3 is applied

$$Tp = aHs^b \quad (3)$$

The parameters a, b, α, ß, γ and λ are obtained for each buoy from "Puertos del Estado" reports [31].

To analyse and construct the data of SL, the procedure followed was very similar to the method explained above but instead of taking the buoy records, tide gauge records were downloaded [32].

Finally, a statistical study was carried out to analyse the correlation of the hydroclimatic extremes. In this part of the study, the methodology was different. First, the monthly extremes of SL and the day and hour when the event happened were collected. Then, a search for the wave occurring at that hour was carried out and the values were plotted. Therefore, the main variable in the correlation analysis conducted was the SL, and the secondary value was the wave height.

3. Results

This section is divided into four subsections. Section 3.1 is the study of Hs and Tp, in which the evolution of the parameters Hs and Tp of each buoy is presented.

Section 3.2 is the study of extreme statistical parameters. Section 3.3 is the study of SL, similar to Section 3.1, but analysing the SL variable. Section 3.4 is the correlation of hydroclimatic extremes, in which the correlation of the extremes of Hs, Tp and SL is studied.

3.1. Study of Hs and Tp Monthly Extremes

The evolution of monthly Hs extreme records is presented in Figures 3–10. The Y-axis is the Hs in m and X-axis the time. All the X-axes of this section have been unified to facilitate the comparison of the different spots. All the figures presented have an X-axis that goes from January 2000 to March 2022 even if there were more records or less than this year frame. As mentioned above, it was decided to unify all graphs in order to facilitate the interpretation and comparison of the results.

Each graph also shows the trend line following the maxima of Hs and its analytical expression (in the colour red). This was carried out by means of a linear adjustment because it was considered to give the most visual and tangible image of the trend followed by the extremes. This will facilitate the analysis of the results and the reaching of the conclusions presented in Sections 4 and 5.

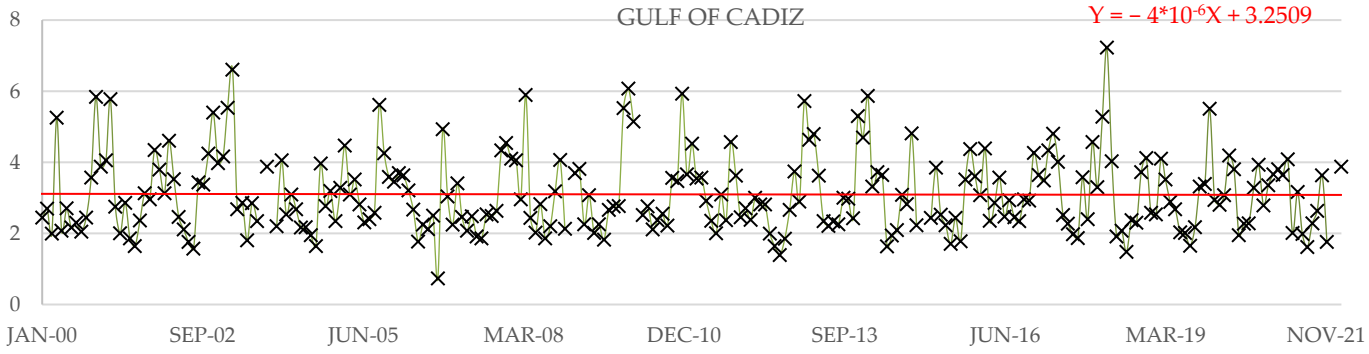


Figure 3. Evolution of Hs maxima in the Gulf of Cadiz.

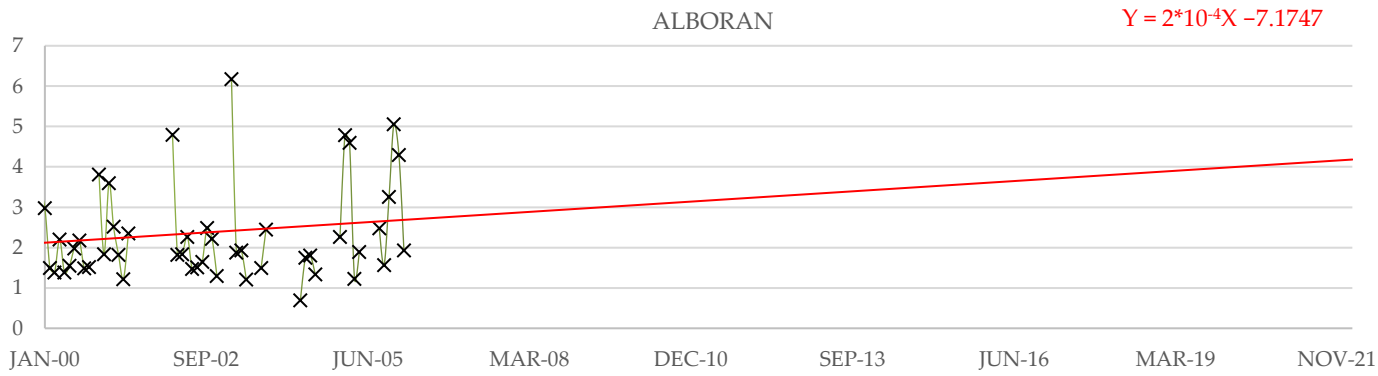


Figure 4. Evolution of Hs maxima in Alboran.

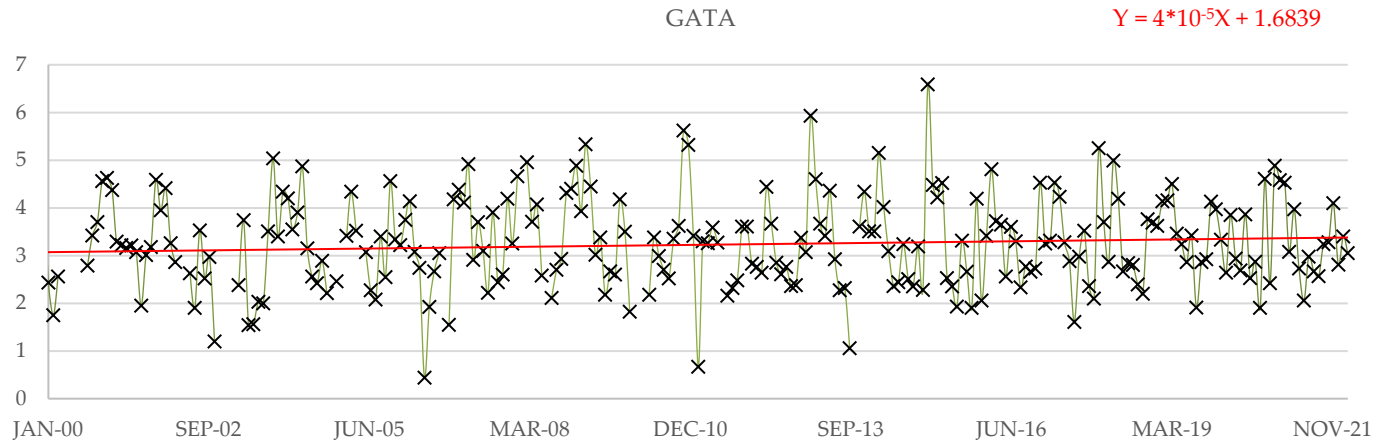


Figure 5. Evolution of Hs maxima in Cabo de Gata.

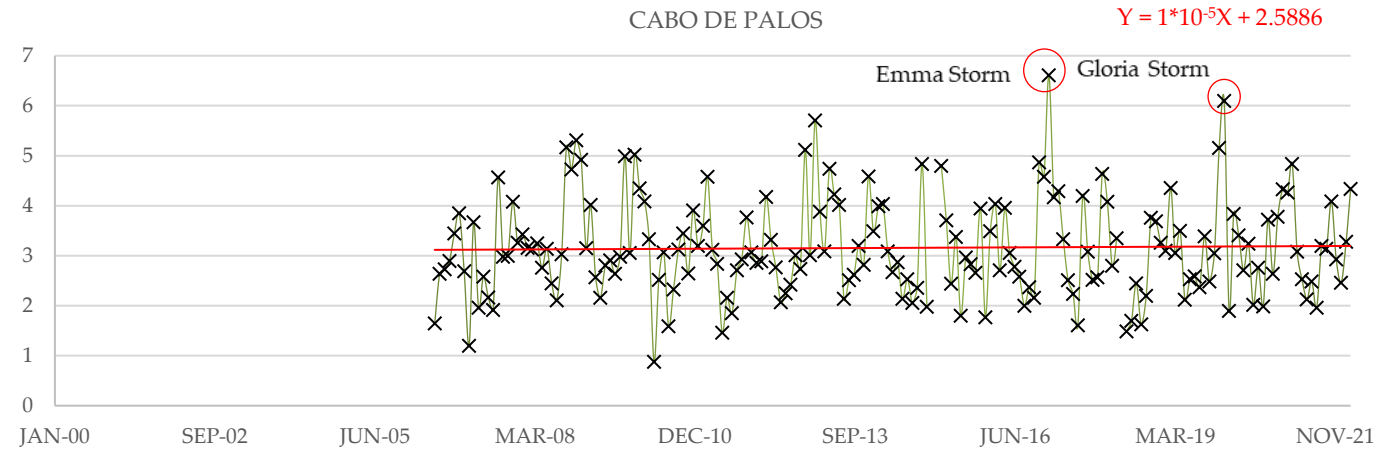


Figure 6. Evolution of Hs maxima in Cabo de Palos.

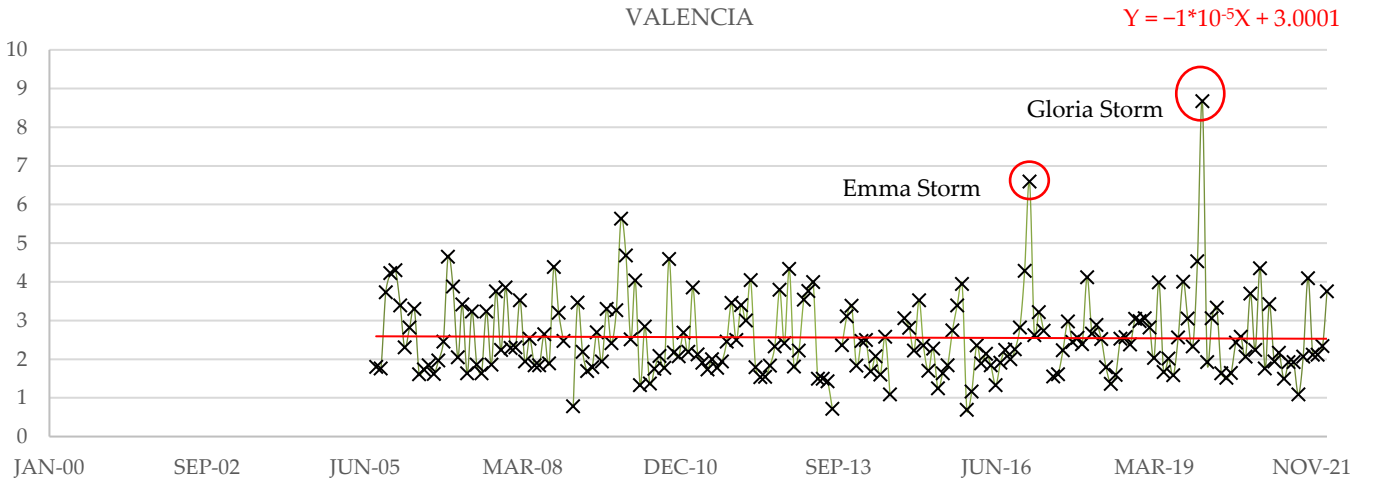


Figure 7. Evolution of Hs maxima in Valencia.

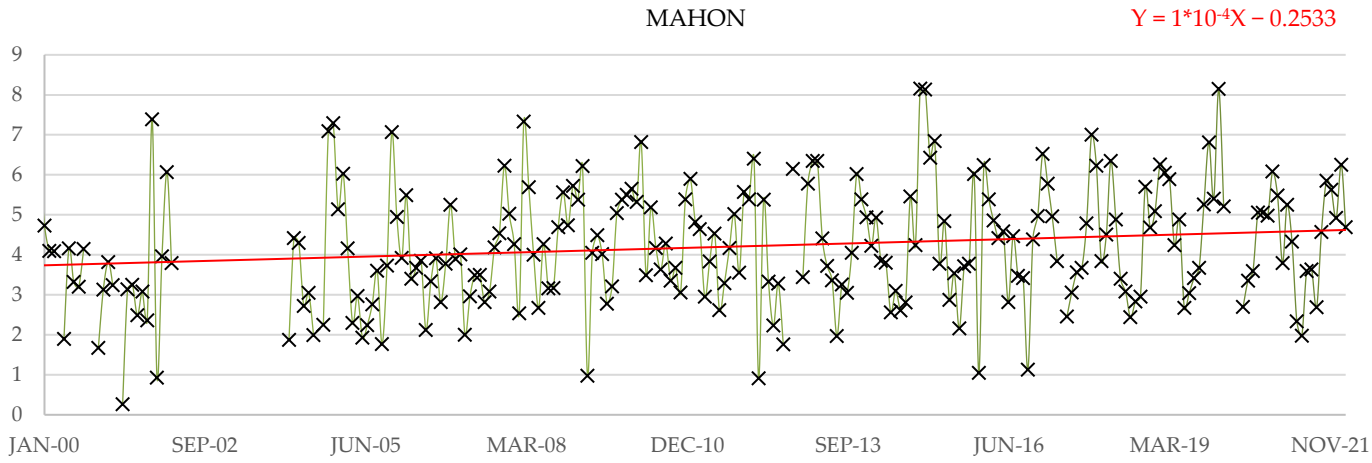


Figure 8. Evolution of Hs maxima in Mahon.

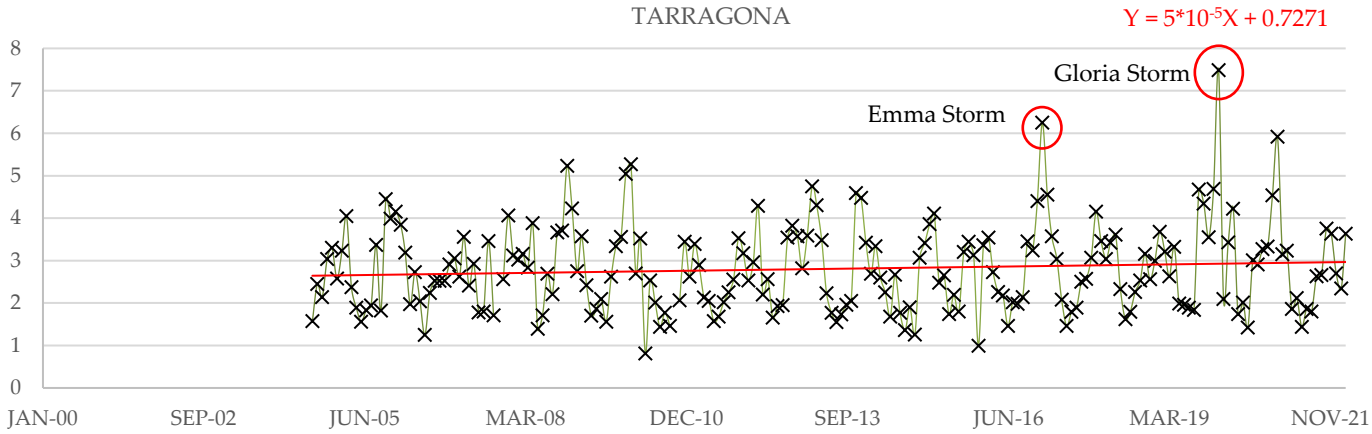


Figure 9. Evolution of Hs maxima in Tarragona.

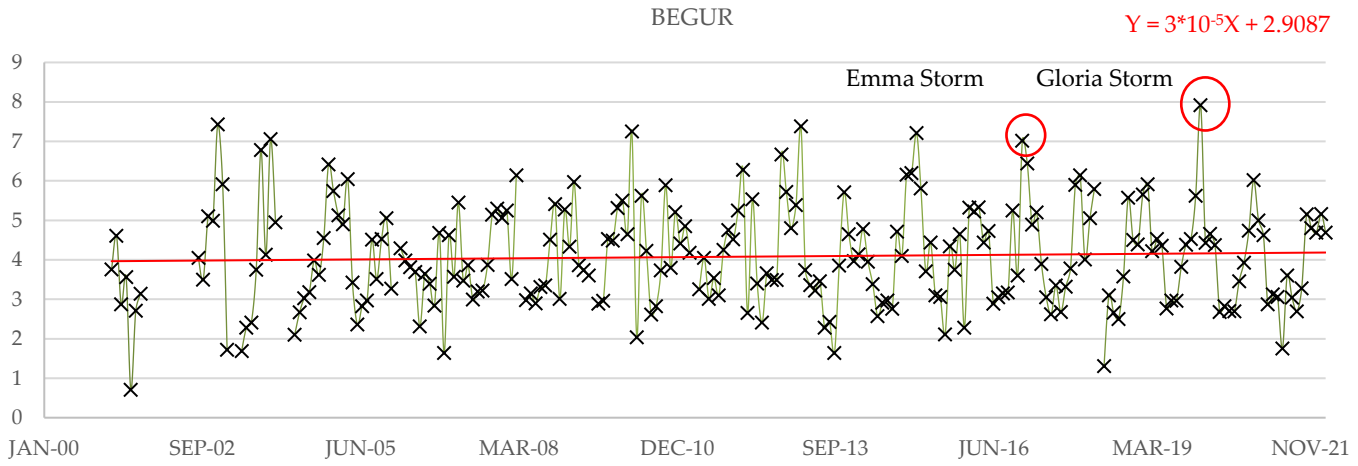


Figure 10. Evolution of Hs maxima in Begur.

The evolution of monthly Tp extremes is presented in Figures 11–18. As was done with Hs, the trend line is presented in the graphs and the adjustment performed is linear due to the same reason (colour red). The Y-axis is the Tp in s, and the X-

axis the time. The X-axis of the graphs has also been unified and all X-axes of Figures 11–18 range from January 2000 to March 2022.

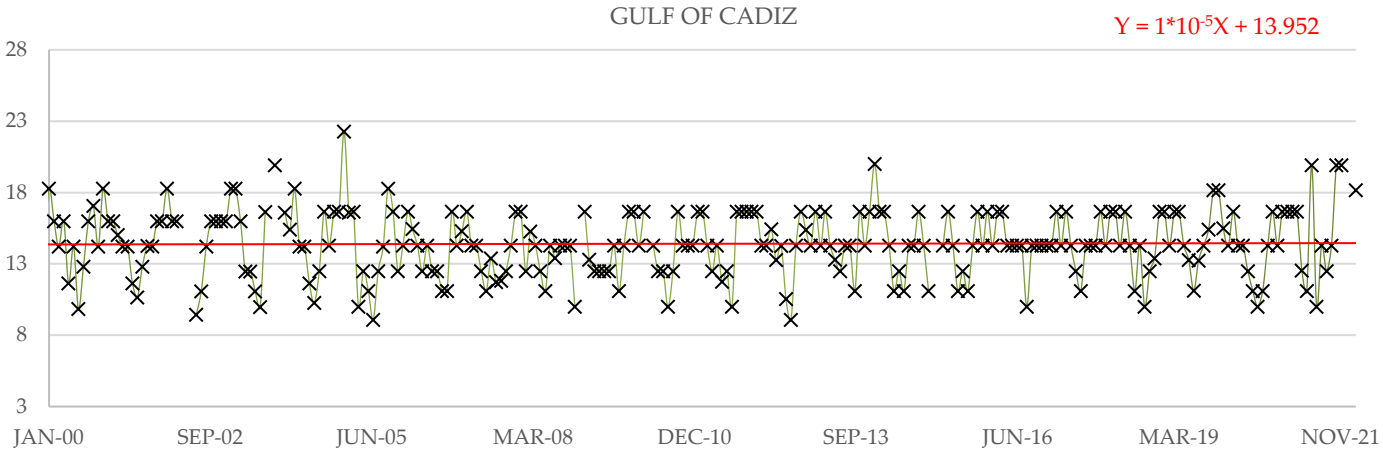


Figure 11. Evolution of Tp maxima in the Gulf of Cadiz.

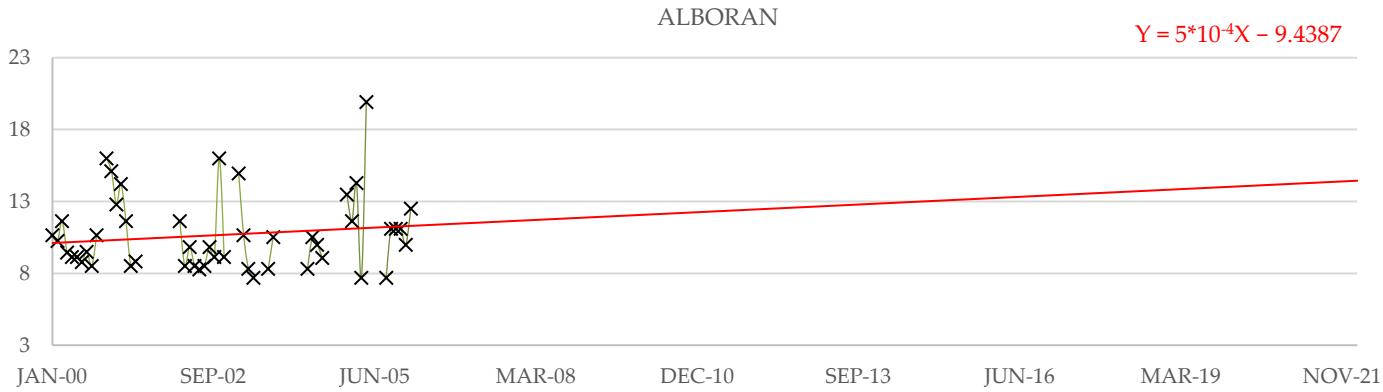


Figure 12. Evolution of Tp maxima in Alboran.

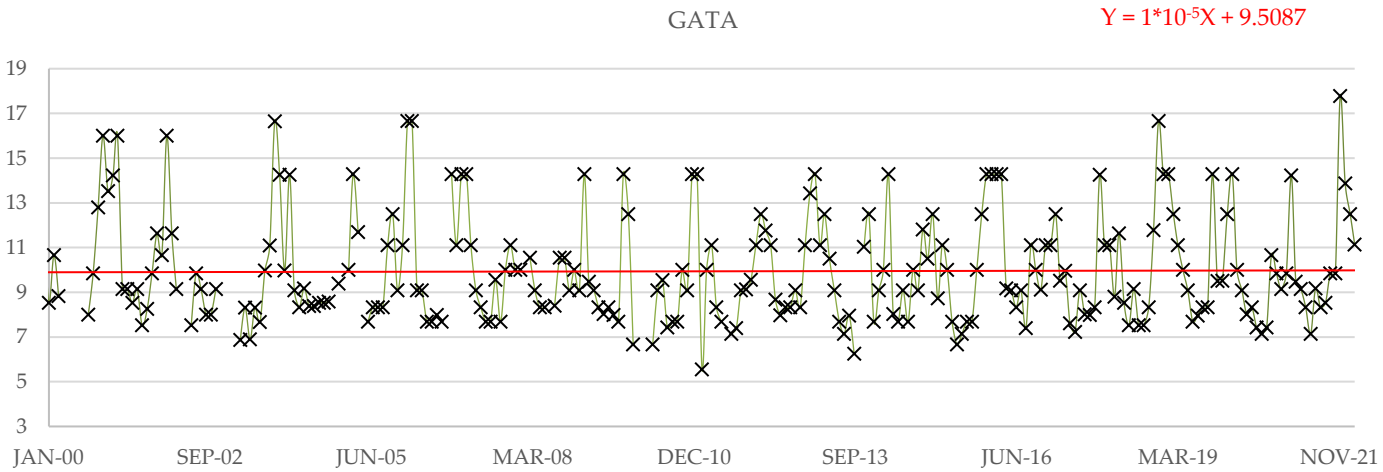


Figure 13. Evolution of Tp maxima in Cabo de Gata.

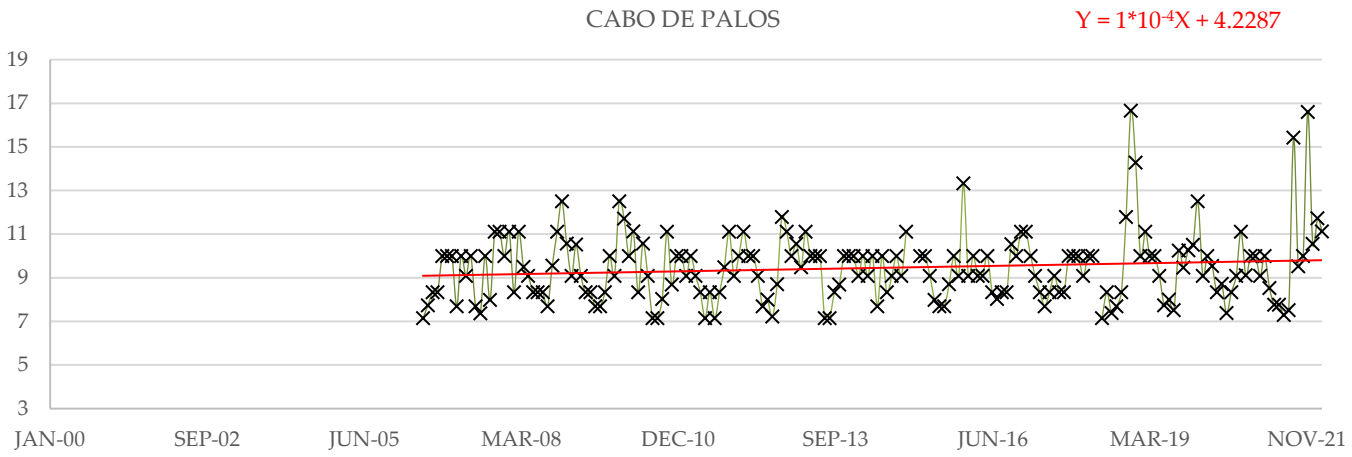


Figure 14. Evolution of Tp maxima in Cabo de Palos.

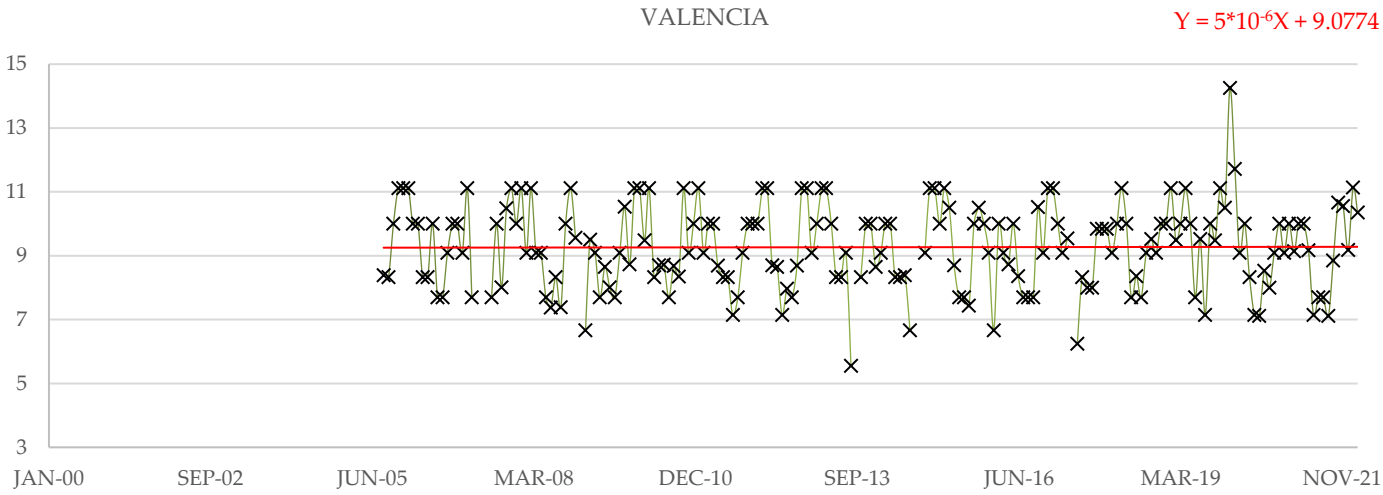


Figure 15. Evolution of Tp maxima in Valencia.

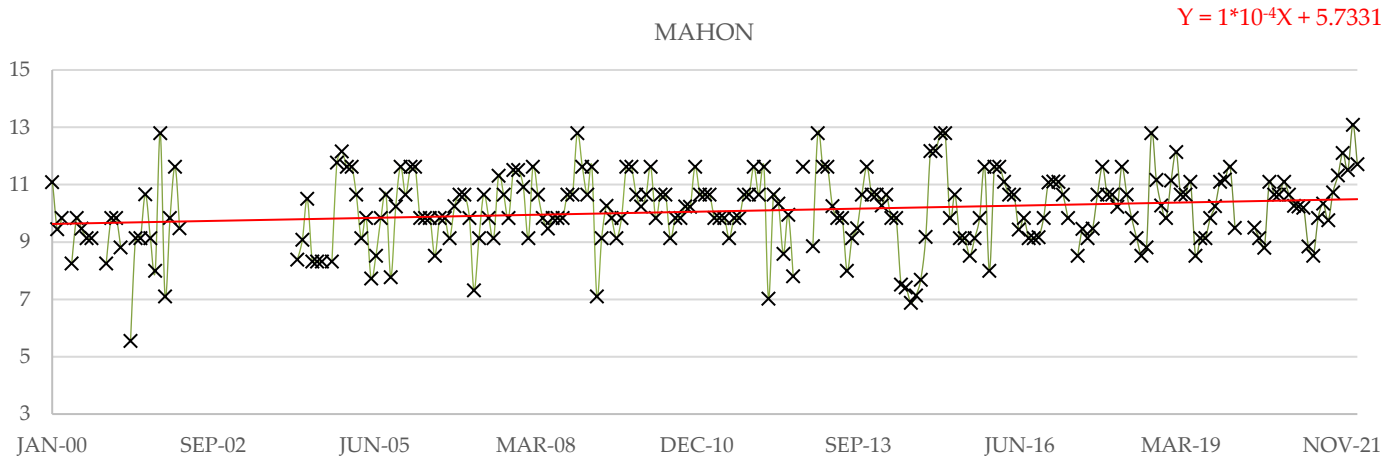


Figure 16. Evolution of Tp maxima in Mahon.

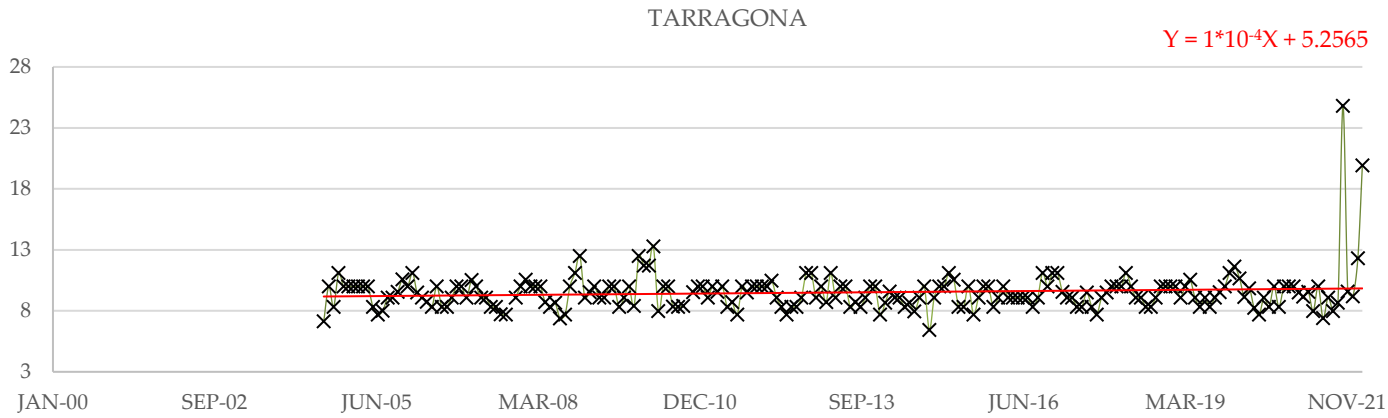


Figure 17. Evolution of Tp maxima in Tarragona.

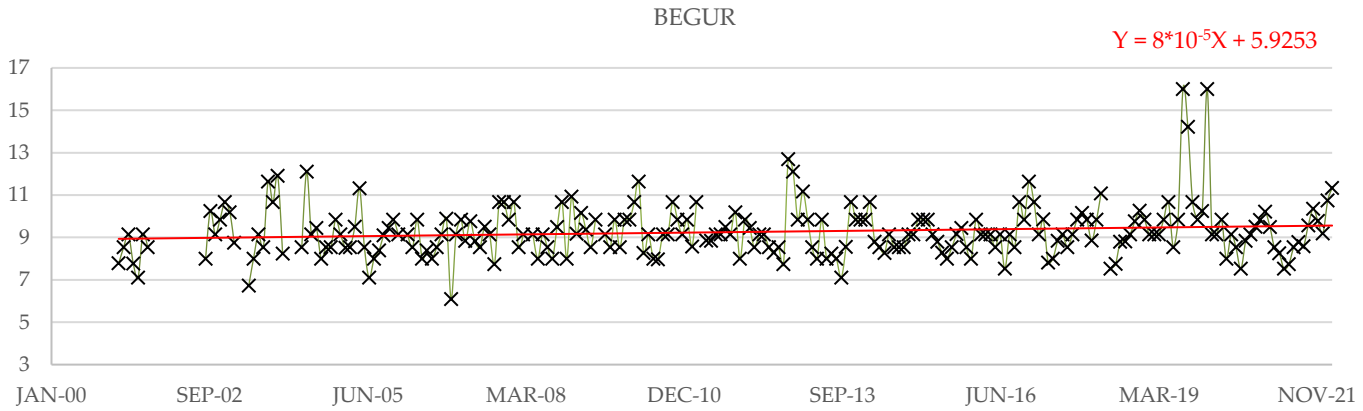


Figure 18. Evolution of Tp maxima in Begur.

Table 3 presents the summary of the slopes of the fitted lines of both the Hs and Tp plots. These results will be discussed later in Section 4, as well as their significance in the context of hydroclimatic extremes and future projections.

Table 3. Summary of Hs and Tp extreme trends for each buoy.

Buoy	Hs Slope	Tp Slope
Gulf of Cadiz	−4x10-6	1x10-5
Alboran	3x10-4	5x10-4
Gata	4x10-5	1x10-5
Cabo de Palos	1x10-5	1x10-4
Valencia	−1x10-5	5x10-6
Mahon	1x10-4	1x10-4
Tarragona	5x10-5	1x10-4
Begur	3x10-5	8x10-5

3.2. Study of Extreme Statistical Parameters

The study of the parameters of the extreme statistical distribution is presented in this section. Figures 19–26 present the number of times Hs has exceeded the exceedance threshold α per year.

The Y-axis is the number of exceedances per month, and the X-axis the time. In this section, all the X-axes have also been unified to facilitate the comparison of the different spots. All the figures presented has an X-axis that ranges from 2000 to 2022, even if there were more records or less than the frame of that year. As mentioned above, it was decided to unify all graphs in order to facilitate the interpretation and comparison of the results.

All the figures also present a linear adjustment to facilitate a subsequent trend study of extremes (colour red).

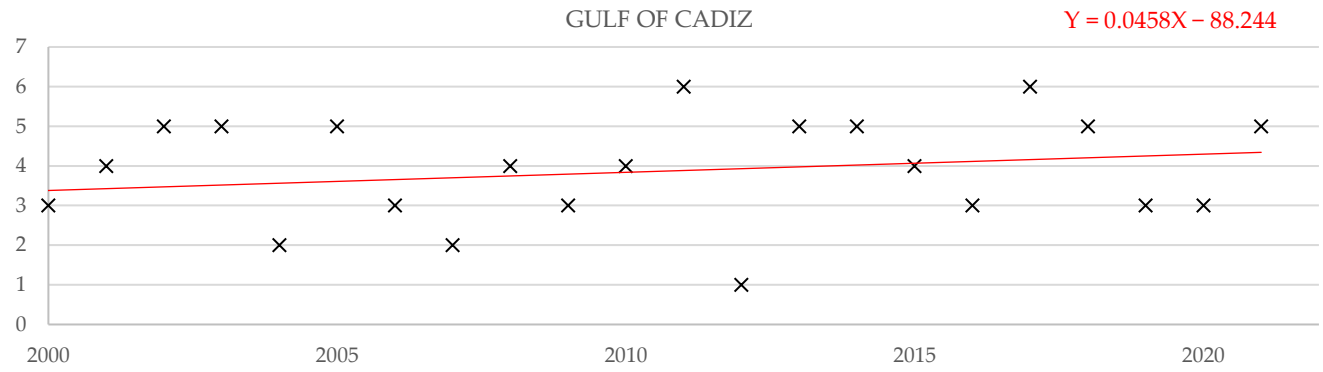


Figure 19. Number of times Hs has exceeded the threshold in the Gulf of Cadiz.



Figure 20. Number of times Hs has exceeded the threshold in Alboran.

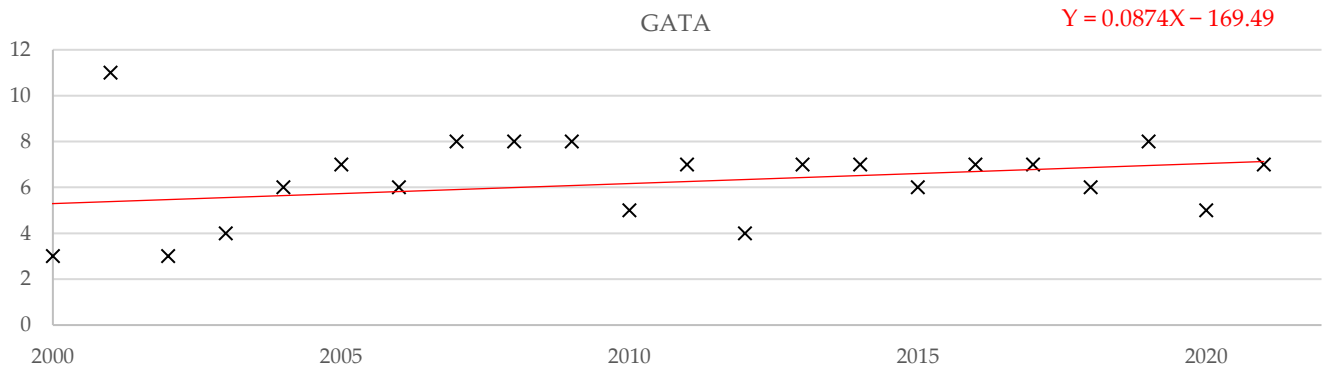


Figure 21. Number of times Hs has exceeded the threshold in Gata.

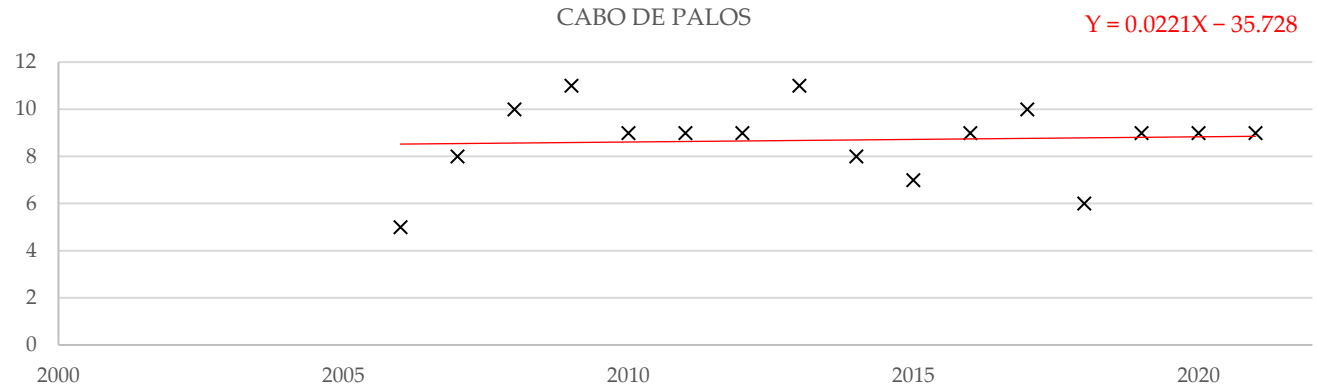


Figure 22. Number of times Hs has exceeded the threshold in Cabo de Palos.

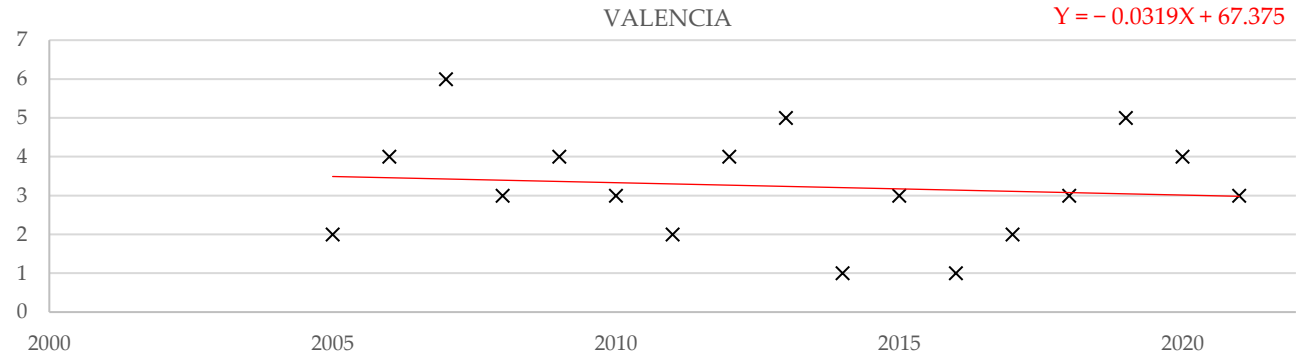


Figure 23. Number of times Hs has exceeded the threshold in Valencia.

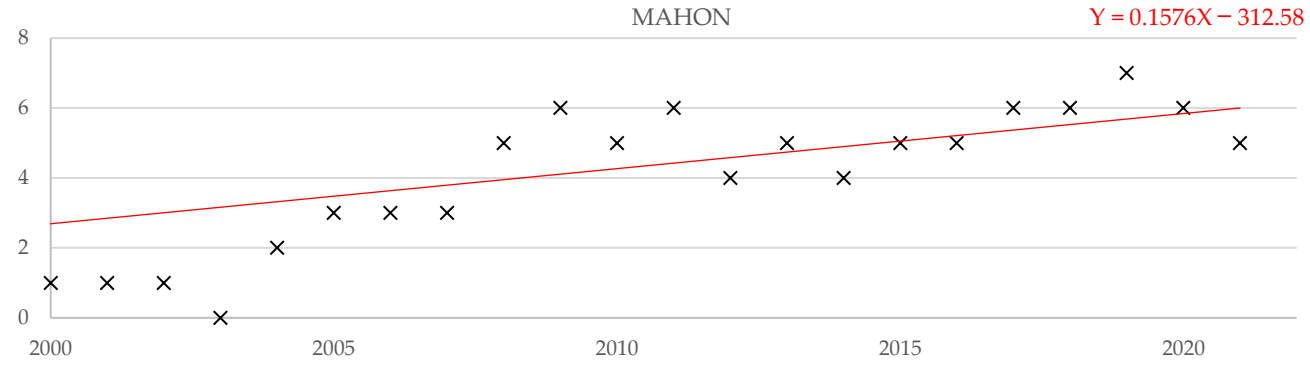


Figure 24. Number of times Hs has exceeded the threshold in Mahon.

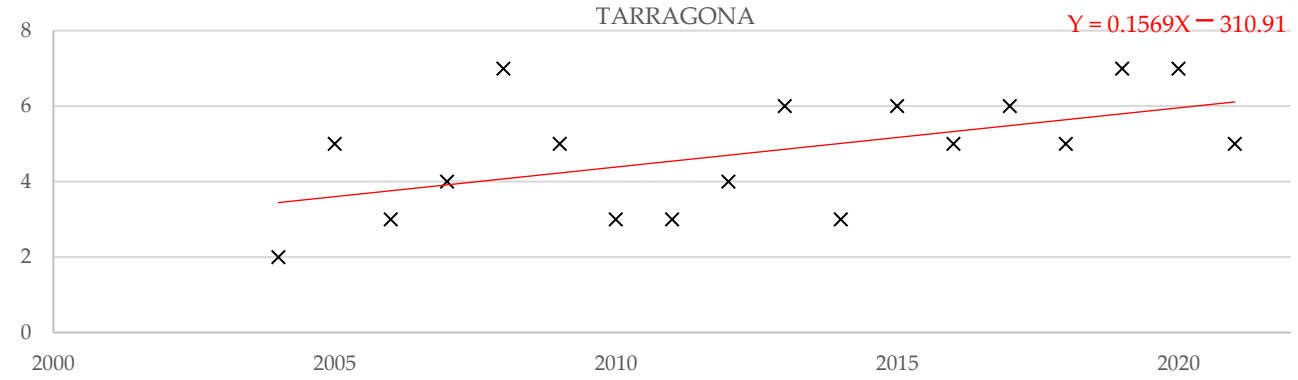


Figure 25. Number of times Hs has exceeded the threshold in Tarragona.

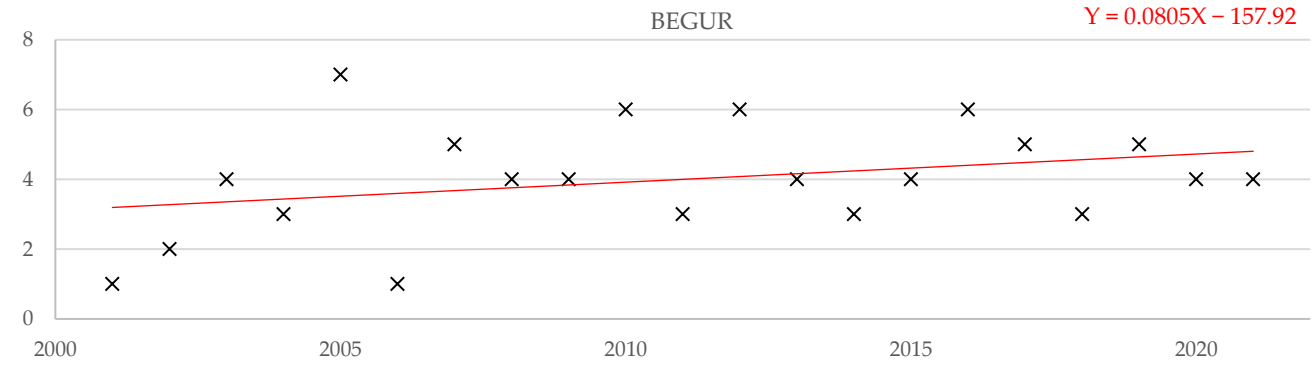


Figure 26. Number of times Hs has exceeded the threshold in Begur.

Figures 27–32 present the times that Tp has exceeded the expected Tp in 100 years. As with Hs, the Y-axis presents the number of exceedances and the X-axis is time, and all X-axes are unified. In addition, all the graphs present a linear adjustment to facilitate the analysis of the trend of the hydroclimatic extremes (colour red). Valencia results are not presented because in the case of this buoy, the Tp in 100 years was only exceeded twice, once in 2019 and once in 2021.

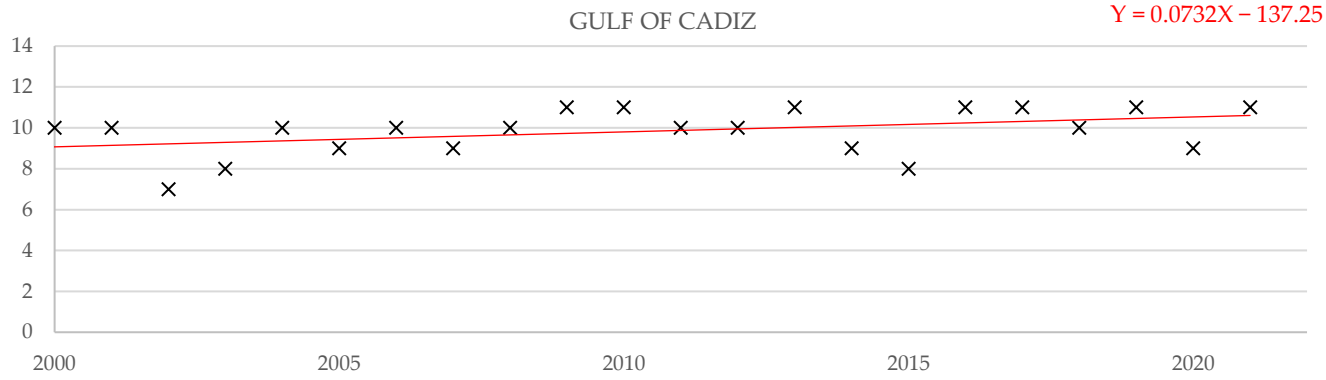


Figure 27. Number of times Tp has exceeded the expected Tp in 100 years in the Gulf of Cadiz.



Figure 28. Number of times Tp has exceeded the expected Tp in 100 years in Alboran.



Figure 29. Number of times Tp has exceeded the expected Tp in 100 years in Gata.

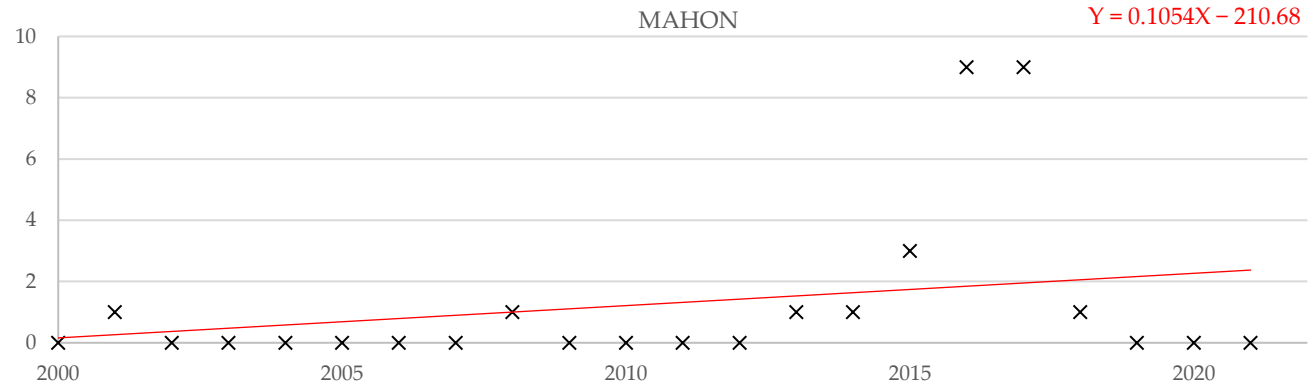


Figure 30. Number of times Tp has exceeded the expected Tp in 100 years in Mahon.

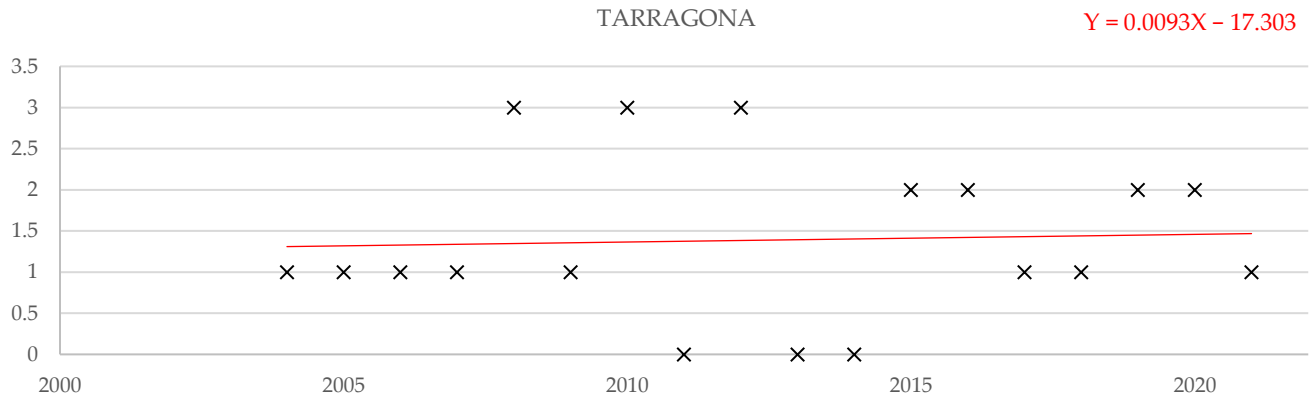


Figure 31. Number of times Tp has exceeded the expected Tp in 100 years in Tarragona.

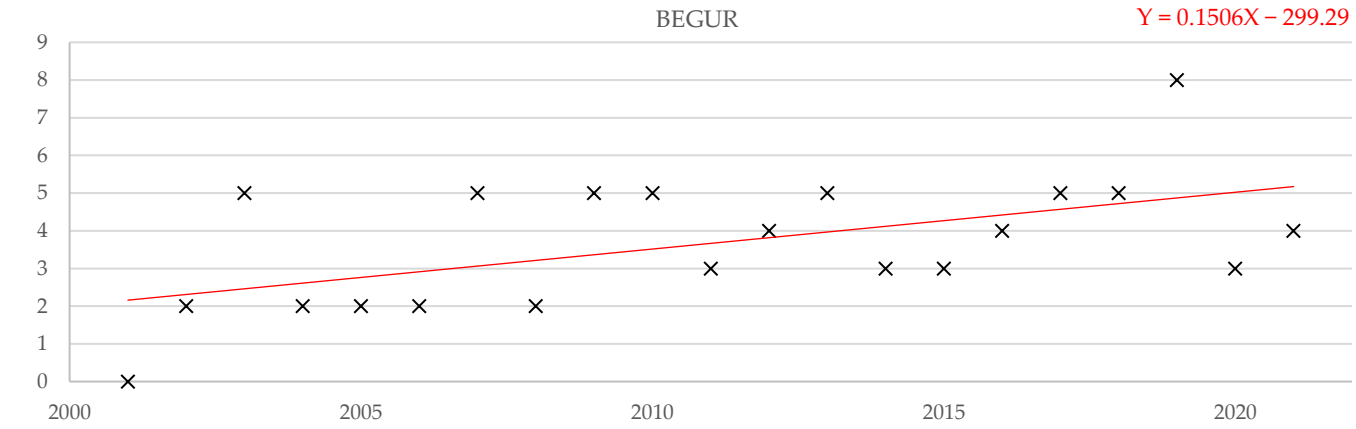


Figure 32. Number of times Tp has exceeded the expected Tp in 100 years in Begur.

These graphs will be analysed in Section 4. In addition, the parameters α, ß, λ and γ per se and the Spanish ports' reports of the years 2010, 2011, 2012, 2014, 2015, 2017 and 2020 have been studied [33]. The results were the following:

In the Gulf of Cadiz, all the parameters remained stable. In Alboran, λ increased and the parameters ß and γ were less than one. In Cabo the Gata, λ was slightly

increasing, while γ was slightly decreasing and β remained stable. In Cabo de Palos, β and γ were decreasing and λ was increasing. In Valencia, β and γ were also decreasing and γ was less than one. In Mahon, γ was stable and β decreased. In Tarragona, both γ and β decreased and γ was less than one too. In Begur, γ was stable and β decreased.

It has to be said that the reports of the different years of Gata and Begur had different initial years, which may alter the final results when comparing reports from one year to the next.

3.3. Study of SL monthly extremes

In this section, the evolution of the monthly maximums of SL in each tide gauge is presented from Figures 33–40. It is structured as the previous sections:

The Y-axis is the SL extremes in m, and the X-axis the time. The X-axes of all the graphs is unified from January 2006 to December 2020. Each graph also shows the trend line (colour red).

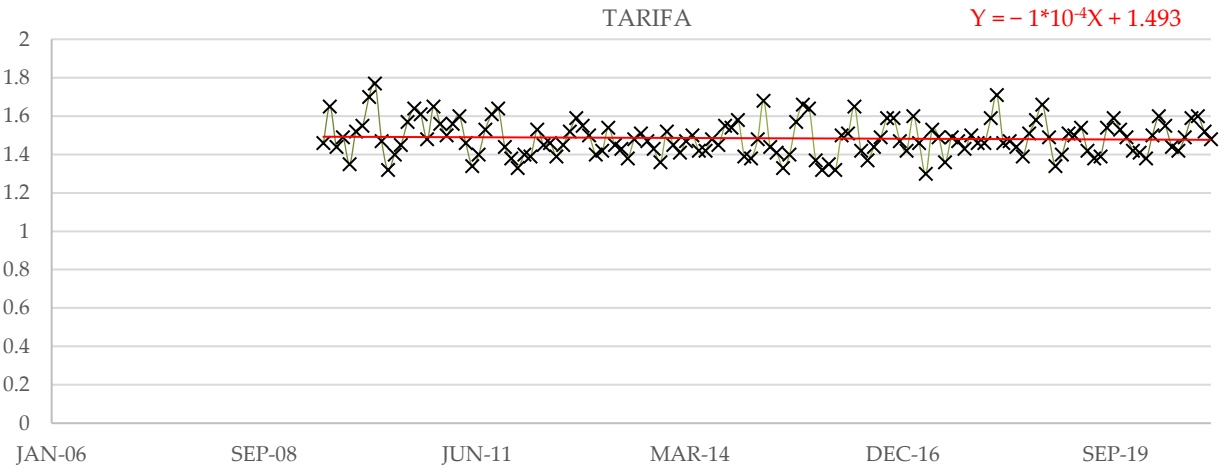


Figure 33. Evolution of SL maxima in Tarifa.

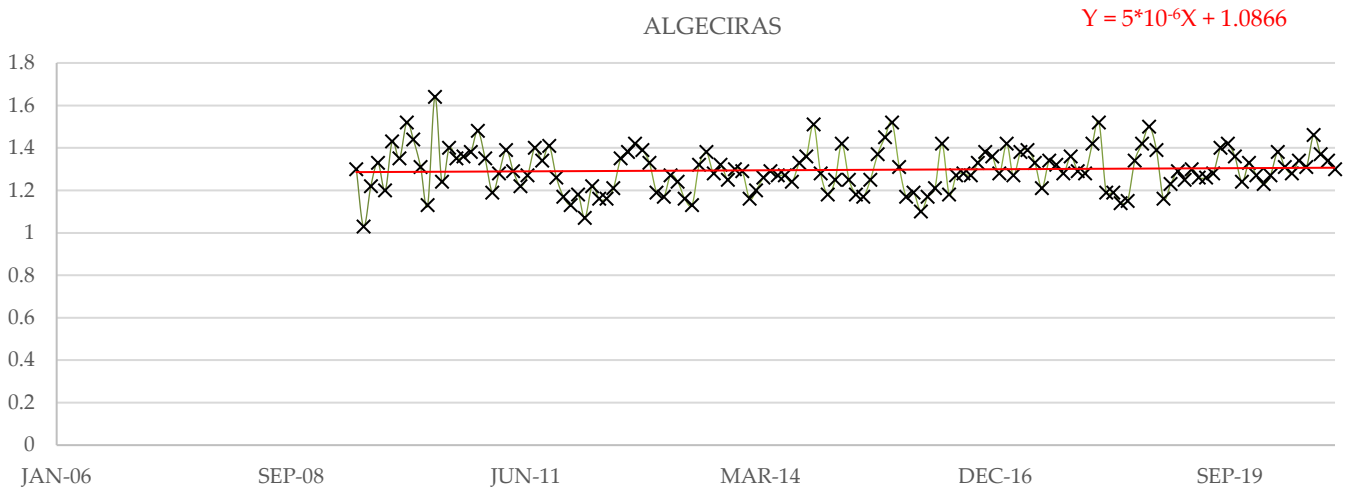


Figure 34. Evolution of SL maxima in Algeciras.

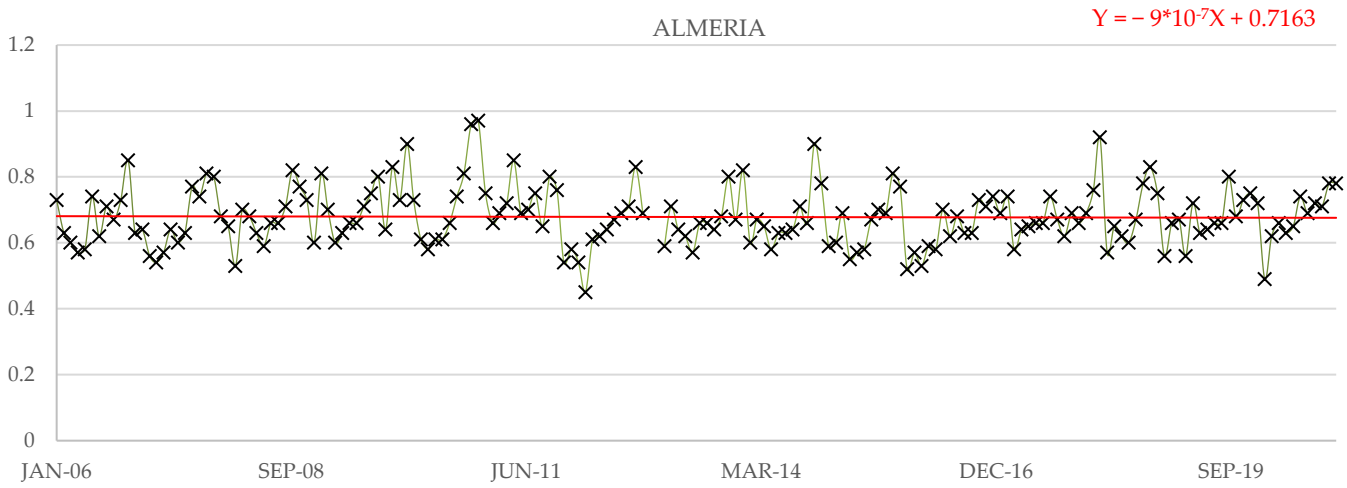


Figure 35. Evolution of SL maxima in Almeria.

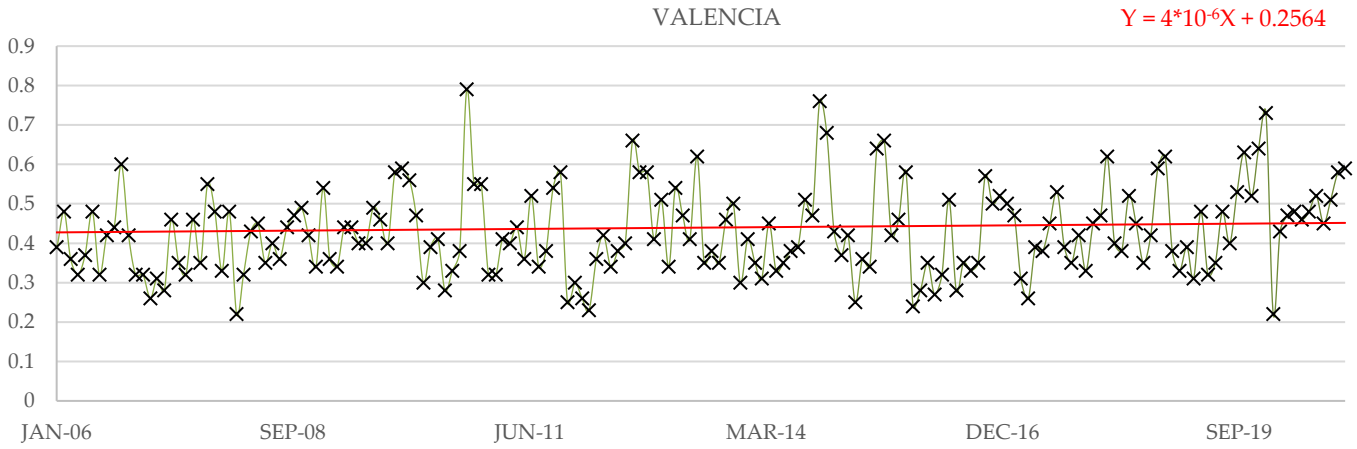


Figure 36. Evolution of SL maxima in Valencia.

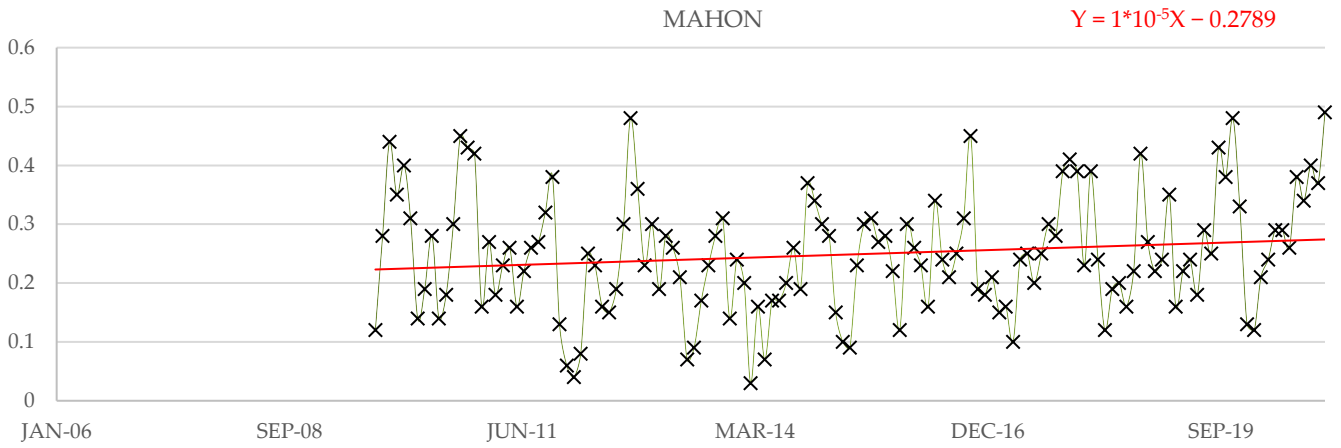


Figure 37. Evolution of SL maxima in Mahon.

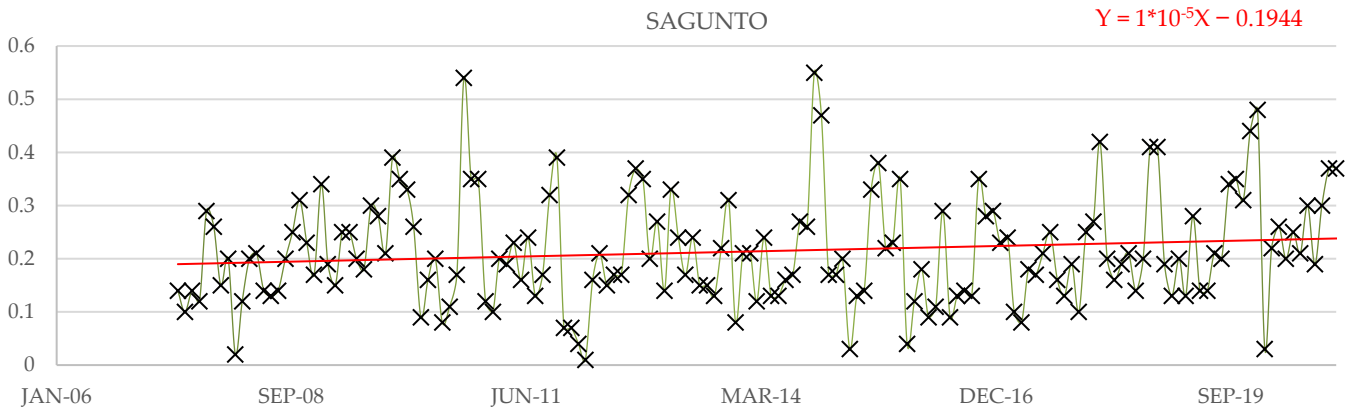


Figure 38. Evolution of SL maxima in Sagunto.

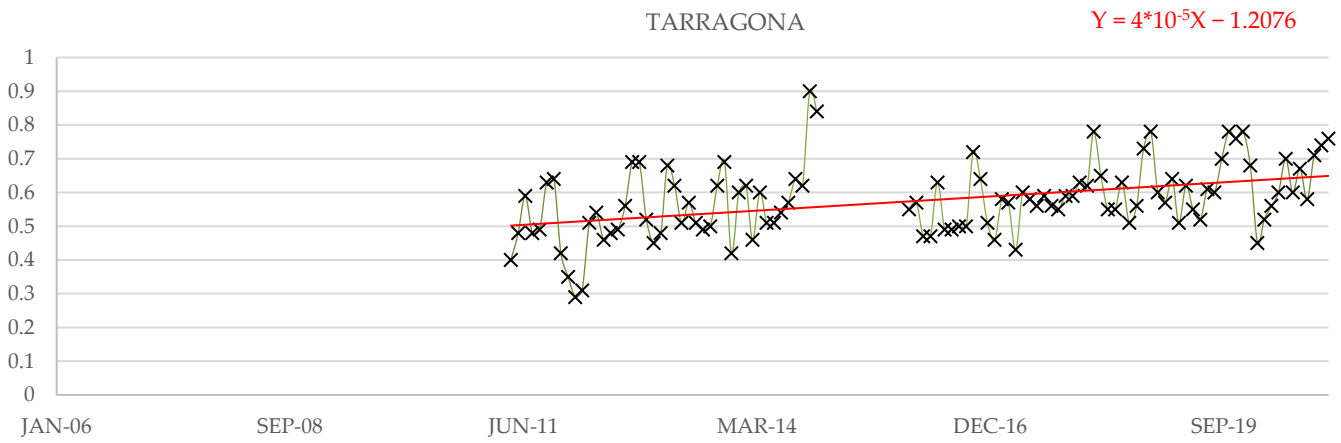


Figure 39. Evolution of SL maxima in Tarragona from 2011 to 2020.

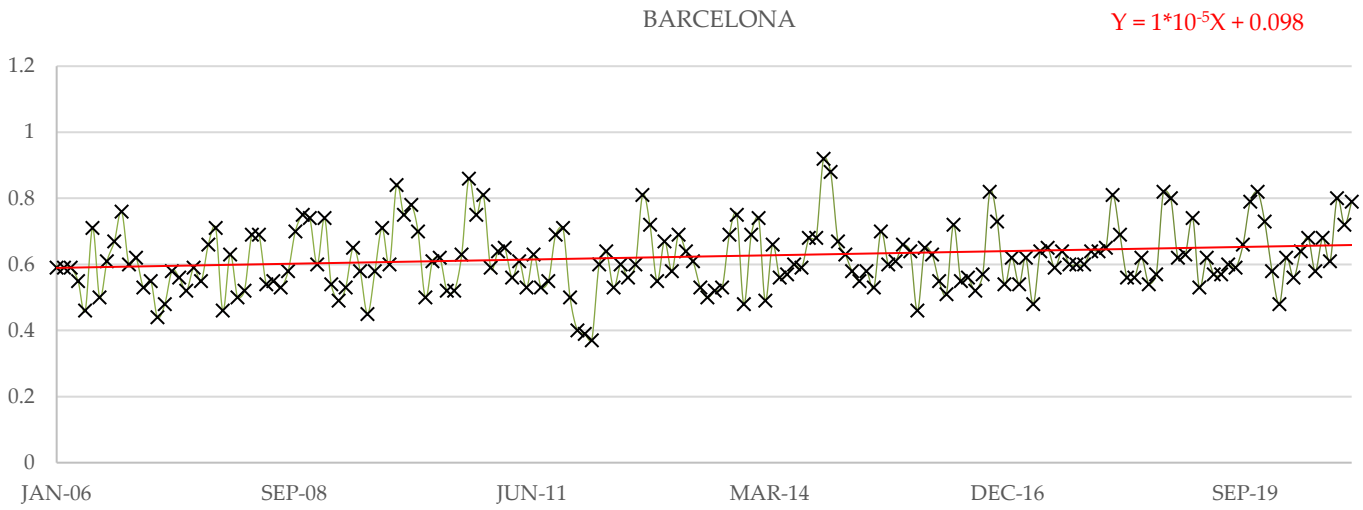


Figure 40. Evolution of SL maxima in Barcelona.

In Table 4, the slope of each tide gauge is presented, which is basically the trend of the SL. This will be further explained in Section 4.

Table 4. Summary of SL slopes.

Tide Gauge	SL Slope
1. Tarifa	−1x10-4
2. Algeciras	5x10-6
3. Almeria	−9x10-7
4. Valencia	4x10-6
5. Mahon	1x10-5
6. Sagunto	1x10-5
7. Tarragona	4x10-5
8. Barcelona	1x10-5

3.4. Correlation of Hydroclimatic extremes

The correlation of SL and Hs maxima has also been studied. Only the correlation in which deep buoys were available has been studied. Therefore, the correlation of maxima in Tarifa and Algeciras have not been analysed, because it was considered that, as the buoys were not in deep water, the results were not reliable. Graphs are presented from Figures 41–46.

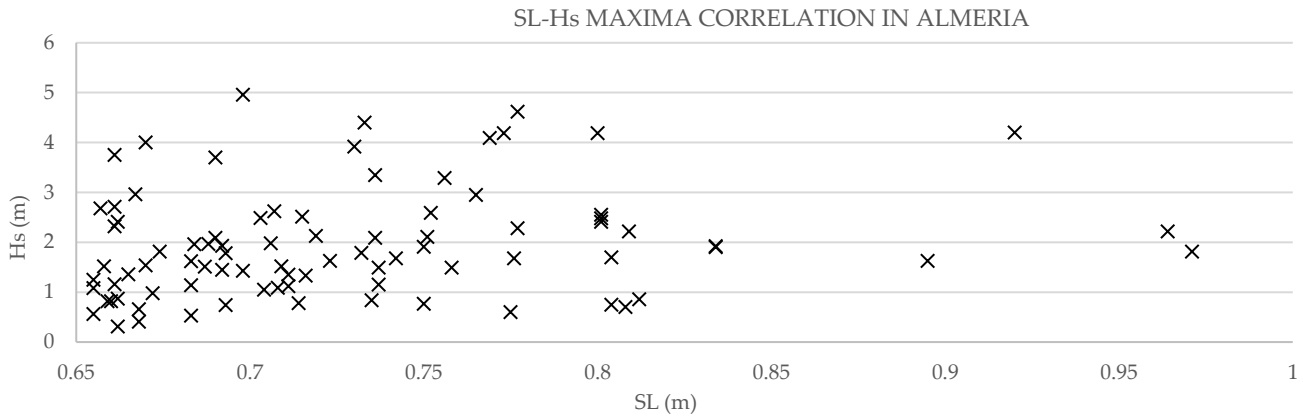


Figure 41. Correlation of extremes in Almeria.

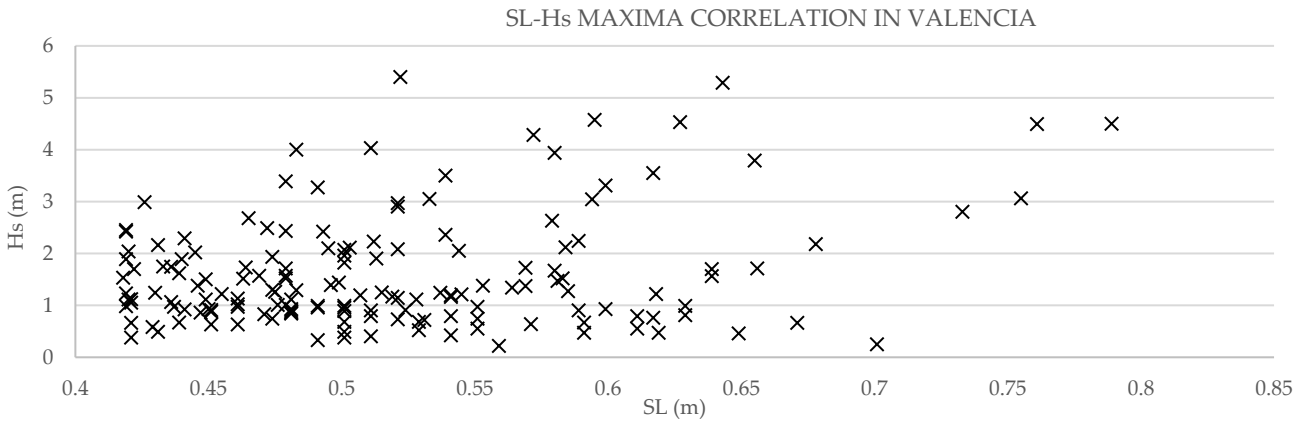


Figure 42. Correlation of extremes in Valencia.

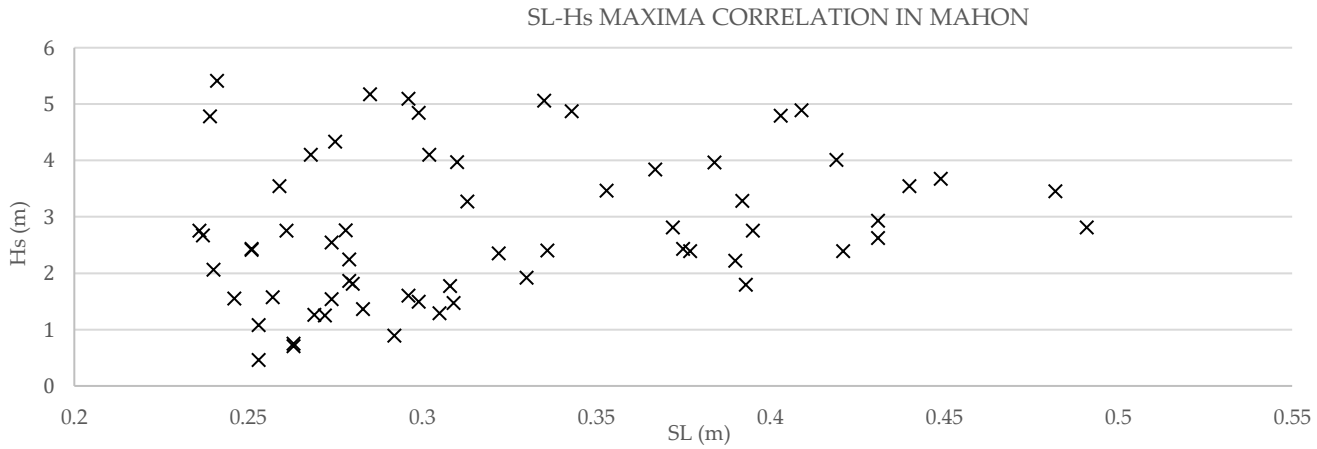


Figure 43. Correlation of extremes in Mahon.

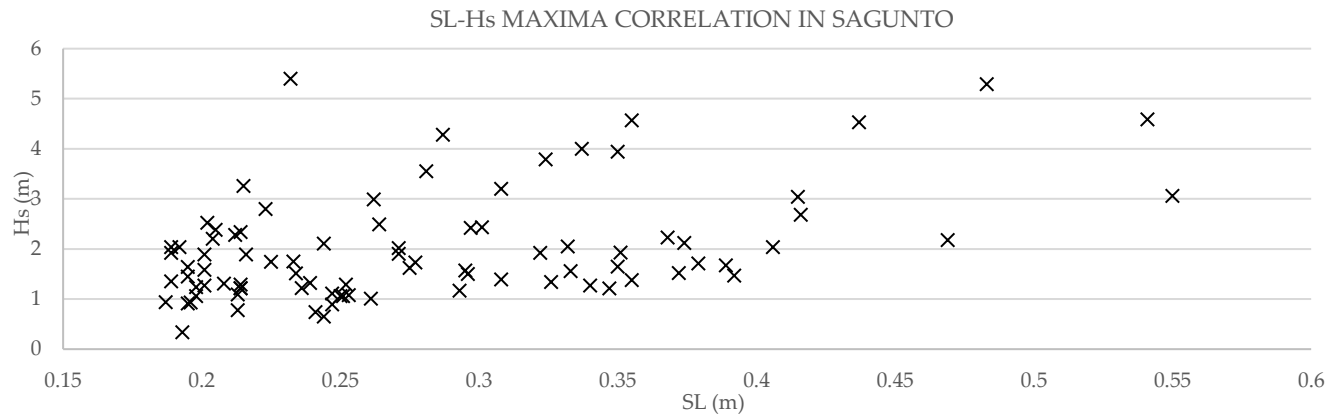


Figure 44. Correlation of extremes in Sagunto.

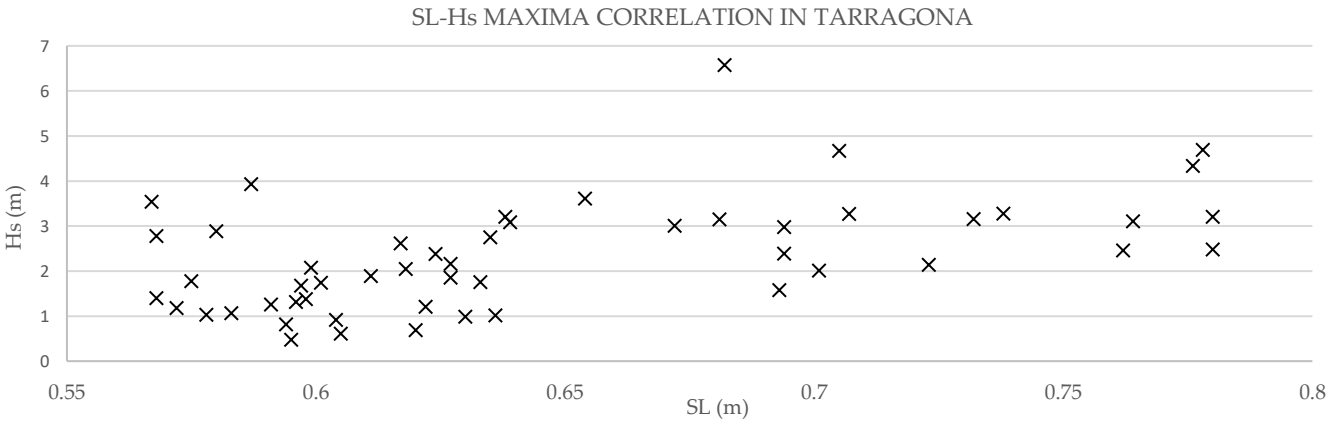


Figure 45. Correlation of extremes in Tarragona.

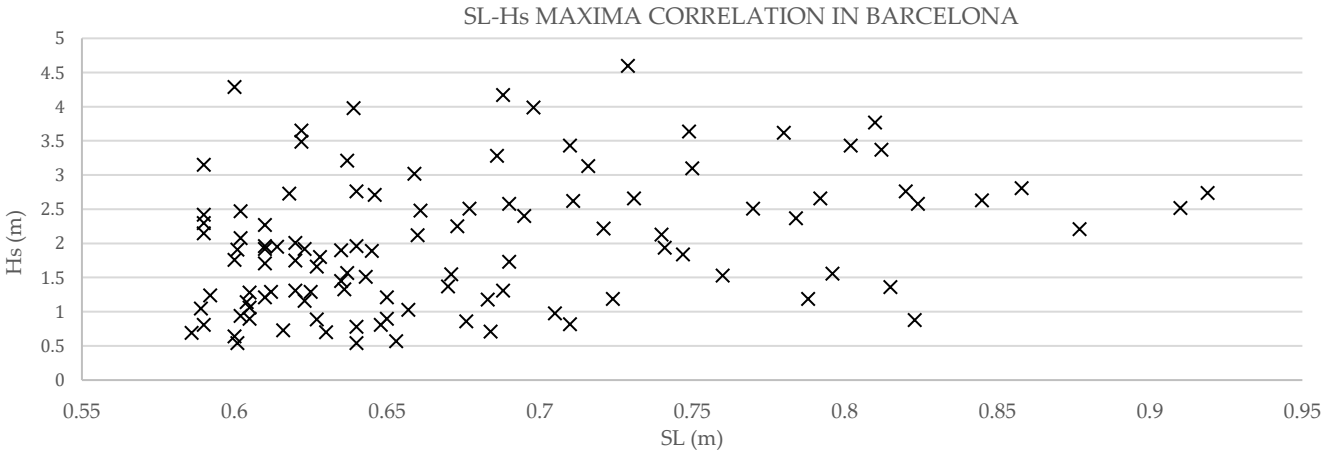


Figure 46. Correlation of extremes in Barcelona.

4. Discussion

4.1. Discussion of Hs and Tp Monthly Extremes

After analysing Hs and Tp plots (Figures 3–18), it can be seen that, referring to Hs extremes, all had an increasing trend, except for the area of the Gulf of Cadiz and Valencia, which had a slightly decreasing tendency. When talking about Tp, all the buoys had also a rising tendency and, the increase of Tp extremes was quicker than the increase of Hs, which is related to the results of Section 3.2.

The Alboran graph (Figure 4) is one of the graphs with the steepest trend line in Hs extremes (0.0003 slope). This is probably because the records of the Alboran Sea buoy are from 1997 to 2006. From 2006 onwards there are no more records. If there were records from 2006 to 2022, the increasing trend line would probably be much smoother. Furthermore, the Alboran Sea is an area affected by local and regional phenomena due to its position. It is an enclosed sea where the wave and wind patterns are very particular, so part of this increasing trend may also be due to these local phenomena and not just to the data series.

Mahon is the other spot where the increasing trend is sharper (0.0001 slope). In this buoy, the problem of the records does not exist. Therefore, this increasing trend can be taken with a high level of certainty, as the records reach from 1993 to 2022. The fact that Hs is increasing more sharply here than in the rest of the Spanish Mediterranean coast is likely due to the wave and wind patterns of this particular area. This clear increasing trend in Hs highlights the need for further studies of climate extremes in this area.

From Cabo de Gata to Begur (Figures 5–7, 9 and 10), the increasing trend of Hs is similar (around 10−4–10−5), excepting the Valencia buoy, which has a slightly decreasing trend. This area may be affected by the same meteorological regional patterns and fluid dynamics. The increasing trend is not as pronounced as in Alboran or Mahon. However, it has to be said that although the trend is less upward, this does not mean that this area is less threatened by extreme events. Although the general trend is less abrupt, this area is prone to very specific but very intense events, as can be seen in the graphs for the 2018 and 2020 storm (Emma and Gloria storm), which far exceeds the rest of the extreme storms. In fact, these “very extreme” events are the most damaging and threatening, and this part of the coast is especially vulnerable to extreme storms. This is also the main

reason why the data of these graphs seem to be very scattered, because in this area hydroclimatic extremes are very occasional and intense.

Referring to the Tp climate variables, the conclusions are the same as with Hs. The only difference, as noted above, is that the increasing trend of Tp is more pronounced than that of Hs. This makes sense since Hs and Tp are related, according to the expression $Tp = a * Hs^b$.

4.2. Discussion of extreme statistical parameters

The wave height threshold was typically exceeded in 30% of cases. The wave heights for 20, 50 and 100 years were exceeded at most once in the entire data series, while with the peak periods, up to 85% of the current data exceeded the expected peak period in 100 years (Gulf of Cadiz). Typically, around 20% of the peak periods of the year exceeded the expected 100-year peak period, which is a representative percentage considering that Tp is the Tp expected in 100 years. The fact that a 20% of the current maxima overpasses the Tp expected in 100 years is proof that the estimated Tp for 100 years already exists in current temperature due to the effect of climate change on extreme hydroclimatic phenomena.

The Hs for 20, 50 and 100 years was rarely exceeded in all the spots studies, but the Tp estimated for 20, 50 and 100 years from now was exceeded in some cases by up to 85% of the monthly maxima. This also shows that a statistical readjustment of this variable Tp is already necessary, as the current values do not accurately represent the extreme Tp. Although the increase in Hs may not be significant, the increase in Tp is significant and an adjustment of the statistical values and parameters is needed to adjust the models to the new reality. If this is not carried out, climate modelling and future projections will become increasingly inaccurate.

Referring to the specific spots studied. The Gulf of Cadiz is one of the buoys that need a more urgent adjustment of the Tp statistical distribution, as the estimated Tp 100 years from now was already exceeded in the 85% of the extreme storms. Valencia, on the contrary, has very well adjusted Tp parameters and the Tp in 20, 50 and 1000 years was hardly exceeded. Tarragona and Mahon are other spots that could be said to be well-adjusted, as the 100-year Tp was only exceeded in 10% of the extreme storms. Finally, the area of Valencia would also need a revision of the statistical parameters but the number of exceedances is not as high as in Cadiz.

4.3. Discussion of SL Monthly Extremes

Referring to SL results, almost all the tide gauges showed an increase in sea level. However, it is striking that the tide gauge at Tarifa shows an appreciable drop in sea level, which may be due to regional effects in the area and would need to be studied in more depth in order to reach reliable conclusions.

As is logical, the area where the maximum sea levels vary the most is the area from Cadiz to Almeria due to the astronomical tides. On the other hand, there is the area of Valencia and Cabo de Palos, whose tidal range is 0. Here, even though the tide is practically nil, it can be seen that sea level peaks are also reached very occasionally, following a similar behaviour to that explained by Hs and Tp extremes in Section 4.1., which increase the vulnerability of these areas to hydroclimatic extreme events. If concomitant extreme events (Hs, Tp and SL) occur in this area, the damage and losses can be devastating, as extreme events in this area of the Mediterranean are very intense. This will be further discussed in the next subsection.

Finally, referring to the area of Barcelona and Tarragona, SL maxima variability is also higher because of the tides.

The general pattern is that SL maxima are increasing, as has been proven by the scientific community. The area where this increasing trend is sharper is the Levantine basin, which makes it more vulnerable to extreme storms, as this paper has already stated.

4.4. Discussion of the Correlation of the Hydroclimatic Extremes

Finally, the correlation of Hs and SL was analysed. As mentioned in the previous subsection, assessing the concomitance and correlation of hydroclimatic extremes is essential to avoid future damage. In addition, the Levant Mediterranean area is particularly vulnerable to occasional, but very intense, extreme events.

In Almeria, extreme SL were associated mostly with medium waves, while extreme waves happened with medium SL.

In Valencia, Mahon and Sagunto, extreme SL occurred with relatively high waves, which should be considered in future designs so as not to underestimate extremes and their concomitance. The fact that extreme events in this area can be concomitant further aggravates their situation, as they already have a devastating effect on their own. If maximums of all three variables were to occur at the same

time, the damage could be even greater, which requires specific studies in the Levante area.

In Barcelona and Tarragona, the concomitance of the hydroclimatic extremes were not as clear as in the Levantine basin with Valencia, Mahon and Sagunto, but it can be appreciated how some of the highest waves and SL also occurred with relatively high SL or waves, respectively. However, it is not as relevant as it is in the Levantine basin.

5. Conclusions

After analysing and discussing all the results, several conclusions can be highlighted.

Mahon and Alboran are the spots in the Spanish Mediterranean were the increasing trend of Hs and Tp extremes is sharper. This increasing trend can be taken with a high level of certainty in Mahon and with a medium level of certainty in Alboran due to the short-period data frame.

Although the increasing trend of Hs maxima in the Levantine basin is not as sharp as in other parts of the Spanish Mediterranean coast, it is subjected to the most intense one-off events. These events are the ones that cause the most damage, making the Spanish Levante region particularly vulnerable to extreme hydroclimatic events.

A readjustment of the Tp statistical parameter is needed. Expected Tp for 100 years is reached in more than 20% of extreme storms in most of the spots studied, which means that the estimated Tp for within 100 years is already happening. The most critical spot is the Gulf of Cadiz with an 85% of exceedance.

SL is increasing along all the Spanish Mediterranean coast, excepting Tarifa, which may be due to local phenomena. The behaviour is very similar to Hs and Tp extreme events but affected by the astronomical tide.

The area with the highest concomitance of extreme events is the Levantine basin. Having SL, Hs and Tp extreme events at the same time would be devastating. Therefore, more in-depth studies on the coast from Valencia to Cabo de Palos are needed, and even more so considering that this is the area of the Spanish coast where the fewest scientific studies have been carried out.

Considering all these conclusions, future research directions should focus on studying the concomitance of hydroclimatic extreme events (Hs, Tp and SL) in the Levantine basin and in the adjustment of statistical Tp parameters.

References

1. Portillo Juan, N.; Negro Valdecantos, V.; del Campo, J.M. Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain. *Sustainability* **2022**, *14*, 7507.
2. Villeta, M.; Valencia, J.L.; Saa, A.; Tarquis, A.M. Evaluation of extreme temperature events in northern Spain based on process control charts. *Theor. Appl. Climatol.* **2018**, *131*, 1323–1335. https://doi.org/10.1007/s00704-017-2038-4.
3. Quereda, J.; Monton, E.; Vázquez, V. La elevación de las temperaturas en el norte de la Comunidad Valenciana: Valor y naturaleza (1950–2016). *Investig. Geográficas* **2018**. https://doi.org/10.14198/INGEO2018.69.03.
4. Lorenzo, N.; Díaz-Poso, A.; Royé, D. Heatwave intensity on the Iberian Peninsula: Future climate projections. *Atmos. Res.* **2021**, *258*, 105655. https://doi.org/10.1016/j.atmosres.2021.105655.
5. Kelebek, M.B.; Batibeniz, F.; Önol, B. Exposure Assessment of Climate Extremes over the Europe–Mediterranean Region. *Atmosphere* **2021**, *12*, 633.
6. Lionello, P.; Cogo, S.; Galati, M.B.; Sanna, A. The Mediterranean surface wave climate inferred from future scenario simulations. *Glob. Planet. Change* **2008**, *63*, 152–162. https://doi.org/10.1016/j.gloplacha.2008.03.004.
7. Kundzewicz, Z.W.; Giannakopoulos, C.; Schwarb, M.; Stjernquist, I.; Schlyter, P.; Szwed, M.; Palutikof, J. Impacts of climate extremes on activity sectors–stakeholders' perspective. *Theor. Appl. Climatol.* **2008**, *93*, 117–132. https://doi.org/10.1007/s00704-007-0327-z.
8. Rueda, A.; Camus, P.; Mendez, F.J.; Tomas, A.; Luceno, A. An extreme value model for maximum wave heights based on weather types. *J. Geophys. Res. -Ocean.* **2016**, *121*, 1262–1273. https://doi.org/10.1002/2015jc010952.
9. Lin-Ye, J.; Garcia-Leon, M.; Gracia, V.; Sanchez-Arcilla, A. A multivariate statistical model of extreme events: An application to the Catalan coast. *Coast. Eng.* **2016**, *117*, 138–156. https://doi.org/10.1016/j.coastaleng.2016.08.002.
10. Vousdoukas, M.I.; Voukouvalas, E.; Annunziato, A.; Giardino, A.; Feyen, L. Projections of extreme storm surge levels along Europe. *Clim. Dyn.* **2016**, *47*, 3171–3190. https://doi.org/10.1007/s00382-016-3019-5.
11. Dentale, F.; Furcolo, P.; Pugliese Carratelli, E.; Reale, F.; Contestabile, P.; Tomasicchio, G.R. Extreme Wave Analysis by Integrating Model and Wave Buoy Data. *Water* **2018**, *10*, 373.
12. Wolff, C.; Vafeidis, A.T.; Muis, S.; Lincke, D.; Satta, A.; Lionello, P.; Jimenez, J.A.; Conte, D.; Hinkel, J. A Mediterranean coastal database for assessing the impacts of sea-level rise and associated hazards. *Sci. Data* **2018**, *5*, 180044. https://doi.org/10.1038/sdata.2018.44.
13. Sayol, J.M.; Marcos, M. Assessing Flood Risk Under Sea Level Rise and Extreme Sea Levels Scenarios: Application to the Ebro Delta (Spain). *J. Geophys. Res. Ocean.* **2018**, *123*, 794–811. https://doi.org/10.1002/2017jc013355.
14. Abdalazeez, A.; Didenkulova, I.; Dutykh, D.; Labart, C. Extreme Inundation Statistics on a Composite Beach. *Water* **2020**, *12*, 1573.

15. de Lalouvière, C.l.H.; Gracia, V.; Sierra, J.P.; Lin-Ye, J.; García-León, M. Impact of Climate Change on Nearshore Waves at a Beach Protected by a Barrier Reef. *Water* **2020**, *12*, 1681.
16. Griggs, G.; Reguero, B.G. Coastal Adaptation to Climate Change and Sea-Level Rise. *Water* **2021**, *13*, 2151.
17. Eguibar, M.Á.; Porta-García, R.; Torrijo, F.J.; Garzón-Roca, J. Flood Hazards in Flat Coastal Areas of the Eastern Iberian Peninsula: A Case Study in Oliva (Valencia, Spain). *Water* **2021**, *13*, 2975.
18. Hernandez-Mora, M.; Meseguer-Ruiz, O.; Karas, C.; Lambert, F. Estimating coastal flood hazard of Tossa de Mar, Spain: A combined model-data interviews approach. *Nat. Hazards* **2021**, *109*, 2153–2171. https://doi.org/10.1007/s11069-021-04914-3.
19. Lopez-Doriga, U.; Jimenez, J.A. Impact of Relative Sea-Level Rise on Low-Lying Coastal Areas of Catalonia, NW Mediterranean, Spain. *Water* **2020**, *12*, 3252. https://doi.org/10.3390/w12113252.
20. Luque, P.; Gomez-Pujol, L.; Marcos, M.; Orfila, A. Coastal Flooding in the Balearic Islands During the Twenty-First Century Caused by Sea-Level Rise and Extreme Events. *Front. Mar. Sci.* **2021**, *8*, 676452. https://doi.org/10.3389/fmars.2021.676452.
21. Carmen Llasat, M. Floods evolution in the Mediterranean region in a context of climate and environmental change. *Cuad. Investig. Geogr.* **2021**, *47*, 13–32. https://doi.org/10.18172/cig.4897.
22. de Alfonso, M.; Lin-Ye, J.; Garcia-Valdecasas, J.M.; Perez-Rubio, S.; Luna, M.Y.; Santos-Munoz, D.; Ruiz, M.I.; Perez-Gomez, B.; Alvarez-Fanjul, E. Storm Gloria: Sea State Evolution Based on in situ Measurements and Modeled Data and Its Impact on Extreme Values. *Front. Mar. Sci.* **2021**, *8*, 646873. https://doi.org/10.3389/fmars.2021.646873.
23. Amarouche, K.; Akpinar, A. Increasing Trend on Storm Wave Intensity in the Western Mediterranean. *Climate* **2021**, *9*, 11. https://doi.org/10.3390/cli9010011.
24. De Leo, F.; Besio, G.; Mentaschi, L. Trends and variability of ocean waves under RCP8.5 emission scenario in the Mediterranean Sea. *Ocean. Dyn.* **2021**, *71*, 97–117. https://doi.org/10.1007/s10236-020-01419-8.
25. Hsu, H.-C.; Chen, Y.-Y.; Chen, Y.-R.; Li, M.-S. Experimental Study of Forces Influencing Vertical Breakwater under Extreme Waves. *Water* **2022**, *14*, 657.
26. Iglesias, I.; Bio, A.; Melo, W.; Avilez-Valente, P.; Pinho, J.; Cruz, M.; Gomes, A.; Vieira, J.; Bastos, L.; Veloso-Gomes, F. Hydrodynamic Model Ensembles for Climate Change Projections in Estuarine Regions. *Water* **2022**, *14*, 1966.
27. Eldeberky, Y.; Metwally, A.; Rakha, K.; Cavaleri, L. Wave Hindcast in the Eastern Mediterranean Sea. In Proceedings of the ASME 2002 21st International Conference on Offshore Mechanics and Arctic Engineering, Oslo, Norway, 23–28 June *2002*; pp. 481–489.
28. Elkut, A.E.; Taha, M.T.; Abu Zed, A.B.E.; Eid, F.M.; Abdallah, M.A. Wind-wave hindcast using modified ECMWF ERA-Interim wind field in the Mediterranean Sea. *Estuar. Coast. Shelf Sci.* **2021**, *252*, 107267. https://doi.org/10.1016/j.ecss.2021.107267.
29. Puertos del Estado. Oceanografía. Availabe online: https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx (accessed on 16 September 2002).
30. Puertos del Estado. Oceanografía. Registros Temporales Oleaje. Availabe online: https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx (accessed on 19 September 2002).
31. Puertos del Estado. Banco de Datos. Informes Extremales. Availabe online: https://bancodatos.puertos.es/BD/informes/extremales/ (accessed on 12 October 2022).
32. Puertos del Estado. Oceanografía. Registros Temporales Mareógrafos. Availabe online: https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx (accessed on 19 September 2022).

33. Puertos del Estado. Banco de Datos. Availabe online: https://bancodatos.puertos.es/BD/informes/extremales/EXT_1_1_1619.pdf (accessed on 20 September 2022).

3.5. Review of the application of Artificial Neural Networks in ocean engineering

Abstract and Keywords

Artificial Neural Networks (ANNs) were firstly used to model ocean engineering problems in the decade of 1990s. Since then, this soft-modelling technique has proved several advantages against traditional approaches. In this article, the application of ANNs in ocean and maritime engineering from its beginnings until today is reviewed. After reviewing more than 90 studies, some general rules to apply in ANNs in ocean engineering are concluded, the most common and advantageous methods are summed up and research gaps in the application of ANNs are identified.

Keywords: Artificial intelligence; Artificial neural networks; Ocean engineering; Wave height forecast; Sea level prediction; Soft-computing

1. Introduction

The concept of Artificial Neural Networks (ANNs) was firstly introduced by McCulloch and Pitts in 1943 (Maier and Dandy, 2000; McCulloch and Pitts, 1943), but it not was until 1986, when Rumelhart et al. developed the backpropagation algorithm, when the use of ANNs became popular (Møller and Moller, 1993; Rumelhart et al., 1986). ANNs are one of the most used techniques in Artificial Intelligence (AI) (Medina et al., 2004). AI can be defined as the discipline or process by which machines learn by themselves. The purpose of IA is to create machines that work like the human brain and that are able "to think" (Standford University Machine Learning Course). Neural Networks (NNs) try to mimic the human brain with computational elements that emulate the neurons (by elements called nodes) and the unions between them (by transfer functions) (Medina et al., 2004; Standford University Machine Learning Course).

NNs are made of different layers: the input layer in which the known data is introduced into de model; the intermediate layers, known as hidden layers; and the output layer in which the final value searched is obtained. Each layer consists of different units (neurons or nodes) (Medina et al., 2004; Standford University Machine Learning Course) and each one is connected to the following layer by a transfer function. In an ANN the output of the layer i-1 is the input of the layer i.

The known data is introduced into the input layer and a bias term is added, then these data are multiplied by an initial weight and summed, this is after transferred through functions to the following layer until the output layer is reached and the final value is obtained (Dixit and Londhe, 2016).

1.1. Transfer functions and types of ANNs

An ANN uses new features that learn from the input and the mapping of the input. To go from one layer of the NN to the next one, transfer functions are used (Standford University Machine Learning Course). The most common transfer functions used in ocean and maritime engineering are presented in Table 1 (Yogitha and Mathivanan, 2018).

Table 1. Most common transfer functions used in ocean engineering problems.

Function	Expression	Application
Linear	$F(x) = x$	Output layers
Logistic sigmoid	$F(x) = \frac{1}{1+e^{-x}}$	Output and hidden layers
Hyperbolic tangent sigmoid	$F(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$	Hidden layers
Radial basis	$F(x) = g(\|\|x-c\|\|)$	Output and hidden layers

The most common type of ANNs is the feed forward multilayer perceptron neural network (FFNN), but there also other network types that proved to perform very well in ocean engineering problems. All these networks are explained below.

The FFNN is the most common NN. It usually has one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons and can be used as a general function approximator (Demuth et al., 1992). The main characteristic of FFNNs is that they only process in formation in one direction, which means that there are no back connections (Teslyuk et al., 2021). The first step to build a FFNN is to define a network object, inputs and outputs are selected and weighted and then, the NN is trained and validated (Demuth et al., 1992). They are the simplest type of NNs, but usually have lower rates of accuracy and fail to model time-varying processes. They are also very sensitive to initial weights.

The Cascade-Feed-Forward network (CFF) is a NN in which each hidden unit forms one layer that receives connections from the net work's original inputs and

the pre-existing hidden units. (Qiao et al., 2016). Therefore, the only difference between a CFFNN and a common FFNN is that each neuron layer is related to all next neuron layers and these additional connections might improve the speed at which the network learns, although the complexity of the NN increases (Pashova and Popova, 2011).

The Feed-Forward Time Delay network (FFTD) is a special feed- forward network with a tapped delay line at the input data set that is introduced to generate dynamics in the static structure. Therefore, FFTDs are used to model dynamic systems and time-varying processes (Yazdizadeh and Khorasani, 2002). Each delay is associated with a weight and its purpose is to provide the network with the capability of representing the relationships between events in time (Pashova and Popova, 2011). The output at time "t" are summed with the outputs at time "t-1", "t-2", etc., allowing the creation of richer and more complex decision surfaces (Menezes and Barreto, 2008). Another difference between the FFTD and the FFNN is in the number of branches that connect any two neurons. In the FFNN, generally, there is only one branch connecting any two neurons, whereas in the FFTD more than one branch may be used (Yazdizadeh and Khorasani, 2002).

The radial basis function (RBF) NN is a network that only uses one hidden layer, the output units form a linear combination of the basis functions computed by hidden units and the basis functions in the hid den layer produce a localised response to the input (Evans et al., 2000). It is segmented in two phases. In the first phase, unsupervised learning of the input information is conducted, it classifies information into clusters, that are used in the second phase to yield the output after a supervised learning process is done (Kalra et al., 2005a). The training phase in a RBF NN consists of finding the centres, widths, and the weights con necting hidden nodes to the output nodes. This not only is more efficient, but can sometimes model the data non-linearities in a better way. Their main problem is that they are highly dependent on the choice of centres and width (Evans et al., 2000).

The General Regression Neural Network (GRNN) is a kind of the radial basis network that is often used for regression problems (Pashova and Popova, 2011). It is a nonlinear RBF NN based on nonparametric estimation (Ding et al., 2013). These networks are composed of a RBF layer and a linear layer and they are based on kernel (nonlinear) regression. As its name indicates, this function is capable of approxi mating any arbitrary function from historical data and it does not require any iterative training process, it draws the estimation functions directly from the

training data (Su et al., 2011). They usually have a higher degree of forecasting accuracy and performs statistically better (Leung et al., 2000). Its main problem is that they do not perform well extrapolating.

The adaptive neuro-fuzzy inference system (ANFIS) is another type of NN that consists of an hybrid scheme which uses the learning capability of the ANN to derive the fuzzy if-then rules. The difference between the common NN and the ANFIS is that while the former captures the underlying dependency in the form of the trained connection weights, the latter does so by establishing the fuzzy language rules (Kalra et al., 2005a). This type of ANN has two main parts called as premise and consequence parts and the training in ANFIS consists of determining the parameters belonging to these parts with an optimization algorithm. It eliminates the disadvantage of ANNs of assigning arbitrary weights without explanation (Karaboga and Kaya, 2019).

Recurrent networks (RNNs) are NNs with internal loops that introduce delayed-activation dependencies across the units of the network (Marhon et al., 2013). The main difference with FFNNs is that in FFNNs the signal is passed only in a forward direction, whereas in RNNs the output obtained is fed back into the network through the input layer (Gopinath and Dwarakish, 2015). Connections between nodes form a time-oriented graphs that allows to exhibit dynamic behaviour over time (Teslyuk et al., 2021), this is the main advantage with respect to other networks, they can represent time-varying processes (Marhon et al., 2013). In addition, they usually perform better when the data is noisy, but they can have difficulties in capturing long-term dependencies because the influence of a given input decays exponentially as it cycles around the network's recurrent connections (vanishing gradient problem). This can be solved including inputs at explicit time lags (NARX networks) or with Long short-term memory (LSTM) NNs (Maier and Dandy, 2000; Graves and Graves, 2012).

Long short-term memory (LSTM) neural networks are developed based on the previously explained RNNs. Instead of the neurons of a conventional NN, they have long and short-term memory cells and forget and outputs gates to write, read and reset operations for the cells (Graves and Graves, 2012; Gao et al., 2021). These gates allows the LSTM NN to store information for a long period of time if it is required, which ends with the vanishing gradient problem. The main advantage with respect to RNNs is that they can be used to classify, process, or predict time series when there are time delays of unknown duration between important events and that they can save information in any length of time (Marhon et al., 2013).

LSTM NNs use historical information to assist with current decisions and, through the mechanism of a forget gate, they can choose to remember or forget long-term information (Fan et al., 2020). Lara-Benítez et al. proved that they are one of the best alternatives (Lara-Benitez et al., 2021).

Determining the ANN architecture is another difficult task. More complex networks usually model processes more accurately but are also more expensive and may have problems in generalisation (Maier and Dandy, 2000).

1.2. Training algorithms

Before an ANN is used for forecasting it has to be trained. The most common procedure is to use the 70–80% of the data to train the network and the remaining 20–30% to test its performance. There are many different training algorithms of which the most common is the back propagation algorithm (Mandal and Prabaharan, 2006). In this section the most popular algorithms are presented.

1.2.1. Gradient descent algorithms

Error backpropagation (BP): It is called backpropagation algorithm because the error is propagated back to update the weights and the biases using an optimization technique (gradient descent), which minimizes the error (Dixit and Londhe, 2016). The backpropagation prop agates the input forward through the network and the sensitivities backward, it updates the network weights and biases in the direction in which the performance function decreases most rapidly, which is the negative of the gradient. Its main drawbacks is that it is a local algorithm, which means that it can reach a local optimum instead of the global optimum, and that it is slow in converging (Demuth et al., 1992; Hagan et al., 1996).

It also exists the variant gradient descent with momentum (GDM), which objective is to faster the convergence of the algorithm. The momentum smooths the oscillations of the algorithm and allows the network to respond not only to the local gradient, but also to recent trends in the error surface, avoiding getting stuck in a local optima (Demuth et al., 1992; Hagan et al., 1996).

One of the main problems of the conventional backpropagation algorithm is that it is too slow. Therefore, there are other algorithms known as “faster training” that avoids these problems (Demuth et al., 1992).

Resilient backpropagation (RBP): ANNs usually use sigmoid functions in the hidden layers. These functions transform values from an infinite range to a range

of 0–1 and of 1 to 1. This can cause the gradient to have a very small magnitude, which can cause small changes in the weights and biases, generating problems when using the gradient descent technique. The objective of this algorithm is to avoid these effects by determining the weights and bias by a separate update value (Demuth et al., 1992).

1.2.2. Conjugate gradient (CG) algorithms

Conjugate gradient descent algorithms are also classified as "faster training". Common gradient descent updates the network weights and biases in the direction in which the performance function decreases most rapidly, which is the negative of the gradient. However, this does not necessarily produce the fastest convergence. Conjugate gradient algorithms also search along conjugate directions. There are different con jugate gradient functions (Demuth et al., 1992; Hagan et al., 1996):

- Fletcher-Reeves (CGF): It computes the previous search direction by multiplying it by a factor that is the ratio of the norm squared of the current gradient to the norm squared of the previous gradient. It guarantees to locate the minimum of any quadratic function of n arguments in at most n iterations (Demuth et al., 1992; Fletcher and Reeves, 1964).
- Polak-Ribiére (CGP): In this case the factor is the inner product of the previous change in the gradient with the current gradient divided by the norm squared of the previous gradient (Demuth et al., 1992; Fletcher and Reeves, 1964).
- Powell-Beale Restarts (CGB): In common conjugate gradient descent algorithms search direction will be reset to the negative direction of the gradient when the number of iterations is equal to the number of network parameters. The CGB technique restarts the direction of searching to the negative of the gradient if there is very little orthogonality left between the current gradient and the previous gradient (Demuth et al., 1992; Powell, 1977).
- Scaled conjugate gradient descent (SCG): All previous functions use line search, which is computationally expensive. Instead of the line search approach, the SCG uses the model-trust region approach (used in the Levenberg-Marquardt algorithm) to scale the step size and combines it with

the conjugate gradient approach (Møller and Moller, 1993; Demuth et al., 1992).

1.2.3. Quasi-Newton's algorithms

Newton's algorithms usually converge faster than gradient methods, but they are much more computationally expensive because they have to compute the Hessian matrix. Quasi-Newton's algorithms are based on Newton's method, but do not require calculation of second derivatives. There are two main algorithms (Demuth et al., 1992):

- Broyden, Fletcher, GolFarb and Shanno (BFGS): It is based on con jugate directions combined with line search. It requires more computation in each iteration and more storage than the conjugate gradient methods, although it generally converges in fewer iterations. It is especially appropriate for small networks (Møller and Moller, 1993; Demuth et al., 1992).
- One step secant algorithm (OSS): It deals with the problem of the storage needed for the BFGS algorithm. This algorithm assumes that at each iteration, the previous Hessian was the identity matrix (Demuth et al., 1992).

1.2.4. Levenberg-Marquardt algorithm (LM)

It is a variation of Newton's method that was designed for minimizing functions that are sums of squares of other nonlinear functions. When the performance function has the form of a sum of squares the Hessian matrix can be approximated as the product of the Jacobian matrix, that is much less complex than computing the Hessian matrix (Demuth et al., 1992; Hagan et al., 1996).

ANNs can perform very well in the training data set, but fail to perform in the test data set, which means that they do not perform well when adding new data. When the ANN is not able to generalize the problem is known as "overfitting". This problem can be solved using regularization or early stopping (Standford University Machine Learning Course; Hagan et al., 1996). Regularization is a technique in which the different features are penalized by multiplying it by a factor λ, smoothing the function and making the network less complex (Stand ford University Machine Learning Course). One of the most common regularization techniques used in ocean engineering is the Bayesian regularization (BR). Mackay (1992) was the first one in applying the Bayesian framework to ANN modelling.

2. ANNs in ocean engineering

Numerical models are the most common used approach to solve ocean and maritime engineering problems. However, with the development of machine learning it has been showed that ANNs are a very good alternative to this traditional approach, they even give better results, especially when the problem is random and includes non-linear patterns (Fan et al., 2020). The differences in the modelling approach between ANNs and traditional statistical models is that the former couple with a lack of strict governing rules, which makes them ideal for problems with a random distribution of variables (Maier and Dandy, 2000). In ocean engineering ANNs have been used since 1995, Mase (1995) was the first one to use them to evaluate the artificial armour layer stability of a breakwater. Since then, ANNs have been widely used. In this paper the following fields of ocean engineering in which ANNs have been applied are distinguished: prediction of wave heights, prediction of the water level, tides, extremes and wave energy and application to ports, pipes and breakwater stability.

2.1. Prediction of wave heights

One of the main applications of ANNs is to predict wave heights, prediction of waves is an uncertain and random process that makes it ideally suited to NN modelling (Deo et al., 2001). Deo and Naidu were the first authors that proved the advantages of ANNs over traditional methods for wave-height forecasting in 1998 (Deo and Sridhar Naidu, 1998). In 2001 Deo et al. predicted significant wave heights and average wave periods from the input of wind speeds (Deo et al., 2001). In 2002 Agrawal and Deo did it from current and previous wave heights (Agrawal and Deo, 2002). They found that when shorter prediction intervals are concerned, the NN had distinct superiority over the stochastic approach. However, this was not the case when a longer interval of prediction was studied. One year later, Deo and Jagdale showed that ANNs could also be used to predict breaking wave heights (Deo and Jagdale, 2003). In 2004 Agrawal and Deo concluded that ANNs outperformed more the statistical regressions when the input–output structure to be established is hazier and more unclear (Agrawal and Deo, 2004).

Also in 2004 Makarynskyy studied how to improve wave predictions with ANNs (Makarynskyy, 2004). A separate NN was developed to forecast each wave parameter at every location. One year later Makarynskyy et al. did a similar study with two approaches (Makarynskyy et al., 2005). They first used one ANN for each parameter (wave height and period) and in the second approach they used a unique

ANN to predict both. They proved that the simplification of the NN sometimes gave better results. In 2005 Karla et al. conducted another two research. In the first one, they used satellite data and saw that this data needed pre-processing before applying it to ANNs (Kalra et al., 2005a). In the second one they also translated offshore satellite wave data and proved that the RBF NN performed better than the FFNN, but for higher waves the FFNN was more accurate (Kalra et al., 2005b).

In 2008 Jain and Deo conducted an experiment on AI tools to fore cast waves in real time (Jain and Deo, 2008). They applied ANNs, genetic programming and model trees and saw that the differences between these 3 techniques were negligible. This same year Malekmohamadi et al. compared the performance of numerical wave modelling (NWM), ANNs and an hybrid model (Malekmohamadi et al., 2008). They concluded that although the NWM leaded to better results, such models were computationally expensive and needed long-time execution. Using artificial patterns for training ANNs leaded to more acceptable results, than those obtained using observed patterns. Also in 2008 Günaydin compared seven different ANN models with traditional regression methods (Gunaydin, 2008). They showed that the best performance was given when all the meteorological factors were included as inputs. Between all the meteorological inputs, wind speed was the most effective. The same year Mahjoobi et al. applied ANNs, Fuzzy Inference Systems (FIS) and ANFIS (Mahjoobi et al., 2008). They implemented a separate network for each parameter, and after one network for the three output variables. Results indicated negligible differences between the two approaches.

In 2009 Tsai and Tsai used ANNs to convert the signal of pressure transducers into significant wave height, wave period, maximum wave height, and spectral peakedness (Tsai and Tsai, 2009). They compared the results with the ones obtained applying a linear wave pressure transfer function and found that for water depths greater than 20 m the NNs worked significantly better.

In 2015 other authors, Gopinath and Dwarakish proved that NARX models needed less data than conventional NNs to obtain the same results (Gopinath and Dwarakish, 2015). In 2016 Londhe et al. used ANNs to predict the errors in wave forecasting and added or subtracted these errors from the valued predicted by the numerical model, improving their predictions (Londhe et al., 2016). Salah et al. discovered that the hybrid model that combined SWAN and ANNs models outperformed both approaches, traditional and soft-computing (Salah et al., 2016). In 2017 Abhigna et al. saw that to predict extreme values of wave heights (both,

high and low) ANNs performed worse (Abhigna et al., 2017). This same year Mudronja et al. stated that ANNs outperformed more traditional models when the number of inputs involved was high (Mudronja et al., 2017) and Sadeghifar et al. proved that the use of NARX networks improved the prediction of wave heights (Sadeghifar et al., 2017).

In 2020 Vieira et al. used ANNs to fill the gaps in wave records (Vieira et al., 2020). The inputs were wave and wind parameters. They performed several FFNNs with 3 (wave parameters) and 5 inputs (wave and wind parameters) and concluded that wind was not and essential parameter in that case, contrary to other studies. This same year, Sta varache et al. studied how to optimize ANNs for wave prediction and proposed to substitute the traditional trial and error approach by genetic algorithms (Stavarache et al., 2020). In 2021 Gao et al. concluded that LSTM networks had several advantages with respect to other soft-computing and conventional models (Gao et al., 2021). They also highlighted that a higher number of hidden neurons did not necessarily imply better results and that this number had to be adjusted to each wave height. These results were in line with Joerges et al. (Joerges et al., 2021).

In 2022 Sadeghifar et al. conducted a study in which they compared the performance of different soft-computing techniques (ANNs, ANFIS, Random Forest and M5P models) and proved that the M5P outperformed the others (Sadeghifar et al., 2022). The studies presented in this section are summed up in Table 2.

Table 2. Wave height studies.

Author	NN type	Architecture	Algorithm	Performance evaluation	Training and test
Deo and Sridhar Naidu (1998)	FFNN	1-5-1	BP	Global errorTime Nºiterations	80% and 20%
		1-2-1	CG		
		1-0-1	**Cascade**[a]		
Deo et al. (2001)	FFNN	2-4-2 4-4-4	**BP**, CG and Cascade	R	80% and 20%
Agrawal and Deo (2002)	FFNN	–	**BP**, CG and Cascade	Correlation coefficient (R) Root mean	80% and 20%

				square error (RMSE)	
Deo and Jagdale (2003)	FFNN	3-5-5-2	BP	R, Maximum, minimum and average error	–
Agrawal and Deo (2004)	FFNN	–	BP and Cascade	R, RMSE	80% and 20%
Makarynskyy (2004)	FFNN	48-97-24	RBP	R, RMSE, SI	52% and 48%
Kalra et al. (2005a)	**RBF**	21-13-1	RBF	R, RMSE, Mean absolute error (MAE), Scatter index (SI), Mean square relative error (MSRE)	80% and 20%
	FFNN	21-9-1	**BP**, SCG, CGB, BFGS, LM		
	ANFIS	–	–		
Kalra et al. (2005b)	FFNN	27-9-1	RBP, SCG, **CGB**, BFGS, LM	R, RMSE, MAE	70% and 30%
	RBF	27-3-1	RBF		
Makarynskyy et al. (2005)	FFNN	8-17-1	BP, RBP, **LM**	R, RMSE, SI	–
		4-9-1			
		2-5-1			
		6-17-4			
		12-21-5			
Jain and Deo (2008)	FFNN	4-4/5-1	LM	R, RMSE, MAE	60% and 40%
Malekmohamadi et al. (2008)	FFNN	8-8-1	BP	R, Total error	–
	Hybrid model	24-24-1	BP		
Gunaydin (2008)	FFNN	3-2-1	LM	R, Mean Square Error (MSE) and	80% and 20%

				Mean Absolute Relative Error (MARE)	
Mahjoobi et al. (2008)	FFNN	4-7-1: Period and height	BP	R, SI, MSE, MSRE, BIAS	–
		4-11-1:Dir log			
		4-8-1:Dir tan			
		4-9-3			
	ANFIS	–	–		
Tsai and Tsai (2009)	FFNN	4-8-2	BP	R, Root Mean Square Deviation (RMSD)	75% and 25%
Gopinath and Dwarakish (2015)	FFNN	52-10-52	LM	R, MSE, RMSE, Sum Squared Error (SSE)	–
		12-5-12			
	NARX	1-6-25	–		
		1-10-12			
Salah et al. (2016)	FFNN	10 hidden nodes	SCG, LM	R, RMSE, SI, BIAS	80% and 20%
Londhe et al. (2016)	FFNN	9-1-1	LM	R, MSE, MAE, RMSE, SI, BIAS, Coefficient of efficiency (CE)	70% and 30%
		4-1-1			
		9-8-1			
Abhigna et al. (2017)	FFNN	–	CG, LM, **BR**	R, MSE	–
	Recurrent NN				
Mudronja et al. (2017)	FFNN	1- [10 to 50]-1	LM, BR	RMSE, MAE, CE, Percent BIAS, Standard Deviation Ratio	–
		2-[10 to 50]-1			

				(RSR) and Persistency Index (PI)	
Vieira et al. (2020)	FFNN	3-[4/8/12]-1 5-[4/8/12]-1	LM, BR	R, RMSE, SI, BIAS	70% and 30%
Stavarache et al. (2020)	FFNN	5-5-9-8-1	LM, Quick Propagation	R, RMSE	–
Gao et al. (2021)	LSTM	Several	BP	R, RMSE, SI MAE	–
Joerges et al. (2021)	LSTM	–	–	R, RMSE, MAE, Mean Arctangent Absolute Percentage Error (MAAPE), Coefficient of Determination R2	70% and 30%
Sadeghifar et al. (2022)	FFNN, ANFIS	4-10-1	–	R, RMSE, MAE, SI, BIAS, CE	70% and 30%

*The algorithms and networks in bold were the ones that performed the best

From all the type of networks, the FFNN was clearly the most used. Although in the last years LSTM neural networks are giving promising results (Fig. 1).

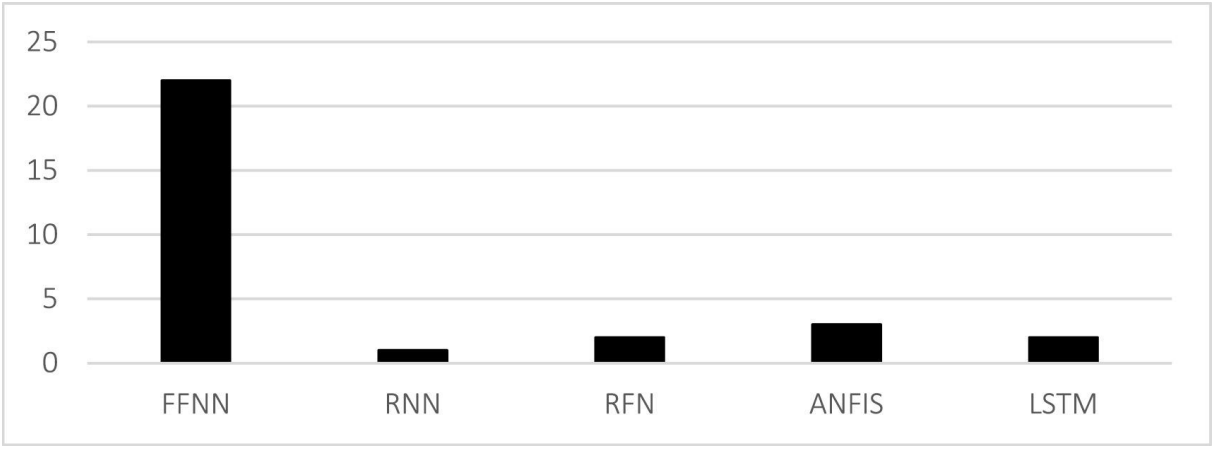


Fig. 1. Frequency of networks used in wave height ANNs studies.

Referring to the architecture of the studies, all of them excepted one opted for a 3-layer neural network because adding more layers did not result in a significant improvement. In ocean engineering, shallow ANNs with one hidden layer are enough to forecast wave-heights. The number of inputs and outputs depends on the physics of the problem and has to be studied in each case. The number of hidden neurons also depends on the problem and its level of complexity. All of them were usually optimized by trial-and-error analysis.

From all the algorithms presented, the BP and the LM were the most used to predict wave height. The efficiency of each algorithm depended on the characteristics of the problem. To choose the best algorithm, each case has to be studied (Fig. 2).

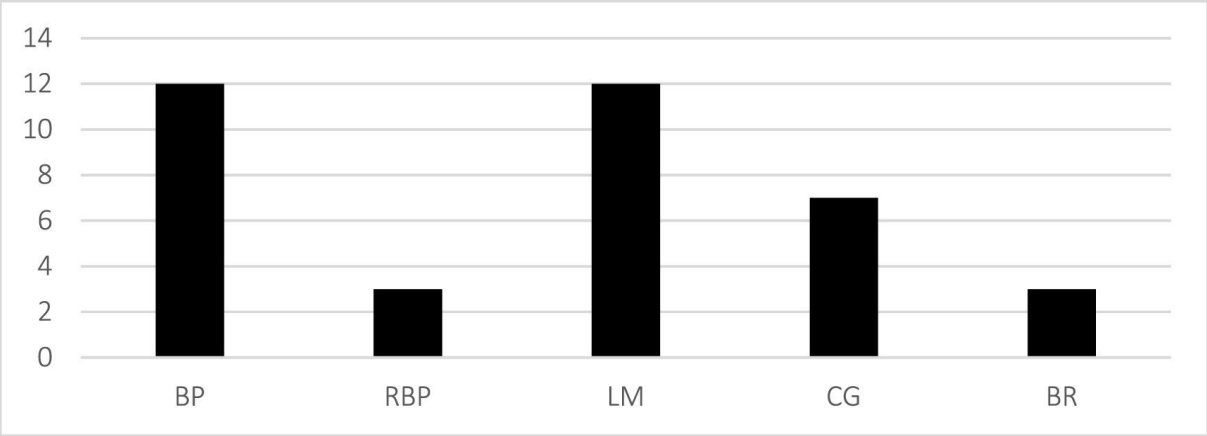


Fig. 2. Frequency of algorithms used in wave height ANNs studies.

With respect to the error measures, 22 out of 24 studies used the correlation index (R) as one of the main variables to optimize the network, 65% of the studies also used the RMSE, and the rest of statistics to quantify the performance of the NN depended on the study and on the authors choice.

2.2. Prediction of water level, tides, extremes and wave power

ANNs have also been used in ocean engineering to predict water level, tides and wave power. Tsai and Lee were one of the first ones to use ANNs to predict tides in 1999 (Tsai and Lee, 1999). Lee et al. also used it two years later, they tried different networks with 0, 1 and 2 hidden layers and concluded that the best performance was given by one hidden layer with BP algorithm (Lee and Jeng, 2002; Lee et al., 2002). In 2004 Lee conducted a similar study, but he did it to predict long-term tides, they proved that ANNs were as accurate as traditional harmonic analysis and needed less data (Lee, 2004). In 2006 Chang and Lin

compared the ANN results with the NAO.99b model and concluded that both methods were accurate and that the NNs could be applied to different locations with similar tide type (Chang and Lin, 2006).

In 2008 Liang et al. performed one-step and multistep methods, concluding that the last one was better to predict water levels (Liang et al., 2008). This same year, Han and Shi predicted water levels and pointed out the usefulness of this technique not only for short-term forecasts, but also for long-term (Han and Shi, 2008). Next year, de Oliveira and Ebecken used a RBF NN and a GRNN to predict sea level and surge. The results indicated that NNs could also be useful as a complement for the standard harmonic model (de Oliveira and Ebecken, 2009).

In 2010 Wenzel and Schröter used ANNs with two main purposes: filling data gaps in the tide gauge time series and estimating the evolution of regional mean sea levels from these tide gauge data (Wenzel and Schroeter, 2010). Next year, Pashova and Popova compared the performance of five different neural networks on predicting water level and concluded that the best performance was given by the FFNN with LM, LM algorithm was the fastest, and BFGS gave the best results in most cases (Pashova and Popova, 2011). In 2012 Filippo et al. also concluded that ANNs were a very interesting tool to predict water levels and also to complement traditional harmonic analysis approach (Filippo et al., 2012). In 2014 Karamouz et al. used two innovative models called ELMAN (special type of the common FFNN with context units) and IDNN (type of FFTD) for sea-level simulation and saw that sea surface temperature was an important variable to consider (Karamouz et al., 2014). This same year Kusnandar et al. estimated the mean sea level change by hybridizing exponential smoothing and a NN, which gave promising results (Kusnandar et al., 2014). Castro et al. also studied the wave energy potential in nearshore coastal areas using ANNs and saw that the most important variable to do these predictions was the wave-height (Castro et al., 2014).

In 2015 El-Diasty and Al-Harbi compared the performance of an ANN; traditional harmonic analysis and a wavelet network, which is an algorithm that connects the NN with the wavelet decomposition. Results showed that the wavelet network was the most accurate (El-Diasty and Al-Harbi, 2015). In 2018 El-Diasty et al. did another research using wavelet technique, but this time hybridizing it with the harmonic analysis technique, which gave even better results (El-Diasty et al., 2018). In 2019 Zhao et al. applied different ANNs to forecast sea level anomalies, they predicted sea level variation with the least-square fitting methodology and

the stochastic residual terms with different NN models. They found that this hybrid approach gave results with relatively high precision (Zhao et al., 2019).

Wang et al. also used wavelet networks to predict tides and sea water level (Wang et al., 2020a, 2020b, 2020c). In 2020 they combined wavelet decomposition and ANFIS models for predicting multi-hour sea levels and highlighted the advantages of these hybrid models (Wang et al., 2020c). After, they improved wavelet networks adding a momentum term (Wang et al., 2020a). This same year, they also improved wavelet networks using a genetic algorithm to optimize the weights and parameters of the NN (Wang et al., 2020b). All these models performed significantly better than the traditional harmonic analysis. In 2020 Avila et al. used Fuzzy systems and ANNs to predict wave energy and both models gave very accurate results without needing a base data of many years (Avila et al., 2020). In this same year Bruneau et al. studied sea level extremes with a deep NN, they used a total of 600 tide gauges around the world to predict the non-tidal residual as well as extremes events and concluded that ANNs outperformed predictions based on multivariate linear regressions (Bruneau et al., 2020).

In 2021 Primo de Siqueira and Paiva used a combined approach of process-based and data-based model to improve the forecasting of sea levels, they used the process-based model to forecast the linear part and the ANNs to forecast the non-linear part and the results obtained were significantly better (Primo de Siqueira and Paiva, 2021). Yang et al. used deep neural networks (NNs with more than two hidden layers) to predict tides and analysed the effects of gaps in data records, they concluded that, among all the methods, deep neural networks performed the best (Yang et al., 2021). Bai and Xu compared RBF and LSTM NNs with an innovative approach: bidirectional LSTM NNs, that is a LSTM NN that combines the information for both, front and back directions of time series, and proved that was the best option (Bai and Xu, 2021). Bento et al. concluded that in the case of extreme unforeseen events (long-term forecasts), the DNN presented a slightly underestimated time-series prediction and more data was needed. The errors of DNNs also depended on the seasonality, winter forecasts were less accurate than summer forecasts due to its greater variability (Bento et al., 2021).

The studies presented in this section are summed up in Table 3. From all the type of networks, the FFNN was clearly the most used. Although in the last years wavelet neural networks are getting more popular (Fig. 3).

Table 3. Sea level, extremes and energy studies.

Author	Network type	Architecture	Algorithm	Performance evaluation	Training and test
Lee and Jeng, 2002	FFNN	4-4-1	BP	RMSE	–
Lee et al. (2002)	FFNN	138-7-69	BP	R, RMSE	–
Lee (2004)	FFNN	138-7-69	BP	R, RMSE	–
Chang and Lin (2006)	FFNN	5 hidden neurons	LM	R, RMSE	–
Han and Shi (2008)	FFNN	3 layers, 10 neurons	LM	R, RMSE, SI	–
Liang et al. (2008)	FFNN	50 hidden neurons	BP	R, RMSE	–
de Oliveira and Ebecken (2009)	RBF	7-9-1	BP for FFNN	R, RMSE, R2	50% and 50%
	FNN	7-11-1			
	GRNN	**7–14–1**			
Wenzel and Schroeter (2010)	FFNN	112-84-56	BP	RMSE	–
Pashova and Popova (2011)	**FFNN**, CFF, FFTD, RBF, and GRNN	11-6-1	BP, **LM**, RBP, BFGS	R, RMSE, SI, R2	79% and 21%
Filippo et al. (2012)	FFNN	4 inputs	BP	R, Regression Ratio, Error	
		1 output			
Karamouz et al. (2014)	FFNN	1 to 20 hidden neurons	BP	RMSE, Volume Error Percent (VE)	–
	ELMAN (FFNN)	19 hidden neurons	LM		

	IDNN (FFTD)	17 hidden neurons	BP		
	ANFIS	–	–		
Castro et al. (2014)	FFNN	3-4-4-1	BP	R, MSE and Normalized MSE (NMSE)	2/3 and 1/3
El-Diasty and Al-Harbi (2015)	ANN	24-12-1	BP	R, RMSE, MAE, Maximum Absolute Error, Percentege of RMSE	80% and 20%
	Wavelet				
El-Diasty et al. (2018)	Hybrid wavelet	24-12-1	–	R, RMSE, MAE,Model error at 95%	80% and 20%
Zhao et al. (2019)	GRNN	5-0.25-1	RBF	R, RMSE	–
	FFNN	5-20-15-1	BP		
	RBF	5-20-1	RBF		
Wang et al. (2020c)	Wavelet	3 layers	LM	RMSE, MAE, MSRE, CE	80% and 20%
	ANFIS				
	Wavelet+ANFIS				
Bruneau et al. (2020)	**FFFNN**	33-48-48-48-1	–	MSE, MAE	2/3 and 1/3
	LSTM	–			
Avila et al. (2020)	FFNN	Input-15-1	GMS, RBP	R, MAE, Mean Relative Error	89% and 11%
Primo de Siqueira and Paiva (2021)	FFNN	10 hidden neurons	LM	R, RMSE	75% and 25%
Bento et al. (2021)	FFNN	MFO iterations	SCG	R, RMSE, MSE, Mean Absolute	75% and 25%

				Log-difference Error (MALE)	
Bai and Xu (2021)	FFNN, RBF, **Bi-LSTM**	80/200 memory cells	BP for FFNN	R, RMSE, MSE, MAE, R2	–
Yang et al. (2021)	FFNN	–[a]	BP	RMSE, MAPE	50–90% and 50-10%

*The algorithms and networks in bold were the ones that performed the best

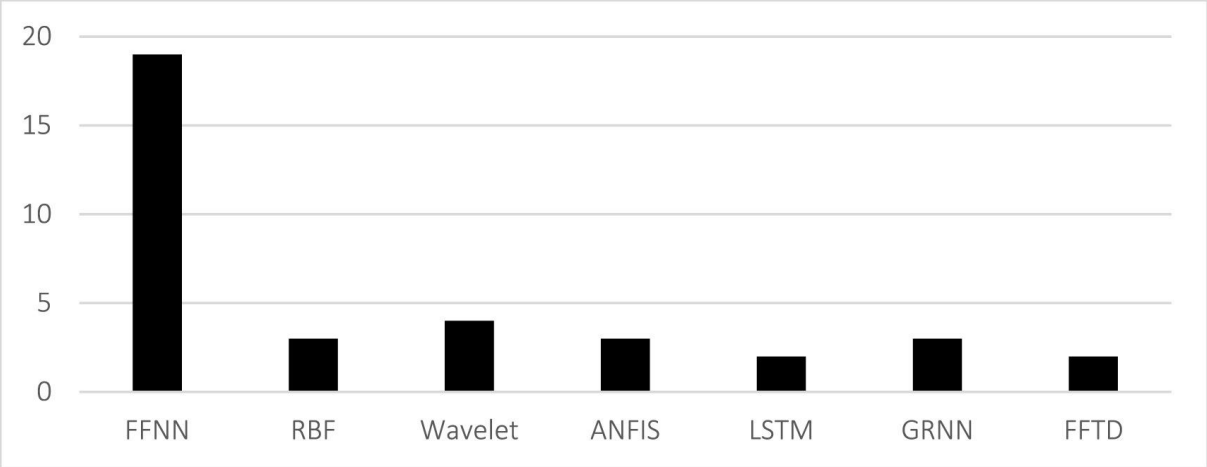


Fig. 3. Frequency of networks used in sea level, energy and extremes ANNs models.

Referring to the network architecture, the conclusions are the same as in Section 2.1.

From all the algorithms presented, the BP was the most used to predict sea level. The efficiency of each algorithm depended on the characteristics of the problem. However, LM algorithm seemed to be the most effective basing on the studies above presented, the majority of them selected the LM algorithm after proving that it was the most effective (Fig. 4).

With respect to the error measures, 18 out of 22 studies (which means more than the 80% of the studies) used the correlation coefficient and the RMSE as the main statistical variables to evaluate the performance of the model.

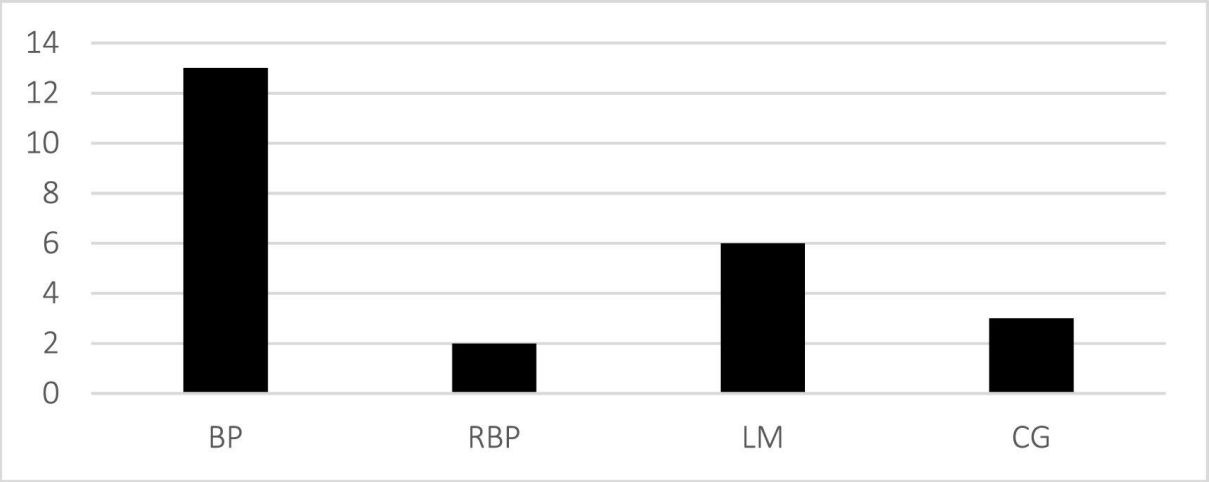


Fig. 4. Frequency of algorithms used in sea level, energy and extremes ANNs models.

2.3. Application to breakwaters, ports and pipes

One of the first applications of ANNs to maritime and ocean engineering was to study breakwater stability and one of the first authors to do it was Mase, who in 1995 used a 3-layer FFNN to estimate the artificial armour layer stability (Mase, 1995; Mase et al., 1995).

In 2005 Yagci et al. used ANNs and fuzzy logic and saw that ANNs can be employed successfully to estimate the breakwater damage ratio (Yagci et al., 2005). This same year Kim and Park compared the performance of 5 different ANNs for stability prediction (Kim and Park, 2005). They concluded that any NN became more effective when using the design parameters as independent network input rather than using artificial parameters. In 2014 Kim et al. highlighted the convenience of including tidal level in this type of studies to not overestimate break water damage (Kim et al., 2014).

In 2008 Iglesias et al. developed a virtual laboratory to estimate breakwater stability (Iglesias et al., 2008). They called the model "a virtual laboratory" because it was capable of estimating the damage caused to a rubble-mound breakwater by waves like a conventional laboratory would do. Two years later, Balas et al. studied breakwater stability with ANNs, fuzzy systems, ANFIS models and ANNs with Principal Component Analysis (PCA) and highlighted that hybrid models performed better than conventional ANNs and captured better nonlinearities (Balas et al., 2010). In 2012 Koç and Balas conducted a similar study in which they used genetic algorithms (GA) and fuzzy NNs, and pointed out that hybrid models needed less data samples (Koç and Balas, 2012). In 2011 Akoz et al. stated that ANNs were an important tool that can be applied for the design of coastal

structures under the action of most critical loading (Akoz et al., 2011). In 2013 Suh et al. combined this soft-modelling technique with SWAN model to reduce the computation time and analysed the effects of seal level rise and wave height increase on caisson sliding (Suh et al., 2013).

Referring to application of ANNs to ports studies, one of the most important authors are López and Iglesias, who in 2013 and 2015 con ducted two different studies to estimate energy and operability conditions in a harbour (López and Iglesias, 2013, 2015). In 2013 they estimated the amount of energy inside the harbour and they used two different NN models, one that directly estimated the energy and another model that did it in two steps. They concluded that ANNs constitute a valid, efficient approach for estimating wave energy inside or outside a harbour and that the direct approach outperformed the model with two steps (López and Iglesias, 2013). To assess harbour operability, they followed a similar methodology and stated that one-hidden layer NN performed better than 2-hidden layer NN for this case, they also carried a sensitivity analysis and saw that wave height and period were the two variables which most influenced wave agitation (L´ opez and Iglesias, 2015).

Previous to these studies Londhe and Deo conducted a research on wave tranquillity inside a harbour (Londhe and Deo, 2003). They used different training algorithms: GDM, RP, CGF, CGP, CGB, SCG, BFGS, OSS, and LM. All the algorithms gave satisfactory results. The CGF algorithm was more memory efficient and the SCG algorithm required the minimum time to train. Kankal and Yüksek. also assessed harbour tranquillity using NNs; they studied the effect of different wave parameters, estimated wave height in the harbour basin and optimized the port layout (Kankal and Yüksek, 2012). In 2020 Zheng at al. used ANNs to estimate waves inside a port and compared it with Boussinesq model performance, they trained the network using numerical models and validated it with observations, proving that ANNs are an efficient tool to forecast wave parameters (Zheng et al., 2020).

In addition, in the last months (March and April 2022) Moslemi et al. have published two articles about the thermal response in pipes containing fluid and gas which has a great application in ocean engineering (Moslemi et al., 2022a, 2022b). In the first study (Moslemi et al., 2022a) they used ANNs to determine the temperature and phase change rate in the extrusion process, they compared the experimental and numerical results and concluded that numerical results can be used in ANNs training and that ANNs can be used to optimize the parameters of

additively manufactured polymers used in ocean pipelines. In the second study (Moslemi et al., 2022b) they applied a FFNN to determine the correlation between the welding pool characteristics as input and the Goldak parameters as output and concluded that ANNs outperform regression models, but they are more complex.

The studies presented in this section are summed up in Table 4.

Table 4. Breakwater and ports studies.

Field	Author	Network type	Architecture	Algorithm	Performance evaluation	Training and test
Breakwaters	Mase et al. (1995)	FFNN	6-12-1	BP	MSE	–
	Yagci et al. (2005)	FFNN	5-3-1	BP	MSE	75% and 25%
		RBF				
		GRNN				
	Kim and Park (2005)	FFNN	[5-7]- 4/5/20-1	BP	Index of agreeement	–
	Iglesias et al. (2008)	FFNN	4-9-8-7-6-5-1	BP	MSE	2/3 and 1/3
	Balas et al. (2010)	FFNN	5-4-1	GDM	R, MSE	80% and 20%
		ANN with PCA	5-4-1			54% and 46%
		ANFIS	10-8-1			–
	Akoz et al. (2011)	FFNN	4-3-3	LM	MSE, MAE, MARE, R2	75% and 25%
	Koç and Balas (2012)	ANFIS	2 hidden layers	–	R, R2, Mean Absolute Percentage Eror (MAPE), CE	–
		Hybrid (GA)	5 inputs and 1 ouput	GA and Climbing Hill		38% and 62%

	Suh et al. (2013)	FFNN	5 to 60 hidden neurons	LM	MSE, Maximum Relative Error (MRE)	70% and 30%
	Kim et al. (2014)	FFNN	2-3-1	BP	RMSE	–
Ports	Londhe and Deo (2003)	FFNN	Module networks [3/4-1-1]	GDM, RP, CGF, CGP, CGB, SCG, BFGS, OSS, and LM	MAE, RMSE, MSRE, MRE, R2, CE, Average error (AE)	80% and 20%
	Kankal and Yüksek (2012)	FFNN	5 inputs 24 outputs. 1–2 hidden layers	BP	Average relative error	80% and 20%
	López and Iglesias (2013)	FFNN	3-2-2-1	BP, BR	R, MSE	–
			3-3-1			
			2-2-1			
	López and Iglesias (2015)	FFNN	15-57-1	LM, BR	MSE, RMSE, Normalized RMSE	–
	Zheng et al. (2020)	FFNN	3-7-8-24	LM, GDM	R, MAE, Relative MAE	85% and 15%
Pipes	Moslemi et al., 2022a, Moslemi et al., 2022b	FFNN	5-7-5	LM	R, MAE, RMSE	80% and 20%
		FFNN	3-3-3-3-4	LM	R, MAE, RMSE	70% and 30%

From all the type of networks, the FFNN was clearly the most used (Fig. 5). Referring to the network architecture, the conclusions are the same as in Section 2.1. and 2.2.

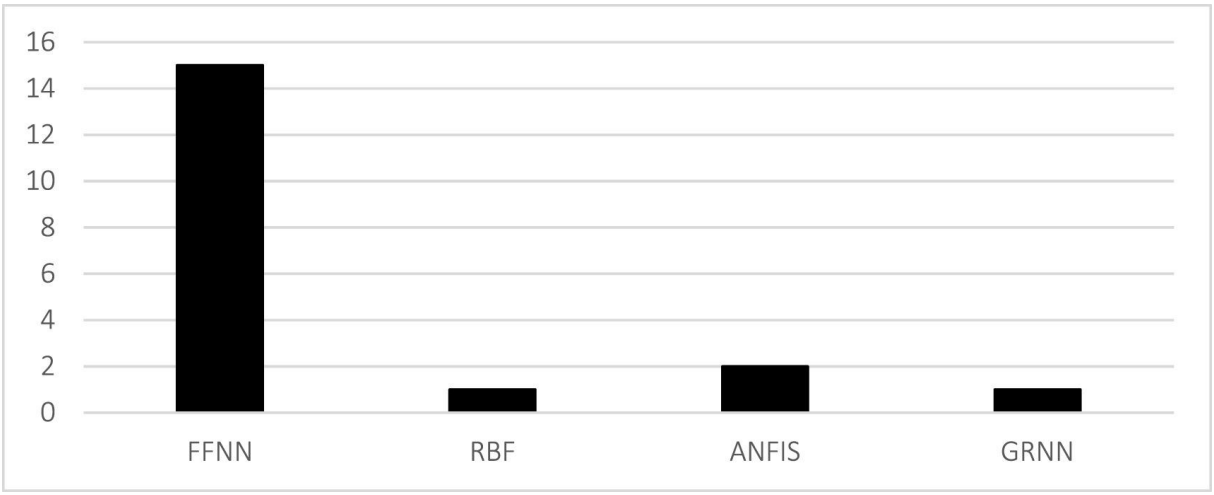


Fig. 5. Frequency of networks used in breakwaters, ports and beaches ANNs models.

From all the algorithms presented, the BP, LM and Conjugant Gradient algorithm were the most used. The efficiency of each algorithm depended on the characteristics of the problem. To choose the best algorithm, each case to be studied (Fig. 6). Finally, with respect to the error measures, in this case the MSE was the statistic variable most used, 48% of studies used it as the performance indicator.

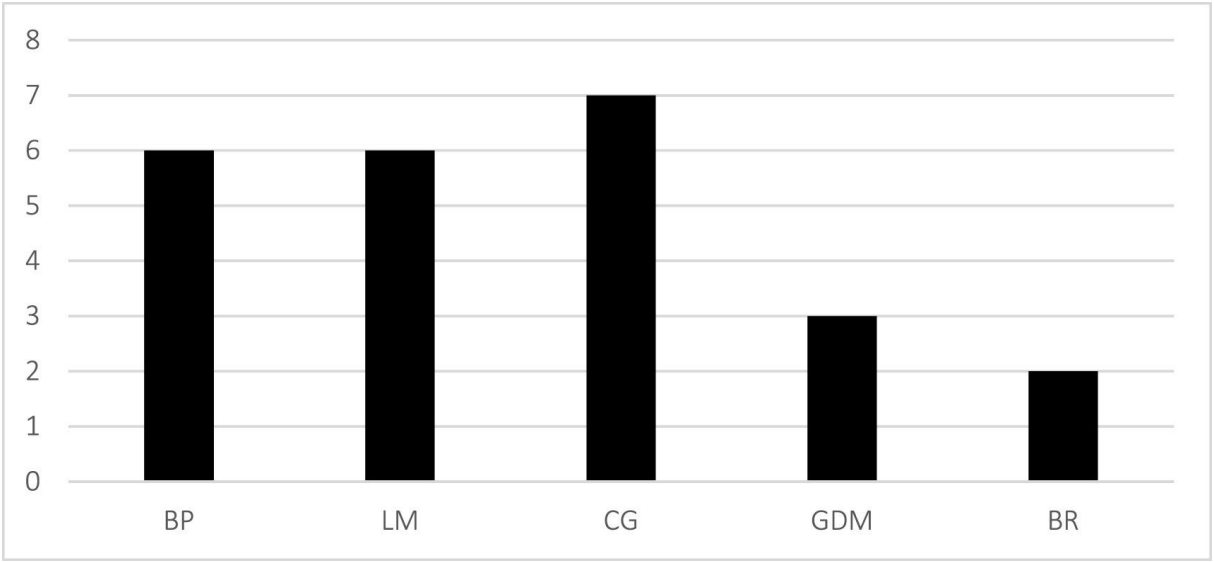


Fig. 6. Frequency of algorithms used in breakwaters, ports and beaches studies.

3. Analysis of the main issues in the application of ANNs in ocean engineering

NNs are a very powerful tool to model complex processes, they have proved to be more accurate and efficient than traditional methods. However, there are still some issues and questions that have to be solved. These are mainly the interpretability of ANNs, the parallel processing power in computation and the amount and quality of data collected to run these models. These three problems are discussed in this section.

3.1. Interpretability and computation of ANNs

One of the main problems of ANNs is that they are "black-box" models, which means that there is not a clear understanding of their operation, they give a final result, but it does not have any explanation or justification. It is a big problem for their application given that in the European Union, the EU General Data Protection Regulation stipulates the legal right of end-users to explanations of algorithmic decisions affecting them (Ivanovs et al., 2021). Therefore, from 2018 until the present the number of studies that try to explain the functioning of ANNs has incredibly increased. In this section a general review of the existing methods to interpret NNs is presented.

He et al. (2020), using Andrews et al. (1995) work, divided the methods to interpret shallow networks (the most used in ocean engineering) into decompositional methods, that extract rule algorithms that work on neuron-level; and pedagogical methods, that extract rule algorithms regardless of the internal structure.

Zhang et al. (2021) proposed another taxonomy for network interpretation methods that had three levels, and the main level was passive vs active models. The passive interpretation process starts from a trained network, with all the weights already learned from the training set. In contrast, active methods require some changes before the training. Fan et al. (2021) also presented a classification of interpretation models. They divided them into Post-hoc models, that explain existing models (named as passive models by Zhang et al., 2021) and Ad-hoc models, that build interpretable models (named as active models by Zhang et al., 2021).

Apart from trying to extract rules from models or building new ones that incorporate this explanation there are other ways to improve the interpretability of NNs. The most common ways are visual representation, widely used in the

medical and clinical field (Lopez-Rubio, 2021, Hicks et al., 2018, Le et al., 2019, Lamy and Tsopra, 2019 among others) and sensitivity analysis, that consists of seeing how changing the inputs or one of the characteristics of the inputs would change the final output.

Most of the studies that deal with the problem of the interpretability and explanation of NNs are from the medical field and the great majority of them try to extract explanation rules form the ANNs model. However, most of the time there are no clear and defined rules and the explanation obtained is not enough to have a clear image of how the ANN works.

In ocean engineering there are no specific studies about how to improve the interpretability of ANNs, the only tool used in scientific literature to improve the interpretability of ocean engineering ANNs is the sensitivity analysis. Some of the models presented in this paper included a sensitivity analysis, but there were no specific studies about how to improve the interpretability of a given model or about how changing the parameters of these models would affect the final results. Therefore, more effort needs to be made in this area.

Referring to the parallel processing power problem, there are some studies in both, ocean engineering and water resources field, that tried to optimize the computing time and resources. Most of them focused on the pre-processing of the data. Cigizoglu (2006) improved the performance of a suspended sediment estimation model by using k-means algorithm. Alagha et al. (2017) did the same with a groundwater salinity model, they pre-processed the datasets grouping them into clusters and conducted one ANN for each cluster. Qiao et al. (2018) proposed an optimization method named Neural Rejuvenation for enhancing resource utilization during training. This new model detects useless neurons and "rejuvenates these dead neurons" by resource reallocation and reinitialization optimizing the model.

There are also models that focus and the other big problem of the ANNs that is the amount and quality of the data. Ma et al. (2021) mixed different sea states to improve the performance of the model and Yan et al. (2019) and Kim and Han (2020) used Data Augmentation technique. This problem will be further discussed in the next section.

3.2. Amount and quality of the collected data

As it has been said in the previous section, one of the main problems, of ANNs is the amount and quality of data needed to train and validate the network. In this section an analysis of the data used to train NNs is presented. Table 5 presents a summary of the data used in the studies presented in this paper.

Table 5. Analysis of the data used in the studies.

Authors	Nº of data to construct the model	Nº of input and output variables	Methods of selection of inputs and outputs	Range of variability of variables	Methods of data acquisition	Predictor
Deo and Sridhar Naidu (1998)	3-hourly records during 16 months	1 input 1 output	Time step analysis	1–3 m	Buoys	Wave
Deo et al. (2001)	3-hourly 5 months (Wave) 3-hourly 18 days (Wave and wind)	Inputs =	Sensitivity analysis	0.5–3 m	Buoys and satellite	Wave and wind
	Weekly 42 months (Satellite)	Outputs		3–7 s		
Agrawal and Deo (2002)	3-hourly records during 16 months	2 inputs 1 output	Statistical analysis	0.2–1.4 m	Buoys	Wave
Deo and Jagdale (2003)	5 datasets	3 inputs 2 outputs	–	9–12 cm 9–15 s 0.033–0.1	Experimental setup and previous studies	Wave and slope
Agrawal and Deo (2004)	Every 20 min during 1 year	2660 patterns	Statistical analysis	–	Buoys	Wave

Makarynskyy (2004)	Hourly records during 22 and 20 months	48 inputs 24 outputs	Time step analysis	0–12 m	Buoys	Wave
Kalra et al. (2005a)	Daily records during 4 years	21 inputs 1 output	Statistical analysis	0-1 (Normal)	Buoys and satellite	Wave
Kalra et al. (2005b)	Daily records during 4 years at 4 spots	27 inputs 1 output	Validation with buoys data	0-1 (Normal)	Satellite	Wave and wind
Makarynskyy et al. (2005)	3-hourly records during 1 month	2 to 12 inputs 1 to 5 outputs	Time step analysis	0.5–3.5 m	Buoys	Wave
Jain and Deo (2008)	3-hourly records from 3 to 7 years at 3 spots	4 inputs 1 output	Time step analysis	1–5 m	Buoys	Wave
Malekmohamadi et al. (2008)	Hourly records	8/24 inputs 1 output	Correlation analysis	0–7 m 4–20 m/s	Buoys and numerical models	Wave and wind
Gunaydin (2008)	Monthly records during 25 years at 3 spots	3 inputs 1 output	Statistical analysis	0.5–4.5 m 1.8–11.5 m/s 10.2–10.5 t/m2 0.2–2	Buoys	Wave, wind, air Temperature
Mahjoobi et al. (2008)	Hourly records during 2 months	4 inputs 1 output	Statistical analysis	0-1 (Normal)	Buoys	Wave and wind
Tsai and Tsai (2009)	37,123 records	4 inputs 2 outputs	Statistical analysis	0.2–0.8 (Normal)	Acoustic wave gauges	Wave

Gopinath and Dwarakish (2015)	Daily, weekly, monthly and yearly data during 3 years	52/12/1 inputs 52/12/25 outputs	Statistical analysis	0.5–4.5 m	Buoys	Wave
Salah et al. (2016)	Hourly records during 5 months at 2 spots	–	–	0.96: height 0.58: wind	Buoys and numerical models	Wave and wind
Londhe et al. (2016)	3-hourly records from 2 to 4 years at 4 spots	4 to 9 inputs 1 output	Average Mutual Information	0.25–2.75 m	Buoys and numerical models	Wave
Abhigna et al. (2017)	Hourly records during 2 years	–	Statistical analysis	–	Buoys and numerical models	Wave and wind
Mudronja et al. (2017)	6-hourly records during 22 years	2 inputs 1 output	Correlation analysis	–	Satellite	Wave and wind
Vieira et al. (2020)	Hourly records during 7 months	3 to 5 inputs 1 output	Statistical analysis	–	Buoys and numerical models	Wave and wind
Stavarache et al. (2020)	Daily record during 1 year	5 inputs 1 output	Neural Power software	–	Buoys	Wave and wind
Gao et al. (2021)	Hourly records during 15 months at 3 spots	Input = 2 times Output	Time step analysis	0-1 (Normal)	Buoys	Wave and wind
Joerges et al. (2021)	Half-hourly records during 11 years at 3 spots (with		Bayesian hyperparameter optimization	0-1 (Normal)	Buoys, Meteorogical Service and surveys	Wave, wind, bathymetrics

	many missing values)					
Sadeghifar et al. (2022)	Hourly records during 7 years	4 inputs 1 output	Statistical Analysis	–	Buoys, databases	Wave and wind
Lee and Jeng (2002)	Hourly records during 1 month at 3 spots	4 inputs 1 output	Statistical analysis	-2-2 m	Tide gauges	Tides
Lee et al. (2002)	Hourly records during 3 years	138 inputs 69 outputs	Weighting-relation analysis	0–3 m	Tide gauges	Tides
Lee (2004)	Hourly records during 3 years	138 inputs 69 outputs	Weighting-relation analysis	0–5 m	Tide gauges	Tides
Chang and Lin (2006)	2-hourly records during 1 year	–	–	-1-1 m	Tide gauges and equation	Tides
Han and Shi (2008)	Hourly records during 3 years at 5 spots	–	Power spectrum analysis	−0.2-3 m	Tide gauges	Tides
Liang et al. (2008)	Hourly records during 1 month at 5 spots	–	Statistical analysis	0-1 (Normal)	Tide gauges	Tides
de Oliveira and Ebecken (2009)	Hourly records during 2 years	7 inputs 1 output	Statistical analysis and low-pass filter	0-1 (Normal)	Gauges	Atm pressure, wind, tide
	Monthly records at 56	112 inputs	Statistical analysis	0-1 (Normal)		Sea level

Wenzel and Schroeter (2010)	tide gauges during more than 50 years (many gaps)	56 outputs			PSMSL database and satellite	
Pashova and Popova (2011)	Daily records during 50 years	11 inputs / 1 output	Statistical analysis	0-1 (Normal)	Tide gauges	Sea level
Filippo et al. (2012)	Hourly records for sea level and 3-hourly records for wind and atm pressure during 2 years at 3 spots	4 inputs / 1 output	Intelligent Problem Solver and Statistical NN function	–	Gauges and data derivation	Atm pressure, wind, tide
Karamouz et al. (2014)	Monthly records during 151 years	–	Correlation analysis	–	Tide gauges, atmospheric databases	Sea level, SLP and SST
Castro et al. (2014)	Hourly records during 13 years at 2 spots	3 inputs / 1 output	Statistical analysis	0.5–10	Buoys	Wave
El-Diasty and Al-Harbi (2015)	Hourly records during one month at 6 spots (3 for water level, 3 for atm data)	24 inputs / 1 output	Statistical analysis	0-1 (Normal)	Tide gauges	Sea level and atm data
El-Diasty et al. (2018)	Hourly records during one month at 4 spots	24 inputs / 1 output	Statistical analysis	0–1.4	Tide gauges	Sea level
		5 inputs			Satellite	Sea level

Zhao et al. (2019)	7-day records during 23 years	1 output	Statistical analysis	0-1 (Normal)		
Wang et al. (2020c)	Hourly records during 4 years at 2 spots	–	Statistical analysis	0-1 (Normal)	Tide gauges, atmospheric databases	Sea level, wind
Bruneau et al. (2020)	Hourly records during 15 years	33 inputs 1 output	Statistical analysis	0-1 (Normal)	GESLA database	Sea level, atm data
Avila et al. (2020)	3-hourly records during 18 years at 2 spots	1 output	Statistical analysis	-1-1 (Normal)	Buoys	Wave
Primo de Siqueira and Paiva (2021)	Hourly records during 2 years at 3 spots	–	Correlation analysis		Tide gauges, atmospheric databases	Wave and wind
Bento et al. (2021)	Hourly data from 13 spots	MFO	Moth-flame optimization	0-1 (Normal)	Buoys	Wave and wind
Bai and Xu (2021)	Hourly data from a day	80 and 200 memory cells	Statistical analysis	–	Tide gauges	Tide
Yang et al. (2021)	Daily and hourly records for 2 years at several spots	–	Statistical analysis	0-1 (Normal)	Tide gauges	Tide
Mase (1995)	–	6 inputs 1 output	–	–	Experimental data	Wave, breakwater
Yagci et al. (2005)	80 wave series to 4 slopes	–	Statistical analysis	0-1 (Normal)	Experimental data	Wave, breakwater

Kim and Park (2005)	641 data samples	5-7 inputs 1 output	Index of agreement	–	Experimental data	Wave, breakwater
Iglesias et al. (2008)	127 data samples	4 inputs 1 output	Statistical analysis	0-1 (Normal)	Experimental data	Wave, breakwater
Balas et al. (2010)	554 data samples (FFNN) 166 data samples (PCA)	5/10 inputs 1 output	Statistical analysis	0-1 (Normal)	Experimental data	Wave, breakwater
Akoz et al. (2011)	42 data samples	4 inputs 3 outputs	Statistical analysis	0.2–0.8 (Normal)	Experimental data	Wave, breakwater
Koç and Balas (2012)	166 data samples	5 inputs 1 output	–	–	Experimental data	Wave, breakwater
Suh et al. (2013)	600 data samples	–	–	–	Numerical model	Wave, seal level
Kim et al. (2014)	49 data samples	2 inputs 1 output	Polynomial regression	8–20 m −0.9-0.9 m	Numerical model	Wave, seal level
Londhe and Deo (2003)	–	3/4 inputs 1 output	Statistical analysis	–	Numerical model	Wave
Kankal and Yüksek (2012)	180 data samples	5 inputs 24 outputs	Statistical analysis	0.1–0.9 (Normal)	Experimental set	Wave, breakwater
López and Iglesias (2013)	Field Campaign	3/2 inputs 1 output	K-fold cross validation	0-1 (Normal)	Pressure and tide gauges	Sea level, pressure, tides

	during 3 months					
López and Iglesias (2015)	Hourly records during 15 years at 2 spots and 20-mins during 7 years in the basin	15 inputs / 1 output	K-fold cross validation and sensitivity analysis	-1-1 (Normal)	Buoys and tide gauges	Wave and wind
Zheng et al. (2020)	133 samples	3 inputs / 24 outputs	K-fold cross validation	-1-1 (Normal)	Numerical models	Wave
Moslemi et al. (2022a)	30 datasets	5 inputs / 5 outputs	Statistical analysis	0-1 (Normal)	Numerical models	–
Moslemi et al. (2022b)	98 models	3 inputs / 4 outputs	Statistical analysis	0-1 (Normal)	Numerical models	–

After analysing these data, the following conclusions are reached. The number of data samples used to train and validate the network depends on many factors, mainly, the prediction period, the type of problem, the type of data and the type of neural network:

- To do short-term predictions, normally datasets around 6000 samples is enough. For long-term predictions, larger datasets are needed (around 25,000 samples), although there are studies that used even more than 150,000 points.

- If the aim of the ANN is to predict extreme values, larger datasets are needed to have a good representation of them. Extreme values are much less frequent and to have enough number of extremes in the training set to be able to predict them with a neural network, longer records are needed.

- If different predictors are used to train the network, for example, atmospheric and ocean predictors, the number of data samples needed is less. When different sources of data are used, the model has a more complete

information and the records of data do not have to be as large as when only the historical wave height is used to predict future values.

- The studies that used experimental laboratory data and not real data needed less data samples.

- The amount of necessary data depended also on the type of neural network used, when more sophisticated models are used (wavelets and hybrid approaches), the number of data samples needed is less.

Referring to the number of inputs and outputs and the methods of selecting them, it highly depended on the type of problem. The most common approach is to select them using a statistical analysis that compares both, the correlation and the errors of the different models. In fact, more than 50% of the studies used a statistical analysis to select the input and output variables.

With respect to the range of variability of the variables used, almost all the studies normalized the variables to avoid numerical problems, especially when they used more than one variable to forecast future values. Around 75% of the studies that used more than one predictor, normalized the variables before training the network.

Finally, most of the studies used buoys or tide gauges records to get the data samples. Many of them complemented this data with atmospheric databases or satellite data points. In fact, satellite measures are much easier to obtain and are a good alternative in countries where there are not buoy or tide gauges records. Many breakwater and ports studies also used laboratory data.

4. Conclusions

ANNs are a soft-computing technique ideal to model non-linear and random problems. Therefore, they have a huge potential in ocean and maritime engineering, especially when the input-output structure is unclear. Although each case has to be carefully studied, after reviewing 93 studies, some general patterns and rules are concluded.

More complex ANNs structures do not involve better results. From all the studies reviewed, it can be inferred that, in general, a 3-layer ANN with a number of hidden nodes between 1 and 15 is enough to model ocean engineering problems. In

all the studies presented, the most common structure was the FFNN; recurrent networks usually gave better performance, but they were more computationally expensive and a simpler structure (FFNN) was more convenient in most of the cases. The algorithms that usually gave better results were the BP and LM algorithms. In addition, algorithms classified as "faster algorithms" usually solved the problem in less time, but in the studies reviewed in this paper, faster training did not show better performance.

Referring to the amount of data needed, normally 6000 data samples were enough for short-term predictions and 25,000 for long-term fore casting. However, it depended a lot on the quality of the data, the type of NN used and the aim of the study.

After reviewing all these studies, it can also be concluded that combining different AI techniques and developing hybrid models that imply ANNs; such as ANFIS models, hybrid wavelet models and models that combine traditional approach and soft-computing are very beneficial. They need less data than conventional ANNs and usually perform better. In the last years wavelet neural networks are giving very good results in predicting water level and tides, and LSTM and ANFIS models are also promising tools to forecast wave heights.

However, there are still some issues about ANNs unsolved, such as if they are good tools for long-term predictions or if artificial patterns are better for training than design patterns. The "black-box" problem is still unsolved and there are many questions about the quality and quantity of the data needed. Therefore, future research lines should focus on these topics.

References

Abhigna, P., Jerritta, S., Srinivasan, R., Rajendran, V., 2017. Ieee. Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in bay of bengal. In proceedings of the name of the conference IEEE international conference on communication and signal processing (ICCSP). Melmaruvathur, INDIA.

Agrawal, J.D., Deo, M.C., 2002. On-line wave prediction. Mar. Struct. 15 (1), 57–74. https://doi.org/10.1016/S0951-8339(01)00014–4.

Agrawal, J.D., Deo, M.C., 2004. Wave parameter estimation using neural networks. Mar. Struct. 17 (7), 536–550. https://doi.org/10.1016/j.marstruc.2005.01.001.

Akoz, M.S., Cobaner, M., Kirkgoz, M.S., Oner, A.A., 2011. Prediction of geometrical properties of perfect breaking waves on composite breakwaters. Appl. Ocean Res. 33 (3), 178–185. https://doi.org/10.1016/j.apor.2011.03.003.

Alagha, J.S., Seyam, M., Said, Md, A, M., Mogheir, Y., 2017. Integrating an artificial intelligence approach with k-means clustering to model groundwater salinity: the case of Gaza coastal aquifer (Palestine). Hydrogeol. J. 25 (8), 2347–2361. https:// doi.org/10.1007/s10040-017-1658-1

Andrews, R., Diederich, J., Tickle, A.B., 1995. Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowl. Base Syst. 8 (6), 373–389. https://doi.org/10.1016/0950-7051(96)81920-81924

Avila, D., Marichal, G.N., Padron, I., Quiza, R., Hernandez, A., 2020. Forecasting of wave energy in canary islands based on artificial intelligence. Appl. Ocean Res. 101 https://doi.org/10.1016/j.apor.2020.102189.

Bai, L.-H., Xu, H., 2021. Accurate estimation of tidal level using bidirectional long short-term memory recurrent neural network. Ocean Eng. 235, 108765 https://doi.org/ 10.1016/j.oceaneng.2021.108765.

Balas, C.E., Koc, M.L., Tur, R., 2010. Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters. Appl. Ocean Res. 32 (4), 425–433. https://doi.org/ 10.1016/j.apor.2010.09.005.

Bento, P.M.R., Pombo, J.A.N., Mendes, R.P.G., Calado, M.R.A., Mariano, S.J.P.S., 2021. Ocean wave energy forecasting using optimised deep learning neural networks. Ocean Eng. 219 https://doi.org/10.1016/j.oceaneng.2020.108372.

Bruneau, N., Polton, J., Williams, J., Holt, J., 2020. Estimation of global coastal sea level extremes using neural networks. Environ. Res. Lett. 15 (7) https://doi.org/10.1088/ 1748-9326/ab89d6

Castro, A., Carballo, R., Iglesias, G., Rabu˜ nal, J.R., 2014. Performance of artificial neural networks in nearshore wave power prediction. Appl. Soft Comput. 23, 194–201. https://doi.org/10.1016/j.asoc.2014.06.031.

Chang, H.-K., Lin, L.-C., 2006. Multi-point tidal prediction using artificial neural network with tide-generating forces. Coast. Eng. 53 (10), 857–864. https://doi.org/10.1016/ j.coastaleng.2006.05.001.

Cigizoglu, Kerem, 2006. H.; Kisi, ¨ O. Methods to improve the neural network performance in suspended sediment estimation. J. Hydrol. 317 (3), 221–238. https://doi.org/ 10.1016/j.jhydrol.2005.05.019.

de Oliveira, M.M.F., Ebecken, N.F.F., 2009. A Model to Predict the Coastal Sea Level Variations and Surge. In Proceedings of the Name of the Conference 1st International Conference on Physical Coastal Processes. Management and Engineering, Malta. Demuth, H., Beale, M., 1992. In: Edition, M. (Ed.), Neural Network Toolbox. Natick, Massachusetts, USA.

Deo, M.C., Jagdale, S.S., 2003. Prediction of breaking waves with neural networks. Ocean Eng. 30 (9), 1163–1178. https://doi.org/10.1016/s0029-8018(02)00086-0.

Deo, M.C., Sridhar Naidu, C., 1998. Real time wave forecasting using neural networks. Ocean Eng. 26 (3), 191–203. https://doi.org/10.1016/S0029-8018(97) 10025–10027.

Deo, M.C., Jha, A., Chaphekar, A.S., Ravikant, K., 2001. Neural networks for wave forecasting. Ocean Eng. 28 (7), 889–898. https://doi.org/10.1016/S0029-8018(00) 00027–5.

Ding, S., Chang, X.H., Wu, Q.H., 2013. A study on approximation performances of general regression neural network. In proceedings of the name of the conference 3rd international

conference on machinery electronics and control engineering (ICMECE 2013). Jinan, PEOPLES R CHINA.

Dixit, P., Londhe, S., 2016. Prediction of extreme wave heights using neuro wavelet technique. Appl. Ocean Res. 58, 241–252. https://doi.org/10.1016/j. apor.2016.04.011.

El-Diasty, M., Al-Harbi, S., 2015. Development of wavelet network model for accurate water levels prediction with meteorological effects. Appl. Ocean Res. 53, 228–235. https://doi.org/10.1016/j.apor.2015.09.008.

El-Diasty, M., Al-Harbi, S., Pagiatakis, S., 2018. Hybrid harmonic analysis and wavelet network model for sea water level prediction. Appl. Ocean Res. 70, 14–21. https:// doi.org/10.1016/j.apor.2017.11.007.

Evans, P., Persaud, K.C., McNeish, A.S., Sneath, R.W., Hobson, N., Magan, N., 2000. Evaluation of a radial basis function neural network for the determination of wheat quality from electronic nose data. Sensor. Actuator. B Chem. 69 (3), 348–358. https://doi.org/10.1016/S0925-4005(00)00485–00488.

Fan, S., Xiao, N., Dong, S., 2020. A novel model to predict significant wave height based on long short-term memory network. Ocean Eng. 205 https://doi.org/10.1016/j. oceaneng.2020.107298.

Fan, F.-L., Xiong, J., Li, M., Wang, G., 2021. On interpretability of artificial neural networks: a survey. IEEE transactions on radiation and plasma medical sciences 5 (6), 741–760. https://doi.org/10.1109/trpms.2021.3066428.

Filippo, A., Rebelo Torres, A., Kjerfve, B., Monat, A., 2012. Application of Artificial Neural Network (ANN) to improve forecasting of sea level. Ocean Coast Manag. 55, 101–110. https://doi.org/10.1016/j.ocecoaman.2011.09.007.

Fletcher, R., Reeves, C.M., 1964. Function minimization by conjugate gradients. Comput. J. 7, 149–154.

Gao, S., Huang, J., Li, Y., Liu, G., Bi, F., Bai, Z., 2021. A forecasting model for wave heights based on a long short-term memory neural network. Acta Oceanol. Sin. 40 (1), 62–69. https://doi.org/10.1007/s13131-020-1680-3.

Gopinath, D.I., Dwarakish, G.S., 2015. Wave prediction using neural networks at new mangalore port along west coast of India. Aquatic Procedia 4, 143–150. https://doi. org/10.1016/j.aqpro.2015.02.020.

Graves, A., 2012. Long short-term memory. In: Graves, A. (Ed.), Supervised Sequence Labelling with Recurrent Neural Networks. Springer, Berlin, Heidelberg, pp. 37–45. Berlin Heidelberg.

Gunaydin, K., 2008. The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. Ocean Eng. 35 (14–15), 1406–1415. https://doi.org/10.1016/j.oceaneng.2008.07.008.

Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. PWS Publishing Boston, MA. Han, G., Shi, Y., 2008. Development of an atlantic Canadian coastal water level neural network model. J. Atmos. Ocean. Technol. 25 (11), 2117–2132. https://doi.org/ 10.1175/2008jtecho569.1.

He, C., Ma, M., Wang, P., 2020. Extract interpretability-accuracy balanced rules from artificial neural networks: a review. Neurocomputing 387, 346–358. https://doi. org/10.1016/j.neucom.2020.01.036.

Hicks, S.A., Riegler, M., Pogorelov, K., Anonsen, K.V., de Lange, T., Johansen, D., Jeppsson, M., Randel, K.R., Eskeland, S., Halvorsen, P., 2018. Dissecting deep neural networks for better medical image classification and classification understanding. In proceedings of the name of the conference 31st IEEE international symposium on computer-based medical systems (CBMS). Karlstad Univ, Karlstad, SWEDEN.

Iglesias, G., Rabu˜ nal, J., Losada, M.A., Pach´ on, H., Castro, A., Carballo, R., 2008. A virtual laboratory for stability tests of rubble-mound breakwaters. Ocean Eng. 35 (11), 1113–1120. https://doi.org/10.1016/j.oceaneng.2008.04.014.

Ivanovs, M., Kadikis, R., Ozols, K., 2021. Perturbation-based methods for explaining deep neural networks: a survey. Pattern Recogn. Lett. 150, 228–234. https://doi.org/10.1016/j.patrec.2021.06.030.

Jain, P., Deo, M., 2008. Artificial intelligence tools to forecast ocean waves in real time. Open Ocean Eng. J. 1, 13–20.

Joerges, C., Berkenbrink, C., Stumpe, B., 2021. Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks. Ocean Eng. 232 https://doi.org/10.1016/j.oceaneng.2021.109046.

Kalra, R., Deo, M.C., Kumar, R., Agarwal, V.K., 2005a. RBF network for spatial mapping of wave heights. Mar. Struct. 18 (3), 289–300. https://doi.org/10.1016/j.marstruc.2005.09.003.

Kalra, R., Deo, M.C., Kumar, R., Agarwal, V.K., 2005b. Artificial neural network to translate offshore satellite wave data to coastal locations. Ocean Eng. 32 (16), 1917–1932. https://doi.org/10.1016/j.oceaneng.2005.01.007.

Kankal, M., Yüksek, ¨ O., 2012. Artificial neural network approach for assessing harbor tranquility: the case of Trabzon Yacht Harbor, Turkey. Appl. Ocean Res. 38, 23–31. https://doi.org/10.1016/j.apor.2012.05.009.

Karaboga, D., Kaya, E., 2019. Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. Artif. Intell. Rev. 52 (4), 2263–2293. https://doi.org/10.1007/s10462-017-9610-2.

Karamouz, M., Kia, M., Nazif, S., 2014. Prediction of sea level using a hybrid data-driven model: new challenges after hurricane sandy. Water Quality, Exposure and Health 6. https://doi.org/10.1007/s12403-014-0119-5.

Kim, H.I., Han, K.Y., 2020. Urban flood prediction using deep neural network with data augmentation. Water 12 (3), 899. Retrieved from. https://www.mdpi.com /2073-4441/12/3/899.

Kim, D.H., Park, W.S., 2005. Neural network for design and reliability analysis of rubble mound breakwaters. Ocean Eng. 32 (11), 1332–1349. https://doi.org/10.1016/j.oceaneng.2004.11.008.

Kim, D.H., Kim, Y.J., Hur, D.S., 2014. Artificial neural network based breakwater damage estimation considering tidal level variation. Ocean Eng. 87, 185–190. https://doi.org/10.1016/j.oceaneng.2014.06.001.

Koç, M.L., Balas, C.E., 2012. Genetic algorithms based logic-driven fuzzy neural networks for stability assessment of rubble-mound breakwaters. Appl. Ocean Res. 37, 211–219. https://doi.org/10.1016/j.apor.2012.04.005.

Kusnandar, D., Mara, M.N., Debataraja, N.N., 2014. Diagnostic models for estimating mean sea level change using hybrid model of exponential smoothing and neural network.

In proceedings of the name of the conference 1st international conference on actuarial science and statistics (ICASS). Bandung, INDONESIA.

Lamy, J.-B., Tsopra, R., 2019. Visual Explanation of Simple Neural Networks Using Interactive Rainbow Boxes. In Proceedings of the Name of the Conference 23rd International Conference on the Information Visualisation (IV). Incorporating Biomedical Visualization, Learning Analytics and Geometric Modelling and Imaging, Paris, FRANCE.

Lara-Benitez, P., Carranza-Garcia, M., Riquelme, J.C., 2021. An experimental review on deep learning architectures for time series forecasting. Int. J. Neural Syst. 31 (3) https://doi.org/10.1142/s0129065721300011.

Le, F., Srivatsa, M., Reddy, K.K., Roy, K., 2019. Using graphical models as explanations in deep neural networks. In: Proceedings of the Name of the Conference 16th IEEE International Conference on Mobile Ad Hoc and Smart Systems (IEEE MASS). Ieee, Monterey, CA.

Lee, T.-L., 2004. Back-propagation neural network for long-term tidal predictions. Ocean Eng. 31 (2), 225–238. https://doi.org/10.1016/S0029-8018(03)00115-X.

Lee, T.L., Jeng, D.S., 2002. Application of artificial neural networks in tide-forecasting. Ocean Eng. 29 (9), 1003–1022. https://doi.org/10.1016/S0029-8018(01)00068–3.

Lee, T.L., Tsai, C.P., Jeng, D.S., Shieh, R.J., 2002. Neural network for the prediction and supplement of tidal record in Taichung Harbor, Taiwan. Adv. Eng. Software 33 (6), 329–338. https://doi.org/10.1016/S0965-9978(02)00043–1.

Leung, M.T., Chen, A.-S., Daouk, H., 2000. Forecasting exchange rates using general regression neural networks. Comput. Oper. Res. 27 (11), 1093–1110. https://doi.org/10.1016/S0305-0548(99)00144–00146.

Liang, S.X., Li, M.C., Sun, Z.C., 2008. Prediction models for tidal level including strong meteorologic effects using a neural network. Ocean Eng. 35 (7), 666–675. https://doi.org/10.1016/j.oceaneng.2007.12.006.

Londhe, S.N., Deo, M.C., 2003. Wave tranquility studies using neural networks. Mar. Struct. 16 (6), 419–436. https://doi.org/10.1016/j.marstruc.2003.09.001.

Londhe, S.N., Shah, S., Dixit, P.R., Nair, T.M.B., Sirisha, P., Jain, R., 2016. A coupled numerical and artificial neural network model for improving location specific wave forecast. Appl. Ocean Res. 59, 483–491. https://doi.org/10.1016/j. apor.2016.07.004.

López, M., Iglesias, G., 2013. Artificial Intelligence for estimating infragravity energy in a harbour. Ocean Eng. 57, 56–63. https://doi.org/10.1016/j.oceaneng.2012.08.009.

López, M., Iglesias, G., 2015. Artificial neural networks applied to port operability assessment. Ocean Eng. 109, 298. https://doi.org/10.1016/j.oceaneng.2015.09.016.

Lopez-Rubio, E., 2021. Throwing light on black boxes: emergence of visual categories from deep learning. Synthese 198 (10), 10021–10041. https://doi.org/10.1007/ s11229-020-02700-5.

Ma, X., Huang, L., Duan, W., Jing, Y., Zheng, Q., 2021. The performance and optimization of ANN-WP model under unknown sea states. Ocean Eng. 239, 109858 https://doi.org/10.1016/j.oceaneng.2021.109858.

Mackay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 4, 415–447. Mahjoobi, J., Etemad-Shahidi, A., Kazeminezhad, M.H., 2008. Hindcasting of wave parameters using

different soft computing methods. Appl. Ocean Res. 30 (1), 28–36. https://doi.org/10.1016/j.apor.2008.03.002.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Software 15 (1), 101–124. https://doi.org/10.1016/S1364-8152(99) 00007–9.

Makarynskyy, O., 2004. Improving wave predictions with artificial neural networks. Ocean Engineering 31 (5), 709–724. https://doi.org/10.1016/j. oceaneng.2003.05.003.

Makarynskyy, O., Pires-Silva, A.A., Makarynska, D., Ventura-Soares, C., 2005. Artificial neural networks in wave predictions at the west coast of Portugal. Comput. Geosci. 31 (4), 415–424. https://doi.org/10.1016/j.cageo.2004.10.005.

Malekmohamadi, I., Ghiassi, R., Yazdanpanah, M.J., 2008. Wave hindcasting by coupling numerical model and artificial neural networks. Ocean Eng. 35 (3), 417–425. https://doi.org/10.1016/j.oceaneng.2007.09.003.

Mandal, S., Prabaharan, N., 2006. Ocean wave forecasting using recurrent neural networks. Ocean Eng. 33 (10), 1401–1410. https://doi.org/10.1016/j. oceaneng.2005.08.007.

Marhon, S.A., Cameron, C.J.F., Kremer, S.C., 2013. Recurrent neural networks. In: Bianchini, M., Maggini, M., Jain, L.C. (Eds.), Handbook on Neural Information Processing. Springer, Berlin, Heidelberg, pp. 29–65 (Berlin Heidelberg).

Mase, H., 1995. Evaluation of artificial armour layer stability by neural network method.

Mase, H., Sakamoto, M., Sakai, T., 1995. Neural network for stability analysis of rubble-mound breakwaters. J. Waterway Port Coastal Ocean Eng. asce 121, 294–299.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas imminent in nervous activity. Bull. Math. Biophys. 115–133. s

Medina, J.R., Garrido, J.M., G´ omez-Martín, M.E., Vidal, C., 2004. Armor damage analysis using neural networks. Coastal structures 236–248.

Menezes, J.M.P., Barreto, G.A., 2008. Long-term time series prediction with the NARX network: an empirical evaluation. Neurocomputing 71 (16), 3335–3343. https://doi. org/10.1016/j.neucom.2008.01.030.

Møller, M., Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural networks 6, 525-533. Neural Network. 6, 525–533. https://doi.org/ 10.1016/S0893-6080(05)80056-5.

Moslemi, N., Abdi, B., Gohari, S., Sudin, I., Atashpaz-Gargari, E., Redzuan, N., Ayob, A., Burvill, C., Su, M., Arya, F., 2022a. Thermal response analysis and parameter prediction of additively manufactured polymers. Appl. Therm. Eng. 212, 118533 https://doi.org/10.1016/j.applthermaleng.2022.118533.

Moslemi, N., Gohari, S., Abdi, B., Sudin, I., Ghandvar, H., Redzuan, N., Hassan, S., Ayob, A., Rhee, S., 2022b. A novel systematic numerical approach on determination of heat source parameters in welding process. J. Mater. Res. Technol. 18, 4427–4444. https://doi.org/10.1016/j.jmrt.2022.04.039.

Mudronja, L., Matic, P., Katalinic, M., 2017. Data-based modelling of significant wave height in the adriatic sea. Transactions on Maritime Science-Toms 6 (1), 5–13. https://doi.org/10.7225/toms.v06.n01.001.

Pashova, L., Popova, S., 2011. Daily sea level forecast at tide gauge Burgas, Bulgaria using artificial neural networks. J. Sea Res. 66 (2), 154–161. https://doi.org/10.1016/j.seares.2011.05.012.

Powell, M.J.D., 1977. Restart procedures for the conjugate gradient method. Math. Program. 12, 241–254.

Primo de Siqueira, B.V., Paiva, A.d.M., 2021. Using neural network to improve sea level prediction along the southeastern Brazilian coast. Ocean Model. 168 https://doi.org/10.1016/j.ocemod.2021.101898.

Qiao, J., Li, F., Han, H., Li, W., 2016. Constructive algorithm for fully connected cascade feedforward neural networks. Neurocomputing 182, 154–164. https://doi.org/10.1016/j.neucom.2015.12.003.

Qiao, S., Lin, Z., Zhang, J., Yuille, A., 2018. Neural Rejuvenation: Improving Deep Network Training by Enhancing Computational Resource Utilization.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. In: McClelland, J.L. (Ed.), Learning Internal Representations by Error Propagation. MIT Press, Cambridge.

Sadeghifar, T., Motlagh, M.N., Azad, M.T., Mahdizadeh, M.M., 2017. Coastal wave height prediction using recurrent neural networks (RNNs) in the south caspian sea. Mar. Geodes. 40 (6), 454–465. https://doi.org/10.1080/01490419.2017.1359220.

Sadeghifar, T., Lama, G.F.C., Sihag, P., Bayram, A., Kisi, O., 2022. Wave height predictions in complex sea flows through soft-computing models: case study of Persian Gulf. Ocean Eng. 245, 110467 https://doi.org/10.1016/j. oceaneng.2021.110467.

Salah, P., Reisi-Dehkordi, A., Kamranzad, B., 2016. A hybrid approach to estimate the nearshore wave characteristics in the Persian Gulf. Appl. Ocean Res. 57, 1–7. https://doi.org/10.1016/j.apor.2016.02.005.

Standford University Machine Learning Course. Available online. https://www.coursera.org/learn/machine-learning/home.

Stavarache, G., Ciortan, S., Rusu, E., 2020. Optimization of artificial neural networks based models for wave height prediction. In proceedings of the name of the conference 5th international conference on advances and clean energy research (ICACER). Electron. Netw.

Su, I.-J., Tsai, C.-C., Sung, W.-T., 2011. Comparison of BP and GRNN Algorithm for Factory Monitoring In Proceedings of the Name of the Conference 1st International Conference on Mechanical Engineering. Phuket, THAILAND.

Suh, K.-D., Kim, S.-W., Kim, S., Cheon, S., 2013. Effects of climate change on stability of caisson breakwaters in different water depths. Ocean Engineering 71, 103–112. https://doi.org/10.1016/j.oceaneng.2013.02.017.

Teslyuk, V., Kazarian, A., Kryvinska, N., Tsmots, I., 2021. Optimal artificial neural network type selection method for usage in smart house systems. Sensors 21 (1), 47. Retrieved from. https://www.mdpi.com/1424-8220/21/1/47.

Tsai, C.-P., Lee, T.-L., 1999. Back-Propagation neural network in tidal-level forecasting. J. Waterw. Port, Coast. Ocean Eng. 125 (4), 195–202. https://doi.org/10.1061/ (ASCE)0733-950X(1999)125:4(195).

Tsai, J.-C., Tsai, C.-H., 2009. Wave measurements by pressure transducers using artificial neural networks. Ocean Eng. 36 (15), 1149–1157. https://doi.org/10.1016/j. oceaneng.2009.08.007.

Vieira, F., Cavalcante, G., Campos, E., Taveira-Pinto, F., 2020. A methodology for data gap filling in wave records using Artificial Neural Networks. Appl. Ocean Res. 98 https://doi.org/10.1016/j.apor.2020.102109.

Wang, X., Yin, J., Wang, H., 2020a. Ieee. Modular tide prediction model based on improved wavelet neural network. In proceedings of the name of the conference 32nd Chinese control and decision conference (CCDC). Hefei, PEOPLES R CHINA.

Wang, H., Yin, J., Wang, X., 2020b. Ieee. Optimization of wavelet neural network model for tide prediction based on genetic algorithm. In proceedings of the name of the conference 32nd Chinese control and decision conference (CCDC). Hefei, PEOPLES R CHINA.

Wang, B., Wang, B., Wu, W., Xi, C., Wang, J., 2020c. Sea-water-level prediction via combined wavelet decomposition, neuro-fuzzy and neural networks using SLA and wind information. Acta Oceanol. Sin. 39 (5), 157–167. https://doi.org/10.1007/ s13131-020-1569-1.

Wenzel, M., Schroeter, J., 2010. Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. Journal of Geophysical Research-Oceans 115. https://doi.org/10.1029/2009jc005630.

Yagci, O., Mercan, D.E., Cigizoglu, H.K., Kabdasli, M.S., 2005. Artificial intelligence methods in breakwater damage ratio estimation. Ocean Eng. 32 (17), 2088–2106. https://doi.org/10.1016/j.oceaneng.2005.03.004.

Yan, Y., Zhang, C., Su, N., 2019. A novel data augmentation method for sea ice scene classification on of arctic aerial images. IEEE Access 7, 104241–104249. https://doi. org/10.1109/access.2019.2929861.

Yang, C.-H., Wu, C.-H., Hsieh, C.-M., Wang, Y.-C., Tsen, I.F., Tseng, S.-H., 2021. Deep learning for imputation and forecasting tidal level. IEEE J. Ocean. Eng. 46 (4), 1261–1271. https://doi.org/10.1109/joe.2021.3073931.

Yazdizadeh, A., Khorasani, K., 2002. Adaptive time delay neural network structures for nonlinear system identification. Neurocomputing 47 (1), 207–240. https://doi.org/ 10.1016/S0925-2312(01)00589–6. Yogitha, R., Mathivanan, G., 2018. Performance Analysis of Transfer Functions in an Artificial Neural Network, 0393–0397.

Zhang, Y., Tiˇno, P., Leonardis, A., Tang, K., 2021. A survey on neural network interpretability. IEEE Transactions on Emerging Topics in Computational Intelligence 5 (5), 726–742. https://doi.org/10.1109/TETCI.2021.3100641.

Zhao, J., Fan, Y., Mu, Y., 2019. sea level prediction in the yellow sea from satellite altimetry with a combined least squares-neural network approach. Mar. Geodes. 42 (4), 344–366. https://doi.org/10.1080/01490419.2019.1626306.

Zheng, Z., Ma, X., Ma, Y., Dong, G., 2020. Wave estimation within a port using a fully nonlinear Boussinesq wave model and artificial neural networks. Ocean Eng. 216 https://doi.org/10.1016/j.oceaneng.2020.108073.

3.6. Uncertainties in the application of Artificial Neural Networks in ocean engineering

Abstract and Keywords

Artificial Neural Networks (ANNs) are becoming more popular to model ocean engineering problems. With the development of Artificial Intelligence, data-driven models are showing better behaviour and performance than the traditional models used in ocean engineering. However, the main limitation of ANNs models is the uncertainty associated to them and their "black box" nature. ANNs models present final results without any uncertainty or explanation and this limits enormously their applicability, especially in decision-making tasks. This research paper tries to deal with this problem. Given the exponential growth that artificial intelligence models are currently experiencing, it is necessary to address these limitations so that the field of ocean engineering does not fall behind. This is the final aim of this review research paper. A review of how ocean engineering studies have dealt with ANNs uncertainties has been carried out with the final objective of proposing new methods to deal with these uncertainties to make the application of this tool less constrained by its "black-box nature" and facilitate the expansion of these models in the field.

Keywords: Artificial neural networks; Uncertainties; Ocean engineering; Artificial intelligence; Soft-computing; Data-driven models

1. Introduction

Artificial Neural Networks (ANNs) are one of the most popular techniques of artificial intelligence (AI) (Medina et al., 2004). They were introduced by McCulloch and Pitts in 1943 (Maier and Dandy, 2000; McCulloch and Pitts, 1943) and since then they have proved to be one of the most efficient techniques to model almost every type of problem. ANNs are data-driven models that base their performance on the input data. They are made of different layers of neurons that pass the information from one layer to another applying different algorithms and transfer functions, trying to simulate the operation of the human brain. ANNs do not have strict governing rules, being an ideal methodology to model random processes (Maier and Dandy, 2000). Ocean engineering problems usually involve random and non-linear patterns, which makes them perfectly suited to ANN modelling (Deo et al., 2001). ANNs started to be used in ocean engineering in 1995,

when Mase (1995) used them to evaluate the artificial armour layer stability of a breakwater. Since then, ANNs have been widely used. However, despite their great potential in ocean engineering and the fact that they have proved to perform even better than traditional numerical models, ANNs have not reached the same level of development in ocean engineering as in other similar fields such as hydraulics or hydrology (Portillo Juan and Negro Valdecantos, 2022).

The main problems of ANNs is that they are "black-box" models (Nabavi-Pelesaraei et al., 2020; Shamshirband et al., 2020), which means that there is not a clear understanding of their operation, they give a final result, but it does not have any explanation or justification (Portillo Juan and Negro Valdecantos, 2022). A physical explanation of the ANN model does not exist, which makes its interpretability very difficult. Moreover, ANNs do not reproduce identical results, as their performance depends on different model parameters that vary each time the ANN is applied (Elshorbagy et al., 2009; Elshorbagy, A. et al., 2010). Results given by ANNs are usually presented without any uncertainty, this is mainly due to the "black box" problem mentioned above and the difficulty of determining the uncertainties related to the model (Coral et al., 2016). This problem limits the potential of ANNs especially in tasks involving decision-making, as there are no clear explanations of the results obtained (Borrego et al., 2008; Cabaneros and Hughes, 2022).

Different authors from several fields have tried to quantify ANNs uncertainties and have developed different methodologies to minimize these uncertainties. (Chryssolouris et al., 1996; Hwang and Ding, 1997; Veaux et al., 1998) are the main references and the authors that started to deal with this problem (Coral et al., 2016). In water resources King ston and Lambert (2005) and Khan and Coulibaly (2006) applied Bayesian Neural Networks (BNN) to deal with uncertainties. Boucher et al. (2009) proposed a novel approach based on the ensemble technique to assess the performance of ANNs. Alvisi and Franchini (2011) used fuzzy neural networks to quantify uncertainties. Chitsazan et al. (2015) developed a hierarchical Bayesian model averaging (HBMA) method to analyze prediction uncertainty and Raheli et al. (2017) developed a hybrid model to quantify the oxygen in a river and the uncertainties associated to these predictions. Kwon et al. (2020) applied BNN to quantify uncertainties in biomedical images segmentation. Weiss et al. (2021) developed a tool (UNCERTAINTY-WIZARD) to quantify uncertainties in ANN models. Finally, Cabaneros and Hughes (2022) recently did

a review on the methods used to handle uncertainties in the application of ANNs in air pollution forecasting.

However, despite some authors' efforts to deal with this problem, the research on uncertainties in ANNs is still very limited and it is still a challenging task for the scientific community, especially in ocean engineering, where, to the authors' knowledge, there are almost no studies that deal with this problem. The aim of this research paper is to tackle this problem and measure the uncertainties in the application of ANNs in ocean engineering, with the final objective of proposing different methodologies to end with these uncertainties.

The growing interest in artificial intelligence techniques and, in particular, in neural networks makes this study particularly relevant. There is a high level of interest in deep learning, machine learning and ANNs, and research is increasingly focusing on optimization techniques (Kariri et al., 2023). In fact, in ocean engineering, the number of scientific studies in indexed journals that apply neural networks has grown exponentially from 2016 and it is expected to continue to grow in the next years (Fig. 1).

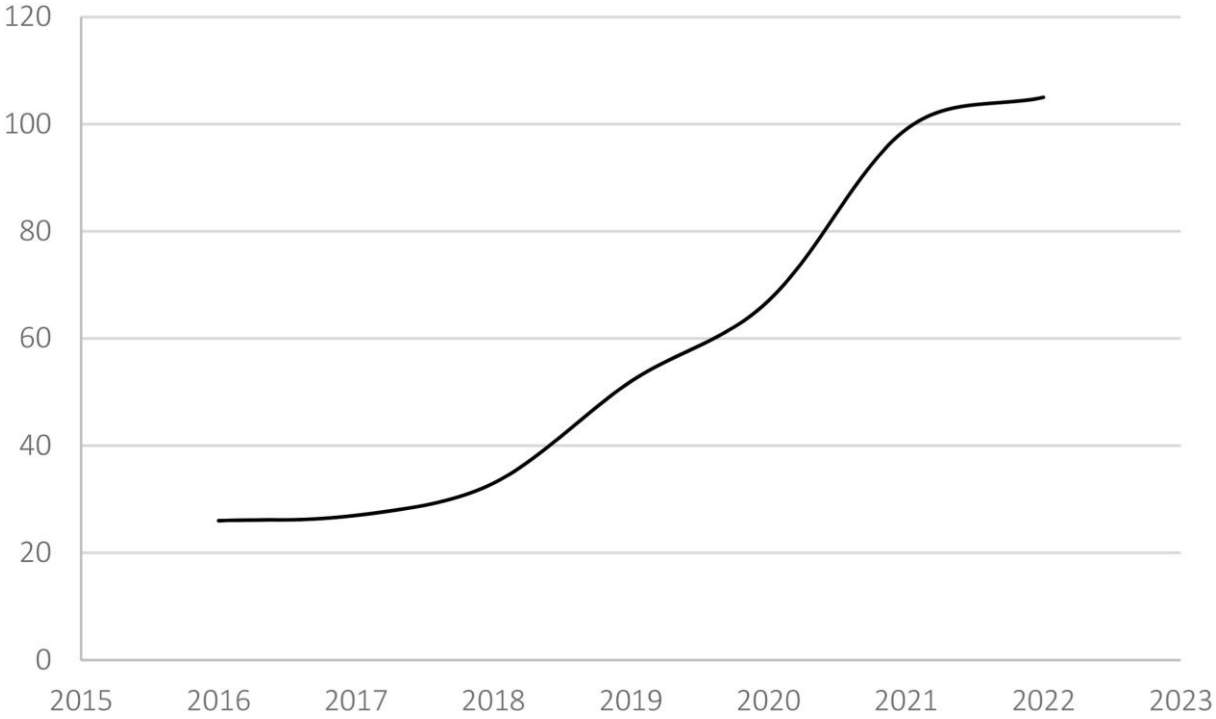


Fig. 1. Evolution of the number of research papers about ANNs in ocean engineering indexed in Web of Science.

It is likely that in the coming years artificial intelligence models will be the most widely used, but first, the problem of uncertainties inherent in these models must be resolved. Due to this fact, new disciplines such as explainable artificial

intelligence (XAI) are emerging to address the main barriers of these models (Samek et al., 2021), and they are growing even with a faster rate that ANNs studies (Fan et al., 2021). However, these disciplines have not yet been applied to ocean engineering. In this context, the authors of the article have seen it as essential to carry out the present study. The aim of this article is to put into context the main uncertainties faced by neural networks in the field of ocean engineering with the objective of being the starting point for the application of XAI and of proposing solutions to the main limitations and drawbacks of ANNs.

To conduct this study, Portillo and Negro review of the application of ANNs in ocean engineering has been taken as the reference document (Portillo Juan and Negro Valdecantos, 2022).

2. Uncertainties

In this section the main sources of uncertainties and the methods to handle them are presented.

2.1. Source of uncertainties

Usually, uncertainties are classified in two main types: aleatoric and epistemic. Aleatoric uncertainties are those uncertainties inherent to the dynamics of the system, such as predictor excitations, noisiness, or im perfections. Epistemic uncertainties are those that are due to the lack of knowledge of the model (Cabaneros and Hughes, 2022). This study will focus on the latter, as epistemic uncertainties are what limits the applicability of ANNs.

Uncertainties when applying ANNs in ocean engineering problems can be divided in two main blocks: uncertainties due to the inputs and the noise associated to these inputs; and model uncertainties, that are caused by the ANN model and its forecasting capacity (Li et al., 2019). Model uncertainty can be divided into the uncertainty related to the parameters of the model and the uncertainty linked to the model structure. In this section all these sources of uncertainties are explained (Fig. 2).

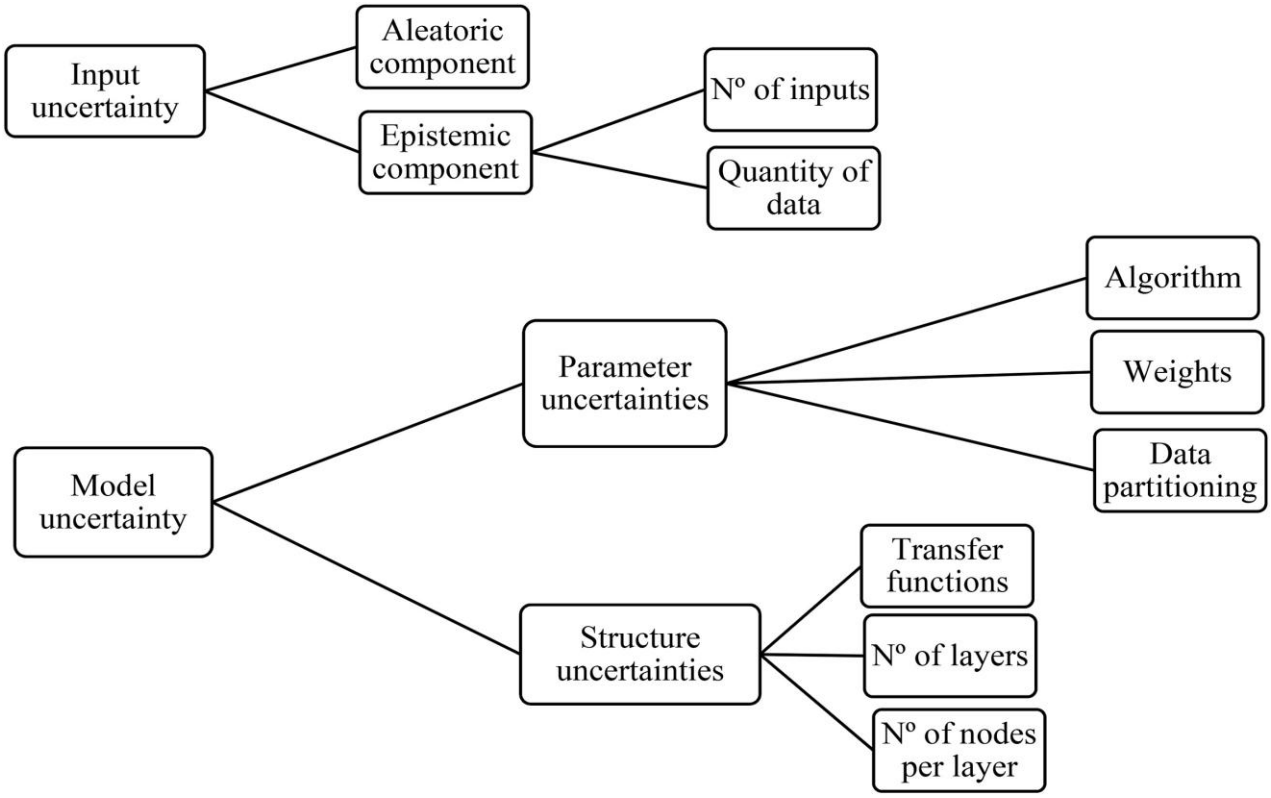


Fig. 2. Classification of ANNs uncertainties in ocean engineering.

Input uncertainties

Input uncertainties can be defined as those that are associated to the variables that are used as inputs of the ANN. Input uncertainties in the application of ANNs in ocean engineering have two main components, one aleatoric and one epistemic. The aleatoric component of input uncertainties is due to the measurement of the variables, to the accuracy of buoys, tide gauges, etc. And to the noise associated to these variables. The epistemic component of input uncertainties is that associated with the ANN model and it is the one on which this article focuses. To develop an ANN, one of the main problems is to select the number of variables to use as an input and to choose the quantity of data to train and validate the model. There are no general rules for identifying the optimal amount of data needed for training and the optimum variables that model the process without incurring unnecessary computational costs (Cabaneros and Hughes, 2022). Some predictors usually model the process better than others, and to develop a neural network model these must be carefully chosen. However, there is still no clear methodology to optimize the number of inputs and the quantity of data to use in ANN models, the great majority of the models do this by trial and error, which has a large computational cost.

Model uncertainties

As explained above, model uncertainties of ANNs can be divided into parameter uncertainties and structure uncertainties.

Parameter uncertainties refer to those uncertainties related to the selection of parameters to develop the model. These parameters are mainly the selection of the algorithms, the initial weights, and the di vision of the data. Although some algorithms have proven to perform better than others, a clear methodology to select the optimum algorithm does not exist (Portillo Juan and Negro Valdecantos, 2022). Another parameter to be chosen when training and validating an ANN are the initial weights and the division of the data. There are different methods to select these parameters such as k-fold cross-validation, randomly, stratified, etc. The selection of the initial weights and data partitioning is directly related to the problem of the impossibility for neural networks to give identical results each time they are trained and used, which limits its applicability and highly increases the uncertainties of these models.

Structure uncertainties refer to the uncertainties related to the architecture of the model. These uncertainties are mainly the selection of the transfer functions between the layers, the number of layers and the number of nodes in each layer. As happened with the rest of uncertainties there are no clear methodologies to optimize these values, and usually a trial and error method is used to obtain the structure of the model, which has a high computational cost.

2.2. Methods for addressing ANNs uncertainties

The main methods to handle the uncertainties in the application of ANNs are explained in this section.

Almost all studies in ocean engineering apply the trial and error technique. This is the most extended approach, especially in early studies. It has also been applied with traditional methods, such as linear regression. It is the most basic, but also computationally expensive solution to handle ANN uncertainties. As its name suggests, it basically consists of developing different models and trying different number of inputs, predictors, ANN parameters and model structures. After ANNs are trained and validated, a statistical analysis is usually conducted to see which of the combinations gives the best performance. As can be assumed, testing all

possible combinations to optimize the model is practically impossible due to the large computational cost that would be associated with it.

Another approach similar to the trial an error is the ensemble approach. In this case, instead of selecting the model that performs better, different models are developed, and the final results are ensembled to create one single output that is the combination of all the models. It is not very common in ocean engineering, as the trial and error approach is usually preferred.

Sensitivity analysis is also an extended technique among ocean engineering studies to account uncertainty. It is usually used to neglect variables that has very little effect on the final result. Minor perturbations are introduced on each of the variables and the effect on the output is examined to see if it varies considerably or not.

There are also other more advanced techniques that although they are not very common among ocean engineers, they have been applied in other fields. They are mainly bayesian techniques, fuzzy methodologies, bootstrapping, Monte Carlo simulations and optimization methods.

Bayesian approach is based on Bayes' theorem and it defines the network weights was a probability density function (PDF). A prior PDF, p0 (w), is assigned to the network parameters which is then updated using the training data and Bayes' theorem to yield the posterior PDF (Cabaneros and Hughes, 2022).

Fuzzy methodologies are based on fuzzy logic that operates using if- then rules (Cabaneros and Hughes, 2022). In ocean engineering is also common to develop Adaptive Neuro-Fuzzy Inference System (ANFIS) models to improve the performance of ANNs. However, this approach cannot estimate model uncertainty and additional uncertainties estimations must be done.

Bootstrapping is an intensive resampling technique with replacement and it is based on developing k networks with the same configuration. Each ANN is trained with a set of n elements randomly chosen (with replacement) from an N element database ($n < N$) (Coral et al., 2016). It is carried out by sampling various realizations of input–output patterns to estimate statistical characteristics (Cabaneros and Hughes, 2022). Although it has been proven to be one of the best methodologies to assess uncertainty in ANNs in similar fields, such as water resources or atmospheric studies, it is not a popular technique when applying ANNs in ocean engineering problems.

Monte Carlo simulations is a probabilistic technique that samples different realizations of model inputs or parameters by assigning the ranges and probability distribution of each variable. It assigns different PDFs to each predictor (Cabaneros and Hughes, 2022; Kasiviswanathan and Sudheer, 2017). It happens the same as with bootstrapping methodologies, although it is one of the best techniques to solve model uncertainties, it is rarely used in ANNs in ocean engineering.

Optimization methods, unlike the other methodologies presented, identify the ANN parameters with an acceptable level of uncertainty rather than training the model and then quantifying the uncertainty (Kasiviswanathan and Sudheer, 2017). The most widely used technique is genetic algorithms (GA).

3. Methodology

The methodology followed in this review paper can be divided into three main blocks. The search for scientific articles, the statistical analysis, and the extrapolation of different methodologies (Fig. 3).

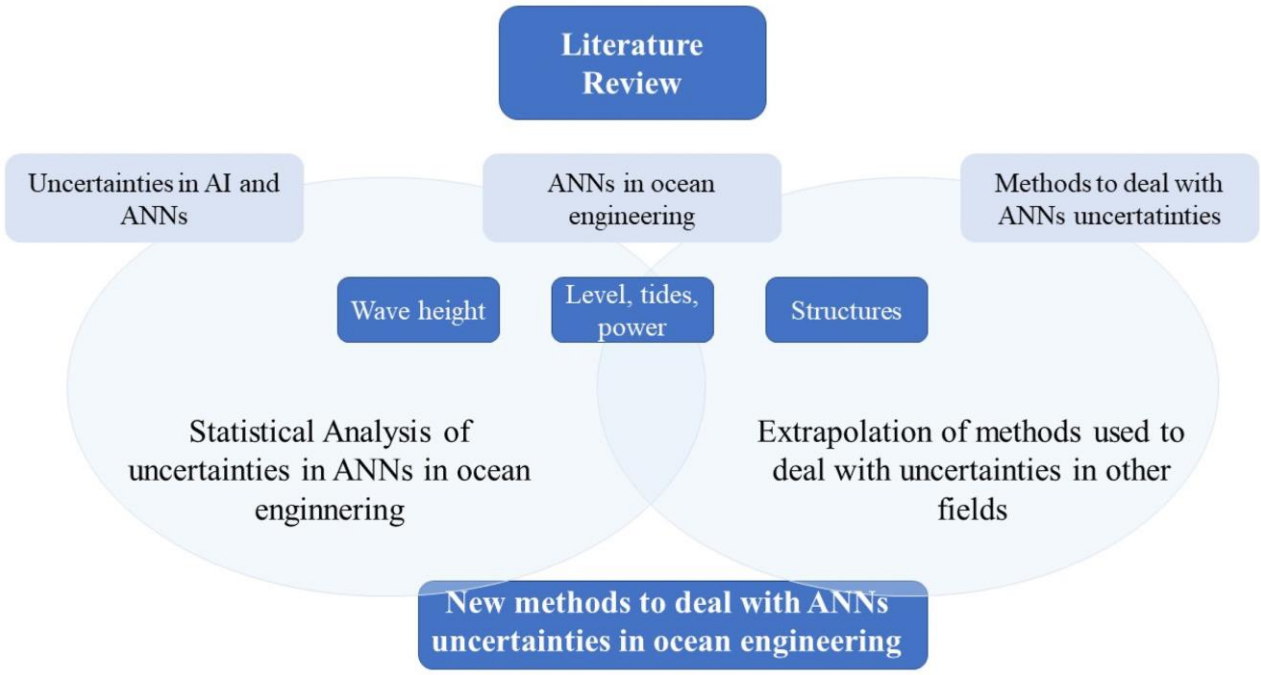


Fig. 3. Methodology.

For the first phase, a structured literature review was conducted, following the same methodology as in Portillo Juan et al. (2022). This methodology has three main stages: planning, review, and information synthesis. In this case, three main thematic blocks could be distinguished; uncertainties in ANNs, ANNs in ocean

engineering and methods to deal with ANNs uncertainties. The research papers were searched by keywords using “Uncertainties Artificial Neural Networks” in the first block, “Artificial Neural Networks Ocean Engineering” in the second block and “Methods uncertainties Artificial Neural Networks” in the third block. The databases used were mainly Web of Science (WOS), Scopus and PubMed. Once all the articles were searched, they were filtered and classified by topic, resulting in a total of 219 research pa pers. The inclusion criterion was that the papers actually dealt either with ANNs uncertainties or that applied neural networks to ocean engineering problems. Priority was given to previous review articles; first all review articles of the topic were summarized and then the research papers.

With all the articles classified and filtered, an article-by-article study was made of how the different uncertainties inherent to neural networks had been dealt with. While in blocks 1 and 3 more importance was given to summarising the main ideas and conclusions of the articles, in block 2 (neural networks in ocean engineering) the analysis and review focused mainly on the methodology of these studies. For the block 2 studies, all data on parameters, model structure, algorithms, etc. were taken for further statistical treatment. With this data, a statistical study based on frequencies and rule extraction was carried out to extract general pat terns of how studies using neural networks in ocean engineering have faced the main limitations of these models.

Finally, a review of how other disciplines have dealt with these uncertainties was conducted and an extrapolation was made for ocean engineering with the aim of improving the implementation of ANNs, pointing out how future studies can improve the handling of these uncertainties.

4. Analysis of ANNs uncertainties in ocean engineering

In this section, the studies of the review presented by Portillo Juan and Negro Valdecantos (2022) are analyzed in terms of uncertainty. The method to deal with each uncertainty is presented for each study and general statistics and patterns are obtained and highlighted. This section is divided into four subsections. Section 4.1. Input uncertainty, Section 4.2. Parameter uncertainty, Section 4.3. Structure uncertainty and Section 4.4 Statistical summary, which presents a summary of the previous subsections.

The results in section 4.1. have been specifically divided by the thematic areas used by Portillo Juan and Negro Valdecantos (2022) in their review, as the uncertainty

associated with the input variables de pends on the area of ocean engineering being modelled and on the topic being addressed. However, in the case of section 4.2. and 4.3., which are about the uncertainties directly related to the model and do not depend on the variables used, the results presented are global and are not distinguished by thematic areas to facilitate their interpretation and understanding. However, with the aim of also providing and image of which type of methodology is used in each type of problem, in the end of each subsection, a table that classifies the most used methodologies to deal with uncertainties according to the type of problem is presented.

4.1. Input uncertainties

The epistemic component of input uncertainty is analyzed in this section. Table 1 presents the main factors related to this uncertainty (number of inputs and quantity of data) and how the studies presented in Portillo Juan and Negro Valdecantos (2022) dealt with this uncertainty.

Table 1. Input uncertainties.

Authors	Data quantity	Nº of inputs	Method to handle input uncertainty
	Wave forecasting studies		
Deo and Sridhar Naidu (1998)	3-hourly records during 16 months	1 input (Wave)	Time step analysis
Deo et al. (2001)	3-hourly 5 months (Wave) 3-hourly 18 days (Wave and wind) Weekly 42 months (Satellite)	Wave and wind	Sensitivity analysis
Agrawal and Deo (2002)	3-hourly records during 16 months	2 inputs (Wave)	Statistical analysis
Deo and Jagdale (2003)	5 datasets	3 inputs Wave and slope	–
Agrawal and Deo (2004)	Every 20 min during 1 year	Wave	Statistical analysis

Makarynskyy (2004)	Hourly records during 22 and 20 months	48 inputs (Wave)	Time step analysis
Kalra et al. (2005b)	Daily records during 4 years	21 inputs (Wave)	Statistical analysis.Fuzzy methods
Kalra et al. (2005a)	Daily records during 4 years at 4 spots	27 inputs (Wave and wind)	Validation with buoys data and sensitivity analysis
Makarynskyy et al. (2005)	3-hourly records during 1 month	2 to 12 inputs (Wave)	Time step analysis
Jain, 2008	3-hourly records from 3 to 7 years at 3 spots	4 inputs (Wave)	Time step analysis
Malekmohamadi et al. (2008)	Hourly records	8/24 inputs (Wave and wind)	Correlation analysis
Gunaydin (2008)	Monthly records during 25 years at 3 spots	3 inputs (Wave, wind and air temperature)	Statistical analysis
Mahjoobi et al. (2008)	Hourly records during 2 months	4 inputs (Wave and wind)	Statistical analysis, Fuzzy methods
Tsai and Tsai (2009)	37,123 records	4 inputs (Wave)	Statistical analysis
Gopinath and Dwarakish (2015)	Daily, weekly, monthly and yearly data during 3 years	52/12/1 inputs (Wave)	Statistical analysis
Salah et al. (2016)	Hourly records during 5 months at 2 spots	Wave and wind	–
Londhe et al. (2016)	3-hourly records from 2 to 4 years at 4 spots	4 to 9 inputs (Wave)	Average Mutual Information
Abhigna et al. (2017)	Hourly records during 2 years	Wave and wind	Statistical analysis
Mudronja et al. (2017)	6-hourly records during 22 years	2 inputs (Wave and wind)	Correlation analysis

Vieira et al. (2020)	Hourly records during 7 months	3 to 5 inputs (Wave and wind)	Statistical analysis
Stavarache and Ciortan, 2020	Daily record during 1 year	5 inputs (Wave and wind)	Neural Power software
Gao et al. (2021)	Hourly records during 15 months at 3 spots	Wave and wind	Time step analysis
Joerges et al. (2021)	Half-hourly records during 11 years at 3 spots (with many missing values)	Wave, wind, bathymetrics	Bayesian hyperparameter optimization
Sadeghifar et al. (2022)	Hourly records during 7 years	4 inputs (Wave and wind)	Statistical Analysis, Fuzzy methods
	Water level, tides, extremes and energy studies		
Lee, T. L. and Jeng (2002)	Hourly records during 1 month at 3 spots	4 inputs (Tides)	Statistical analysis
Lee, T. L. et al. (2002)	Hourly records during 3 years	138 inputs (Tides)	Weighting-relation analysis
Lee, T.-L., 2004	Hourly records during 3 years	138 inputs (Tides)	Weighting-relation analysis
Chang and Lin (2006)	2-hourly records during 1 year	Tides	–
Han and Shi (2008)	Hourly records during 3 years at 5 spots	Tides	Power spectrum analysis
Liang et al. (2008)	Hourly records during 1 month at 5 spots	Tides	Statistical analysis
de Oliveira and Ebecken (2009)	Hourly records during 2 years	7 inputs (Atm pressure, wind, tide)	Statistical analysis and low-pass filter
Wenzel, 2010	Monthly records at 56 tide gauges during more	112 inputs (Sea level)	Statistical analysis and k-fold cross validation

	than 50 years (many gaps)		
Pashova and Popova (2011)	Daily records during 50 years	11 inputs (Sea level)	Statistical analysis
Filippo et al. (2012)	Hourly records for sea level and 3-hourly records for wind and atm pressure during 2 years at 3 spots	4 inputs (Atm pressure, wind, tide)	Intelligent Problem Solver and Statistical NN function
Karamouz et al. (2014)	Monthly records during 151 years	Sea level, SLP and SST	Correlation analysis. Selection methods
Castro et al. (2014)	Hourly records during 13 years at 2 spots	3 inputs (Wave)	Statistical analysis
El-Diasty and Al-Harbi, 2015	Hourly records during 1 month at 6 spots (3 for water level, 3 for atm data)	24 inputs (Sea level and atm data)	Statistical analysis
El-Diasty et al. (2018)	Hourly records during 1 month at 4 spots	24 inputs (Sea level)	Statistical analysis
Zhao et al. (2019)	7-day records during 23 years	5 inputs (Sea level)	Statistical analysis
Wang, B. et al. (2020)	Hourly records during 4 years at 2 spots	Sea level, wind	Statistical analysis. Fuzzy methods K-fold cross validation
Bruneau et al. (2020)	Hourly records during 15 years	33 inputs (Sea level, atm data)	Statistical analysis
Avila et al. (2020)	3-hourly records during 18 years at 2 spots	Wave	Statistical analysis. Fuzzy methods
Primo de Siqueira and Paiva, 2021	Hourly records during 2 years at 3 spots	Wave and wind	Correlation analysis

Bento et al. (2021)	Hourly data from 13 spots	Wave and wind	Moth-flame optimization (MFO)
Bai and Xu (2021)	Hourly data from a day	Tide	Statistical analysis
Yang et al. (2021)	Daily and hourly records for 2 years at several spots	Tide	Statistical analysis
	Breakwaters, ports and pipes studies		
Mase et al. (1995)	–	6 inputs (Wave, breakwater)	–
Yagci et al. (2005)	80 wave series to 4 slopes	Wave, breakwater	Statistical analysis
Kim, D. H. and Park (2005)	641 data samples	5-7 inputs (Wave, breakwater)	Index of agreement
Iglesias et al. (2008)	127 data samples	4 inputs (Wave, breakwater)	Statistical analysis
Balas et al. (2010)	554 data samples (FFNN) 166 data samples (PCA)	5/10 inputs (Wave, breakwater)	Statistical analysis. Fuzzy methods
Akoz et al. (2011)	42 data samples	4 inputs (Wave, breakwater)	Statistical analysis
Koç and Balas (2012)	166 data samples	5 inputs (Wave, breakwater)	Optimization (GA) and Fuzzy methods
Suh et al. (2013)	600 data samples	Wave, sea level	–
Kim, D. H. et al. (2014)	49 data samples	2 inputs (Wave, sea level)	Polynomial regression
Londhe and Deo (2003)	–	3/4 inputs (Wave)	Statistical analysis
Kankal and Yüksek (2012)	180 data samples	5 inputs (Wave, breakwater)	Statistical analysis

López, M. and Iglesias (2013)	Field Campaign during 3 months	3/2 inputs (Sea level, pressure tides)	K-fold cross validation
López and Iglesias, 2015	Hourly records during 15 years at 2 spots and 20-mins during 7 years in the basin	15 inputs (Wave and wind)	K-fold cross validation and sensitivity analysis
López and Iglesias, 2015	133 samples	3 inputs (Wave)	K-fold cross validation
Moslemi et al. (2022a)	30 datasets	5 inputs	Statistical analysis
Moslemi et al. (2022b)	98 models	3 inputs	Statistical analysi

To deal with input uncertainty most of the studies, regardless of the type of problem, used trial an error methodology. For wave heights studies, trial and error or statistical analysis followed by time-step analysis were the main methodologies used. For wave level, tides, extremes and wave power, trial and error or statistical analysis method ology were also the most popular, and, finally, for studies that involved maritime structures statistical analysis and k-fold cross validation were the most extended methodologies.

Wave height forecasting studies

As stated in Section 2. Uncertainties, one of the main uncertainties when using ANN models in ocean engineering is the selection of the predictor that best model the process. In the case of wave height fore casting studies, the main decision is to include or not wind data and other complimentary data, such as air or sea temperature. Including wind data and other variables can improve the model performance, but it increases considerably the computational cost. Fig. 4 shows the number of studies that included other type of data apart from wave climate data.

Most of studies only used or wave data or wave and wind data. The decision of including or not the wind data mostly depended on the quantity of data that was available. The quantity of data available was usually inversely correlated to the number of predictors and inputs used in the model. In general, when a minimum

of 4000 data samples was available, wave predictor was enough to train the ANN and wind data was not used. If, in addition, other data sources beyond instrumental records such as satellites or numerical models were used, the minimum data samples for not including wind as a predictor was around 2000.

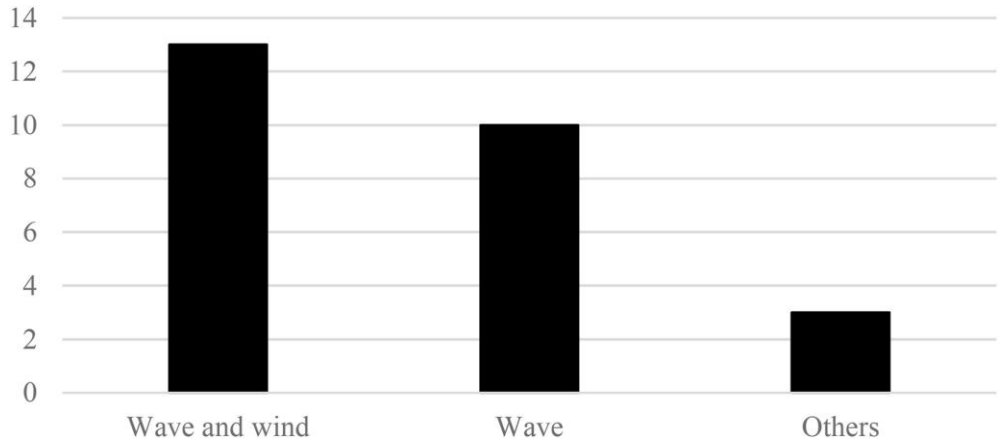


Fig. 4. Predictors used in wave height forecasting models.

Referring to the method to handle these uncertainties, more than 80% of the studies used trial and error approaches, which supposes a high-computational cost. Fig. 5 shows the methodology used to deal with input uncertainties.

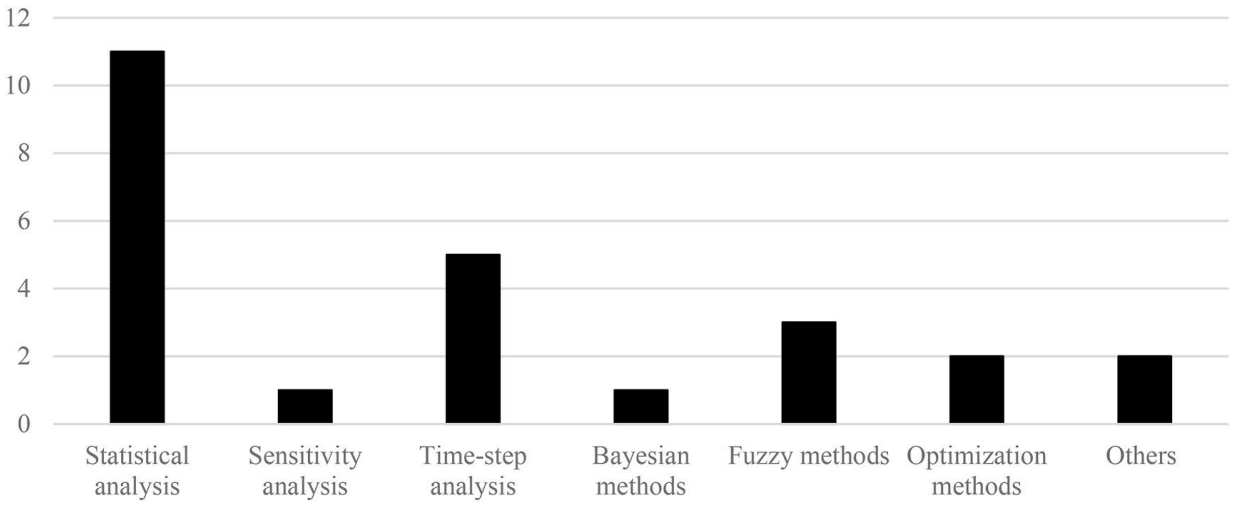


Fig. 5. Methodologies to handle input uncertainty in wave height forecasting studies.

Water level, tides, extremes and wave power studies

In this case, most of the studies used sea level as the main predictor. As happened with wave height forecasting studies, when the quantity of data was limited, other data, such as pressure level or wave parameters were added as inputs to the model.

In this case, in general, when a minimum of 2000 data samples was available, sea level or tide could be used as the only predictor. To handle input uncertainty most of the studies used statistical analysis. This method implies assessing the performance of each model evaluating different statistical parameters, which means a trial and error methodology. Fig. 6 shows the number of studies that used each methodology to handle input uncertainty.

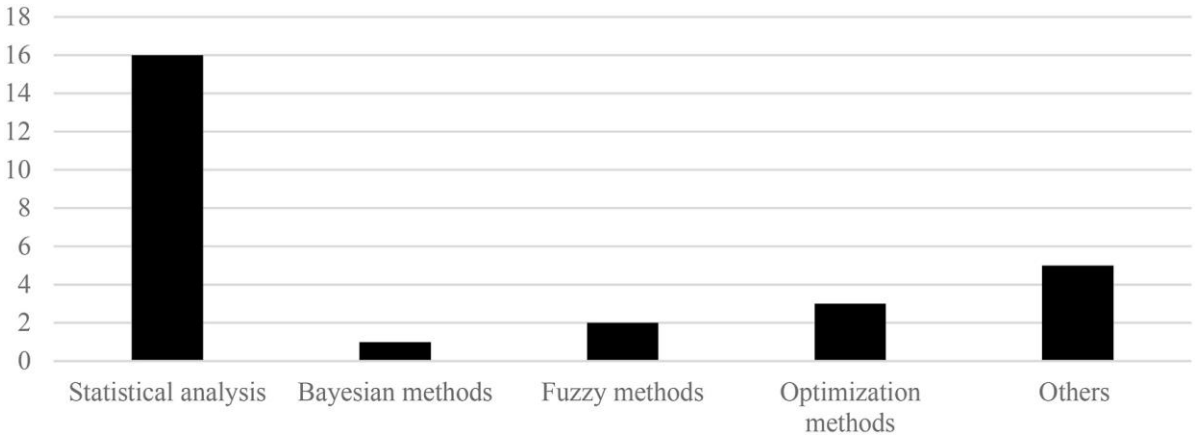


Fig. 6. Methodologies to handle input uncertainty in sea level, extremes, and energy studies.

Application to breakwaters, ports, and pipes

In the case of the application of ANNs to breakwaters, ports and pipes the main predictors used were wave parameters and all the parameters related to the structure. Fig. 7 shows the methods used to handle input uncertainties in these studies, and as happened in previous sections, the great majority used a trial and error approach.

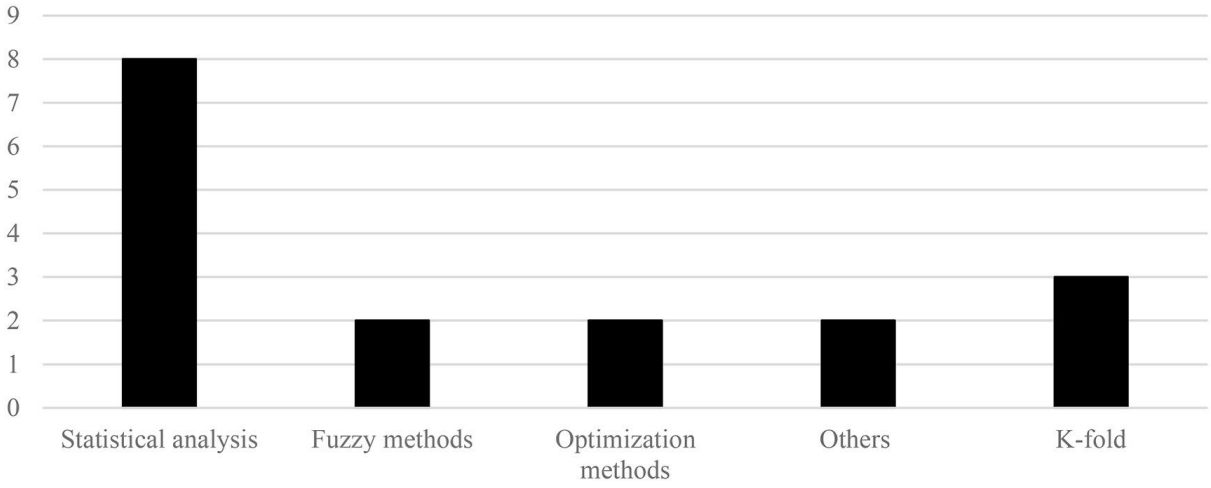


Fig. 7. Methodologies to handle input uncertainty in breakwater, ports and pipe studies.

4.2. Parameter uncertainties

In this section the data related to parameter uncertainties of the ANNs is presented. The algorithm used in the models, the method to handle its uncertainty and the method to handle input weights and data partitioning uncertainties are showed in Table 2.

Table 2. Parameter uncertainties.

Authors	Algorithm	Method to handle algorithm uncertainty	Method to handle initial weights uncertainty	Method to divide the data	Method to handle data partitioning uncertainty
		Wave forecasting studies			
Deo and Sridhar Naidu (1998)	Cascade	Trial and error	Trial and error	–	None
Deo et al. (2001)	BP	Trial and error	Trial and error	–	None
Agrawal and Deo (2002)	BP	Trial and error	Trial and error	Contiguous blocks (First for training)	None
Deo and Jagdale (2003)	BP	Trial and error	Combination of deterministic component with random values	Validation with laboratory experiments	None
Agrawal and Deo (2004)	BP and Cascade	Trial and error	Trial and error	Random selection	None
Makarynskyy (2004)	RBP	None	Different datasets for training	Contiguous blocks (First for training)	None

Kalra et al. (2005b)	ANFIS, RBF and BP	None	None	Clustering	None
Kalra et al. (2005a)	CGB	None	Trial and error	Contiguous blocks (First for training)	None
Makarynskyy et al. (2005)	LM	Trial and error	Trial and error	Clustering	None
Jain, 2008	LM	Trial and error	Trial and error	Contiguous blocks (First for training)	None
Malekmohamadi et al. (2008)	BP	None	Trial and error	–	–
Gunaydin (2008)	LM	None	None	Contiguous blocks (First for training)	None
Mahjoobi et al. (2008)	ANFIS, BP	None	None	–	–
Tsai and Tsai (2009)	BP	None	None	Clustering	None
Gopinath and Dwarakish (2015)	LM	None	None	–	–
Salah et al. (2016)	SCG, LM	Trial and error	None	Contiguous blocks (First for training)	None
Londhe et al. (2016)	LM	None	None	Clustering	None
Abhigna et al. (2017)	LM, BR	Trial and error	None	–	–
Mudronja et al. (2017)	LM, BR	Trial and error	None	–	–

Vieira et al. (2020)	LM, BR	Trial and error	Trial and error	Clustering	None
Stavarache and Ciortan, 2020	LM, Quick Propagation	Optimization (GA)	None	–	–
Gao et al. (2021)	BP	None	None	–	–
Joerges et al. (2021)	–	None	None	Resampling methodology from (Pirhooshyaran et al., 2020) and chronological split	Ensemble approach
Sadeghifar et al. (2022)	ANFIS	None	None	Clustering	None
Water level, tides, extremes and energy studies					
Lee, T. L. and Jeng (2002)	BP	None	None	–	–
Lee, T. L. et al. (2002)	BP	None	None	–	–
Lee, T.-L., 2004	BP	None	None	–	–
Chang and Lin (2006)	LM	None	None	–	–
Han and Shi (2008)	LM	None	None	–	–
Liang et al. (2008)	BP	None	None	–	–
de Oliveira and Ebecken (2009)	BP	Trial and error	None	Clustering	None

Wenzel, 2010	BP	Trial and error	None	Contiguous blocks and random	K-fold cross validation
Pashova and Popova (2011)	LM	Trial and error	Trial and error	Clustering	None
Filippo et al. (2012)	BP	Intelligent Problem Solver and Statistical NN function	Intelligent Problem Solver and Statistical NN function	Intelligent Problem Solver and Statistical NN function	Intelligent Problem Solver and Statistical NN function
Karamouz et al. (2014)	LM, BP	Trial and error	None	–	–
Castro et al. (2014)	BP	Trial and error	Trial and error	Contiguous blocks (First for training)	None
El-Diasty and Al-Harbi, 2015	BP	None	None	Contiguous blocks (First for training)	None
El-Diasty et al. (2018)	–	–	None	Contiguous blocks (First for training)	None
Zhao et al. (2019)	RBF, BP	Trial and error	None	Contiguous blocks (First for training)	None
Wang, B. et al. (2020)	LM	Trial and error	None	Contiguous blocks (First for training)	K-fold cross validation
Bruneau et al. (2020)	–	Ensemble approach	Ensemble approach	Contiguous blocks (First for training)	Ensemble approach

Avila et al. (2020)	RBP	None	None	Contiguous blocks (First for training)	None
Primo de Siqueira and Paiva, 2021	LM	Trial and error	None	Contiguous blocks (First for training)	None
Bento et al. (2021)	SCG	Optimization methods (MFO)	Optimization methods (MFO)	–	Optimization methods (MFO)
Bai and Xu (2021)	BP	Trial and error	None	Clustering	None
Yang et al. (2021)	BP	None	Convergence techniques	Clustering	Trial and error
		Breakwaters, ports and pipes studies			
Mase et al. (1995)	BP	None	Trial and error	Random selection	None
Yagci et al. (2005)	BP	None	None	Clustering	Trial and error
Kim, D. H. and Park (2005)	BP	None	None	Random selection	None
Iglesias et al. (2008)	BP	None	None	Random selection	None
Balas et al. (2010)	ANFIS, GDM	Trial and error	None	Random selection	None
Akoz et al. (2011)	LM	None	None	Random selection	None
Koç and Balas (2012)	ANFIS, GA and Climbing Hill	Optimization (GA)	Optimization (GA)	Random selection	Optimization (GA)

Suh et al. (2013)	LM	MonteCarlo Simulation	MonteCarlo Simulation	Clustering	MonteCarlo Simulation
Kim, D. H. et al. (2014)	BP	MonteCarlo Simulation	MonteCarlo Simulation	–	–
Londhe and Deo (2003)	LM	Trial and error	Trial and error	Division by characteristics	Trial and error
Kankal and Yüksek (2012)	BP	Trial and error	None	Random selection	None
López, M. and Iglesias (2013)	BP, BR	Trial and error	Trial and error	Contiguous blocks (First for training)	None
López and Iglesias, 2015	LM, BR	Trial and error	Trial and error	Contiguous blocks (First for training)	None
Zheng et al. (2020)	LM, GDM	Trial and error	None	Contiguous blocks (First for training)	None
Moslemi et al. (2022a)	LM	None	None	Random selection	None
Moslemi et al. (2022b)	LM	None	None	Random selection	None

Referring to the algorithms, the following abbreviations have been used: Backpropagation (BP), Resilient Backpropagation (RBP), Adaptive Neuro-Fuzzy Inference System (ANFIS), Radial Basis Function (RBF), Powell-Beale Restarts Algorithm (CGB), Levenberg-Marquardt (LM), Scaled Conjugated Gradient (SCG) and Gradient Descent Momentum (GDM).

Figs. 8–10 show the statistics of how the studies managed the uncertainties of the algorithm, of initial weights and of data partitioning, respectively.

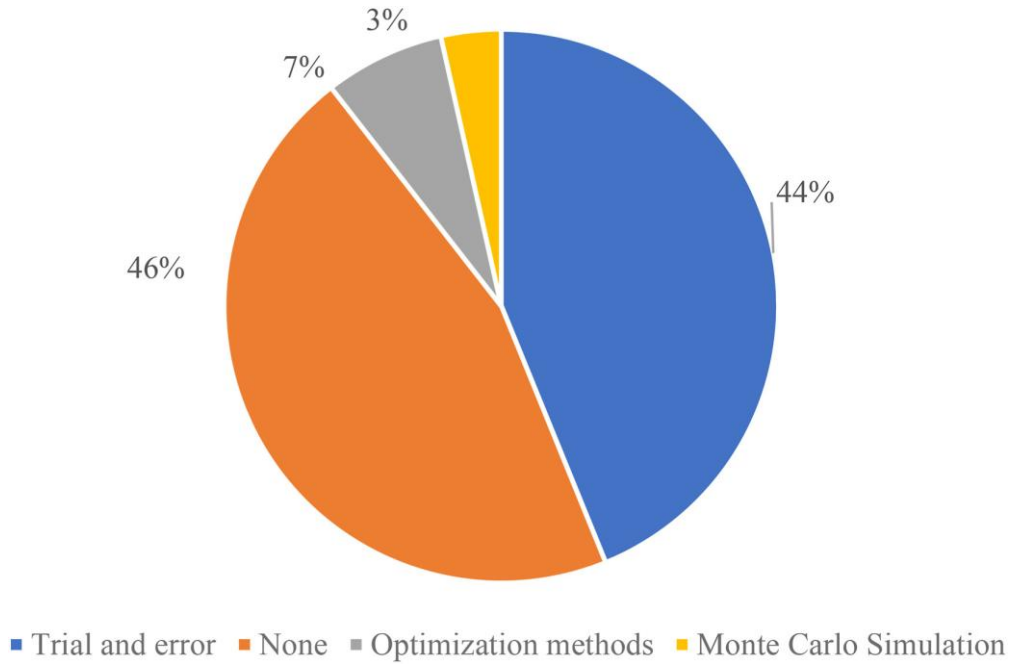


Fig. 8. Methodology used to deal with algorithm uncertainty.

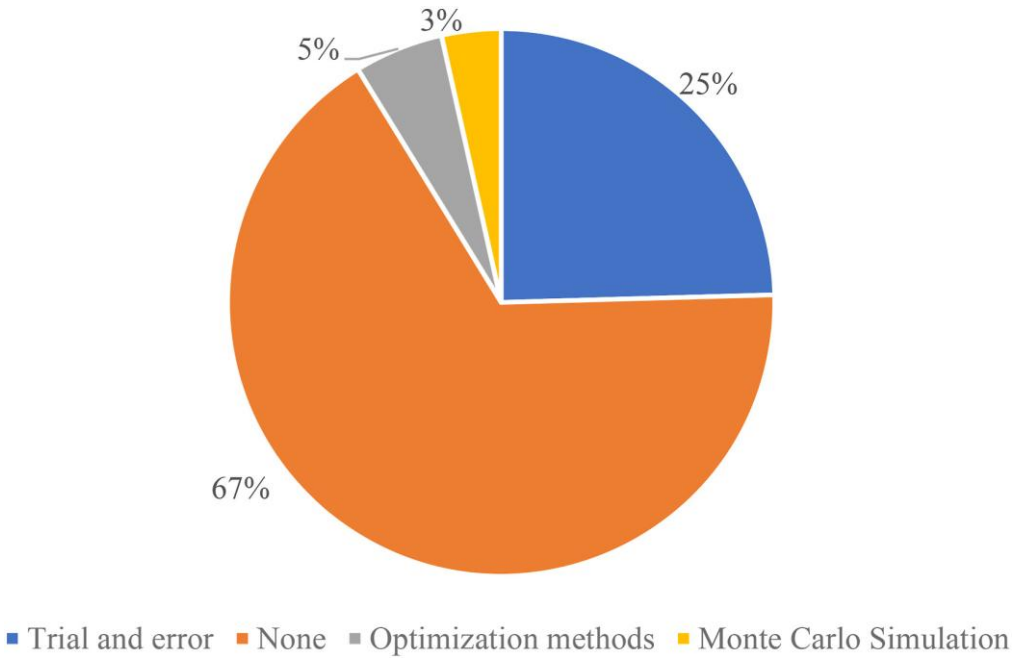


Fig. 9. Methodology used to deal with initial weights uncertainty.

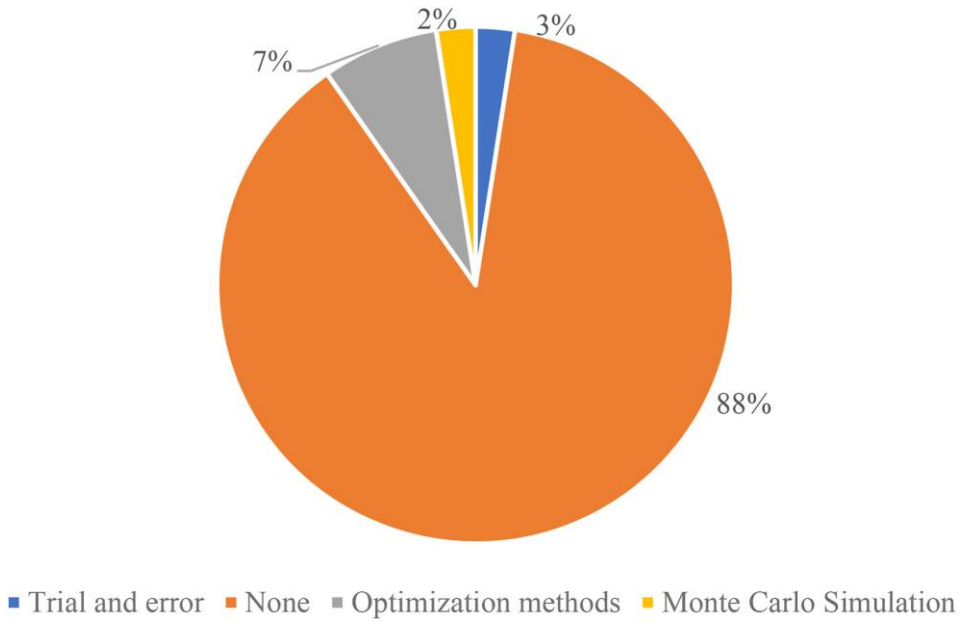


Fig. 10. Methodology used to deal with data partitioning uncertainty.

As can be seen from Fig. 8, almost all studies or imposed the algorithm to be used in the neural network or used the trial and error approach to select the algorithm that gave the best performance. Only a 10% of the studies used more sophisticated techniques, such as optimization methods or Monte Carlo simulations.

The case of the uncertainties related to initial weights is worse. More than 65% of the studies did not consider the uncertainties related to these initial weights, which can affect the performance of the ANN, as results vary considerably depending on the initial weights. There can be variations of the order of a unit of tenths in both, correlation coefficients and model errors, especially with certain algorithms such as BP that are more sensitive to these initial weights.

The same is true for the uncertainty associated to data partitioning. Such partitioning can be done randomly, temporally, by blocks, etc. And this also influences the performance of the model. However, hardly any of the studies of this review have carried out an analysis of these uncertainties.

If those studies that did deal with parameter uncertainty are considered, the methodologies most used in relation to the type of problem are the following (Table 3).

Table 3. Categorization of methodologies to deal with parameter uncertainties basing on the type of problem.

Type of problem	Most common methodology to deal with uncertainties (considering the studies that do not apply any method)		
	Algorithm	Initial weights	Data partitioning
Wave height	Trial and error	Trial and error	–
Water level, tides	Trial and error	Trial and error	Trial and error
			K-fold
Structures	Trial and error	Trial and error	Trial and error
	Monte Carlo simulation	Monte Carlo simulation	Monte Carlo simulation

4.3. Structure uncertainties

Table 4 presents the methods used to handle structure uncertainty, the uncertainties related to the ANN architecture and its transfer functions.

Table 4. Structure uncertainties.

Authors	Architecture	Method to handle architecture uncertainty	Transfer function	Method to handle transfer function uncertainty
		Wave forecasting studies		
Deo and Sridhar Naidu (1998)	1-20-1	Trial and error	Sigmoid	None
Deo et al. (2001)	2-4-2	Trial and error	Sigmoid	None
Agrawal and Deo (2002)	–	Trial and error	Sigmoid	None

Deo and Jagdale (2003)	3-5-5-2	Trial and error	Sigmoid	None
Agrawal and Deo (2004)	–	Trial and error	Sigmoid	None
Makarynskyy (2004)	48-97-24	Empirical expression from (Huang and Foo, 2002)	Logistic sigmoid: hidden Linear: output	None
Kalra et al. (2005b)	21-13-1 21-9-1	Trial and error	Gaussian and sigmoid	None
Kalra et al. (2005a)	27-3-1	Trial and error	Gaussian and sigmoid	None
Makarynskyy et al. (2005)	Several	Saliency analysis (Sensitivity)	Logistic sigmoid: hidden Linear: output	None
Jain, 2008	4-4-1 4-5-1	Trial and error	Logistic sigmoid: hidden Linear: output	None
Malekmohamadi et al. (2008)	8-8-1	Trial and error	–	–
Gunaydin (2008)	3-2-1	Trial and error	Hyperbolic tangent sigmoid: hidden linear:output	None
Mahjoobi et al. (2008)	4-7-1 4-11-1 4-8-1 4-9-3	Trial and error	Logistic and Hyperbolic tangent sigmoid: hidden linear:output	None
Tsai and Tsai (2009)	4-8-2	Trial and error	–	None

Gopinath and Dwarakish (2015)	52-10-52 12-5-12	Trial and error	Logistic sigmoid	None
Salah et al. (2016)	10 hidden nodes	Sensitivity analysis	Logistic sigmoid	Trial and error
Londhe et al. (2016)	9-1-1 4-1-1 9-8-1	Trial and error	Logistic sigmoid: hidden linear: output	None
Abhigna et al. (2017)	–	Bayesian methods	–	–
Mudronja et al. (2017)	–	Bayesian methods	Sigmoid	None
Vieira et al. (2020)	1- [10 to 50]-1 2- [10 to 50]-1	Bayesian methods	Logistic sigmoid: hidden Linear: output	None
Stavarache and Ciortan, 2020	3- [4/8/12]-1 5-[4/8/12]-1	Optimization (GA)	Logistic and Hyperbolic tangent Sigmoid	Optimization (GA)
Gao et al. (2021)	–	Sensitivity analysis	Sigmoid	None
Joerges et al. (2021)	–	Bayesian methods	Logistic and Hyperbolic tangent Sigmoid	None
Sadeghifar et al. (2022)	4-10-1	Trial and error	–	–
		Water level, tides, extremes and energy studies		
Lee, T. L. and Jeng (2002)	4-4-1	Trial and error	Sigmoid	None
Lee, T. L. et al. (2002)	138-7-69	Trial and error	Sigmoid	None

Lee, T.-L., 2004	138-7-69	Trial and error	Sigmoid	None
Chang and Lin (2006)	5 hidden neurons	Trial and error	Hyperbolic tangent sigmoid	None
Han and Shi (2008)	3 layers, 10 neurons	Trial and error	Hyperbolic tangent: hidden Linear: Output	None
Liang et al. (2008)	50 hidden neurons	Empirical expression	Sigmoid	None
de Oliveira and Ebecken (2009)	7-14-1	Trial and error	Hyperbolic tangent sigmoid	None
Wenzel, 2010	112-84-56	Trial and error	Hyperbolic tangent: hidden Linear: Output	Trial and error
Pashova and Popova (2011)	11-6-1	Trial and error and empirical expression	Hyperbolic tangent: hidden Linear: Output	Trial and error
Filippo et al. (2012)	1 to 20 hidden neurons	Intelligent Problem Solver and Statistical NN function	–	–
Karamouz et al. (2014)	17 to 19 hidden neurons	Trial and error and selection methods	Hyperbolic tangent: hidden Linear: Output	None
Castro et al. (2014)	3-4-4-1	Trial and error	Hyperbolic tangent: hidden Linear: Output	None
El-Diasty and Al-Harbi, 2015	24-12-1	Trial and error	Sigmoid, Wavelet	Empirical methods
El-Diasty et al. (2018)	24-12-1	Trial and error	Wavelet	Empirical methods
Zhao et al. (2019)	5-20-15-1 5-20-1	Trial and error	–	–

Wang, B. et al. (2020)	3 layers	Trial and error	Hyperbolic tangent: hidden Linear: Output	None
Bruneau et al. (2020)	–	Ensemble approach	Leaky ReLU and sigmoid	None
Avila et al. (2020)	15 hidden neurons	Trial and error	Hyperbolic tangent: hidden Linear: Output	None
Primo de Siqueira and Paiva, 2021	10 hidden neurons	Trial and error	Hyperbolic tangent: hidden Linear: Output	None
Bento et al. (2021)	–	Optimization methods (MFO)	–	Optimization methods (MFO)
Bai and Xu (2021)	–	Sensitivity analysis	–	Sensitivity analysis
Yang et al. (2021)	–	Optimization methods	–	–
		Breakwaters, ports and pipes studies		
Mase et al. (1995)	6-12-1	Trial and error	–	–
Yagci et al. (2005)	5-3-1	Trial and error	–	–
Kim, D. H. and Park (2005)	[5–7]-4-1 [5–7]-5-1 [5–7]-20-1	Trial and error	Sigmoid	None
Iglesias et al. (2008)	4-9-8-7-6-5-1	Trial and error	Hyperbolic tangent sigmoid	None
Balas et al. (2010)	5-4-1 10-8-1	Trial and error	Sigmoid: hidden Linear: output	None
Akoz et al. (2011)	4-3-3	Trial and error	Sigmoid: hidden Linear: Output	Trial and error

Koç and Balas (2012)	2 hidden layers	Optimization (GA)	–	–
Suh et al. (2013)	5 to 60 hidden neurons	MonteCarlo Simulation	Hyperbolic tangent: hidden Linear: Output	MonteCarlo Simulation
Kim, D. H. et al. (2014)	2-3-1	MonteCarlo Simulation	Sigmoid: hidden Linear: Output	MonteCarlo Simulation
Londhe and Deo (2003)	3-1-1 4-1-1	Modular trial and error	–	–
Kankal and Yüksek (2012)	1-2 hidden layers	Trial and error	Sigmoid functions	None
López, M. and Iglesias (2013)	3-2-2-1 3-3-1 2-2-1	K-fold cross validation	Hyperbolic tangent: hidden Linear: Output	None
López and Iglesias, 2015	15-57-1	K-fold cross validation	Hyperbolic tangent: hidden Linear: Output	None
Zheng et al. (2020)	3-7-8-24	K-fold cross validation	Hyperbolic tangent: hidden Linear: Output	None
Moslemi et al. (2022a)	5-7-5	Trial and error	–	–
Moslemi et al. (2022b)	3-3-3-3-4	Trial and error	–	–

From Fig. 11, almost all the studies optimized the architecture of the ANNs with the trial and error approach, which implies the development of a large number of models and a high computational cost. Only around 15% opted for other techniques. Referring to the transfer function, almost no research studied the uncertainty associated with them, virtually all studies chose to select the transfer functions that have been shown to work best throughout the history of neural networks, without conducting any additional analysis.

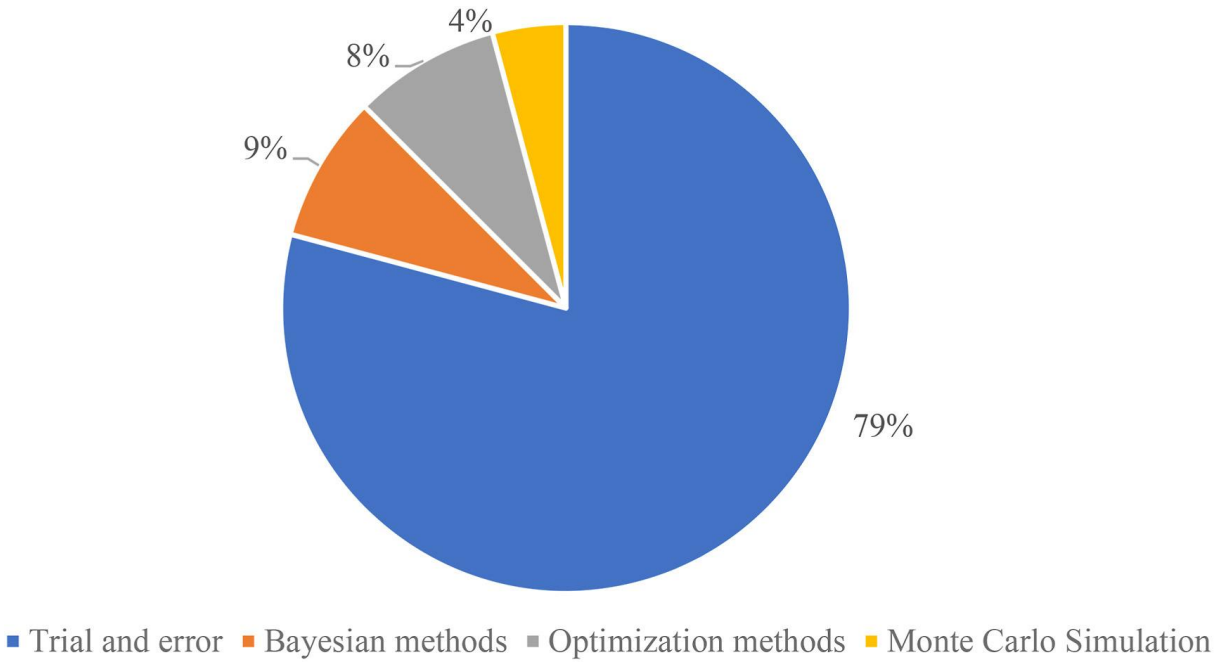


Fig. 11. Data on methodology used to deal with architecture uncertainties.

If these methodologies used to deal with structure uncertainties are categorized based on the type of the problem, the results are the following (Table 5).

Table 5. Categorization of methodologies to deal with structure uncertainties basing on the type of problem.

Type of problem	**Most common methodology to deal with uncertainties (considering the studies that do not apply any method)**	
	Architecture	**Transfer function**
Wave height	Trial and error	–
Water level, tides	Trial and error	Trial and error
Structures	Trial and error	Monte Carlo simulation
	K-fold cross validation	
	Monte Carlo simulation	

4.4. Statistical summary

In this section, a general overview of the situation of how ANNs studies in ocean engineering are dealing with the uncertainty inherent in these models is presented.

Table 6 presents the number and percentage of studies that applied other methodologies apart from the trial and error approach to quantify ANNs uncertainty.

Table 6. Number and percentage of studies trat treated ANNs uncertainties.

Uncertainty	N° of studies			% of studies		
	Wave height	Tides	Structures	Wave height	Tides	Structures
Input	14	11	9	56%	41%	43%
	34			**51%**		
Parameter	1	2	3	4%	9%	19%
	6			**10%**		
Structure	9	7	6	36%	31%	42%
	22			**35%**		

As per the table, most of ocean engineering studies only treated input uncertainty (51% of the studies). Only a 10% of models treated uncertainty related to parameters (algorithm, initial weights, and data partitioning), which clearly shows the need to implement new techniques that address the limitations of neural networks in ocean engineering, especially considering the influence that these parameters can have in the final performance of the model with certain algorithms. A clear example of this is the sensitivity of one of the most used algorithms in ocean engineering, the BP algorithm, to initial weights. Although it is known that this algorithm is very sensitive to the initial weights of the neural network, virtually no studies have done specific studies on the influence of these initial weights on the development of the models. As already mentioned in the introduction, given the great development of artificial intelligence techniques, it is necessary to start addressing all these uncertainties so that their application in ocean engineering does not lag behind.

5. Discussion

In this section new methodologies that can be applied to deal with the uncertainty of ANNs in ocean engineering are proposed and discussed. Table 7 shows the most common methodologies used in ocean engineering to deal with ANNs uncertainty versus the new methodologies proposed on this research paper based on the latest developments in artificial intelligence and what is being done in other fields. Finally, some guidelines are given for the application of these new methodologies. The explanation of the proposed methodologies to be implemented in ocean engineering studies is presented in the following subsections.

Table 7. Most common methodologies to deal with ANNs uncertainty vs new methodologies proposed and application guidelines.

Uncertainty	Most used methodology to date	Methodologies proposed	Application guidelines
Input	Trial and error – Statistical analysis	General rules: Wave studies: Less than 4000 data samples. Wave + Wind data Tide studies: Less than 2000 data samples. Water level + Complimentary data	Preliminary studies
		Bootstrap methodology	If only input uncertainty is considered
		Bayesian methodology	Especially for small datasets
		Monte Carlo simulations	Complex processes that involve many variables that may be correlated.
			Ocean engineering studies that imply the interaction with maritime structures.
Model	None or Trial and error	Bayesian methodology	Especially for small datasets
		Genetic algorithm	Mainly for parameter optimization
		Monte Carlo simulations	Complex processes that involve many variables that may be correlated.

			Ocean engineering studies that imply the interaction with maritime structures.
Black box nature	None	Rule extraction	Using OAS and CDRPs
		Surrogate models	Replace decision functions by self-explanatory models

5.1. Input uncertainty

Input uncertainty is one of the main sources of variance in ANNs (Chitsazan et al., 2015). However, the majority of the studies in ocean engineering that apply ANNs deal with input uncertainty using a trial and error approach and applying statistical analysis, which is an early methodology that has a high computational cost and is not very accurate.

After analyzing all these studies, some general rules have been extracted. For wave height studies when a minimum of 4000 data samples is available, wind data is not strictly necessary. If, in addition, other data sources beyond instrumental records such as satellites or numerical models are used, the minimum data samples for not including wind as a predictor is around 2000. For water level and tides studies a minimum of 2000 data samples is enough to use sea level or tide as the only predictor.

In addition to these general rules that can be applied in preliminary studies, the bootstrapping methodology, Bayesian techniques and Monte Carlo simulations are proposed as the preferred methodologies to deal with input uncertainty. Bootstrapping is the most used technique in water resources and hydrological studies to deal with input uncertainty. Water resources and hydrology are fields close to ocean engineering and which can serve as a reference when it comes to defining advances in artificial intelligence, since the application of soft-computing techniques in these disciplines is slightly more advanced than in ocean engineering. Jeong and Kim (2005); Jia and Culver (2006); Tiwari and Chatterjee (2010); Kasiviswanathan et al. (2016) are some of the authors that applied Bootstrap Based Artificial Neural Networks (BTANNs) in hydrological studies.

The bootstrap technique is based on data resampling with replacement (Efron and Tibshirani, 1994). It is one of the simplest approach to deal with uncertainty since it does not require complex computations of derivatives and Hessian-matrix

inversion involved in linear methods or the Monte Carlo solutions of the integrals involved in the Bayesian approach (Tiwari and Chatterjee, 2010). This methodology is a methodology that allows the quantification of input uncertainty by means of confidence intervals and has a lower computational cost than more sophisticated techniques like the Bayesian approach or Monte Carlo simulations. In BTANNs each bootstrapped sample is used to develop an ANN, and many ANNs are developed and combined to build a relation between the model input and output with their uncertainties from ANNs included (Wang and y Ma, 2023). The confidence interval represents the uncertainty in the predictions. BTANNs use a range of weight sets instead of a single set and the confidence interval shows the limits to which the predictions could have varied based on the weight set used (Tiwari and Chatterjee, 2010).

Apart from the BTANNs there are also other approaches, such as Bayesian Artificial Neural Networks (BANNs) that usually narrows the uncertainty band compared to BTANNs, but that have a slightly higher computational cost (Kasiviswanathan et al., 2016). They are especially recommended for small datasets (Wang and y Ma, 2023; Weytjens and y De Weerdt, 2022) and if model uncertainties are considered too. Monte Carlo simulations and Genetic Algorithms (GA) are also a good alternative to BTANN if all ANNs (input and model) uncertainties are considered (Section 5.2.).

Apart from these most extended methods to deal with input uncertainty, some authors developed other innovative approaches that could also be applied in ocean engineering. Kasiviswanathan et al. (2018) proposed a method that combined probabilistic and ensemble simulation approach and that applied GA. Kim, T. et al. (2020) also used an ensemble approach. Sildir et al. (2022) proposed a mixed integer nonlinear programming (MINLP) formulation to detect the best features and connections among the ANN elements. These latter innovative approaches could also be applied in ocean engineering.

5.2. Model uncertainty

As with input uncertainty, the hydrological field has been taken as the main reference to propose new methodologies to deal with model uncertainty, as it is the closest field to ocean engineering with slightly more development in neural network methodologies. After reviewing different studies, Bayesian methods, Genetic algorithms and Monte Carlo simulations have been considered as the preferred methodologies to deal with model uncertainty.

Bayesian methodologies are based on the definition of probability density functions (PDF). It defines the network weights as a PDF and it is assigned to the network parameters which is then updated using the training data and Bayes' theorem to yield the posterior PDF (Cabaneros and Hughes, 2022). It is the best method when dealing with small datasets and it is also used to reduce input uncertainty. Chitsazan et al. (2015); Kasiviswanathan et al. (2016); Kasiviswanathan and Sudheer (2016) and Weytjens and y De Weerdt (2022) are some of the authors that used Bayesian techniques to quantify uncertainty in ANN models. This approach is especially recommended when only small datasets are available.

Genetic algorithm (GA) is an optimization technique that bases its behaviour on the functioning of human genes and on the field of genetics. The GA initialises a set of possible solutions (chromosomes), and then the solutions are set through the process of natural selection. The different options are evaluated through a fitness function which ranks the chromosomes. Once the ranking is done, the parent chromosomes, that are the solutions that fit the best the fitness function, are selected for the crossover and the mutation. During the crossover, an exchange of information between the two parents solutions selected occurs. In the mutation stage, the genes of the chromosomes of the crossed offspring are changed. The entire process is carried out until a certain condition is met. Kasiviswanathan and Sudheer (2017), Kasiviswanathan et al. (2018) applied GA to model input uncertainty. Wu et al. applied it to obtain different sets of appropriate model parameters, including the ANN weights, bias, transfer functions and adjusting factors. Finally, Jiang et al. (2022) also used GA to obtain the optimal values of the model parameters. This methodology is mainly used to obtain the optimal ANN parameters.

Monte Carlo simulations is one of the most extended methodologies used to deal with model uncertainty. It defines the main variables and generates random samples assigning PDFs to each of these main variables and models the dependent variables in relation to these main ones. These PDFs are propagated through the network obtaining a final result with its uncertainty. Coral et al. (2016); She and You (2019); Lee, J. W. et al. (2020) and Mokhtari et al. (2021) are some of the authors that applied Monte Carlo simulations to deal with model uncertainty. It is a methodology especially recommended for complex processes that involve many different variables that may be correlated. In the case of ocean engineering, it is especially recommended for studies that involve maritime structures.

5.3. Black-box nature

The black box nature of data-driven models and ANNs is one of the biggest problems of these models and least advanced in methodologies that deal with it. To the authors knowledge, there are almost no studies either in ocean engineering or in hydrological resources that deal with the black-box nature of ANNs. Research that addresses the black-box nature of ANNs are from specific studies of artificial intelligence and soft-computing.

There are two main approaches to explain ANNs models, post-hoc interpretability models and ad-hoc. The former explains the existing models, while the latter constructs interpretable ones (Fan et al., 2021). This section will focus on post-hoc techniques as they do not involve developing new models and are the easiest to be implemented in the existing ocean engineering models. One of the easiest way to start implementing this techniques is using rule extraction strategies with different algorithms such as the ordered-attribute search (OAS) algorithm and identifying the critical nodes with Critical Data Routing Paths (CDRPs) (Lee, H. and Kim, 2021). The other main technique that could be applied in ocean engineering studies without involving a huge effort and major advances in the state of the art is using surrogate models that replace the decision function and the parts of the model that are not interpretable with a local surrogate model that is structured in a way that it is self-explanatory (Samek et al., 2021).

6. Conclusions

ANNs are an excellent tool to model ocean engineering problems and in the last decades its application in maritime engineering has reached a high level of development. However, its use is associated with a large number of uncertainties due to its black box nature that can limit their applicability. Addressing these limitations is a matter of urgency in view of the development that these methodologies have been undergoing in recent years, so that the field of ocean engineering does not fall behind. Hence, in this research paper, the uncertainties of these models have been analyzed and new methodologies to deal with them have been proposed.

There are clear differences between the uncertainties associated with the input variables and the uncertainties associated with the model itself. While almost all the models studied covered input uncertainty (more than 90% of the studies),

model uncertainty was studied in less research papers, which may be due to its high complexity.

The most studied uncertainties associated with the model are those related to the algorithm and the network architecture. However, uncertainties associated with initial weights, data partitioning and transfer functions have been largely unaddressed in ocean engineering models. From these three blocks, the least studied are the uncertainties related to transfer functions. However, this is also not of great importance as the choice of these functions is based on scientific evidence from years of experience with neural networks. Virtually all studies use sigmoid and linear functions, which have been shown to perform best historically. What is more important is to study the uncertainties associated with the initial weights and the partitioning of the data.

Initial weights uncertainties were only studied in a 30% of the research papers presented and data partitioning uncertainties in less than 20%. The uncertainty associated with the initial weights can lead to changes in the model and its performance of the order of tenths. Errors of the order of tenths can lead to the choice of one model or another, so this uncertainty can lead to a model being discarded when it should not have to be. However, few studies reported having done an analysis of the influence of these initial weights. Moreover, the initial weights have a greater or lesser influence depending on the algorithm used and are especially important when the algorithm is BP, one of the most widespread in ocean engineering. The same applies to the way in which the data is divided up to develop the model. Only 12% of the studies did an analysis of how splitting the data in one way or another affected the training, validation, and testing of the model. Considering that neural networks are data-driven models and base their entire performance on the dataset used to train, validate and test them, this partitioning should be analyzed and optimized.

Referring to the methodology used to study these uncertainties, around 80% of the studies used the trial and error approach to study the uncertainty, which is an early methodology that is very time consuming and has a high computational cost associated. After analyzing how uncertainties are addressed in other fields, some recommendations have been done. Bootstrapping is the suggested methodology to deal with input uncertainty, it is the methodology that gives the best performance with the lowest computational cost. However, if model uncertainty is also considered other methodologies are more suitable. Bayesian technique is the preferred technique when only small datasets are available. Genetic algorithm is

the ideal methodology to deal with parameters uncertainties and Monte Carlo simulations are highly recommended in ocean engineering studies that model maritime structures problems or involve many ocean variables that may be correlated. Finally, for the interpretability of ANNs, it has been suggested to use rule extraction methodologies or to use surrogate self-explanatory models in those parts of the model that are not interpretable.

Future research lines should focus on working more on model uncertainties, especially on those associated to data partitioning and initial weights. Given that ANNs are data-driven models, these two uncertainties are of particular interest and have a great influence on the final outcome of the model. However, they have not been studied sufficiently so far. In addition, models-based methodologies and hybrid approaches (Bayesian and optimization methodologies, Monte Carlo simulations and fuzzy methodologies) should be studied more deeply and be applied in future research to optimize time and computing.

References

Abhigna, P., Jerritta, S., Srinivasan, R., Rajendran, V. y, Ieee, 2017. Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in bay of bengal, 2017 Apr 06-08. In: Paper Presented at the IEEE International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, INDIA.

Agrawal, J. D. y, Deo, M.C., 2002. On-line wave prediction. Mar. Struct. 15 (1), 57–74. https://doi.org/10.1016/S0951-8339(01)00014-4.

Agrawal, J. D. y, Deo, M.C., 2004. Wave parameter estimation using neural networks. Mar. Struct. 17 (7), 536–550. https://doi.org/10.1016/j.marstruc.2005.01.001. Akoz, M.S., Cobaner, M., Kirkgoz, M. S. y, Oner, A.A., 2011. Prediction of geometrical properties of perfect breaking waves on composite breakwaters. Appl. Ocean Res. 33 (3), 178–185. https://doi.org/10.1016/j.apor.2011.03.003.

Alvisi, S. y, Franchini, M., 2011. Fuzzy neural networks for water level and discharge forecasting with uncertainty. Environ. Model. Software 26 (4), 523–537. https://doi. org/10.1016/j.envsoft.2010.10.016.

Avila, D., Marichal, G.N., Padron, I., Quiza, R. y, Hernandez, A., 2020. Forecasting of wave energy in canary islands based on artificial intelligence. Appl. Ocean Res. 101 https://doi.org/10.1016/j.apor.2020.102189. Bai, L.-H. y, Xu, H., 2021. Accurate estimation of tidal level using bidirectional long short-term memory recurrent neural network. Ocean Eng. 235, 108765 https://doi. org/10.1016/j.oceaneng.2021.108765.

Balas, C.E., Koc, M. L. y, Tur, R., 2010. Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters. Appl. Ocean Res. 32 (4), 425–433. https://doi.org/ 10.1016/j.apor.2010.09.005.

Bento, P.M.R., Pombo, J.A.N., Mendes, R.P.G., Calado, M.R.A., y Mariano, S.J.P.S., 2021. Ocean wave energy forecasting using optimised deep learning neural networks. Ocean Eng. 219 https://doi.org/10.1016/j.oceaneng.2020.108372.

Borrego, C., Monteiro, A., Ferreira, J., Miranda, A.I., Costa, A.M., Carvalho, A.C., y Lopes, M., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. Environ. Int. 34 (5), 613–620. https://doi.org/10.1016/j. envint.2007.12.005.

Boucher, M.-A., Perreault, L., y Anctil, F., 2009. Tools for the assessment of hydrological ensemble forecasts obtained by neural networks. Journal of Hydroinformatics - J HYDROINFORM 11. https://doi.org/10.2166/hydro.2009.037.

Bruneau, N., Polton, J., Williams, J. y, Holt, J., 2020. Estimation of global coastal sea level extremes using neural networks. Environ. Res. Lett. 15 (7) https://doi.org/ 10.1088/1748-9326/ab89d6.

Cabaneros, S. M. y, Hughes, B., 2022. Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting. Environ. Model. Software 158. https://doi.org/10.1016/j.envsoft.2022.105529.

Castro, A., Carballo, R., Iglesias, G. y, Rabu˜nal, J.R., 2014. Performance of artificial neural networks in nearshore wave power prediction. Appl. Soft Comput. 23, 194–201. https://doi.org/10.1016/j.asoc.2014.06.031.

Chang, H.-K. y, Lin, L.-C., 2006. Multi-point tidal prediction using artificial neural network with tide-generating forces. Coast. Eng. 53 (10), 857–864. https://doi.org/ 10.1016/j.coastaleng.2006.05.001.

Chitsazan, N., Nadiri, A. A. y, Tsai, F.T.C., 2015. Prediction and structural uncertainty analyses of artificial neural networks using hierarchical Bayesian model averaging. J. Hydrol. 528, 52–62. https://doi.org/10.1016/j.jhydrol.2015.06.007.

Chryssolouris, G., Lee, M. y, Ramsey, A., 1996. Confidence interval prediction for neural network models. IEEE Trans. Neural Network. 7 (1), 229–232. https://doi.org/ 10.1109/72.478409.

Coral, R., Flesch, C.A., Penz, C.A., Roisenberg, M. y, Pacheco, A.L.S., 2016. A Monte Carlo-based method for assessing the measurement uncertainty in the training and use of artificial neural networks. Metrol. Meas. Syst. 23 (2), 281–294. https://doi. org/10.1515/mms-2016-0015.

de Oliveira, M. M. F. y, Ebecken, N.F.F., 2009. A model to predict the coastal sea level variations and surge, 2009 Sepp. 14–16. In: Paper Presented at the 1st International Conference on Physical Coastal Processes, Management and Engineering, Malta. Deo, M. C. y, Jagdale, S.S., 2003. Prediction of breaking waves with neural networks. Ocean Eng. 30 (9), 1163–1178. https://doi.org/10.1016/s0029-8018(02)00086-0.

Deo, M.C., Jha, A., Chaphekar, A. S. y, Ravikant, K., 2001. Neural networks for wave forecasting. Ocean Eng. 28 (7), 889–898. https://doi.org/10.1016/S0029-8018(00) 00027-5.

Deo, M. C. y, Sridhar Naidu, C., 1998. Real time wave forecasting using neural networks. Ocean Eng. 26 (3), 191–203. https://doi.org/10.1016/S0029-8018(97)10025-7.

Efron, B. y, Tibshirani, R., 1994. An Introduction to the Bootstrap. El-Diasty, M. y, Al-Harbi, S., 2015. Development of wavelet network model for accurate water levels prediction with meteorological effects. Appl. Ocean Res. 53, 228–235. https://doi.org/10.1016/j.apor.2015.09.008.

El-Diasty, M., Al-Harbi, S., y Pagiatakis, S., 2018. Hybrid harmonic analysis and wavelet network model for sea water level prediction. Appl. Ocean Res. 70, 14–21. https://doi.org/10.1016/j.apor.2017.11.007.

Elshorbagy, A., Corzo, G., Srinivasulu, S., y Solomatine, D.P., 2009. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: concepts and methodology. Hydrol. Earth Syst. Sci. 14, 1931–1941.

Elshorbagy, A., Corzo, G., Srinivasulu, S., y Solomatine, D.P., 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: concepts and methodology. Hydrol. Earth Syst. Sci. 14 (10), 1931–1941. https://doi.org/10.5194/hess-14-1931-2010.

Fan, F.-L., Xiong, J., Li, M. y, Wang, G., 2021. On interpretability of artificial neural networks: a survey. IEEE transactions on radiation and plasma medical sciences 5 (6), 741–760. https://doi.org/10.1109/trpms.2021.3066428.

Filippo, A., Rebelo Torres, A., Kjerfve, B. y, Monat, A., 2012. Application of Artificial Neural Network (ANN) to improve forecasting of sea level. Ocean Coast Manag. 55, 101–110. https://doi.org/10.1016/j.ocecoaman.2011.09.007.

Gao, S., Huang, J., Li, Y., Liu, G., Bi, F., y Bai, Z., 2021. A forecasting model for wave heights based on a long short-term memory neural network. Acta Oceanol. Sin. 40 (1), 62–69. https://doi.org/10.1007/s13131-020-1680-3.

Gopinath, D. I. y, Dwarakish, G.S., 2015. Wave prediction using neural networks at new mangalore port along west coast of India. Aquatic Procedia 4, 143–150. https://doi.org/10.1016/j.aqpro.2015.02.020.

Gunaydin, K., 2008. The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. Ocean Eng. 35 (14–15), 1406–1415. https://doi.org/10.1016/j.oceaneng.2008.07.008.

Han, G. y, Shi, Y., 2008. Development of an atlantic Canadian coastal water level neural network model. J. Atmos. Ocean. Technol. 25 (11), 2117–2132. https://doi.org/10.1175/2008jtecho569.1.

Huang, W. y, Foo, S., 2002. Neural network modeling of salinity variation in Apalachicola River. Water Res. 36 (1), 356–362. https://doi.org/10.1016/s00431354(01)00195-6

Hwang, J. T. G. y, Ding, A.A., 1997. Prediction intervals for artificial neural networks. J. Am. Stat. Assoc. 92 (438), 748–757. https://doi.org/10.1080/01621459.1997.10474027.

Iglesias, G., Rabunal, J., Losada, M.A., Pach´ on, H., Castro, A. y, Carballo, R., 2008. A virtual laboratory for stability tests of rubble-mound breakwaters. Ocean Eng. 35 (11), 1113–1120. https://doi.org/10.1016/j.oceaneng.2008.04.014.

Jain, P. y Deo M., 2008. Artificial intelligence tools to forecast ocean waves in real time. Open Ocean Eng. J. 1, 13–20.

Jeong, D.-I. y, Kim, Y.-O., 2005. Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrol. Process. 19 (19), 3819–3835. https://doi.org/10.1002/hyp.5983.

Jia, Y. y, Culver, T.B., 2006. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. J. Hydrol. 331 (3), 580–590. https://doi.org/10.1016/j.jhydrol.2006.06.005.

Jiang, W.H., Liu, J.F., Peng, A.B., Liu, G. D. y, Zhang, R., 2022. Quantifying the contribution of uncertainty sources of artificial neural network models using ANOVA for

reservoir power generation. Nord. Hydrol 53 (12), 1455–1467. https://doi.org/10.2166/nh.2022.052

Joerges, C., Berkenbrink, C. y, Stumpe, B., 2021. Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks. Ocean Eng. 232 https://doi.org/10.1016/j.oceaneng.2021.109046.

Kalra, R., Deo, M.C., Kumar, R. y, Agarwal, V.K., 2005a. Artificial neural network to translate offshore satellite wave data to coastal locations. Ocean Eng. 32 (16), 1917–1932. https://doi.org/10.1016/j.oceaneng.2005.01.007.

Kalra, R., Deo, M.C., Kumar, R. y, Agarwal, V.K., 2005b. RBF network for spatial mapping of wave heights. Mar. Struct. 18 (3), 289–300. https://doi.org/10.1016/j.marstruc.2005.09.003.

Kankal, M. y, Yüksek, ¨ O., 2012. Artificial neural network approach for assessing harbor tranquility: the case of Trabzon Yacht Harbor, Turkey. Appl. Ocean Res. 38, 23–31. https://doi.org/10.1016/j.apor.2012.05.009.

Karamouz, M., Kia, M. y, Nazif, S., 2014. Prediction of sea level using a hybrid data- driven model: new challenges after hurricane sandy. Water Quality, Exposure and Health 6. https://doi.org/10.1007/s12403-014-0119-5.

Kariri, E., Louati, H., Louati, A. y, Masmoudi, F., 2023. Exploring the advancements and future research directions of artificial neural networks: a text mining approach. Applied Sciences-Basel 13 (5). https://doi.org/10.3390/app13053186.

Kasiviswanathan, K. S. y, Sudheer, K.P., 2016. Comparison of methods used for quantifying prediction interval in artificial neural network hydrologic models. Modeling Earth Systems and Environment 2 (1). https://doi.org/10.1007/s40808016-0079-9.

Kasiviswanathan, K. S. y, Sudheer, K.P., 2017. Methods used for quantifying the prediction uncertainty of artificial neural network based hydrologic models. Stoch. Environ. Res. Risk Assess. 31 (7), 1659–1670. https://doi.org/10.1007/s00477-0161369-5

Kasiviswanathan, K.S., Sudheer, K. P. y, He, J., 2016. Quantification of prediction uncertainty in artificial neural network models. In: Shanmuganathan, S., Samarasinghe, S. (Eds.), Artificial Neural Network Modelling, vol. 628, pp. 145–159.

Kasiviswanathan, K.S., Sudheer, K. P. y, He, J.X., 2018. Probabilistic and ensemble simulation approaches for input uncertainty quantification of artificial neural network hydrological models. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 63 (1), 101–113. https://doi.org/10.1080/02626667.2017.1393686

Khan, M. S. y, Coulibaly, P., 2006. Bayesian neural network for rainfall-runoff modeling. Water Resour. Res. 42 (7) https://doi.org/10.1029/2005WR003971

Kim, D.H., Kim, Y. J. y, Hur, D.S., 2014. Artificial neural network based breakwater damage estimation considering tidal level variation. Ocean Eng. 87, 185–190. https://doi.org/10.1016/j.oceaneng.2014.06.001

Kim, D. H. y, Park, W.S., 2005. Neural network for design and reliability analysis of rubble mound breakwaters. Ocean Eng. 32 (11), 1332–1349. https://doi.org/10.1016/j.oceaneng.2004.11.008

Kim, T., Shin, J.Y., Kim, H. y, Heo, J.H., 2020. Ensemble-based neural network modeling for hydrologic forecasts: addressing uncertainty in the model structure and input variable selection. Water Resour. Res. 56 (6) https://doi.org/10.1029/2019wr026262

Kingston, G.B., Lambert, M. F. y, Maier, H.R., 2005. Bayesian training of artificial neural networks used for water resources modeling. Water Resour. Res. 41.

Koç, M. L. y, Balas, C.E., 2012. Genetic algorithms based logic-driven fuzzy neural networks for stability assessment of rubble-mound breakwaters. Appl. Ocean Res. 37, 211–219. https://doi.org/10.1016/j.apor.2012.04.005

Kwon, Y., Won, J.-H., Kim, B. J. y, Paik, M.C., 2020. Uncertainty quantification using Bayesian neural networks in classification: application to biomedical image segmentation. Comput. Stat. Data Anal. 142 https://doi.org/10.1016/j.csda.2019.106816.

Lee, H. y, Kim, H., 2021. Uncertainty of rules extracted from artificial neural networks. Appl. Artif. Intell. 35 (8), 589–604. https://doi.org/10.1080/08839514.2021.1922845.

Lee, J.W., Kim, C.G., Lee, J.E., Kim, N. W. y, Kim, H., 2020. Medium-Term rainfall forecasts using artificial neural networks with monte-carlo cross-validation and aggregation for the han river basin, korea. Water 12 (6). https://doi.org/10.3390/w12061743.

Lee, T.-L., 2004. Back-propagation neural network for long-term tidal predictions. Ocean Eng. 31 (2), 225–238. https://doi.org/10.1016/S0029-8018(03)00115-X.

Lee, T. L. y, Jeng, D.S., 2002. Application of artificial neural networks in tide-forecasting. Ocean Eng. 29 (9), 1003–1022. https://doi.org/10.1016/S00298018(01)00068-3

Lee, T.L., Tsai, C.P., Jeng, D. S. y, Shieh, R.J., 2002. Neural network for the prediction and supplement of tidal record in Taichung Harbor, Taiwan. Adv. Eng. Software 33 (6), 329–338. https://doi.org/10.1016/S0965-9978(02)00043-1

Li, L., Gao, Z. y, Yuan, Z.-M., 2019. On the sensitivity and uncertainty of wave energy conversion with an artificial neural-network-based controller. Ocean Eng. 183, 282–293. https://doi.org/10.1016/j.oceaneng.2019.05.003.

Liang, S.X., Li, M. C. y, Sun, Z.C., 2008. Prediction models for tidal level including strong meteorologic effects using a neural network. Ocean Eng. 35 (7), 666–675. https:// doi.org/10.1016/j.oceaneng.2007.12.006

Londhe, S. N. y, Deo, M.C., 2003. Wave tranquility studies using neural networks. Mar. Struct. 16 (6), 419–436. https://doi.org/10.1016/j.marstruc.2003.09.001.

Londhe, S.N., Shah, S., Dixit, P.R., Nair, T.M.B., Sirisha, P. y, Jain, R., 2016. A coupled numerical and artificial neural network model for improving location specific wave forecast. Appl. Ocean Res. 59, 483–491. https://doi.org/10.1016/j.apor.2016.07.004.

López, M. y, Iglesias, G., 2013. Artificial Intelligence for estimating infragravity energy in a harbour. Ocean Eng. 57, 56–63. https://doi.org/10.1016/j.oceaneng.2012.08.009.

López, M. y, Iglesias, G., 2015. Artificial neural networks applied to port operability assessment. Ocean Eng. 109, 298. https://doi.org/10.1016/j.oceaneng.2015.09.016.

Mahjoobi, J., Etemad-Shahidi, A. y, Kazeminezhad, M.H., 2008. Hindcasting of wave parameters using different soft computing methods. Appl. Ocean Res. 30 (1), 28–36. https://doi.org/10.1016/j.apor.2008.03.002.

Maier, H. R. y, Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Software 15 (1), 101–124. https://doi.org/10.1016/S1364-8152(99)00007- 9.

Makarynskyy, O., 2004. Improving wave predictions with artificial neural networks. Ocean Eng. 31 (5), 709–724. https://doi.org/10.1016/j.oceaneng.2003.05.003

Makarynskyy, O., Pires-Silva, A.A., Makarynska, D. y, Ventura-Soares, C., 2005. Artificial neural networks in wave predictions at the west coast of Portugal. Comput. Geosci. 31 (4), 415–424. https://doi.org/10.1016/j.cageo.2004.10.005.

Malekmohamadi, I., Ghiassi, R. y, Yazdanpanah, M.J., 2008. Wave hindcasting by coupling numerical model and artificial neural networks. Ocean Eng. 35 (3), 417–425. https://doi.org/10.1016/j.oceaneng.2007.09.003.

Mase, H., 1995. Evaluation of Artificial Armour Layer Stability by Neural Network Method.

Mase, H., Sakamoto, M. y, Sakai, T., 1995. Neural network for stability analysis of rubble-mound breakwaters. Journal of Waterway Port Coastal and Ocean Engineering-asce 121, 294–299.

McCulloch, W. S. y, Pitts, W., 1943. A logical calculus of the ideas imminent in nervous activity. Bull. Math. Biophys. 115–133.

Medina, J.R., Garrido, J.M., G´omez-Martín, M. E. y, Vidal, C.s., 2004. Armor damage analysis using neural networks. Coastal structures 236–248.

Mokhtari, I., Bechkit, W., Rivano, H. y, Yaici, M.R., 2021. Uncertainty-aware deep learning architectures for highly dynamic air quality prediction. IEEE Access 9, 14765–14778. https://doi.org/10.1109/ACCESS.2021.3052429.

Moslemi, N., Abdi, B., Gohari, S., Sudin, I., Atashpaz-Gargari, E., Redzuan, N., Arya, F., 2022a. Thermal response analysis and parameter prediction of additively manufactured polymers. Appl. Therm. Eng. 212, 118533 https://doi.org/10.1016/j.applthermaleng.2022.118533.

Moslemi, N., Gohari, S., Abdi, B., Sudin, I., Ghandvar, H., Redzuan, N., Rhee, S., 2022b. A novel systematic numerical approach on determination of heat source parameters in welding process. J. Mater. Res. Technol. 18, 4427–4444. https://doi.org/10.1016/j.jmrt.2022.04.039.

Mudronja, L., Matic, P. y, Katalinic, M., 2017. Data-based modelling of significant wave height in the adriatic sea. Transactions on Maritime Science-Toms 6 (1), 5–13. https://doi.org/10.7225/toms.v06.n01.001.

Nabavi-Pelesaraei, A., Rafiee, S., Hosseini-Fashami, F. y, Chau, K.W., 2020. Artificial neural networks and adaptive neuro-fuzzy inference system in energy modeling of agricultural products. In: Predictive Modelling for Energy Management and Power Systems Engineering, pp. 299–334.

Pashova, L. y, Popova, S., 2011. Daily sea level forecast at tide gauge Burgas, Bulgaria using artificial neural networks. J. Sea Res. 66 (2), 154–161. https://doi.org/10.1016/j.seares.2011.05.012.

Pirhooshyaran, M., Scheinberg, K. y, Snyder, L.V., 2020. Feature engineering and forecasting via derivative-free optimization and ensemble of sequence-to-sequence networks with applications in renewable energy. Energy 196, 117136. https://doi.org/10.1016/j.energy.2020.117136.

Portillo Juan, N. y, Negro Valdecantos, V., 2022. Review of the application of artificial neural networks in ocean engineering. Ocean Eng. 259, 111947 https://doi.org/10.1016/j.oceaneng.2022.111947.

Portillo Juan, N., Negro Valdecantos, V. y, del Campo, J.M., 2022. Review of the impacts of climate change on ports and harbours and their adaptation in Spain. Sustainability 14 (12), 7507. Retrieved from. https://www.mdpi.com/2071-1050/14/12/7507.

Primo de Siqueira, B. V. y, Paiva, A.d.M., 2021. Using neural network to improve sea level prediction along the southeastern Brazilian coast. Ocean Model. 168 https://doi.org/10.1016/j.ocemod.2021.101898.

Raheli, B., Aalami, M.T., El-Shafie, A., Ghorbani, M. A. y, Deo, R.C., 2017. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. Environ. Earth Sci. 76 (14) https://doi.org/10.1007/s12665-017-6842-z.

Sadeghifar, T., Lama, G.F.C., Sihag, P., Bayram, A., y Kisi, O., 2022. Wave height predictions in complex sea flows through soft-computing models: case study of Persian Gulf. Ocean Eng. 245, 110467 https://doi.org/10.1016/j.oceaneng.2021.110467.

Salah, P., Reisi-Dehkordi, A., y Kamranzad, B., 2016. A hybrid approach to estimate the nearshore wave characteristics in the Persian Gulf. Appl. Ocean Res. 57, 1–7. https://doi.org/10.1016/j.apor.2016.02.005.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J., y Muller, K.R., 2021. Explaining deep neural networks and beyond: a review of methods and applications. Proc. IEEE 109 (3), 247–278. https://doi.org/10.1109/jproc.2021.3060483.

Shamshirband, S., Mosavi, A., Rabczuk, T., Nabipour, N. y, Chau, K.-w., 2020. Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. Engineering Applications of Computational Fluid Mechanics 14 (1), 805–817. https://doi.org/10.1080/19942060.2020.1773932.

She, L. y, You, X.Y., 2019. A dynamic flow forecast model for urban drainage using the coupled artificial neural network. Water Resour. Manag. 33 (9), 3143–3153. https://doi.org/10.1007/s11269-019-02294-9.

Sildir, H., Sarrafi, S. y, Aydin, E., 2022. Uncertainty propagation based MINLP approach for artificial neural network structure reduction. Processes 10 (9). https://doi.org/10.3390/pr10091716.

Stavarache, G., Ciortan, S. y, Rusu, E., 2020. Optimization of artificial neural networks based models for wave height prediction, 2020Apr 23-25. In: Paper Presented at the 5th International Conference on Advances and Clean Energy Research (ICACER), Electr Network.

Suh, K.-D., Kim, S.-W., Kim, S. y, Cheon, S., 2013. Effects of climate change on stability of caisson breakwaters in different water depths. Ocean Eng. 71, 103–112. https://doi.org/10.1016/j.oceaneng.2013.02.017.

Tiwari, M. K. y, Chatterjee, C., 2010. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). J. Hydrol. 382 (1), 20–33. https://doi.org/10.1016/j.jhydrol.2009.12.013.

Tsai, J.-C. y, Tsai, C.-H., 2009. Wave measurements by pressure transducers using artificial neural networks. Ocean Eng. 36 (15), 1149–1157. https://doi.org/10.1016/j.oceaneng.2009.08.007.

Veaux, R.D.D., Schweinsberg, J., Schumi, J., y Ungar, L.H., 1998. Prediction intervals for neural networks via nonlinear regression. Technometrics 40, 273–282.

Vieira, F., Cavalcante, G., Campos, E. y, Taveira-Pinto, F., 2020. A methodology for data gap filling in wave records using Artificial Neural Networks. Appl. Ocean Res. 98 https://doi.org/10.1016/j.apor.2020.102109.

Wang, B., Wang, B., Wu, W., Xi, C., y Wang, J., 2020. Sea-water-level prediction via combined wavelet decomposition, neuro-fuzzy and neural networks using SLA and wind information. Acta Oceanol. Sin. 39 (5), 157–167. https://doi.org/10.1007/s13131-020-1569-1.

Wang, W.H., y Ma, W.M., 2023. Bootstrapped artificial neural network model for uncertainty analysis in MELCOR simulation of severe accident. Prog. Nucl. Energy 157. https://doi.org/10.1016/j.pnucene.2022.104556.

Weiss, M., Tonella, P. y, Ieee Comp, S. O. C., 2021. UNCERTAINTY-WIZARD: fast and user-friendly neural network uncertainty quantification, 2021Apr 12-16. In: Paper Presented at the 14th IEEE Conference on Software Testing, Verification and Validation (ICST), Electr Network.

Wenzel, M., y Schroeter, J., 2010. Reconstruction of regional mean sea level anomalies from tide gauges using neural networks. Journal of Geophysical Research-Oceans 115. https://doi.org/10.1029/2009jc005630.

Weytjens, H., y De Weerdt, J., 2022. Learning uncertainty with artificial neural networks for predictive process monitoring. Appl. Soft Comput. 125 https://doi.org/10.1016/j.asoc.2022.109134.

Wu, S. J.;Hsu, C. T. y Chang, C. H. Stochastic modeling of artificial neural networks for real-time hydrological forecasts based on uncertainties in transfer functions and ANN weights. Nord. Hydrol. doi:10.2166/nh.2021.030.

Yagci, O., Mercan, D.E., Cigizoglu, H.K., y Kabdasli, M.S., 2005. Artificial intelligence methods in breakwater damage ratio estimation. Ocean Eng. 32 (17), 2088–2106. https://doi.org/10.1016/j.oceaneng.2005.03.004.

Yang, C.-H., Wu, C.-H., Hsieh, C.-M., Wang, Y.-C., Tsen, I. F. y, Tseng, S.-H., 2021. Deep learning for imputation and forecasting tidal level. IEEE J. Ocean. Eng. 46 (4), 1261–1271. https://doi.org/10.1109/joe.2021.3073931.

Zhao, J., Fan, Y. y, Mu, Y., 2019. Sea level prediction in the yellow Sea from satellite altimetry with a combined least squares-neural network approach. Mar. Geodesy 42 (4), 344–366. https://doi.org/10.1080/01490419.2019.1626306.

Zheng, Z., Ma, X., Ma, Y. y, Dong, G., 2020. Wave estimation within a port using a fully nonlinear Boussinesq wave model and artificial neural networks. Ocean Eng. 216 https://doi.org/10.1016/j.oceaneng.2020.108073.

3.7. Comparison of SIMAR-WANA, ERA-5 and Waverys databases for maritime climate estimations and implications for coastal protection structures

Abstract and Keywords

Estimating maritime climate accurately is essential to ensure the safety of marine structures. Most reliable data are in-situ measurements, but due to the lack of this data in both, time and space, other datasets from reanalysis and wave models have to be used. In this research article the three main databases of the Spanish Mediterranean Sea used for the design of maritime structures (ERA5, Waverys and SIMAR-WANA) are compared with the buoys of Valencia, Alicante and Cabo de Palos. A statistical analysis using Python, Matlab and AMEVA tool has been carried for both, the whole historic records (20,362,032 data) and the extreme distribution (95^{th} percentile). Once the statistical study was carried out, an analysis was made to assess the consequences of using these datasets in the design of coastal protection and their reliability. Waverys showed a slightly better performance when compared to in situ measurements and SIMAR-WANA showed the least underestimation of the extreme values of Hs. Regarding the consequences on the design of coastal protection structures, the estimation of extreme events proved to be of particular interest, the databases typically underestimate the values of extreme events, which was found to have the potential to lead to errors in the design of coastal protection structures up to 60%.

Keywords: Dataset comparison; coastal protection; extreme events; ERA5; Copernicus; Spanish Mediterranean Sea

1. Introduction

Estimating maritime climate accurately is essential for many human activities, such as, shipping, port operation or naval architecture. Reducing errors or uncertainties in wind and wave estimation is key to ensuring the operation and preventing the failure of marine and coastal structures (Ardhuin et al., 2007; Campos and Guedes Soares, 2016). Data for the estimation of maritime climate is typically in the design process of coastal or port structures. In-situ measurements are usually the most reliable data. However, these measurements are very limited in both, time, and space; there are also time gaps caused by extreme storms; and

some of these data are affected by refraction, local winds, and shoaling. The number of available instrumental records is far from what is needed to properly model the maritime climate, especially for extreme events modelling, that is one of the main concerns in structural design. Therefore, it is necessary to use data from reanalysis and numerical models that can cover larger areas and time spans (Agarwal et al., 2013; Bromwich and Wang, 2008; Campos and Guedes Soares, 2016; Stopa and Cheung, 2014). The output from hindcast or reanalysis are often used as boundary conditions in the models and if the data feeding the model already has errors in it, these errors will be propagated to the final results. Therefore, ensuring the quality of hindcast and reanalysis is essential for the correct design of maritime structures.

To data, there are several studies dedicated to the improvement of wave numerical models. The first studies were conducted around the 2000's (Sebastiao et al., 2000). In 2006 the problem of predicting extreme events accurately was studied more deeply (Stephens and Gorman, 2006). Ponce de León and Guedes Soares (2015) studied the extreme event "Hercules" hindcast using WAM models. Akpinar and Ponce de León (2016) assessed the wind reanalysis for extreme events in the Black Sea and obtained the best SWAN configuration to model these events.

Since then, different studies to improve wave estimation have been carried: the improvement of WAM models (Bhatt et al., 2004), the evaluation of ERA-Interim and North American Regional Reanalysis (NARR) (NCEP) wind products (Appendini et al., 2013), the study of WaveWatch III models (Alday, M. et al., 2022; Boudière et al., 2013), the development of new parameters to evaluate the accuracy of different databases (Catterson et al., 2016), the comparison of HIPOCAS and ERA reanalysis datasets in the North Atlantic (Campos and Guedes Soares, 2016), the evaluation of different models in different areas of the oceans (Bjorkqvist et al., 2020; Umesh and Swain, 2018), the comparison of model hindcasts forced by different wind reanalyzes (Sharmar et al., 2021), etc.

In the case of the Mediterranean Sea, as it is an enclosed basin, the evaluation of these datasets needs special treatment. Wind and wave patterns are much more complex because they are affected by land and differ from those of open oceans (Cavaleri and Bertotti, 2004). Ardhuin et al. (2007) performed one of the first studies to compare wave and wind measurements in the Mediterranean Sea and concluded that WaveWatch III models underestimate extreme wave heights and that Jason altimeter wind speeds are on average consistent with the corresponding buoy data. Two years later, Ponce de Leon and Soares (2008) analyzed the

sensitivity of the models in the Western Mediterranean Sea and highlighted the difficulty of wave hindcasting in areas surrounded by land. Aydogan and Ayat (2021) conducted a similar study in a different enclosed basin, the Black Sea, and the conclusions were along the same line.

However, in the area of the Spanish Mediterranean, the number of studies is very limited, and more effort should be put into modelling this area (Eldeberky et al., 2002; Elkut et al., 2021).

The aim of this study is to address the importance of improving the estimations of the wave climate in the Spanish Mediterranean Sea. For this purpose, this research paper analyzes the 3 main long-term databases used in the design of maritime structures in the Spanish Mediterranean Sea (ERA5, Waverys and SIMAR-WANA) and compares them with the buoys of Valencia, Alicante and Cabo de Palos, to verify which are the most accurate in the long-term and extreme distributions. In addition, the consequences of using different databases for the design of coastal protection structures are studied, which to the authors' knowledge, has never been done before.

2. Area of study and datasets

2.1. Area of study

The area of study is located offshore SE Spain, in the coast that extends from Valencia to Cabo de Palos. This area is known as Western Mediterranean, the Mediterranean Sea is a semi-closed deep ocean basin surrounded by continents with different orographic and land use characteristics, whose climate is strongly influenced by land-ocean interactions and by its orography (Ciardini et al., 2016). Due to its geography, its wave and atmospheric patterns are affected by regional and local phenomena. The Western Mediterranean is one of the areas that is going to be most affected by climate change (Portillo Juan et al., 2023), so that the importance of studying its wave climate and of correctly characterizing the main wave climate variables (Hs, Tp and Dir).

2.2. Buoy data

In this area there are three buoys from Puertos del Estado external network, which is the Spanish authority in charge of obtaining historical maritime climate records. The main characteristics of these buoys are presented in Table 1.

Table 1. Buoy characteristics.

Buoy	Longitude	Latitude	Depth	Records
Valencia	0.20 E	39.51 N	260	15/09/2005 - 01/10/2022
Alicante	-0.417 E	38.25 N	50	26/09/1985 - 15/01/2014
Cabo de Palos	-0.31 E	37.65 N	230	18/07/2006 - 01/10/2022

The main differences between the 3 buoys are their location, depth, and the amount of data available. Depending on where they are located, they are used to model one area of the coast or another. With respect to the depth, the most problematic buoy is the Alicante buoy, that it is 60 m deep and could be affected by refraction. Finally, regarding the data sets, all of them give hourly data and the way to obtain historical records is the same for the three buoys, the one established by the Spanish authority Puertos del Estado (Puertos del Estado. Oceanografía. Registros históricos de oleaje). Whether the data sets are longer or shorter depends on when Puertos del Estado installed the buoy and when the authority decided to stop using it (in the case of Alicante). Figure 1 shows the location of each buoy and the bathymetry.

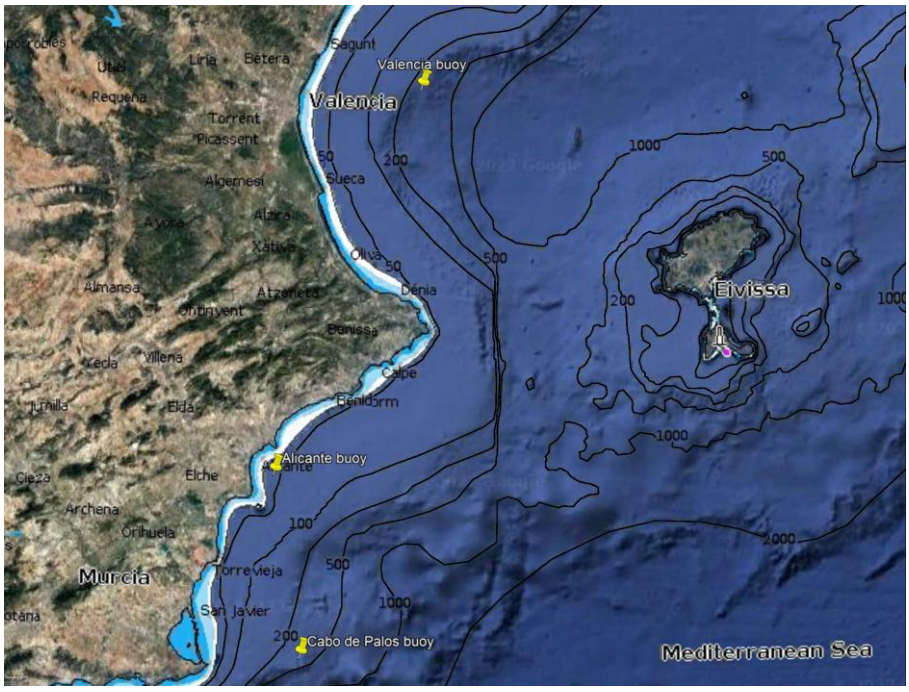


Fig.1. Location of the buoys and use over time.

Referring to the Alicante buoy, although it is the buoy with the longest data time series (almost 30 years), it is very shallow and is affected by refraction, local winds, and shoaling, which may affect the records taken by the buoy. Data from this buoy

has been historically used to check results and as supporting and complimentary data. In addition, it has to be said that Alicante buoy records changed in 04/09/2007 and from this date onwards the records were taken with another buoy 2 m deeper, but practically in the same position as the previous one.

Figure 2 shows the historical records of the three buoys used in this paper, the area where there are common records in the three buoys is highlighted in red and corresponds to the time period between 18/07/2006 to 15/01/2014. From the graph, it can be seen that wave height historical records are slightly higher in the area of Cabo de Palos. On the other hand, in Alicante, wave heights historical records are lower than in Valencia and Cabo de Palos, which, may be due to refraction effects. The most significant statistical parameters of the historical records of the buoys are presented in Table 2. It should be noted that the Spanish port buoys are composed of scalar buoys of the Waverider (Datawell) type and directional buoys of the Triaxys™ (Axys) type.

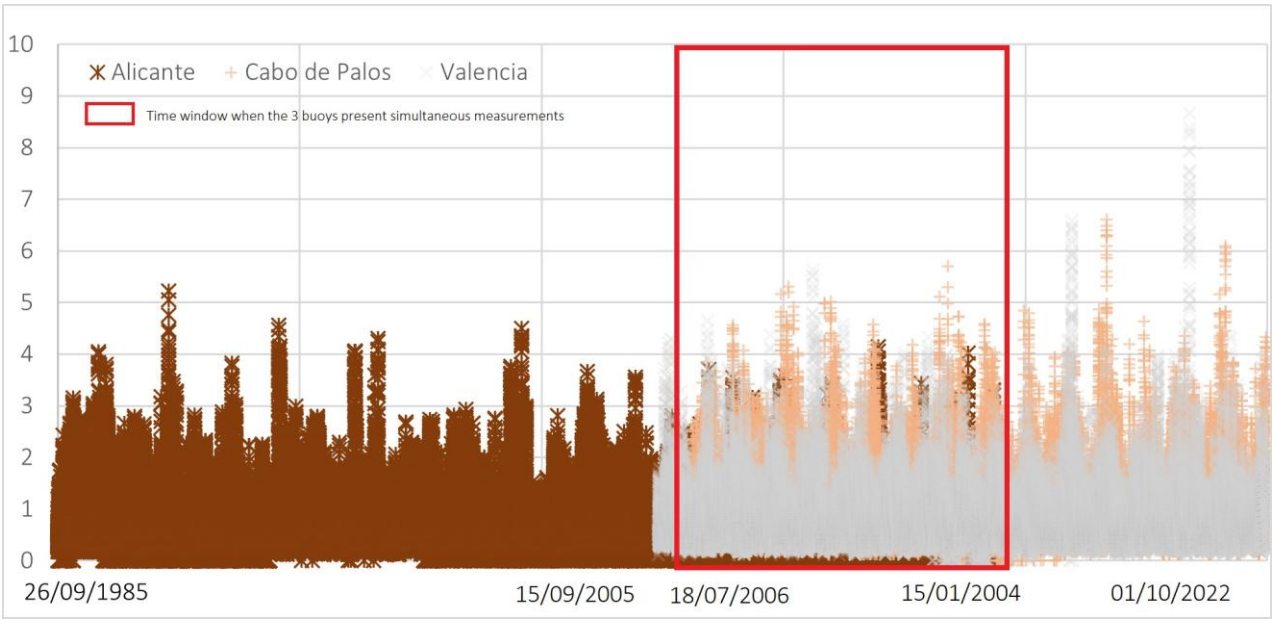


Fig.2. Wave measurements over time from the three buoys studied.

Table 2. Buoy statistical parameters.

	Valencia		Alicante		Cabo de Palos	
	Hs (m)	Tp (s)	Hs (m)	Tp (s)	Hs (m)	Tp (s)
Mean	0.82	5.11	0.69	5.49	1.00	5.57
Standard Deviation (SD)	0.53	1.57	0.40	1.57	0.63	1.45
Minimum	0.13	1.7	0.13	1.84	0.11	1.74
Maximum	8.67	11.77	5.23	14.95	6.61	16.66

2.3. Wave datasets

The records from these 3 buoys have been used to compare Waverys, ERA5 and SIMAR-WANA databases in the Spanish Mediterranean Sea. The location and depth of wave data extraction from each database (SIMAR-WANA, ERA-5 and Waverys) are presented in Table 3.

Table 3. Location and depth of data extraction points of the three databases.

	ERA5			Waverys			SIMAR		
	Latitude (º)	Longitude (º)	Depth (m)	Latitude (º)	Longitude (º)	Depth (m)	Latitude (º)	Longitude (º)	Depth (m)
Valencia	39.51	0.20	260	39.60	0.20	145	39.50	0.167	240
Alicante	38.20	-0.40	55	38.20	-0.40	55	38.25	-0.42	40
Cabo de Palos	37.65	-0.31	230	37.60	-0.40	440	37.67	-0.33	260

ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather that replaces the previous ERA-Interim reanalysis. Data is available from 1950, with Climate Data Store entries for 1950-1978 and from 1959 onwards (Hersbach et al., 2020). ERA5 uses the 4D-Var data assimilation IFS model cycle 41r2. The atmospheric model in the IFS is coupled to a land-surface model (HTESSEL), and an ocean wave model (WAM). This database has a hourly time resolution and a spatial resolution of 0.25º (ECMWF, 2018).

Waverys is based on the MFWAM model (Aouf et al., 2019), a third generation wave model. The model uses an irregular grid with a 1/5º spatial resolution. It considers oceanic currents from the GLORYS12 physical ocean reanalysis (Jean-Michel et al., 2021) and assimilates significant wave height observed from historical altimetry missions and directional wave spectra from Sentinel 1 SAR from 2017 onwards (CMEMS, 2021). It has a 3-hourly time resolution and a spatial resolution of 0.2º and covers the period 1993–2019. It assimilates data from satellite HY-2A (Law-Chune et al., 2021).

The SIMAR-WANA dataset refers to two subsets, SIMAR from 1958 to 2005 and WANA from 2006 to the present. The SIMAR ensemble consists of simulations from atmospheric and wave models that cover the entire Spanish coastal environment. The wave simulation was carried out by Puertos del Estado, using data from the AEMET atmosphere simulation. The WANA data come from the sea state prediction system that Puertos del Estado has developed in collaboration

with AEMET. WANA data are diagnostic and analysis data with hourly assimilation. This means that for each instant the model provides wind and pressure fields consistent with the previous evolution of the modelled parameters and with the observations made. Both datasets, SIMAR and WANA, use WaveWatch III models and give hourly data with a spatial resolution going from 0.25º to 0.05º in the points developed specifically for ports (Estado, 2020) due to the nested models developed in this specific points. WANA subset also uses WAM model.

The main differences of the three datasets are presented in Table 4.

Table 4. Summary of the main characteristics of ERA5, Waverys and SIMAR-WANA datasets.

	ERA5	Waverys	SIMAR	WANA
Organism	ECMWF	ECMWF (CMEMS)	Spanish Ports	Spanish Ports
Wave model	ECWAM third-generation model with IFS Cy41r2	MFWAM V4	WaveWatch III	WaveWatch III and WAM.
Wave Spatial resolution	0.25º	0.2º	0.125º and 0.25º at the edges	0.125º to 0.05º
Wind model	McRad atmosphere model (Morcrette et al., 2008). Atmospheric Motion Vectors winds and satellites.	ERA5	Dynamic downscaling based on RCA3.5 model (Samuelsson et al., 2009) with ERA-Interim and ERA40 data	HIRLAM (Bosssing Christensen et al., 2007) until 2018 and HARMONIE-AROME (Bengtsson et al., 2017) from 2018
Wind Spatial resolution	0.25º	0.2º	0.2º	0.1º to 0.02º
Time resolution	Hourly	3-Hourly	Hourly	Hourly
Time series	1959-2020	1993-2021	1958-2005	2005-2022
Bathymetry	ETOPO2/NOAA	ETOPO2/NOAA	ETOPO2/NOAA	ETOPO2/NOAA

Interactions	Coupled atmospheric model	Wave-current interaction included	Coupled atmospheric model	Coupled atmospheric model
Physics model	ECWAM-IFS41R2 physics	ST4-like physics	-	-

3. Methodology

The methodology followed in this study can be divided into three main steps: the downloading and pre-processing of the data, the statistical analysis of the data and the application of the results obtained to the design of coastal protection structures.

Once all the data was obtained (Copernicus; ECMWF; *Puertos del Estado. Oceanografía. Registros históricos de oleaje*), a pre-processing of all the data and compilation of all the measurements to evaluate each dataset for the three sites were done. As there were 3 different locations and variable record lengths and transmission rates, a time-pairing of the four databases in the 3 locations was conducted.

The pre-processing and time-pairing was conducted applying Python (Rossum, 1989), using the own scripts developed by the authors. There were time gaps in the buoy records that had to be eliminated in order to compare the different databases with the existing records. To address this problem, an algorithm to detect these time gaps and eliminate them was developed. After eliminating these time gaps, the matching of the 3 databases with the buoy measurements was done using a time-pairing methodology, each simulated data of each database was paired with another simulated data of the other 2 databases and with the buoy observations. Finally, all the data was filtered to eliminate false values and errors. All values of wave heights less than 0.1 m and periods less than 1 s were removed, as they were considered unreliable. These errors may be mainly due to noise in the measurements.

Table 5 shows the amount of records and time period used in each location and database. Whenever possible, the same comparison period has been used for the 3 databases (2006-2020), so that the comparisons are consistent. However, in the case of the Alicante buoy this has not been possible, as the SIMAR-WANA data

were only available for a two-year period (2012-2014), which is why the conclusions obtained for the SIMAR-WANA data sets at this location should be taken with particular caution. Referring to the amount of data available, all SIMAR-WANA and ERA5 databases have the same number of data, while Waverys has one third of this data as the time resolution of this database is 3-hourly and not hourly.

Table 5. Database matching.

	Trans. rate	Record length			Time match record length for evaluation			Nº of observations		
		Valencia	Alicante	Cabo de Palos	Valencia	Alicante	Cabo de Palos	Valencia	Alicante	Cabo de Palos
Buoy	Hourly	2005 - 2022	1985-2014	2006 - 2022		-		117,112	102,223	118,601
Waverys	3-hourly		1993-2021		2005-2020	1993-2006	2006-2020	39,037	34,074	39,534
ERA5	Hourly		1959-2020					117,112	102,223	118,601
SIMAR-WANA	Hourly	1958-2022	2012-2022	1958-2022		2012-2014			11,071	

The second step in the methodology followed in this paper was the statistical analysis. To do this part of the study, the software MATLAB (Hunter, 2003), AMEVA (IHC, 2013) and Python (Rossum, 1989) were used. The values of correlation coefficient (r), Root Mean Square Error (RMSE), Bias, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) for the whole dataset and the 95th percentile of the variables were obtained. The 95th percentile was obtained from the cumulative density functions (CDFs) of Hs, obtaining the 95th percentile of Tp basing on the Tp results specific to individual 95th Hs values. Three variables were analyzed, the significant wave height (Hs), the peak period (Tp) and the wave direction defined at the spectral peak frequency (Dir). The direction has been calculated in degrees, being the reference point in all the data the north (0º) and calculating all the directions in a clockwise direction.

The definition of the statistical metrics used is presented from Equation 1 to 5.

$$r = \frac{\sum_{i=1}^{n}(o_i-\bar{o})(u_i-\bar{u})}{\sqrt{\sum_{i=1}^{n}(o_i-\bar{o})^2}\sqrt{\sum_{i=1}^{n}(u_i-\bar{u})^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i-u_i)^2} \quad (2)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(o_i-u_i) \quad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|o_i - u_i| \quad (4)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|o_i - u_i|}{u_i} \quad (5)$$

Where o_i is the output of the hindcast or reanalysis database, u_i is the time-paired buoy record, $\bar{o}$ is the mean of the outputs of the database in each location, $\bar{u}$ the mean of the historical records of the buoy, and n the number of observations of each database in each location.

The 95th percentile of the records has been used to define the extreme distribution because it corresponds to the return period of 20 years, that it is the return period that the Spanish consortium uses to define extreme events. Relating the 95th percentile with the return period of the structures is the most common approach in the design of coastal maritime structures. The DNV (Det Norske Veritas) standards (Hopstad et al., 2018) and Puertos del Estado reports for the extreme distribution (*Informes anuales Oceanografía Puertos del Estado*) were used. The DNV-ST-0119 proposes a table that relates different percentiles with different return periods (page 57, table 4-3), following this table the return period that corresponds to the 95th percentile is 20 years, which agrees with the Equation proposed by Puertos del Estado (Equation 6).

$$T_r = \frac{1}{P(H_r)} \quad (6)$$

Where T_r is the return period and $P(H_r)$ the probability of exceedance of H_r, and H_r the wave height associated to T_r. In this case, the 95th percentile is the one that is being used. Therefore, the probability of exceedance is 5%, being the return period 20 years. This value is of particular importance because is the return period that the Spanish consortium (which is the organism that insures public works) establishes to define extreme events and above which it covers the costs of the damage caused in maritime structures. Therefore, all calculations presented in this section have a return period of 20 years.

In addition, in each of the subsections Taylor diagrams are presented. It has been decided to use this type of diagram because it is the most appropriate and widespread when analyzing and assessing which model or dataset best fits measurements. These diagrams are widely used to study atmospheric weather, climate and other fields of engineering.

When all the data was analyzed, it was compared to the annual reports of Puertos del Estado (*Informes anuales Oceanografía Puertos del Estado*). As a mismatch of

the values of Tp obtained and the ones presented in these reports was found, a thorough study of this variable was done. The 95th percentile of Tp and the peak period related to the 95th percentile of Hs were compared to the expected Tp for a return period of 20 years for Cabo de Palos and Valencia location. Finally, the Tp results specific to individual 95th Hs values was plotted against the 20 years expected Tp.

Finally, a level I deterministic empirical design approach was used for the analysis of the coastal protection structures. For this purpose, the empirical expressions commonly used in breakwater design were taken and expressed as a function of Hs and Tp leaving the rest of the variables as a constant c with the final aim of conducting a sensitivity study of how errors in the measurement of Hs and Tp can affect the final design of coastal protection structures. For the simplicity of the study, it has been considered that the errors of the output of the hindcast and reanalysis databases in deep water, that are the errors studied in this paper, are propagated through the models to nearshore location.

The main limitation of this study is that it has been assumed that the errors in the input of wave propagation models, commonly SWAN (Booij et al., 1999) or statistical models, are transmitted to the final results and, consequently, to the final design of coastal protection structures. Wave propagation modelling to shallow (coastal) areas has not been carried because it was considered that it would exceed the scope of the article and the final article would far exceed the length of a scientific article, but future research lines could study more in depth how these errors in the input of wave propagation models could affect their results.

4. Results

In this section the results obtained are presented. First, a statistical analysis to assess the performance of Waverys, ERA5 and SIMAR-WANA is presented at the three buoy sites. Then, all the results are combined to determine the quality of these wave databases in the Spanish Mediterranean Sea. Finally, the analysis of the peak period, which has been considered of special interest, is presented.

4.1. Results per buoy

Valencia

Dataset

The statistical analysis for the time-paired measurement to the four long-term archives using the period of record of the buoy data in Valencia is presented in Table 6. The data analyzed go from September 2005 through December 2020.

Table 6. Statistical analysis of Valencia buoy.

	Hs (m)			Tp (s)			Dir (º)		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	0.64	0.66	0.78	6.39	5.90	6.29	117.46	122.04	116.31
r	0.93	0.94	0.81	0.69	0.69	0.60	0.68	0.71	0.63
RMSE	0.27	0.25	0.33	1.42	1.39	1.73	55.00	55.58	63.60
Bias	-0.16	-0.14	-0.03	0.50	0.10	0.46	-4.58	0.57	-5.73
MAE Mean	0.20	0.18	0.22	0.96	0.91	1.20	29.79	27.50	34.95
MAE Max	3.27	2.60	6.16	18.36	19.01	21.94	358.10	359.82	356.95
MAPE Mean	0.26	0.21	0.28	0.22	0.19	0.26	0.66	0.77	0.54
MAPE Max	9.75	11.27	14.81	0.79	0.78	0.82	354.69	358.63	357

In Valencia buoy, SIMAR-WANA was the database that performed the worst in the three variables studied, Hs, Tp and Dir. It has the lowest r (0.81, 0.60 and 0.63 for Hs, Tp and Dir) for the three variables and the highest RMSE, MAE and MAPE (0.33 m, 0.22 m and 0.28 for Hs). However, SIMAR-WANA bias is the closest to 0, while Waverys and ERA5 bias are -0.16 m and -0.14 m, respectively. This means that SIMAR-WANA is the database that shows the least underestimation for Hs, while Waverys and ERA5 may estimate wave heights 0.16 and 0.14 m lower than in the measurements, which supposes an underestimation of the 20% of the wave height. It is also remarkable that the correlation coefficient is good for Hs for the three databases (around 0.9) but decreases considerably for the variables Tp and Dir (around 0.6).

95th percentile

Data have also been studied in the extreme distribution. For this purpose, the 95th percentile was computed.

In Table 7 we present the values of the statistical parameters for the 95th percentile (P_95) of Hs and Tp in Valencia. Columns 2 to 4 show results for 95th percentile of Hs and columns 5 to 7 show results for Tp associated to 95th percentile of Hs. Hs correlation coefficients decrease (0.86, 0.86 and 0.62 for Waverys, ERA5 and

SIMAR-WANA), while Tp correlation coefficient increase (0.83, 0.80 and 0.68 for Waverys, ERA5 and SIMAR-WANA), in comparison to the statistical results obtained for the whole dataset (0.93, 0.94 and 0.81 for Hs and 0.69, 0.69 and 0.60 for Tp). The directional correlation coefficients remain more or less unchanged. This suggests that the three databases are less accurate in the simulation of Hs related to extreme events, which is not the case with the peak periods. It would not be the case if instead of considering the Tp associated to these 95th percentile of Hs, the 95th percentile of Tp was considered. In this case, the accuracy of the databases to estimate extreme Tp would get much worse, with correlation coefficients around 0.2.

Table 7. Statistical analysis of the 95th percentile of Valencia buoy.

	Hs (P_95-1.79 m)			**Tp (P_95_Hs_1.79 m)**			**Dir (P_95_Hs_1.79 m)**		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	1.94	1.90	2.12	7.88	7.56	8.00	113.41	118.42	114.22
r	0.86	0.86	0.62	0.83	0.80	0.68	0.70	0.70	0.62
RMSE	1.39	1.38	1.46	1.20	1.30	2.33	33.11	58.26	65.94
Bias	-0.47	-0.52	-0.30	0.34	0.04	0.44	-2.35	-1.96	-7.28
MAE Mean	0.49	0.54	0.58	0.77	0.74	1.11	10.10	28.94	35
MAE Max	3.27	2.60	6.07	5.96	11.61	10.84	342.35	351.99	350
MAPE Mean	0.21	0.23	0.24	0.11	0.10	0.16	0.17	0.67	0.46
MAPE Max	0.69	0.77	3.28	1.19	1.45	1.72	42.65	288.78	108.3

For the extreme distribution, the errors increase, reaching maximums of 6 m for wave heights and 11.6 s for peak periods. These large errors in the recording of extreme events may be due to various causes, such as the failure of the buoys due to extreme phenomena, the difficulty of the databases in detecting and predicting this type of event, or the possible time lag that leads to precision failures. We note that, for all analyzed datasets, there is a negative bias when comparing with buoy data (-0.47 m, -0.52 m and -0.30 m for Waverys, ERA5 and SIMAR-WANA). This means that the databases may slightly underestimate the values of extreme wave heights (up to 25% of the wave height). The database that performs the best in this aspect is SIMAR-WANA. To illustrate better these results, Quantile-Quantile plots

of Hs and Tp data are presented in Figure 3 and 4. The 95th percentile is the area between the two red lines. The difference between the databases and the actual historical records has been shaded in different colors. Yellow corresponds to SIMAR-WANA data, red to ERA5 and grey to Waverys. The term "Ref" refers to the reference line, that corresponds to buoy wave measurements. As already mentioned in the previous section, the fit of the Hs quantiles is much better than that of Tp. It is also remarkable how until the 95th percentile the fit of the distributions is very accurate, but for the 95th percentile and, consequently, for extreme events, it significantly decreases. In this aspect, SIMAR-WANA is the database with the best performance. However, to select the best database, other statistical parameters have also to be considered apart from the bias. This is the main reason why Taylor diagrams have also been plotted (See section below).

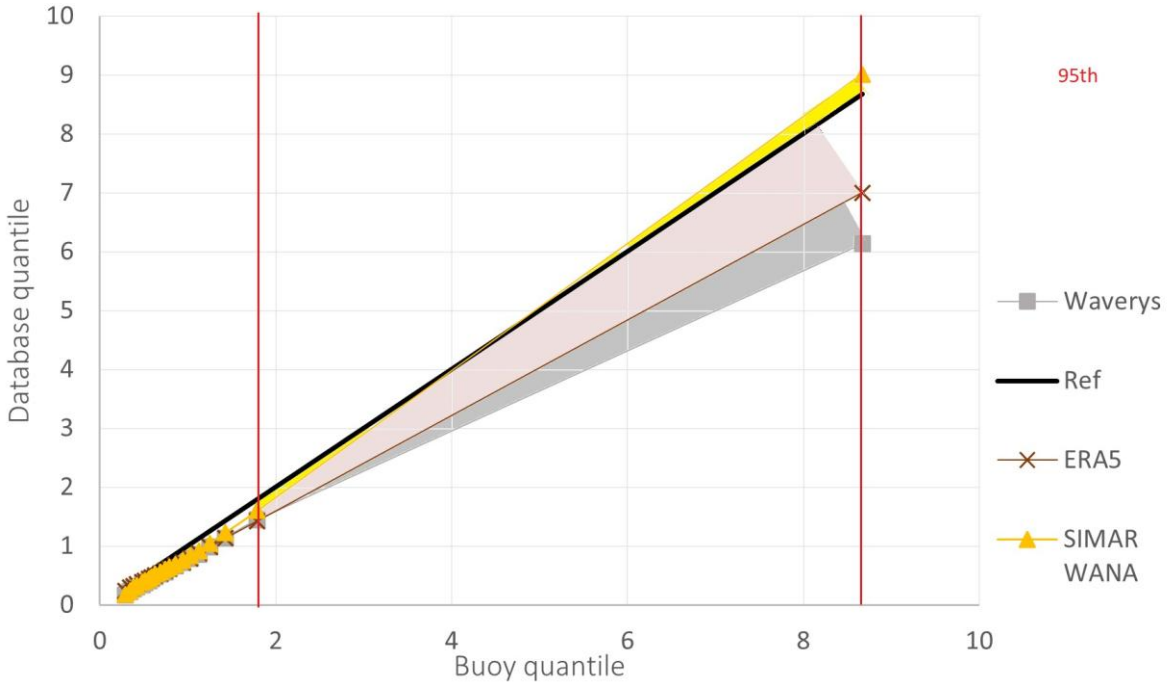


Fig.3. Quantile-Quantile plot of Hs in Valencia.

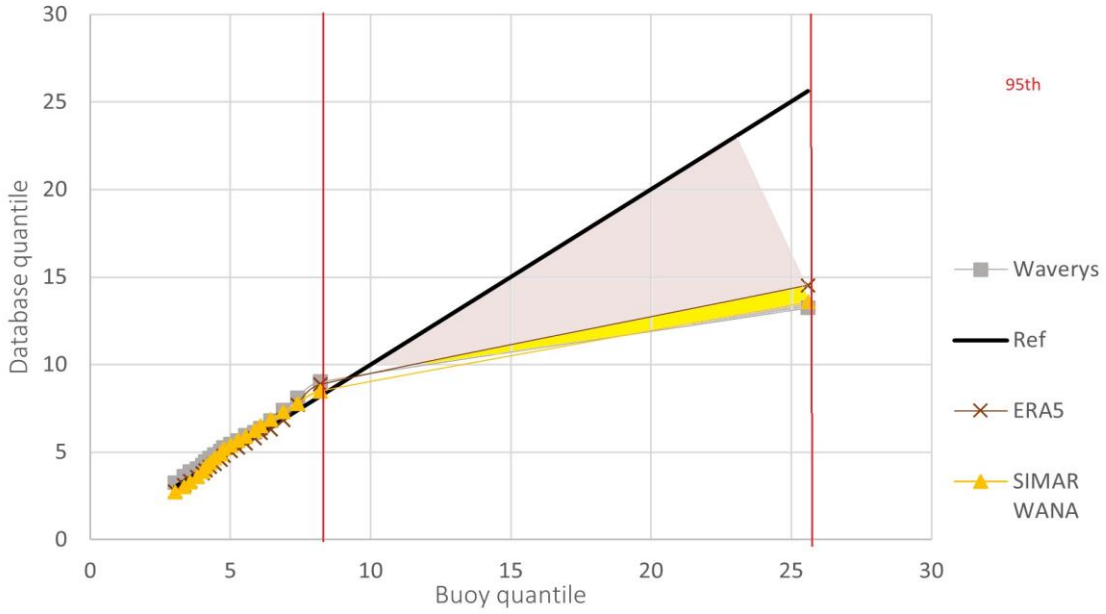


Fig.4. Quantile-Quantile plots of Tp in Valencia.

Taylor diagrams

In this section the Taylor diagrams for Hs and Tp for Valencia buoy are presented in Figure 5 and Figure 6. In all the Taylor diagrams presented in this research, the long-term distribution is represented by an asterisk and the 95th percentile is represented by a lower case 'e' and by pentagrams. 'WR' refers to Waverys database and 'SW' to SIMAR-WANA database. The straight lines of the diagrams are the correlation coefficients, the arches in a dotted line are the standard deviations (SD) and the arches in continuous lines are the RMSEs.

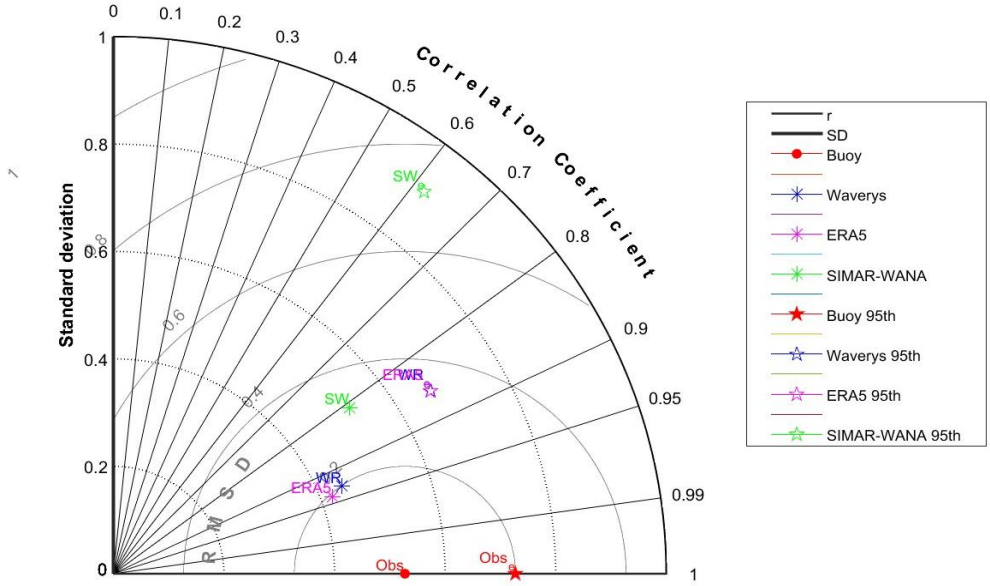


Fig.5. Taylor diagram for Hs in Valencia.

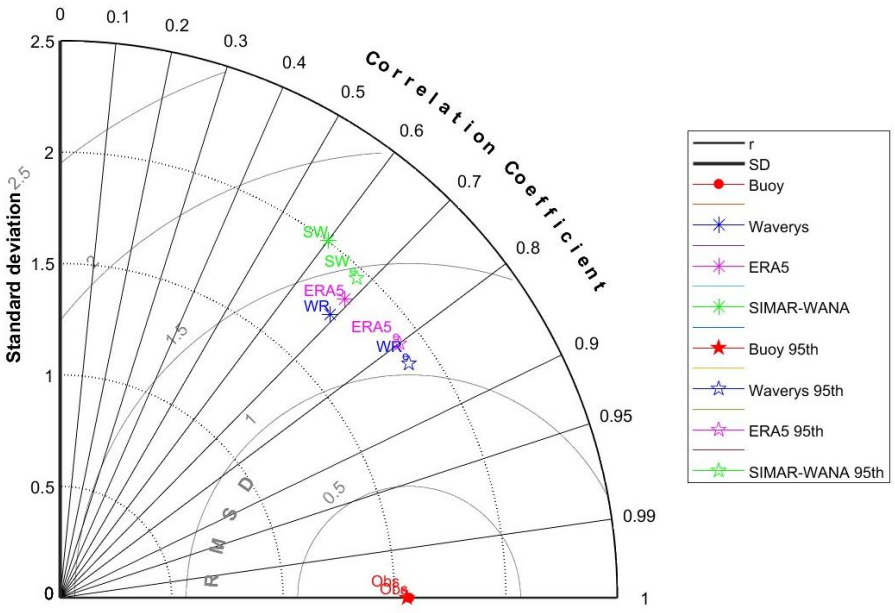


Fig.6. Taylor diagram for Tp in Valencia.

From Figure 5, Waverys and ERA5 perform very similar for Hs estimations, while SIMAR-WANA shows a slightly worse performance, especially in the extreme distribution (r of 0.62 for Hs and RMSE of 1.46 m). In the case of Tp, the results are in line with those obtained for Hs, with correlation coefficients of 0.8 for Waverys and ERA5 and of 0.62 for SIMAR-WANA and differences of 1 s in the RMSE. From these graphs, it can be seen that the most accurate databases are Waverys and ERA5.

Alicante

Dataset

Analyzes are based on consistent time paired observations where the time series go from 1993 to 2006 for Waverys and ERA5 and from 2012 to 2014 for SIMAR-WANA to retain consistency.

The values of the statistical analysis are presented in Table 8. In this case, unlike the Valencia buoy, the most accurate database is that of SIMAR-WANA. It has the lowest MAE for Hs and Tp (0.13 m and 1.18 s) and the lowest RMSE for Hs (0.20 m). However, the SIMAR WANA results have only been analyzed for 2 years, so it is not possible to say anything conclusive for longer periods of time or regarding extreme events.

Table 8. Statistical analysis of Alicante buoy.

	Hs (m)			Tp (s)			Dir (º)		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	0.61	0.65	0.73	5.63	5.39	5.73	Scalar Buoy		
r	0.88	0.86	0.90	0.69	0.63	0.65			
RMSE	0.20	0.21	0.20	1.29	1.40	1.78			
Bias	-0.06	-0.02	0.05	0.15	-0.09	0.42			
MAE Mean	0.14	0.14	0.13	0.85	0.90	1.18			
MAE Max	1.88	2.12	1.93	22.58	22.79	12.25			
MAPE Mean	0.22	0.22	0.22	0.18	0.18	0.23			
MAPE Max	2.33	6.90	2.67	3.06	3.43	3.41			

The same as in Valencia buoy happens again with the estimation of Tp. While Hs has correlation coefficients around 0.9, Tp correlation coefficients are around 0.7.

In addition, the bias of the three databases is very close to 0 when estimating Hs (-0.06, -0.02 and 0.05 m for Waverys, ERA5 and SIMAR-WANA). For Tp estimation, Waverys and SIMAR-WANA slightly overestimates the values of the peak period (bias = 0.42 s).

95th percentile

Similar to what was found with the Valencia buoy, the accuracy of the databases to estimate extreme waves decrease (Table 9). The database that least underestimates the extreme values of Hs is SIMAR-WANA, with a bias very close to 0 for the 95th percentile of Hs (0.09 m). However, other statistical parameters have also to be considered to evaluate the performance of the different databases. Waverys has the highest r (0.82 for Hs and 0.81 for Tp), the lowest RMSE (0.47 m for Hs and 0.93 for Tp) and the lowest MAE for Tp (0.68 s). Therefore, it could be concluded that, in general terms, Waverys is the database that shows the highest accuracy, although it may underestimate extreme events.

Table 9. Statistical analysis of the 95th percentile of Alicante buoy.

	Hs (P_95-1.48 m)		Hs (P_95-1.49 m)	Tp (P_95_Hs_1.48 m)		Tp (P_95_Hs_1.49 m)
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	1.65	1.64		7.37	7.20	7.95
r	0.82	0.80	0.78	0.81	0.78	0.70
RMSE	0.47	0.49	0.48	0.93	1.10	1.82
Bias	-0.33	-0.35	0.09	-0.21	-0.37	0.49
MAE Mean	0.40	0.41	0.36	0.68	0.73	0.92
MAE Max	1.88	2.12	1.93	5.16	7.17	7.72
MAPE Mean	0.20	0.21	0.18	0.09	0.10	0.13
MAPE Max	0.89	1.05	0.85	1.15	1.25	1.69

Figures 7 and 8 show the quantile-quantile plots of Hs and Tp of Alicante. The plots are very similar to that of Valencia location, with SIMAR-WANA data showing the best performance.

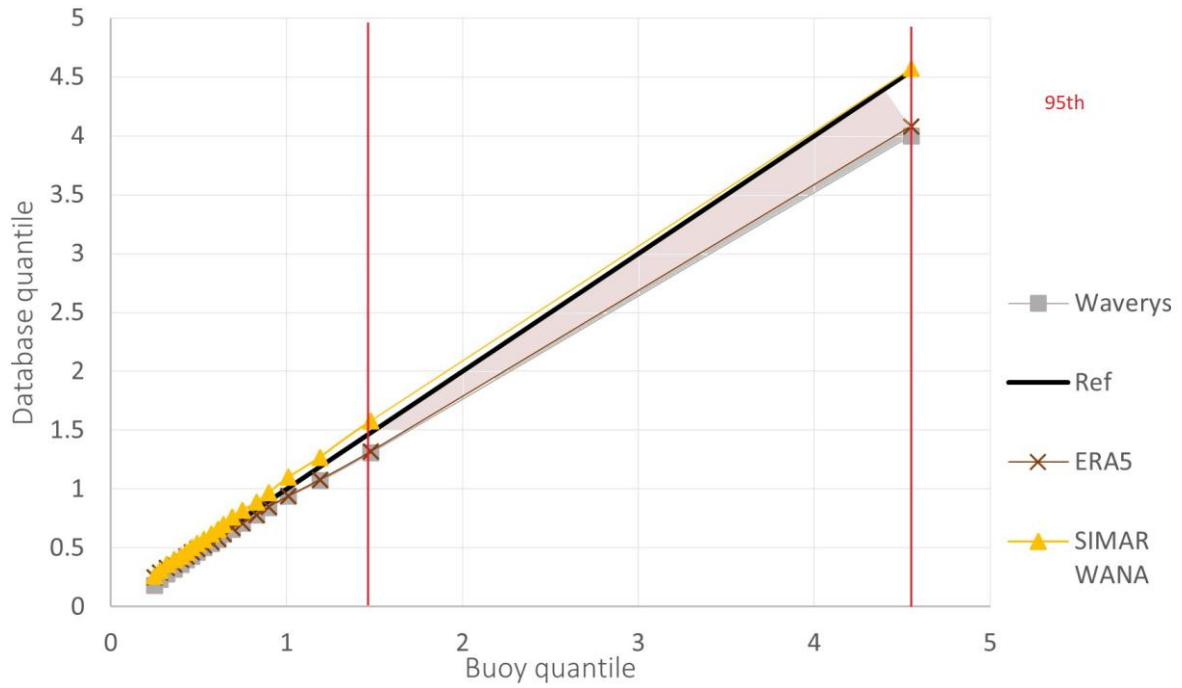


Fig.7. Quantile-Quantile plot of Hs in Alicante.

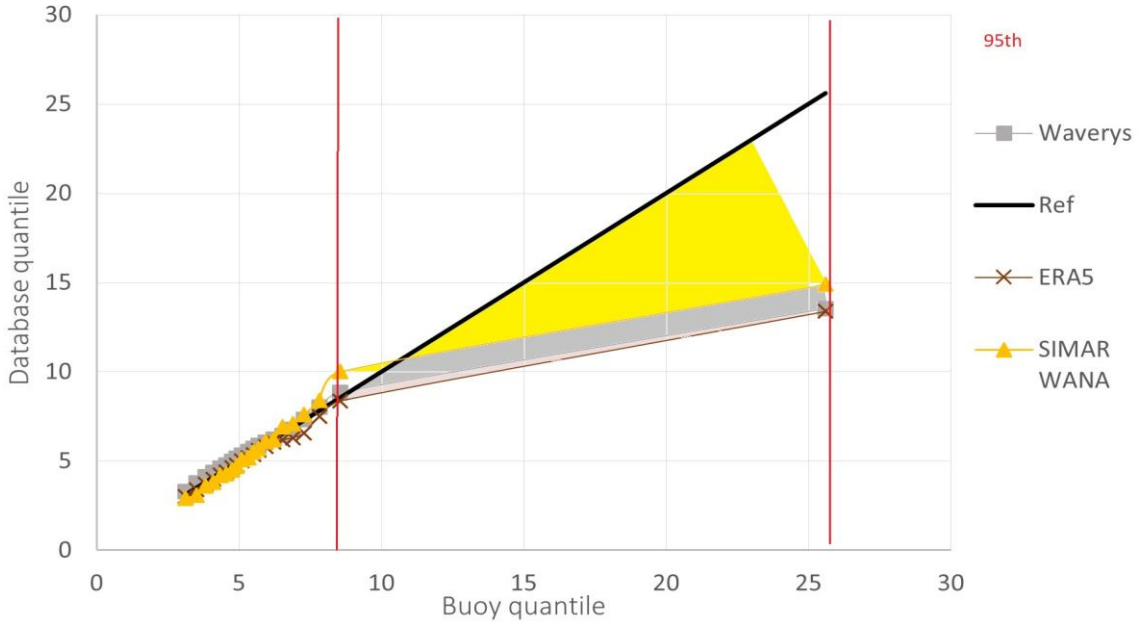


Fig.8. Quantile-Quantile plot of Tp in Alicante.

Taylor diagrams

As done with Valencia buoy, Figures 9 and 10 present the Taylor diagrams at Alicante for Hs and Tp, respectively. Waverys and ERA5 performance is very similar, especially in the case of Hs estimations. SIMAR-WANA differs a little from the other two databases. From these diagrams it could be said that, in general terms, Waverys and ERA5 are the databases that best reflect the measurements of the buoys.

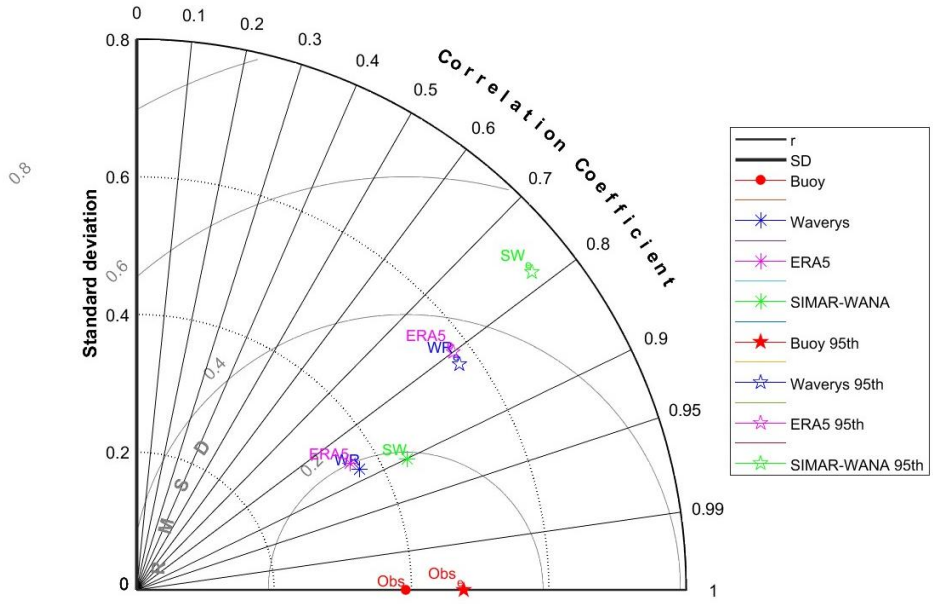


Fig.9. Taylor diagram for Hs in Alicante.

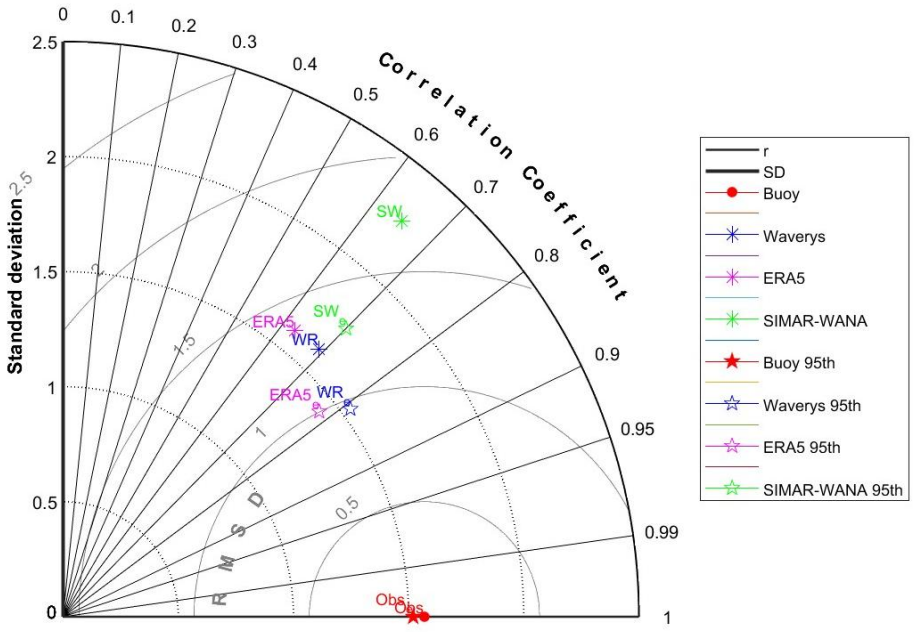


Fig.10. Taylor diagram for Tp in Alicante.

Cabo de Palos

Dataset

The time window with model-buoy paired data covers from July 2007 to 31 December 2020.

The values of the statistical analysis in Cabo de Palos are presented in Table 10. Similar to what was found with the Valencia buoy, Waverys shows the highest accuracy, but SIMAR-WANA has a bias closer to 0. Waverys correlation coefficients are the highest for the three variables studied (0.94, 0.73 and 0.79 for Hs, Tp and Dir) and it also has the lowest RMSE for Hs and Dir (0.27 m and 50.99º). In Cabo de Palos, Tp and Dir are estimated with lower accuracy than in Alicante and Valencia.

Table 10. Statistical analysis of Cabo de Palos buoy.

	Hs (m)			Tp (s)			Dir (º)		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	0.86	0.84	0.98	5.85	5.55	5.98	129.49	134.65	128.92
r	0.94	0.94	0.91	0.73	0.73	0.67	0.79	0.77	0.75
RMSE	0.27	0.29	0.26	1.21	1.15	1.53	50.99	53.86	59.59
Bias	-0.16	-0.18	-0.04	0.31	-0.02	0.44	-5.73	-0.57	-6.88

MAE Mean	0.19	0.21	0.18	0.81	0.75	1.05	25.21	26.36	30.37
MAE Max	2.18	2.53	3.28	11.44	13.74	13.47	357.53	358.10	359.24
MAPE Mean	0.20	0.19	0.20	0.16	0.14	0.20	0.57	0.57	0.57
MAPE Max	2.42	4.05	4.10	4.81	4.61	3.61	298.72	357.04	358.00

95th percentile

The results for the 95th Percentile are presented in Table 11. As in the previous cases, lower accuracy can be observed for the extreme values in Cabo de Palos for all datasets, but in this case, the decrease of accuracy is higher than in the other two buoys. SIMAR-WANA is the database that shows the least underestimation of the extreme values of Hs. (bias = -0.21 m) On contrary, Waverys and ERA5 bias reach values lower than - 0.6 m, where the 0.6 m error is more than 25% of the total wave height represented. The quantile-quantile plot of Hs and Tp are presented in Figure 11 and 12, respectively.

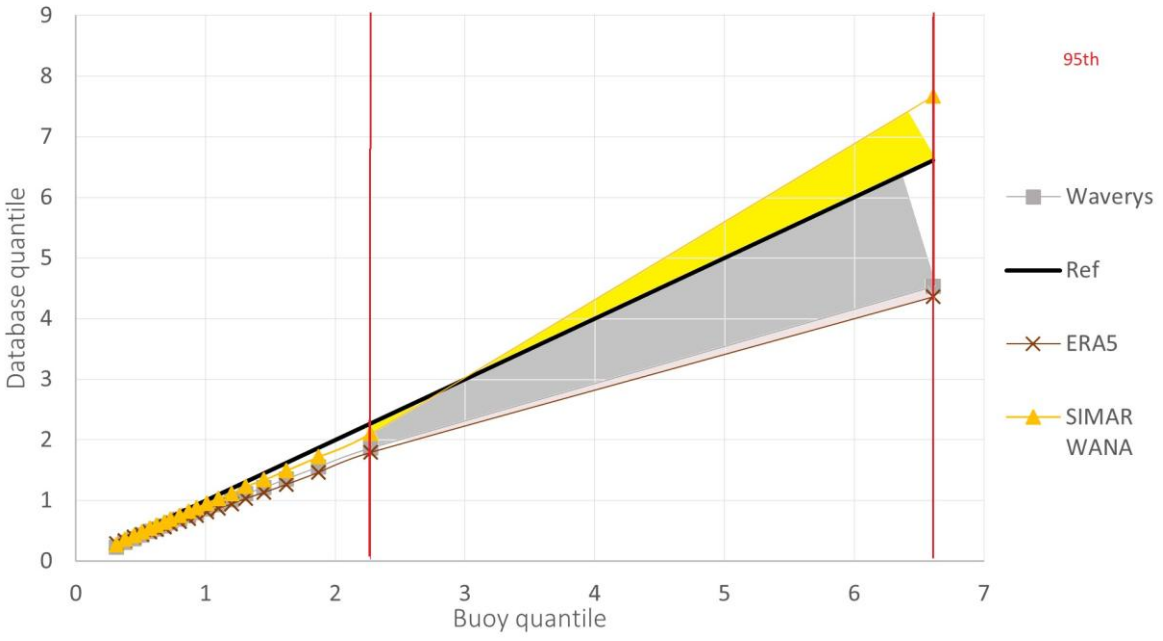


Fig.11. Quantile-Quantile plots of Hs in Cabo de Palos.

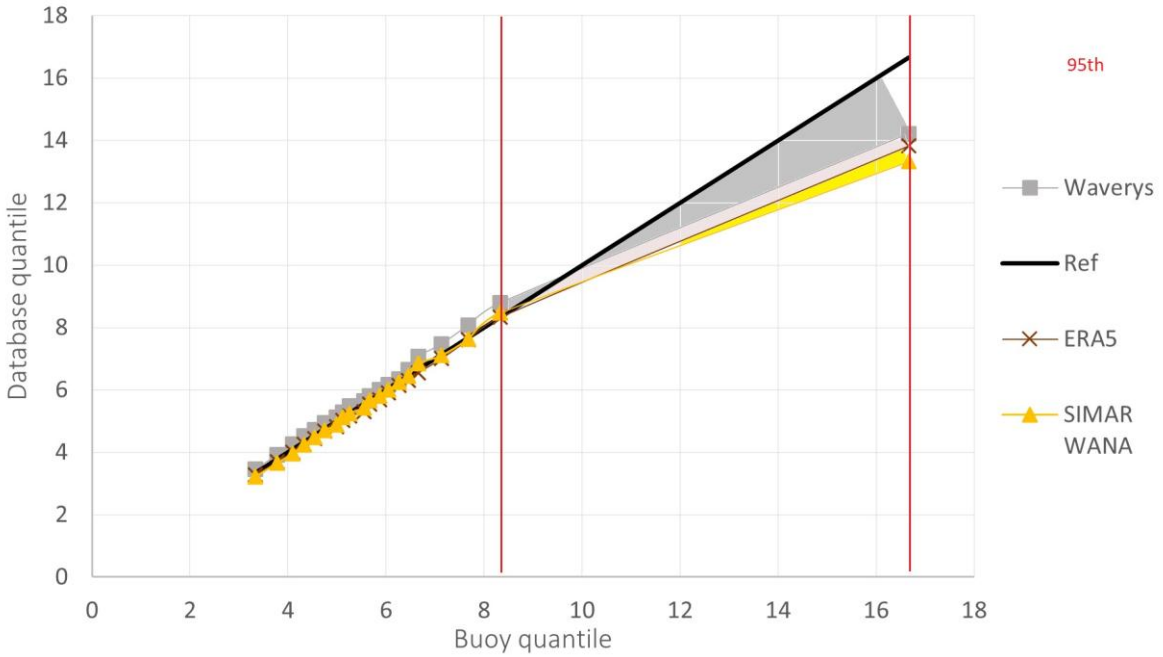


Fig.12. Quantile-Quantile plots of Tp in Cabo de Palos.

Table 11. Statistical analysis of the 95th Percentile of Cabo de Palos buoy

	Hs (P_95-2.27 m)			Tp (P_95_Hs_2.27 m)			Dir (P_95_Hs_2.27 m)		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	2.25	2.19	2.62	7.86	7.68	8.19	140.68	145.77	141.51
r	0.75	0.74	0.70	0.71	0.64	0.63	0.74	0.72	0.72
RMSE	0.70	0.74	0.59	0.89	0.95	1.78	35.70	64.14	67.47
Bias	-0.58	-0.63	-0.21	0.24	0.04	0.55	-0.96	2.47	-1.93
MAE Mean	0.60	0.65	0.46	0.65	0.66	0.98	10.89	33.99	36.94
MAE Max	2.18	2.53	3.28	4.95	5.71	6.57	348.95	344.17	352.00
MAPE Mean	0.21	0.23	0.16	0.09	0.09	0.13	0.18	0.56	0.47
MAPE Max	0.62	0.65	1.21	0.95	0.92	1.26	80.91	168.37	168.50

For Cabo de Palos location, the fit of the 95th percentile distribution of the databases concerning the variable Tp is better than for the rest of locations, and SIMAR-WANA slightly overestimates it (bias = 0.55 s).

Referring to the other statistical parameters studied, Waverys is the database that presents the highest correlation coefficient for the three variables studied (0.75,

0.71 and 0.74 for Hs, Tp and Dir) It also presents the lowest RMSE and MAE for Tp and Dir (0.89 and 0.65 s and 35.70 and 10.89 º). In the case of the variable Hs, SIMAR-WANA is the database with the lowest errors (RMSE of 0.59 m and MAE of 0.46 m).

Taylor diagrams

Figures 13 and 14 show the Taylor diagrams for Hs and Tp in Cabo de Palos buoy, respectively. The conclusions that can be drawn from these two figures are in line with those obtained for the other two buoys. ERA5 and Waverys perform very similar, especially for Hs and Waverys is the database that performs most accurately in almost all the cases, (highest r and lowest RMSE and MAE as explained before).

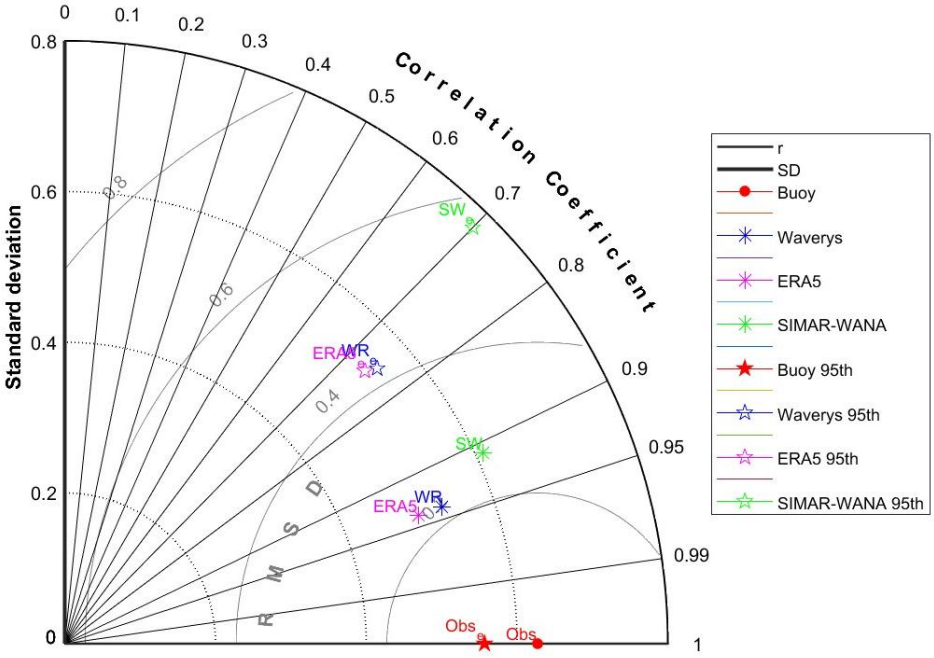


Fig.13. Taylor diagram for Hs in Cabo de Palos.

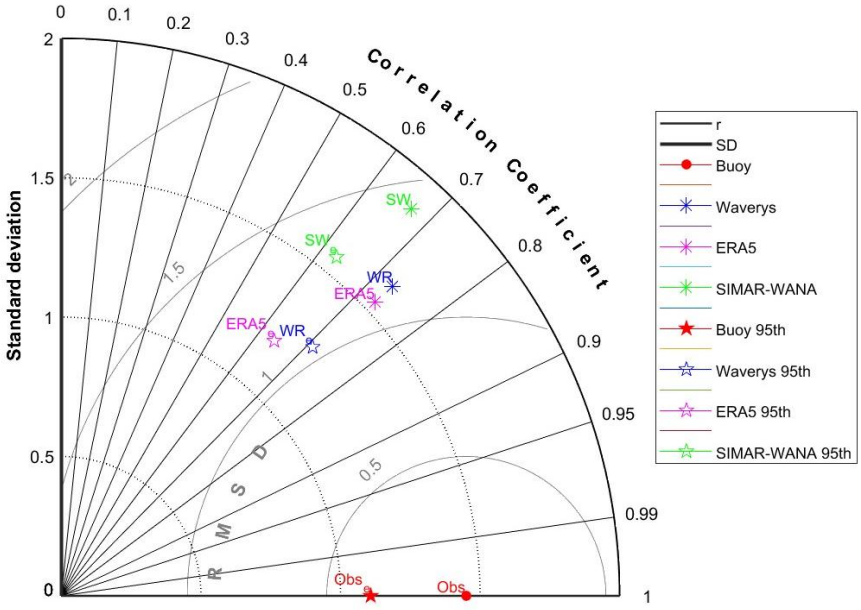


Fig.14. Taylor diagram for Tp in Cabo de Palos.

4.2. Results for the Spanish Mediterranean Sea

In this section the results presented in the previous one (Section. Results per buoy) are combined to obtain general conclusions about the performance of the three databases studied in the Spanish Mediterranean Sea.

<u>Dataset</u>

Values of the global statistical analysis for the Spanish Mediterranean are presented in Table 12. As could be deduced from the statistical analysis per buoy, the three databases show similar performance in the Spanish Mediterranean, but Waverys has slightly more precise statistical values and shows a better performance. It has the highest correlation coefficients for the three variables studied (0.93, 0.70 and 0.70 for Hs, Tp and Dir).and the lowest RMSE (0.25 m, 1.30 s and 52.71º). On the other hand, SIMAR-WANA is the database that shows the least underestimation of the value of Hs (bias = -0.03 m), which is also of great importance for the design of marine structures.

Table 12. Statistical analysis in the Spanish Mediterranean Sea,

	Hs			**Tp**			**Dir (º)**		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	0.71	0.72	0.87	5.69	5.37	5.74	132.79	130.33	123.53
r	0.93	0.92	0.87	0.70	0.68	0.63	0.75	0.74	0.70

RMSE	0.25	0.26	0.29	1.30	1.30	1.62	52.71	53.86	61.88
Bias	-0.14	-0.12	-0.03	0.32	0.00	0.43	-5.16	0.00	-6.30
MAE Mean	0.18	0.18	0.18	0.85	0.83	1.10	28.07	28.07	33.23
MAE Max	3.27	2.60	6.16	22.58	22.79	21.94	358.10	359.82	359.24
MAPE Mean	0.23	0.21	0.23	0.18	0.17	0.22	18.91	22.35	15.47
MAPE Max	2.42	6.90	8.54	4.81	4.61	3.61	354.69	358.63	358.00

95th percentile

The statistical analysis for the extreme distribution is presented in Table 13. As with the whole dataset, in the case of the extreme distribution, the three databases show similar performance. SIMAR-WANA presents a slightly lower correlation coefficient in the case of extreme Hs estimation (0.68 compared to 0.84 and 0.83 of Waverys and ERA5), but Puertos del Estado itself warns about this issue when analyzing extreme storms (*Puertos del Estado. Ministerio de Transportes, movilidad y agenda urbana. Conjunto de Datos SIMAR*, 2020). However, although SIMAR-WANA is the database with the lowest r is also the database with the bias closer to 0, which suggests that its use is adequate for the design of coastal protection structures. Considering other statistical parameters, it could be said that Waverys is the database with the best performance, with the highest r (0.84, 0.79 and 0.72 for Hs, Tp and Dir), lowest RMSE and MAE (0.61 m, 1.01 s, 34.43º and 0.51 m, 0.70 s and 10.49º) and lowest MAPE for Tp and Dir (0.10 and 0.18).

Table 13. Statistical analysis for the extreme distribution in the Spanish Mediterranean Sea,

	Hs (P_95-2.27 m)			**Tp (P_95_Hs_2.27 m)**			**Dir (P_95_Hs_2.27 m)**		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Mean	1.98	1.94	2.36	7.73	7.50	8.09	127.03	132.03	127.98
r	0.84	0.83	0.68	0.79	0.75	0.66	0.72	0.72	0.68
RMSE	0.61	0.64	0.70	1.01	1.12	2.06	34.43	61.26	66.70
Bias	-0.48	-0.51	-0.24	0.15	-0.07	0.49	-1.66	0.25	-4.62
MAE Mean	0.51	0.55	0.52	0.70	0.71	1.05	10.49	31.45	35.97
MAE Max	3.27	2.60	6.07	5.96	11.61	10.84	348.95	351.99	352.00
MAPE Mean	0.21	0.22	0.20	0.10	0.10	0.14	0.18	0.62	0.47

MAPE Max	0.89	1.05	3.28	1.19	1.45	1.72	80.91	288.79	168.50

4.3. Analysis of peak period

After analyzing the Tp associated to extreme storms (Hs 95th percentile) and the 95th percentile of Tp, it seems that the peak period in extreme storms needs further study and it will be discussed in the present section. Table 14 has three columns. The first column shows the 95th current percentile of the variables Tp based on historic records of the buoy, the second column shows the mean of the Tp associated to extreme storms (Hs 95th percentile), which was obtained calculating the average of the Tp associated to the 95th percentile of Hs. Finally, the third columns shows the Tp expected in 20 years estimated by Puertos del Estado in their annual reports (Informes anuales Oceanografía Puertos del Estado). As it can be seen in Table 14, the peak period estimated for 20 years is already present in current storms. In fact, in Cabo de Palos this Tp has been exceeded in more than 35% of the extreme storms and in Valencia in more than 18%, which shows that a study of these peak periods should be carried out to adjust them to the current storms. Figures 15 and 16 show how the Tp estimated in 20 years is exceeded in quite a number of extreme storms (95th Percentile Hs). It can be concluded that a statistical adjustment of the variable Tp is needed because current values do not represent the maritime climate precisely enough. The recommendation of the authors is to use the buoy measurements of Hs to adjust Tp and obtain the Tp for different return periods in each specific area.

Table 14. 95th percentile current Tp vs Tp associated to 95th Hs vs Tp expeted in 20 years.

Buoy	95th Percentile Tp	Tp associated to 95th Hs	Tp 20 years
Valencia	8.20 s	7.6 s	8.22 s
Alicante	8.56 s	7.6 s	10.27 s
Cabo de Palos	8.33 s	7.6 s	8.58 s

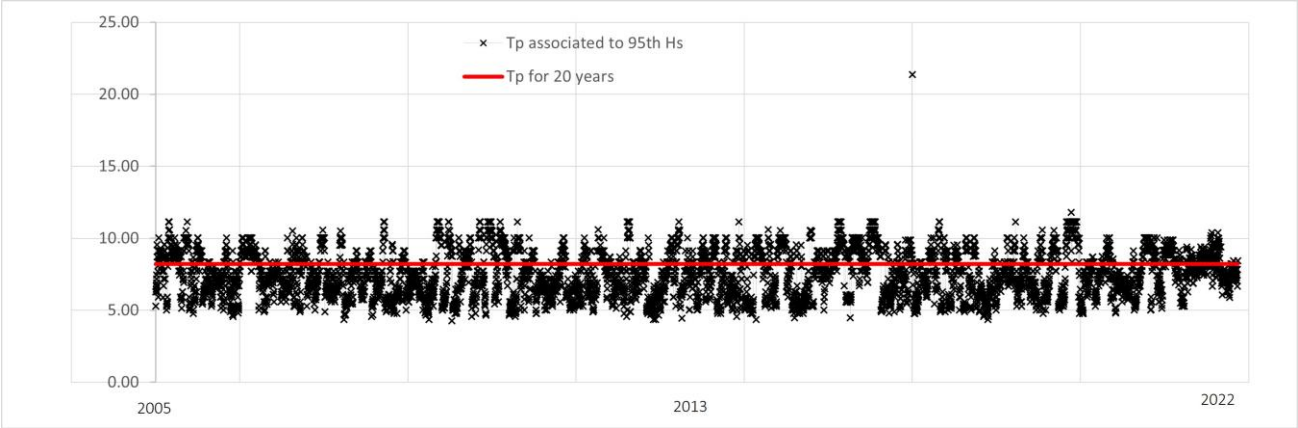


Fig.15. Tp associated to 95th Hs vs Tp expected in 20 years in Valencia.

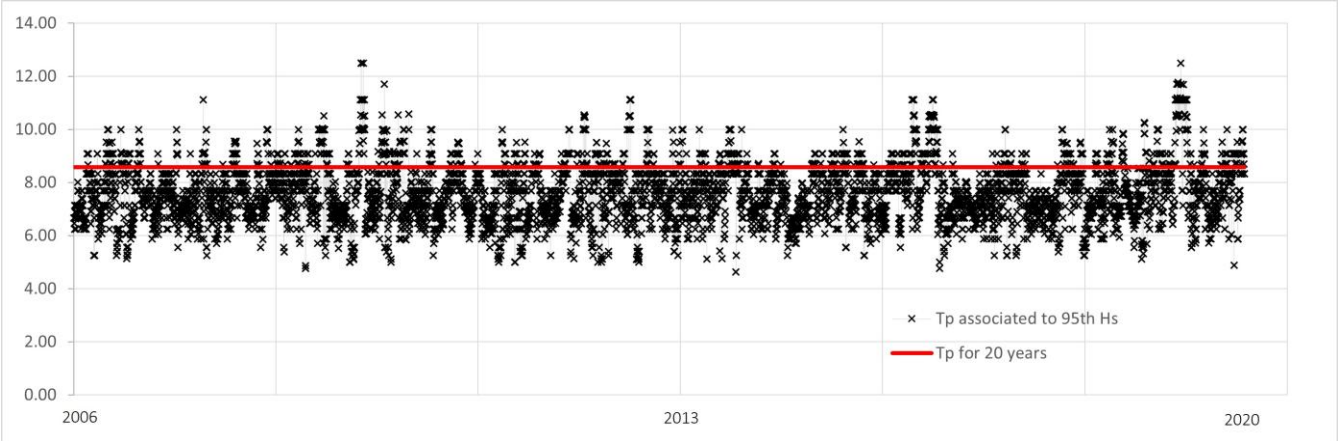


Fig.16. Tp associated to 95th Hs vs Tp expected in 20 years in Cabo de Palos.

5. Discussion

In this section, the results obtained in the statistical analysis of the three main databases of the Spanish Mediterranean Sea are going to be discussed in relation to the design and their use in reliability of coastal protection structures. In particular, the discussion will focus on the analysis of the 3 main typologies: rubble-mound and vertical breakwaters and coastal defense (groynes, detached breakwaters, reef breakwaters and submerged breakwaters). In general, it has been seen that the three databases have a good performance, with correlations ranging from 0.65 to 0.95, and RMSEs ranging from 0.30 m to 1.40 m for Hs and 1.20 s to 2.33 s for Tp. However, a statistical adjustment for the variable Tp is needed and the three databases, especially Waverys and ERA5, may underestimate the value of Hs, especially in the extreme distribution. In the present section it will be analyzed how errors in wave climate estimations may affect the design of coastal protection structures. For the analysis presented, the

values of the extreme distribution in the Spanish Mediterranean Sea, obtained from the study of the historical records of the buoys, are going to be used (Table 15).

Table 15. Values for the structural analysis.

	Hs			Tp		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
Threshold		1.9 m			7.6 s	
Bias	-0.48	-0.51	-0.24	0.15	-0.07	0.49

Rubble-mound breakwaters

The design of rubble-mound breakwaters is totally conditioned by the wave height. The wave height is the main variable for the design of the bank, the crown-height, and the design of the armour, which in turn influences the design of the filter and core. The expressions used for the design of the armour (Hudson, 1953; Iribarren, 1938; Losada and Giménez-Curto, 1982) depends on wave height cubed (Equation 7).

$$W_{50} = f(H^3) = k\gamma_w H^3 \quad (7)$$

Where W_{50} is the average weight of the armour, H the wave height and $k\gamma_w$ a constant that depends on the design approach used (Hudson, 1953; Iribarren, 1938; Losada and Giménez-Curto, 1982). The average armour weight was calculated for the 95th percentile threshold and for the 95th percentile threshold plus the bias. The results are presented in Table 16 and 17.

Table 16. Armour weight values.

	Hs			Hs+Bias		
	Waverys	ERA5	SIMAR	Waverys	ERA5	SIMAR
$\boldsymbol{W_{50}}$		6.86 γ_w		2.86 γ_w	2.69 γ_w	4.57 γ_w

Table 17. Errors in the design of the armour of rubble-mound breakwaters.

		Waverys	ERA5	SIMAR
$\boldsymbol{W_{50}}$	MAE	4.00 γ_w	4.17 γ_w	2.29 γ_w
	MAPE	58.3%	60.8%	33.4%

A calculation of the weight of the armour of a rubble-mound breakwater has been done considering the mean bias of the extreme long-term distribution of the three

databases analyzed. As can be seen from Table 17, assuming that the Hs and Tp errors are propagated to shallow waters, these errors can lead to design errors up to 60%, which is a structural safety hazard. As the weight of the main layer units of a rubble mound breakwater is a function of the design wave height third power (H_D^3) (Hudson, 1953; Van der Meer, 1988), the variability of the action parameter produces a very pronounced alteration in weight with the sensitivity of the structure in its stability to incident wave actions.

On the other hand, the energy absorbed, reflected, and transmitted by a rubble-mound breakwater is sensitive to the wave period, since the reflection coefficient is a function of the square of the Iribarren number (Battjes, 1974), (Equation 8) or the quotient of Iribarren numbers (Allsop – Hetiarachchi (Allsop and Hettiarachchi, 1989), (Equation 9) , and this in turn is function of the wave period (Equation 10 and 11).

$$Reflection\ Coefficient = K_r = 0.1\xi^2 \quad (8)$$

$$K_r = \frac{a\xi^2}{b+\xi^2} \quad (9)$$

$$\xi = 1.25\, T \frac{\tan\alpha}{\sqrt{H}} \quad (10)$$

$$K_r = f(T^2) \quad (11)$$

Where ξ is the Iribarren number, a and b are parameters taken from (Allsop and Hettiarachchi, 1989), T the period, and $tan\alpha$ the slope of the land.

The energy reflected by the breakwater has been analyzed in Table 18.

Table 18. Errors in the reflection coefficient of rubble-mound breakwaters.

		Waverys	ERA5	SIMAR
K_r	MAE	$11.9c$	$10.7c$	$9c$
	MAPE	39%	35%	29.6%

In this case, the errors that can be incurred in the reflection coefficient of the breakwater are not as significant as in the design of the armour, but still reach up to 39% in the case of the Waverys database, which can also cause problems in the safety of the structure.

For the same reason, variations in wave heights produce different run-up and overtopping behaviour, depending on the Iribarren number, the wave height and therefore the crest height of the breakwater (Equation 12 and 13 and Table 19).

$$\frac{Ru}{Hs} = f(\xi) \quad (12)$$

$$\frac{q}{\sqrt{gHs^3}} = 8 \times 10^{-5} \exp\left(3.10\left(\frac{Ru_{\xi 0}^{2\%} - Rc}{Hs}\right)\right) \quad (13)$$

Where Ru is the run-up, Hs the significant wave height, g the gravity, $Ru_{\xi 0}^{2\%}$ the run-up for a likelihood of overrun of 2%, Rc the freeboard and q the overtopping. Based on Table 19, errors in run-up and overtopping can reach 25% and 37% respectively (considering only the influence of the wave height), which could endanger the structural safety of the rubble-mound breakwater.

Table 19. Errors in run-up and overtopping.

		Waverys	ERA5	SIMAR
Ru	MAE	$0.48c$	$0.51c$	$0.24c$
	MAPE	25.3%	26.8%	12.6%
q	MAE	$0.93c$	$0.98c$	$0.48c$
	MAPE	35.5%	37.4%	18.3%

<u>Vertical breakwaters</u>

In the case of the design of vertical breakwaters, the wave height also influences the crown height, wide and pressures on the breakwater. A scheme of the pressures in a vertical breakwater is presented in Figure 17 and the analytical expression of each pressure from Equation 14 to 17.

$$P_1 = f(H) = (\alpha_1 + \alpha_2)\rho_w g H \quad (14)$$

$$P_2 = f(P_1) = \frac{P_1}{ch\ 2\pi\frac{h}{L}} \quad (15)$$

$$P_3 = f(H) = \alpha_1 \alpha_3 \rho_w g H \quad (16)$$

$$P_4 = f(P_1) = \frac{\eta^* - hP_1}{\eta^*} \quad (17)$$

Where α_1, α_2 and α_3 are auxiliar parameters that depend on the depth of the breakwater, wavelength, and design wave height, ρ_w is the water density, h is the depth of the breakwater and η the surface water elevation. All pressures depend on wave height. Pressure is calculated for the threshold and for the threshold (Hs= 1.9 m and Tp=7.6 s) accounting the bias of each database. The results are presented in Table 20. In the case of vertical breakwaters, the errors incurred are smaller than in the case of rubble-mound breakwaters, as the power of wave height is

smaller, 1 compared to 3 in the case of rubble-mound. However, the errors can still reach 27%, which is high enough to jeopardise structural safety.

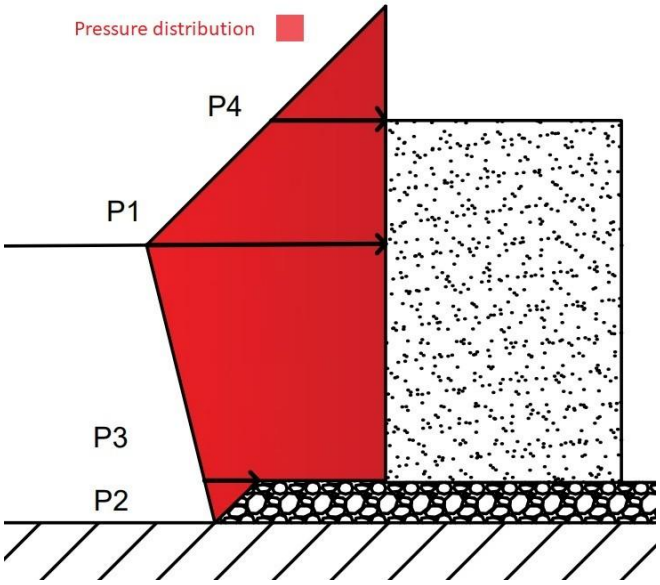


Fig.17. Pressure scheme in a vertical breakwater.

Table 20. Errors in the calculation of pressure in a vertical breakwater.

		Waverys	ERA5	SIMAR
P_i	MAE	$0.48c$	$0.51c$	$0.24c$
	MAPE	25.3%	26.8%	12.6%

Coastal defense

The design of coastal defense structures depends on both, wave height (H) and wavelength (L). On the other hand, using the dispersion relation from the linear wave theory, the wavelength is always related to the local depth., at the same time the wave celerity C=L/T. (Equation 18 and 19).

$$W = f(H^2 L) \quad (18)$$

$$L = f(T) \quad (19)$$

Table 21 presents the results of the structural analysis for coastal defense, applying the expression derived from Equation 18 and 19 ($W = c * H^2 T$) Coastal defense structures are in an intermediate situation between vertical and rubble-mound breakwaters, as the wave height is of order 2. The errors can reach up to 46%, which again endangers the safety of the structures. Table 22 shows the results considering the wavelength influence.

Table 21. Errors in the structural calculation of coastal defense.

		Waverys	ERA5	SIMAR
W	MAE	$1.59c$	$1.68c$	$0.85c$
	MAPE	44.0%	46.5%	23.5%

Table 22. Errors in the structural calculation of coastal defense considering wavelength.

		Waverys	ERA5	SIMAR
W	MAE	$11.8c$	$12.89c$	$5.14c$
	MAPE	43.0%	53.0%	18.7%

6. Conclusions

The aim of this paper was to evaluate the performance of the three main databases (Waverys, ERA5 and SIMAR-WANA) in the Spanish Mediterranean Sea for the design of marine structures. After carrying a detailed statistical study, it has been found that Waverys provides the highest quality estimates in Hs and Tp. It has also been illustrated in the Taylor diagrams presented that the Waverys database is the closest to the wave measurements of the buoys, and therefore, in principle, the one that is most accurate. On the other hand, SIMAR-WANA is the database that shows the least underestimation, which is of particular interest for the safety of maritime structures.

Waverys database uses altimetry and satellite data to improve the numerical model. The main difference with the ERA5 database is that it also considers wave-current interaction, which helps to increase the accuracy of the dataset (Alday, Matias et al., 2021; Marechal and Ardhuin, 2020) and that its spatial resolution is slightly higher than that of ERA5. Therefore, it can be said that considering wave-current interaction can improve the performance of the models in the Western Mediterranean. However, the most important element is the higher resolution of Waverys, which allows to partially improve the refraction effects due to a better representation of the bathymetric features. This is of special interest on those areas similar to the Spanish Mediterranean Sea, close basins governed by local atmospheric phenomena (mainly local winds) and wave patterns. In such areas global numerical models lose accuracy and their performance worsens. Therefore, assimilating satellite data and higher resolution models can considerably improve the performance of the models.

Regarding marine structures and extreme events, it seems that the estimated values given by Waverys, ERA5 and SIMAR-WANA are usually underestimated for extreme storms. The one that performs best in this regard is SIMAR-WANA database. The fact that extreme events are underestimated can have serious consequences for the design of coastal protection structures, with errors of up to 60% in their safety design. The marine structures most affected are rubble-mound breakwaters where the wave height is elevated to the cube and the errors in wave height estimations become more significant.

The authors of this article consider that Waverys database gives the most accurate results in the Spanish Mediterranean area, but that if extreme event data are to be used, they should first be validated with another data source. Therefore, whenever possible, efforts should be made to use wave measurements to predict extreme maritime weather. In addition, in the case of predicting extreme events, special attention is needed, especially in the design of those marine structures where the wave height is of order greater than 1, as is the case of rubble-mound breakwaters. The authors of the article recommend using different databases and methods, instead of only one source of data (no matter which one) if extremal values are going to be used, since, as has been shown, errors in structural stability of up to 60% can be achieved, leading to very high failure risks.

The main shortcoming of this study is that it has been assumed that errors in boundary conditions of numerical models are propagated to final results. Therefore, future research lines could study more in deep how these errors in boundary conditions affect final results.

References

Agarwal, A.;Venugopal, V. y Harrison, G. P. (2013). The assessment of extreme wave analysis methods applied to potential marine energy sites using numerical model data. *Renewable and Sustainable Energy Reviews, 27*, 244-257. doi:https://doi.org/10.1016/j.rser.2013.06.049

Akpinar, A. y Ponce de León, S. (2016). An assessment of the wind re-analyses in the modelling of an extreme sea state in the Black Sea. *Dynamics of Atmospheres and Oceans, 73*, 61-75. doi:https://doi.org/10.1016/j.dynatmoce.2015.12.002

Alday, M.;Accensi, M.;Ardhuin, F. y Dodet, G. (2021). A global wave parameter database for geophysical applications. Part 3: Improved forcing and spectral resolution. *Ocean Modelling, 166*, 101848. doi:https://doi.org/10.1016/j.ocemod.2021.101848

Alday, M.;Ardhuin, F.;Dodet, G. y Accensi, M. (2022). Accuracy of numerical wave model results: application to the Atlantic coasts of Europe. *Ocean Sci., 18*(6), 1665-1689. doi:10.5194/os-18-1665-2022

Allsop, N. W. H. y Hettiarachchi, S. S. L. (1989). Wave reflections in harbours; the design, construction and performance of wave absorbing structures. *Hydraulics Research, Wallingford.* .

Aouf, L.;Dalphinet, A.;Hauser, D.;Delaye, L.;Tison, C.;Chapron, B., . . . Tourain, C. (2019, 28 July-2 Aug. 2019). *On the Assimilation of CFOSAT Wave Data in the Wave Model MFWAM : Verification Phase.* Paper presented at the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium.

Appendini, C.;Torres-Freyermuth, A.;Oropeza, F.;Salles, P.;López, J. y Mendoza, E. (2013). Wave modeling performance in the Gulf of Mexico and Western Caribbean: Wind reanalyses assessment. *Applied Ocean Research, 39*, 20-30. doi:10.1016/j.apor.2012.09.004

Ardhuin, F.;Bertotti, L.;Bidlot, J.-R.;Cavaleri, L.;Filipetto, V.;Lefevre, J.-M. y Wittmann, P. (2007). Comparison of wind and wave measurements and models in the Western Mediterranean Sea. *Ocean Engineering, 34*(3-4), 526-541. doi:10.1016/j.oceaneng.2006.02.008

Aydogan, B. y Ayat, B. (2021). Performance evaluation of SWAN ST6 physics forced by ERA5 wind fields for wave prediction in an enclosed basin. *Ocean Engineering, 240*. doi:10.1016/j.oceaneng.2021.109936

Battjes, J. A. (1974). SURF SIMILARITY. *Coastal Engineering Proceedings, 1*(14), 26. doi:10.9753/icce.v14.26

Bengtsson, L.;Andrae, U.;Aspelien, T.;Batrak, Y.;Calvo, J.;de Rooy, W., . . . Køltzow, M. Ø. (2017). The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System. *Monthly Weather Review, 145*(5), 1919-1935. doi:10.1175/mwr-d-16-0417.1

Bhatt, V.;Sarkar, A.;Kumar, R.;Basu, S. y Agarwal, V. K. (2004). Impact of Oceansat-I MSMR data on analyzed oceanic winds and wave predictions. *Ocean Engineering, 31*(17-18), 2283-2294. doi:10.1016/j.oceaneng.2004.03.011

Bjorkqvist, J. V.;Vaha-Piikkio, O.;Alari, V.;Kuznetsova, A. y Tuomi, L. (2020). WAM, SWAN and WAVEWATCH III in the Finnish archipelago - the effect of spectral performance on bulk wave parameters. *Journal of Operational Oceanography, 13*(1), 55-70. doi:10.1080/1755876x.2019.1633236

Booij, N.;Ris, R. C. y Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions 1. Model description and validation. *Journal of Geophysical Research: Oceans, 104*(C4), 7649-7666. doi:10.1029/98JC02622

Bosssing Christensen, O.;Drews, M.;Hesselbjerg Christensen, J.;Dethof, K.;Ketelsen, K.;Hebestadt, I. y Rinke, A. (2007). *The HIRHAM Regional Climate Model. Version 5 (beta).* Retrieved from Denmark. Danish Meteorological Institute. : http://www.dmi.dk/dmi/tr06-17

Boudière, E.;Maisondieu, C.;Ardhuin, F.;Accensi, M.;Pineau-Guillou, L. y Lepesqueur, J. (2013). A suitable metocean hindcast database for the design of Marine energy converters. *International Journal of Marine Energy, 3-4*, e40-e52. doi:https://doi.org/10.1016/j.ijome.2013.11.010

Bromwich, D. H. y Wang, S.-H. (2008). A review of the temporal and spatial variability of Arctic and Antarctic atmospheric circulation based upon ERA-40. *Dynamics of Atmospheres and Oceans, 44*(3), 213-243. doi:https://doi.org/10.1016/j.dynatmoce.2007.09.001

Campos, R. M. y Guedes Soares, C. (2016). Comparison of HIPOCAS and ERA wind and wave reanalyses in the North Atlantic Ocean. *Ocean Engineering, 112*, 320-334. doi:10.1016/j.oceaneng.2015.12.028

Catterson, V. M.;McMillan, D.;Dinwoodie, I.;Revie, M.;Dowell, J.;Quigley, J. y Wilson, K. (2016). An economic impact metric for evaluating wave height forecasters for

offshore wind maintenance access. *Wind Energy, 19*(2), 199-212. doi:10.1002/we.1826

Cavaleri, L. y Bertotti, L. (2004). Accuracy of the modelled wind and wave fields in enclosed seas. *Tellus A, 56*(2), 167-175. doi:https://doi.org/10.1111/j.1600-0870.2004.00042.x

Ciardini, V.;Contessa, G. M.;Falsaperla, R.;Gómez-Amo, J. L.;Meloni, D.;Monteleone, F., . . . di Sarra, A. (2016). Global and Mediterranean climate change: a short summary. *Ann Ist Super Sanita, 52*(3), 325-337. doi:10.4415/ann_16_03_04

CMEMS. (2021). Global Ocean Waves Reanalysis WAVERYS. doi:10.48670/moi-00022

Copernicus. *Wave Reanalysis Dataset*. Retrieved from: https://www.copernicus.eu/en/access-data

ECMWF. *ERA 5*. Retrieved from: https://ecmwf-models.readthedocs.io/en/latest/download.html#:~:text=ERA5%20%28and%20ERA5-Land%29%20data%20can%20be%20downloaded%20manually,at%20the%20CDS%20and%20setup%20the%20CDS%20key.

ECMWF. (2018). ECMWF Model Documentation. Retrieved from https://confluence.ecmwf.int/display/CKB/ECMWF+Model+Documentation Acessed on 13/09/2022

Eldeberky, Y.;Metwally, A.;Rakha, K. y Cavaleri, L. (2002). *Wave Hindcast in the Eastern Mediterranean Sea.* Paper presented at the ASME 2002 21st International Conference on Offshore Mechanics and Arctic Engineering.

Elkut, A. E.;Taha, M. T.;Abu Zed, A. B. E.;Eid, F. M. y Abdallah, M. A. (2021). Wind-wave hindcast using modified ECMWF ERA-Interim wind field in the Mediterranean Sea. *Estuarine Coastal and Shelf Science, 252*. doi:10.1016/j.ecss.2021.107267

Estado, P. d. (2020). Conjunto de datos SIMAR-WANA. Retrieved from https://bancodatos.puertos.es/BD/informes/INT_8.pdf Acessed on

Hersbach, H.;Bell, B.;Berrisford, P.;Hirahara, S.;Horanyi, A.;Munoz-Sabater, J., . . . Thepaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999-2049. doi:10.1002/qj.3803

Hopstad, A. L. H.;Argyriadis, K.;Manjock, A.;Goldsmith, J. y Ronold, K. O. (2018). *DNV GL Standard for Floating Wind Turbines.* Paper presented at the ASME 2018 1st International Offshore Wind Technical Conference.

Hudson, R. Y. J., R.A. (1953). *Stability of Rubble-Mound Breakwaters*. Vicksburg, MS, USA: Waterways Experiment Station, CERC

Hunter, J. D. (2003). Matlab (Version 2021b). Retrieved from https://login.mathworks.com/embedded-login/landing.html?cid=getmatlab&s_tid=gn_getml

IHC. (2013). Análisis Matemático y Estadístico de Variables Medioambientales (Version v.1.3.2). Retrieved from https://ihcantabria.com/software-y-servicios-tic/software/

Informes anuales Oceanografía Puertos del Estado. Retrieved from https://www.puertos.es/es-es

Iribarren, R. (1938). *Una Fórmula Para El Cálculo de Los Diques De Escollera*. Pasajes, Spain: M. Bermejillo Usabiaga.

Jean-Michel, L.;Eric, G.;Romain, B.-B.;Garric, G.;Angélique, M.;Drevillon, M., . . . Traon, P.-Y. (2021). The Copernicus Global 1/12° Oceanic and Sea Ice GLORYS12 Reanalysis. *Frontiers in Earth Science, 9*, 698876. doi:10.3389/feart.2021.698876

Law-Chune, S.;Aouf, L.;Dalphinet, A.;Levier, B.;Drillet, Y. y Drévillon, M. (2021). WAVERYS: a CMEMS global wave reanalysis during the altimetry period. *Ocean Dynamics, 71*, 357 - 378.

Losada, M. A. y Giménez-Curto, L. A. (1982). Mound breakwaters under oblique wave attack; a working hypothesis. *Coastal Engineering, 6*(1), 83-92. doi:https://doi.org/10.1016/0378-3839(82)90017-5

Marechal, G. y Ardhuin, F. (2020). *Surface currents and significant wave height gradients: matching numerical models and high-resolution altimeter wave heights in the Agulhas current region*.

Morcrette, J. J.;Barker, H. W.;Cole, J. N. S.;Iacono, M. y Pincus, R. (2008). Impact of a New Radiation Package, McRad, in the ECMWF Integrated Forecasting System. *Monthly Weather Review, 136*, 4773-4798.

NCEP. North American Regional Reanalysis. Retrieved from https://www.ncei.noaa.gov/products/weather-climate-models/north-american-regional#:~:text=The%20North%20American%20Regional%20Reanalysis%20%28NARR%29%20is%20a,wind%2C%20moisture%2C%20soil%2C%20and%20dozens%20of%20other%20parameters. Acessed on 14/09/2022

Ponce de León, S. y Guedes Soares, C. (2015). Hindcast of the Hércules winter storm in the North Atlantic. *Natural Hazards, 78*(3), 1883-1897. doi:10.1007/s11069-015-1806-7

Ponce de Leon, S. y Soares, C. G. (2008). Sensitivity of wave model predictions to wind fields in the Western Mediterranean sea. *Coastal Engineering, 55*(11), 920-929. doi:10.1016/j.coastaleng.2008.02.023

Portillo Juan, N.;Olalde Rodríguez, J.;Negro Valdecantos, V. y Iglesias, G. (2023). Data-driven and physics-based approach for wave downscaling: A comparative study. *Ocean Engineering, 285*, 115380. doi:https://doi.org/10.1016/j.oceaneng.2023.115380

Puertos del Estado. Ministerio de Transportes, movilidad y agenda urbana. Conjunto de Datos SIMAR. (2020). Retrieved from https://bancodatos.puertos.es/BD/informes/INT_8.pdf

Puertos del Estado. Oceanografía. Registros históricos de oleaje. Retrieved from https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx

Rossum, G. V. (1989). Python (Version 3.10). Retrieved from https://www.python.org/downloads/windows/

Samuelsson, P.;Willén, U.;Kjellström, E. y Jones, C. (2009). *Applying the Rossby Centre Regional Climate Model (RCA3.5) over the ENSEMBLES-AMMA region: Sensitivity studies and future scenarios*.

Sebastiao, P.;Soares, C. G. y Booij, N. (2000). Wave hindcasting off the coast of Portugal. *Coastal Engineering, 40*(4), 411-425. doi:10.1016/s0378-3839(00)00021-1

Sharmar, V. D.;Markina, M. Y. y Gulev, S. K. (2021). Global Ocean Wind-Wave Model Hindcasts Forced by Different Reanalyzes: A Comparative Assessment. *Journal of Geophysical Research: Oceans, 126*(1), e2020JC016710. doi:https://doi.org/10.1029/2020JC016710

Stephens, S. A. y Gorman, R. M. (2006). Extreme wave predictions around New Zealand from hindcast data. *New Zealand Journal of Marine and Freshwater Research, 40*(3), 399-411. doi:10.1080/00288330.2006.9517431

Stopa, J. E. y Cheung, K. F. (2014). Intercomparison of wind and wave data from the ECMWF Reanalysis Interim and the NCEP Climate Forecast System Reanalysis. *Ocean Modelling, 75*, 65-83. doi:https://doi.org/10.1016/j.ocemod.2013.12.006

Umesh, P. A. y Swain, J. (2018). Inter-comparisons of SWAN hindcasts using boundary conditions from WAM and WWIII for northwest and northeast coasts of India. *Ocean Engineering, 156*, 523-549. doi:10.1016/j.oceaneng.2018.03.029

Van der Meer, J. W. (1988). *Rock slopes and gravel beaches under wave attack*. TU Delft and Deltares.

3.8. Data-driven and physics-based approach for wave downscaling: A comparative study

Abstract and Keywords

The physics-based approach is the most common framework for modelling ocean engineering processes; however, in the last years data-driven models are becoming more and more popular. In the case of wave downscaling, the number of studies involving data-driven models is scarce. In this paper, both approaches are developed and compared on a case study in the Mediterranean Sea. Two Feed Forward Multilayer Perceptron Neural Networks are trained with the Levenberg-Marquardt algorithm and compared with a SWAN model. The behaviour of both models turns out to be similar with respect to the wave height. However, with respect to the peak period, neural networks outperform the SWAN model, which slightly overestimates the peak period. Neural networks show a correlation coefficient and error of 0.96 and 1.00 s, respectively, versus 0.72 and 6.2 s for the SWAN model. Regarding the computational cost, the data-driven approach clearly outperforms the physics-based approach. Therefore, it may be concluded that at those sites where a minimum of 25,000 data samples are available, data-driven models are of interest – not least in semi-enclosed basins governed by local patterns, where large-scale or global physics-based models typically struggle.

Keywords: Data-driven approach; Physics-based approach; Wave downscaling; Artificial Neural Networks; SWAN; Mediterranean Sea

1. Introduction

Predicting maritime climate accurately is essential for ocean engineering. Conventionally, maritime climate is derived from offshore buoy records and then it is propagated to the area of study (Romano-Moreno et al., 2022). Coastal buoy records are normally affected by refraction and shoaling phenomena. This is the main reason why the general practice is to take records from deep-water buoys and propagate them to the study site. This process is known as wave downscaling, which can be defined as the method used to obtain wave climate at local scales (higher resolution) from data at global scales (lower resolution). Downscaling is an everyday part of maritime and ocean engineering and an essential step in the development of any project – hence the importance of optimizing this process.

There are two main approaches to downscaling: dynamical and statistical downscaling. Dynamical downscaling uses physically-based Regional Climate Models (RCMs), with boundary conditions provided by a Global Circulation Model (GCM) (Zhang et al., 2019) and applies numerical models. By contrast, statistical downscaling relates climate variables at global scales with climate variables at local scales by mathematical algorithms (Romano-Moreno et al., 2022).

Traditionally, wave downscaling has been undertaken by means of numerical models (dynamical downscaling, i.e., the physics-based approach). When the maritime climate is propagated from deep water to the site of interest with numerical models, two phases can be distinguished. In the first, the propagation from deep water to coastal waters is calculated with models based on the energy balance equation, which allow large areas to be covered, but with relatively high error levels. The second phase of climate propagation is done at a local level with more sophisticated models (phase-resolving models), but also with a much higher computational cost (Section 1.1. Numerical models).

The first phase of the numerical modelling has its own significant computational cost, but its time and space resolution are still limited. Therefore, this is a phase of maritime climate propagation modelling that can be optimized. With the advent of artificial intelligence (AI), data-driven models are gaining visibility and popularity. The question arises as to which approach is better in terms of both accuracy and computational cost (Boukabara and Hoffmanb, 2022; Boukabara et al., 2019; Gao et al., 2023; Huang et al., 2022; Minuzzi and Farina, 2023). Given the characteristics of wave downscaling, data-driven models may constitute a good alternative to energy balance equation models. Within data-driven models, Artificial Neural Networks (ANNs) are the most widespread in ocean engineering and, therefore, are selected for this work as the data-driven model. More details about physics-based (numerical) models and data-driven models are provided below.

1.1. Numerical models

Physics-based models have historically been used to solve ocean engineering problems and are one of the main methodologies used to downscale waves. Numerical models used to propagate waves from deep to shallow water can be classified into two main groups: Phase-averaged models, which are based on the energy balance equation, and phase-resolving models, which are based on vertically integrated, time-dependent mass and momentum balance equations

(Camus et al., 2011). Phase-resolving models implement Boussinesq-type equations and are defined with respect to the elementary equations of waves, with a highly precise approximation to the physics of the problem. They provide an excellence performance in terms of accuracy but at a high computational cost, which limits their application to local nearshore studies, typically less than 10 km (Liu and Losada, 2002), and short-term studies, which usually means less than 5 years (Umesh and Behera, 2021).

On the other hand, phase-averaged models are based on the energy balance equation, which defines the wave spectrum evolution by the advection and refraction phenomena, and by a source function "S" that aggregates other physical processes that affect the wave spectrum (Janssen, 2008). Some examples of these models are WaveWatch III (WWIII) (Tolman, 1991), WAM (Komen, Gerbrand J et al., 1996) or SWAN (Booij et al., 1999).

Energy-balance (phase-averaged) models solve the energy balance equation considering the refraction and advection phenomena, and aggregating other physical processes that affect wave propagation in the term "S". The term "S" usually includes nonlinear wind input, wave interactions and wave dissipation by bottom friction, white-capping and breaking (Umesh and Behera, 2020). This approach has less computational cost than phase-resolving models, which allows it to be applied to larger areas, but it is also less accurate and has many uncertainties which limit its application, as it cannot cover all spatial and time scales. Energy-balance equation models are highly dependent on the quality of wind data. In addition, there is a great deal of uncertainty associated to the term "S", with many unknows regarding the modelling of bottom friction and white-capping, which can increase modelling errors (Wang, N. et al., 2022). Different formulations for these processes have been implemented to better model source term packages and obtain the most accurate and computational efficient results (Kazeminezhad and Siadatmousavi, 2017). In spite of these efforts, the uncertainties have not been fully resolved yet. Moreover, although they are not as expensive as phase-resolving models, phase-averaged models still have a significant computational cost.

Of these models, the one most commonly used in nearshore processes and wave downscaling is SWAN; therefore, it will be used in this work. It is especially suited for coastal scales and nearshore locations because of its implicit way of solving the energy balance equation (Bingolbali et al., 2019). It simulates the evolution of the wave spectrum following Equation 1 (Salah et al., 2016).

$$\frac{\partial}{\partial t}N + \frac{\partial}{\partial x}c_x N + \frac{\partial}{\partial y}c_y N + \frac{\partial}{\partial \sigma}c_\sigma N + \frac{\partial}{\partial \theta}c_\theta N = \frac{S}{\sigma}, \quad (1)$$

Where N is wave action density (equal to energy density divided by relative frequency); σ is the wave frequency; t is time; θ is wave direction; c_x, c_y, c_σ and c_θ are the wave group velocities in x, y, σ, and θ space, respectively; and S is the sum of all source terms.

SWAN has been used to forecast wave conditions along the coast (Allahdadi et al., 2019; Alomar et al., 2014; Rusu et al., 2008), to propagate waves (Bhowmick et al., 2011), to model ship navigation (Chen et al., 2020) or to model ocean energy (Rute Bento et al., 2015). Other authors, on the other hand, have focused on the evaluation of these models (Amarouche et al., 2019; Amrutha et al., 2016; Aydogan and Ayat, 2021; Umesh and Behera, 2020; Umesh et al., 2018) and their improvement (Elkut et al., 2021; Umesh and Behera, 2021; Yang et al., 2021).

1.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) are data-driven models that began to be used in the 1990s. ANNs try to mimic the human brain and are made of different layers that pass the information from the first (input) layer to the last (output) layer (Medina et al., 2004; "Standford University Machine Learning Course,") using transfer functions. ANNs are trained and validated with large datasets and once they are tested, they can be used for forecasting. To train and validate ANNs different algorithms can be used depending on the purpose of the neural network. In addition, there are different types of neural networks. Those that perform best, both in terms of results and optimization of computational time and resources in ocean engineering, are Feed Forward Multilayer Perceptron Neural Networks (FFNNs) (Portillo Juan and Negro Valdecantos, 2022).

Neural networks make it possible to handle large amounts of data in record time. Their great advantage over other approaches is their low computational cost. However, one of their main disadvantages is precisely that they need a large amount of quality data to work that it is not always available. Many times, there are not enough data to develop an ANN model or the data available has not enough quality (many gap times or not homogeneously distributed, among others).

The other main drawback of ANNs is their uncertainty. They are “black-box” models, which means that there is not a clear understanding of their operation, they give a final result, but it does not have any explanation or justification. A physical explanation of the ANN model does not exist, which makes its

interpretability very difficult. Moreover, ANNs do not reproduce identical results, as their performance depends on different model parameters that vary each time the ANN is applied, the quantification of the uncertainty related to these models is still a big problem of this methodology (Portillo Juan et al., 2023).

Scientific literature on ANNs and ocean modeling problems is also large. ANNs have been applied to a variety of ocean and coastal engineering problems, from the effects of wave diffraction (Iglesias et al., 2010; Iglesias et al., 2009b) and tidal range on the coastline (Iglesias et al., 2009a) to the efficiency of wave energy converters (López, I. and Iglesias, 2014).

Focusing on wave-height forecasting, in 1978 the advantages of ANNs were proved for the first time (Deo and Sridhar Naidu, 1998). Since then, several authors have used ANNs to predict maritime climate (Abhigna et al., 2017; Agrawal and Deo, 2004; Gunaydin, 2008; Jain and Deo, 2008; Londhe et al., 2016; López and Iglesias, 2015; Vieira et al., 2020). In the last years, many studies have also focused on optimizing ANNs models; referring to their architecture (Joerges et al., 2021), comparing different models (Sadeghifar et al., 2022), using hybrid approaches (Campos et al., 2020) or using new models of ANNs (Hu et al., 2021; Minuzzi and Farina, 2023).

In this research paper, these two main approaches, physics-based and data-driven approach, are developed and compared for wave-downscaling in the Spanish Mediterranean Sea. Since physics-based models with an energy balance approach have the drawbacks of numerical models (high computational cost) but are still limited in their accuracy, it was considered necessary to carry out a comparative study of this part of the propagation of the maritime climate with other alternatives. Data-driven models, of particular importance due to their high level of development in recent years, could constitute an excellent alternative to energy balance numerical models, as they would save a great deal of computational time without losing accuracy in the results. Of the data-driven models, ANNs are the most used in ocean engineering and, therefore, were selected for this research.

A multilayer perceptron neural network was designed and optimized, analysing the influence of different data, algorithms, and architectures. In addition, the quantification of the uncertainty of the model has also been done. Both uncertainties have been analysed, the uncertainties related to the input data and the uncertainties related to the model itself. This model was compared with a SWAN model. Conclusions were drawn as to which approach is better for wave

downscaling from waves in deep water to nearshore locations and general rules to apply ANNs for wave downscaling problems have been proposed. To the authors' knowledge, it is the first time that the shortcomings of numerical models based on the energy balance equation have been highlighted and compared with data models in order to propose an optimized process for maritime climate propagation and it is also the first time that some general rules to develop surrogate data-driven models for wave downscaling are proposed. Moreover, this has been done at a regional scale in the specific area of the SE of Spain, where specific studies are lacking (Portillo Juan et al., 2022a).

2. Materials and methods

This section is divided into three subsections: Subsection 2.1., Study area and data; Subsection 2.2., in which the methodology followed to develop the SWAN model is explained; and Subsection 2.3., in which the development of the ANN model is presented.

2.1. Study area and data

The area of study is located off SE Spain. Due to its geography and orography its wave and atmospheric patterns are affected by regional and local phenomena (CAVALERI and BERTOTTI, 2004). In addition, it is one of the areas that are expected to be most affected by climate change (Portillo Juan et al., 2022b). Most of the studies on this area focus on specific parts, such as the Ebro Delta or the area of Catalonia, leaving many areas without climatic studies (Portillo Juan et al., 2022a). This study focuses, precisely, on the part of the Spanish coast where scientific studies are less abundant, the coast that goes from Valencia to Cabo de Palos.

To conduct the study the deep-water and coastal buoys of Valencia and Cabo de Palos were used (Figure 1). Historical hourly wave height (H_s), peak period (T_p) and Direction (*Dir*) records were downloaded from the portal of Spain's State Ports (*Puertos del Estado)* ("Puertos del Estado. Oceanografía. Registros históricos de oleaje,"), as they are the main variables that characterize maritime climate. Moreover, wind parameters obtained from the same source were also taken into consideration as secondary parameters for the SWAN model: this is wind velocity (W_v) and wind direction (W_{dir}). The Alicante buoy has been discarded for the development of the model because it is at shallow depth, which may affect the results.

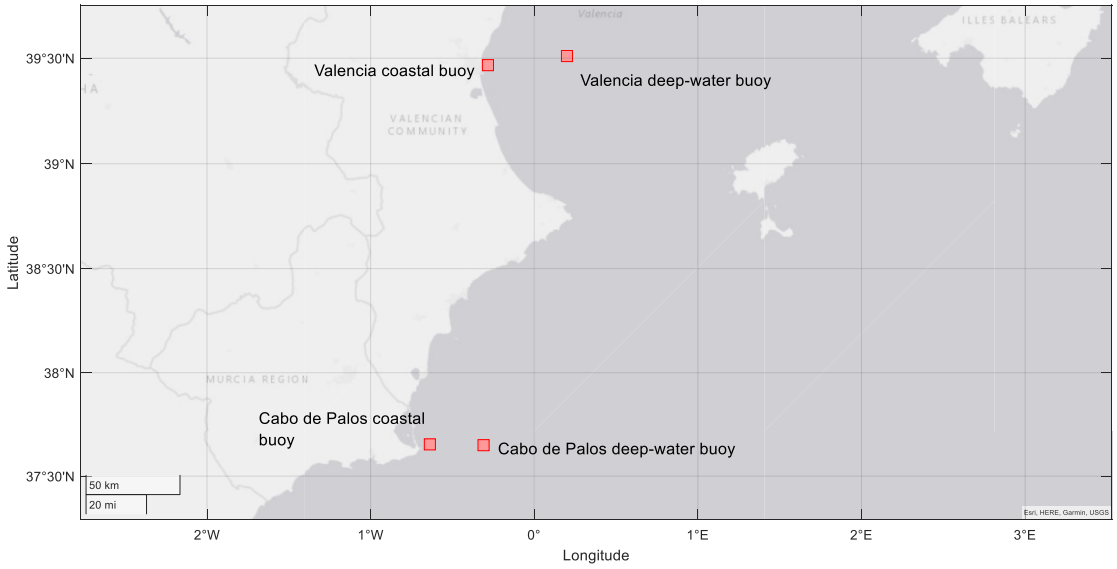


Figure 1. Location of the buoys used in the study.

Wave and wind data were obtained from buoys belonging to *Puertos del Estado*. These devices provide hourly data series of the main variables in the study (H_s, T_p and *Dir*) but also complementary variables (W_v and W_{Dir}) in case of deep-water wave characterization. Only in one buoy (Coastal buoy of Cabo de Palos) the direction of the waves was not available since this device only measures scalar parameters. As a summary, the available parameters of the buoys are presented in Table 1.

Table 1. Main available variables of the studied buoys.

Buoy	**Available data**							
	Wave data					**Wind data**		
	Depth	H_s	T_p	Dir	Series availability	W_v	W_{dir}	Series availability
Valencia	260 m	X	X	X	15/09/2005 to 14/07/2022	X	X	15/09/2005 to 19/07/2022
Coastal – Valencia	20 m	X	X	X	08/06/2005 to 30/10/2013	-	-	-
Cabo de Palos	230 m	X	X	X	18/07/2006 to 14/07/2022	X	X	18/07/2006 to 19/07/2022

Coastal – Cabo de Palos	67 m	X	X	-	14/11/1985 to 18/06/2012	-	-	-

The model bathymetry was obtained from The General Bathymetric Chart of the Ocean (GEBCO). The GEBCO_2022 Grid is a continuous terrain model for ocean and land with a spatial resolution of 15 arc-seconds interval grid. The Chart Datum in GEBCO is referred to the Lowest Astronomical Tide (LAT), that is the lowest water level which can be predicted to occur under average meteorological conditions.

2.2. SWAN model

For the SWAN model, the user interface developed by the IHC (IHC, 2013) was used. A DTM (Digital Terrain Model) of the coastline with the bathymetries mentioned in Section 2.1_was developed, generating a point cloud with XYZ information. The geodetic system used was the WGS84/UTM Zone 30N.

SWAN uses regular grids to compute propagation. Two model grids were used for the present study, an outer and a nested grid. The use of more than one grid has the aim of improving the resolution and accuracy of the model. The outer grid, which includes the deep-water wave data, has a resolution of 200 m with square elements; dimensions of 46400x39800 m and 232 x-nodes and 199 y-nodes in Valencia, and dimensions of 51600x42200 m and 258 x-nodes and 211 y-nodes in Cabo de Palos. The nested grid, which has a resolution of 40 m with square elements, has dimensions of 11000x9000 m and 225 x-nodes and 275 y-nodes in Valencia and 8000x7600 m and 200 x-nodes and 190 y-nodes in Cabo de Palos.

The SWAN model version was 41.31AB and operated using an implicit integration scheme. During simulations, the model was set to run in stationary model with the third-generation option (GEN3). The physics included whitecapping parametrization with $C_{ds} = 2.36 \cdot 10^{-5}$ (Komen, G. J. et al., 1984); depth limited wave breaking with $\gamma = 0.73$ and $\alpha = 1$ (Battjes and Janssen, 1978); bottom friction with $C_b = 0.038\, m^2 \cdot s^{-3}$ using JONSWAP spectra (Hasselmann, K. et al., 1973); quadruplet interactions with $\lambda = 0.25$ and $C_{nl45} = 3 \cdot 10^7$ using the Discrete Interaction Approximation (DIA) (Hasselmann, S. and Hasselmann, 1985) and refraction as default values. The spectrum of the model considered 36 frequencies (0.0356 Hz to 1.00 Hz) and 72 directions. The shape of the spectra was defined with

JONSWAP spectrum ($\gamma = 10, \sigma = 10$ / $\gamma = 3.3, \sigma = 20$) depending on the energy of the incident waves (in function of the wave height).

The forcing functions of the wave data were applied uniformly over the entire model domain and in all cases a sea level of +0.00 m (above LAT) and no currents have been considered.

Finally, the model output (in this case only Hs, Tp and Dir) values were extracted and processed.

2.3. ANN model

The first step to develop a data-driven model is to preprocess and evaluate the quality of the available data. First of all, there were time gaps in the buoy records, which could affect the performance of the ANN. To address this problem, an algorithm with the software PYTHON (Rossum, 1989) to detect these time gaps and eliminate them was developed. Then, all the data were filtered to eliminate false values and errors. Thus, all wave height values under 0.1 m and periods under 1 s were removed. Once the data was preprocessed, the ANN was designed and optimized.

To develop the ANN model, the Deep Learning toolbox of Matlab software version: 9.11.0 (R2021b) was used (MathWorks, 2021). First, the network architecture was optimized and then the algorithm was optimized.

A feedforward multilayer perceptron neural network (FFNN) with three layers was chosen for the data-driven model. The transfer functions used were the hyperbolic tangent sigmoid function (Equation 2) for the first layer and the linear transfer function (Equation 3) for the second layer. These two parameters, number of layers and transfer functions, were selected basing on scientific evidence on previous studies that proved that 3 layers and sigmoidal and linear functions are usually the optimum to model ocean engineering processes without increasing computational cost (Portillo Juan and Negro Valdecantos, 2022).

$$F(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (2)$$

$$F(x) = x \quad (3)$$

To optimize the rest of parameters, a similar methodology to previous studies was followed to train, validate, and test the ANNs (Iglesias et al., 2008; López and Iglesias, 2013; López and Iglesias, 2015). The first parameter to be optimized was the ANN architecture. The numbers of neurons in the input and output layers was

already determined based on the physics of the problem. The number of input neurons of the ANN was 3: wave height, peak period and direction in the deep water buoys $\left(H_s^0, T_p^0, \theta^0\right)$ and the number of output neurons was 2: wave height and peak period in the coastal buoys $\left[H_s^{dw},\ T_p^{dw}\right]$ (Equation 4).

$$H_s^{dw},\ T_p^{dw} = f^{ANN}\left(H_s^0, T_p^0, \theta^0\right) \quad (4)$$

Where the superscript dw means downscaled and the superscript 0 means in deep water.

It has to be said that wind parameters were not included as inputs in the model because the aim of this study is to carry out a comparison of ANNs versus energy balance numerical models to see if ANNs can be used as an alternative for modelling wave propagation in order to save computational time. Since most of the time spent on data-driven models is in the pre-processing of the data, it was decided not to include these additional data. Although including wind data would increase the accuracy of the model, it would also incur a much higher computational cost, which is what it is trying to be avoided.

The parameter to be optimized was the number of hidden neurons. The optimization process was conducted starting with 5 hidden neurons for each buoy and increasing the number of nodes by $5i$, being i the number of iterations, until the optimum number of nodes was reached (Equation 5 and 6). The different options were trained a total of 40 times. The reason for training each case 40 times was to remove the influence of the initialization weights and biases from the analysis.

$$Nº\ hidden\ neurons = 5i \quad (5)$$

$$If\ Error\ f^{ANN}(5\mathrm{i}) < Error\ f^{ANN}[5(\mathrm{i}-1)];\ Nº\ hidden\ neurons = 5(\mathrm{i}-1) \quad (6)$$

The division of the data for training, validating, and testing was 70%, 15% and 15%. The partition was done in contiguous blocks, taking the first records to train the ANN, and the remaining 30% to validate and test the model. 2005 to 2011 data records were used to train the Valencia ANN, and 2011 to 2013 data records were used to validate and test it. In Cabo de Palos, 2006 to 2010 data records were used to train the network and 2010 to 2012 data records were used to validate and test it (Figure 2).

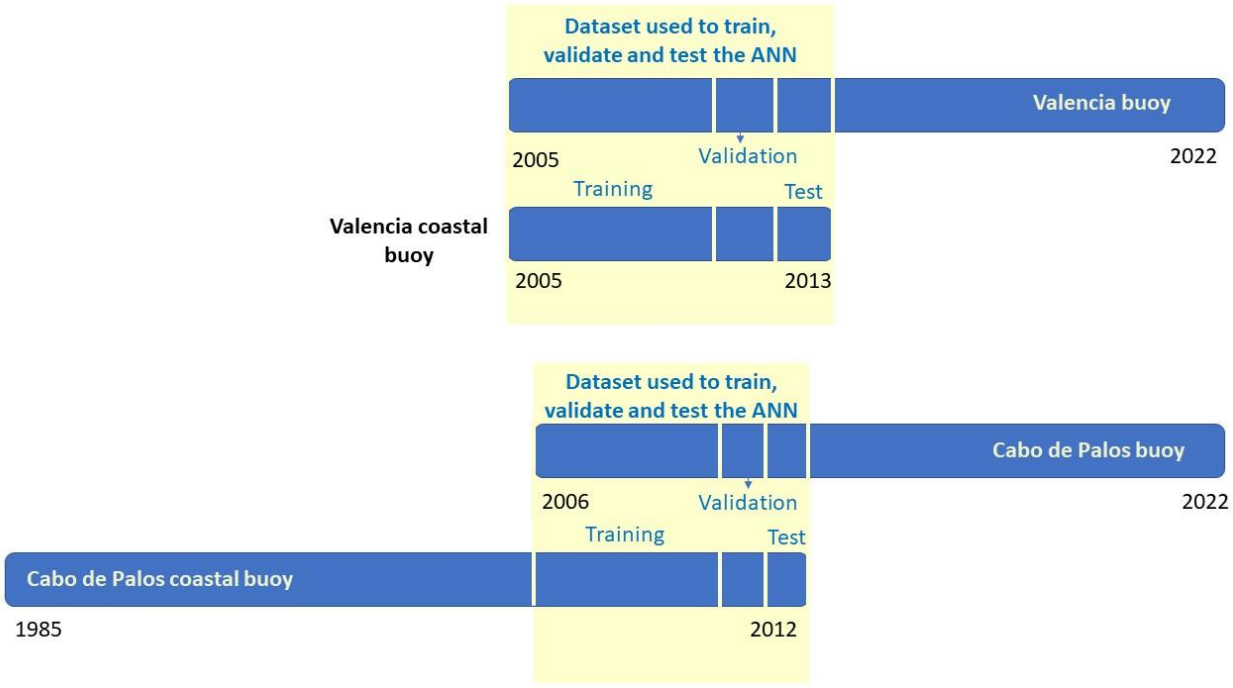


Figure 2. Data partition.

Once the optimal ANN architecture was selected for each buoy, the training algorithm was also optimized following a similar method. Each algorithm was trained a total of 40 times to remove the influence of the initialization weights and biases from the analysis. The algorithms tested were: a gradient descent algorithm (RBP), two conjugated gradient algorithms (SCG, BFGS), Bayesian Regularization (BR) and a Quasi-Newton's algorithm (LM). As with the ANN architecture, the criteria for the selection of the algorithms was based on their results reported in the literature (Portillo Juan and Negro Valdecantos, 2022).

Finally, the limitation of the interpretability and generalization of data-driven models was addressed. A statistical, trial and error and sensitivity analysis study were conducted to quantify the uncertainty of the ANN model and a Matlab code was developed to extract general rules to build ANNs for wave downscaling, testing different ANNs architectures, algorithms and transfer functions.

An ANN with LM algorithm and a [3-35-2] architecture for Valencia buoy and an ANN with LM algorithm and a [3-20-2] architecture for Cabo de Palos were selected as the optimum ANNs (Figure 3).

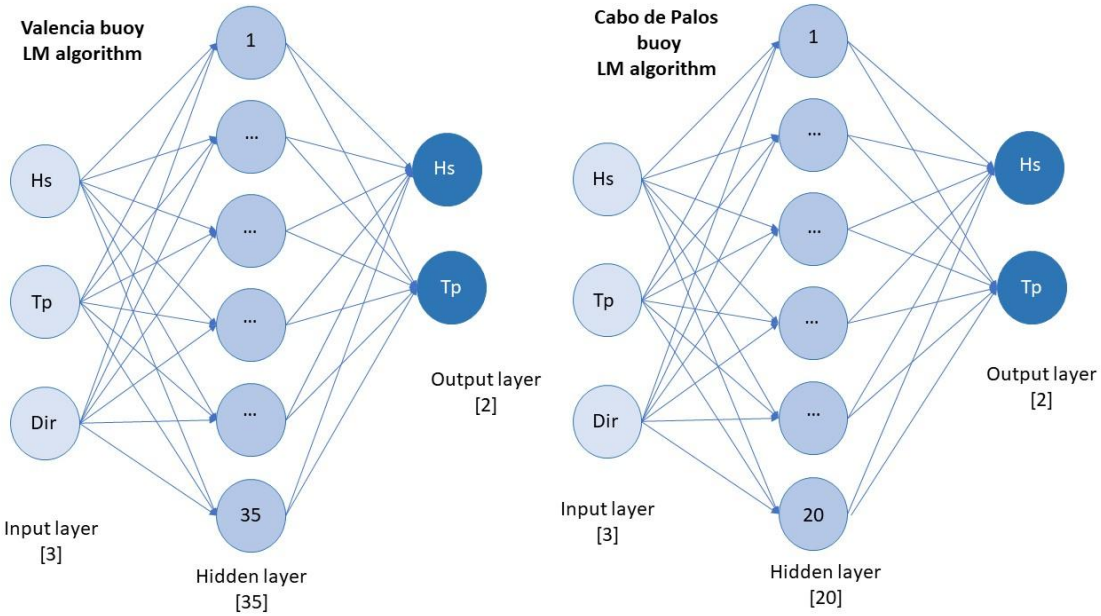


Figure 3. ANN model.

To see further details of the ANN model, please refer to Appendix A.

3. Results

Once the ANN model had been designed, trained and tested, and the SWAN model run, results were compared. To compare the performance of the models, the values of the correlation coefficient (R), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Bias were used (Equation 7 to 12).

$$R = \frac{\sum_{i=1}^{n}(o_i-\bar{o})(u_i-\bar{u})}{\sqrt{\sum_{i=1}^{n}(o_i-\bar{o})^2}\sqrt{\sum_{i=1}^{n}(u_i-\bar{u})^2}} \quad (7)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(o_i - u_i)^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - u_i)^2} \quad (9)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|o_i - u_i| \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|o_i-u_i|}{u_i} \quad (11)$$

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(o_i - u_i) \quad (12)$$

Figures 4 and 5 show the regression plot of the ANN models in Valencia and Cabo de Palos, respectively. The performance of the data-driven model is very accurate, with a correlation coefficient of approx. 0.95. The Cabo de Palos case shows a slightly better performance than the Valencia case, which may be related to the quality of the data available in this location.

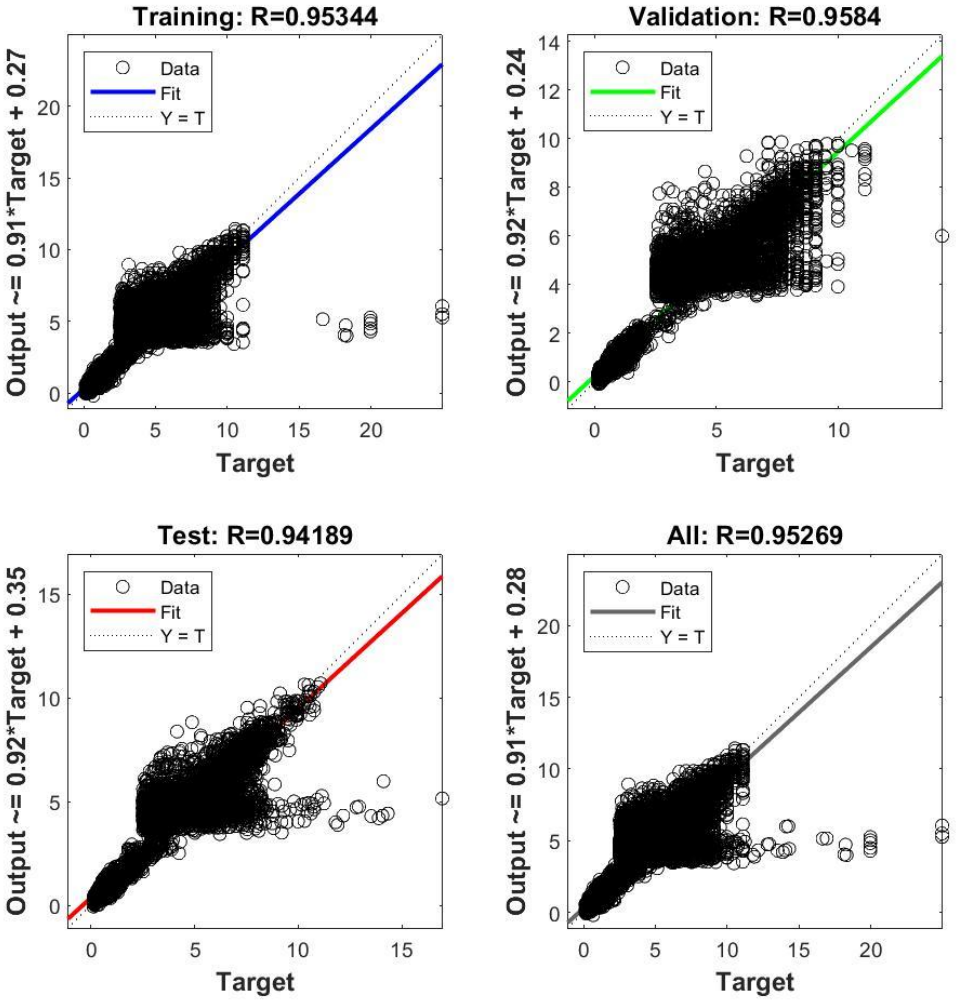


Figure 4. Regression plot for the ANN model in Valencia.

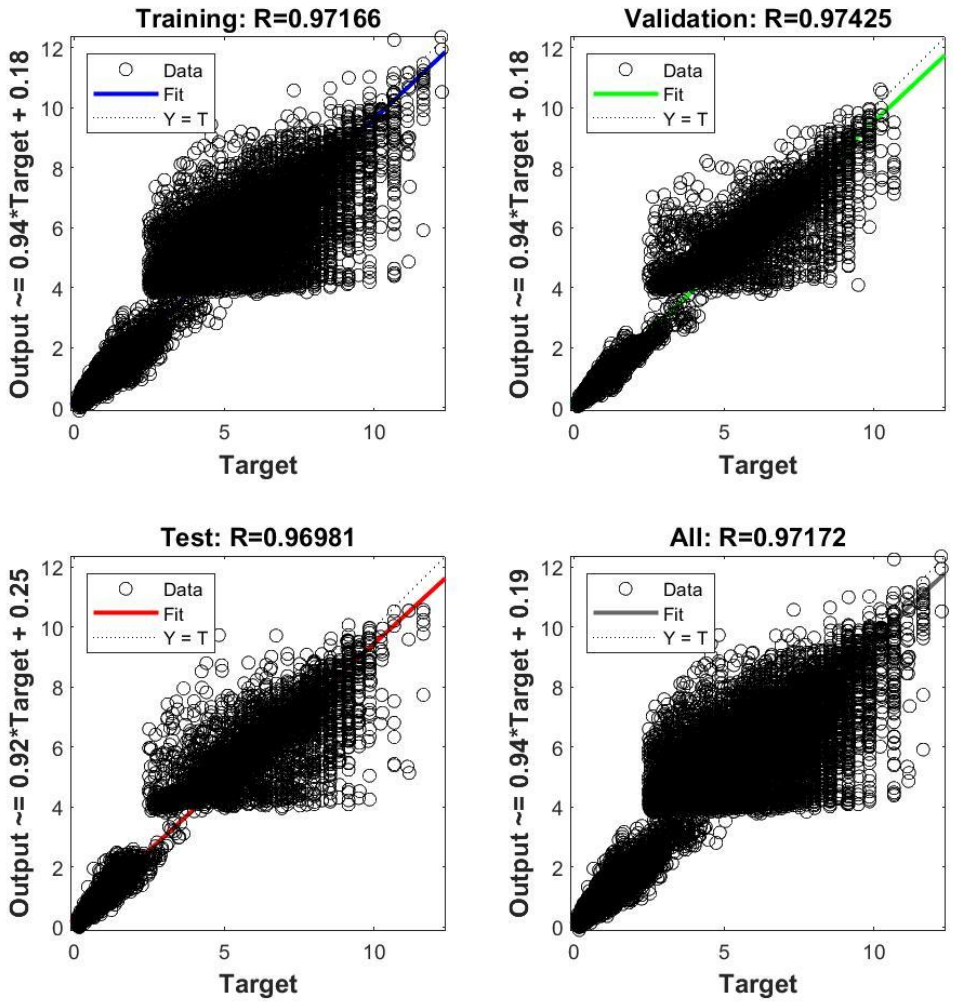


Figure 5. Regression plot for the ANN model in Cabo de Palos.

Figures 6 and 7 present the error histogram of the ANN in Valencia and Cabo de Palos, respectively. Error histograms represent in the X-axis the difference between the target values and the results obtained, and in the Y-axis the number of data samples that have these errors. In the case of Valencia around 85,000 data samples obtained with the ANN model have an overprediction of 0.07303 (main bar in Figure 7), while in Cabo de Palos around 60,000 data samples obtained with the ANN model have an overprediction of 0.1558 (main bar in Figure 8). The remaining bars of both graphs show the number of samples that present higher errors.

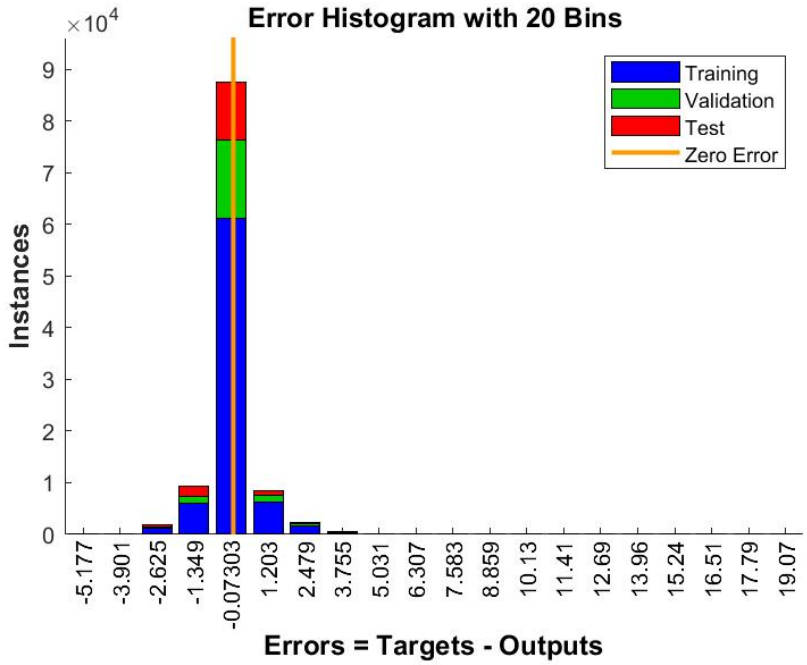


Figure 6. Error histogram for the ANN model in Valencia.

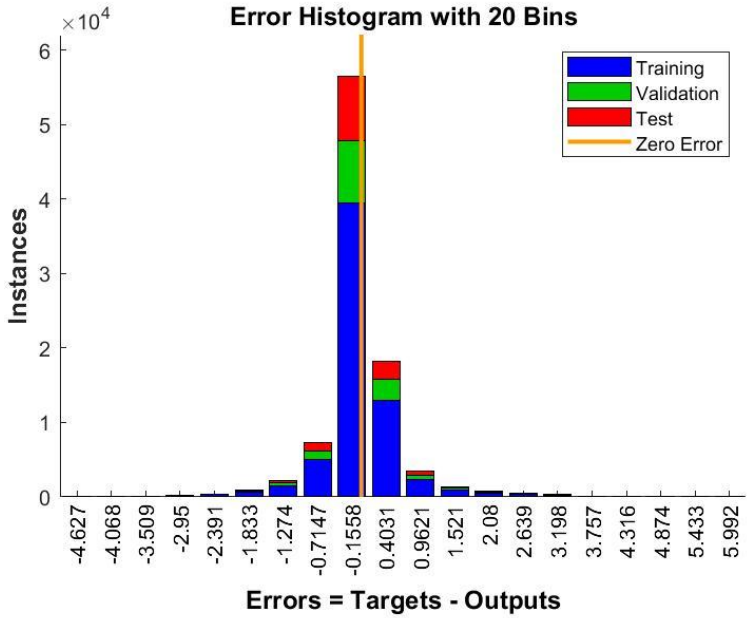


Figure 7. Error histogram for the ANN model in Cabo de Palos.

Figure 8 and Figure 9 present the results for the SWAN model. From these figures it can already be seen that the accuracy in the prediction of periods in SWAN is significantly lower.

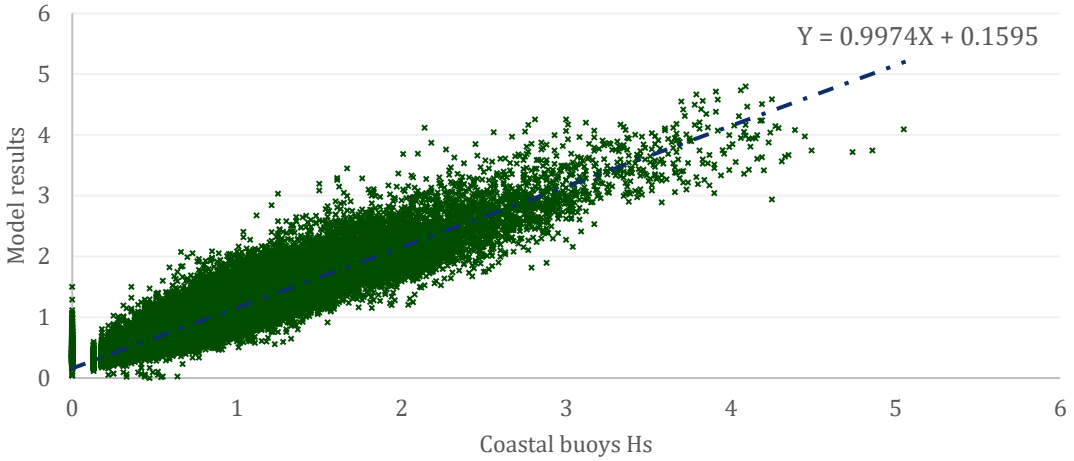


Figure 8. Regression plot for Hs in SWAN model

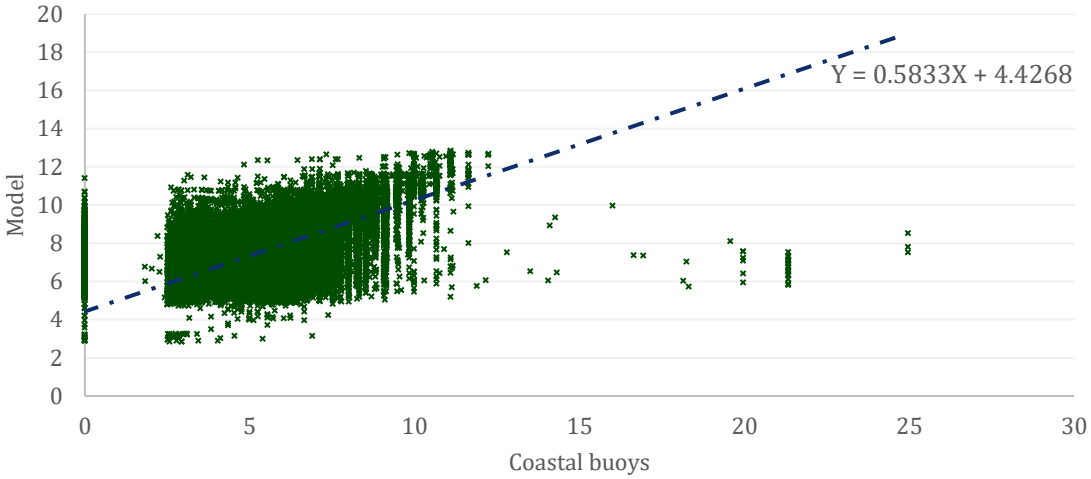


Figure 9. Regression plot for Tp in SWAN model.

Table 2 presents the values of the statistical analysis for both variables analysed, Hs and Tp, and both models. As can be seen, the ANN model performs slightly better than SWAN for Hs downscaling, but significantly better for Tp forecasting.

Table 2. Statistical analysis for Hs downscaling.

	Hs (m)		**Tp (s)**	
	SWAN	**ANN**	**SWAN**	**ANN**
R	0.94	0.96	0.72	0.96
MSE	0.06	0.023	6.2	1.00
RMSE	0.24	0.15	2.49	1.00

MAE	0.18	0.11	2.31	0.68
MAPE	0.33	0.05	0.49	0.05
Bias	0.15	0.002	2.26	0.03

4. Discussion

After implementing both approaches for wave downscaling, the physics-based approach with SWAN model and the data-driven approach with Feed Forward Multilayer Perceptron Neural Networks, it has been proved that both approaches perform similarly for Hs wave downscaling, but that the data-driven approach significantly outperforms the SWAN model for the Tp variable. Models with a physical approach lose accuracy in modelling the peak period of the waves due to the greater randomness that characterises this variable. This randomness makes Tp much more suitable for modelling with neural networks, which are adept at capturing random patterns.

Moreover, in the case of areas such as the Spanish Mediterranean, enclosed basins, this difference between models is even greater, as these areas are governed by local wind and wave patterns which global or larger-scale models struggle to capture. These global models are subsequently used to feed the propagation models. Their errors are thus carried over to the propagation models, causing them to lose accuracy. On the other hand, data-driven models take the data directly from instrumental records, thus avoiding these possible initial errors.

The only thing that models with a data-driven approach require is precisely large amounts of data and consistent series. These models base their performance on the data that feed them, so that the more information available, the better they tend to perform. In this case hourly records of 6 and 8 years were available, i.e., around 60,000 data samples. 25,000 samples are usually sufficient for a neural network to predict accurately in the long term, and 6,000 for the short term (Portillo Juan and Negro Valdecantos, 2022). In this case, the number of data samples available was double than the minimum required for a neural network to perform well. The availability of enough quality data added to the fact that the Spanish Mediterranean is a closed sea governed by local phenomena that are difficult to model with a physical-approach model resulted in the data-driven model outperforming the SWAN model for Tp wave downscaling.

As for the computational cost, it is widely known that models with a data-driven approach require much less time and computational resources than physical models. This is, in fact, one of the great advantages of these models and one of the main reasons why they are selected over traditional models. The only drawback is that although this data processing is much faster, it is also much less intuitive. Neural networks are "black-box" models whose results are difficult to explain and analyse. Furthermore, although there are different ways of optimizing these models, the most common, trial and error, was used in this study. This is a tedious process, which could be accused of being untechnical, as it may not ensure that the optimal model is reached. In spite of these drawbacks, in the present study neural networks have shown a better performance than the SWAN model with much less computational cost.

With regard to the input parameters used in the models, wave and wind parameters have been used for SWAN, while only wave parameters have been used for the ANN because maritime climate data alone were sufficient to achieve the desired results. Even using fewer data than SWAN, neural networks have shown a greater ability to model the Tp variable.

In this case buoy records have been used; it would be interesting, however, to see how ANNs perform with reanalysis data instead of instrumental data, not least for applications where buoy measurements are not available.

A comparative study of physics-based and data-driven approaches has been conducted. The final aim of the research was to optimize the modelling process and reduce the computational time without losing accuracy. As the final objective was to optimize the computational effort, there are some limitations and shortcomings of this study that must be stated. First of all, the datasets used for the ANN model were defined by the wave height, wave period and wave direction without considering other parameters such as wind, currents or sea level because, as stated in Section 2. Methodology, the most time-consuming part of a data-driven model is the pre-processing of the data, and adding more variables to pre-process significantly increases the computational cost, which is contrary to the objective pursued. However, not including these data may affect the accuracy with which extreme events are modelled (Vanem et al., 2022). In addition, wave states were modelled using the bulk wave parameters of the largest energy peak, and other secondary modes were disregarded because doing otherwise would imply handling huge databases – which is virtually prohibitive from the standpoint of computational cost. Not considering these secondary peaks may also affect the

modelling of extreme events, as most energetic wave states are usually composed of different frequencies (Hegermiller et al., 2017). In general, available ocean wave models are designed for time and length scales larger than those needed to properly model individual extreme waves (Breunung and Balachandran, 2023), and ANN do not solve this forecasting problem. Forecasting involves extrapolation, and there is still a great deal of uncertainty concerning ANNs and extrapolation. Therefore, to forecast correctly extreme events in the future, more sophisticated models with higher computational cost are needed, which are typically Computational Fluid Dynamics (CFD) models (Wang, R. et al., 2019).

5. Conclusions

Data-driven models are gaining popularity in ocean engineering problems, and the present research has demonstrated their usefulness in wave downscaling. They are an excellent alternative to balance energy approach models. Using ANNs to downscale wave climate from deep waters to coastal locations allows ocean and coastal engineers to considerably reduce the computational time of the whole process of wave downscaling, which is an essential part of most projects. Using data-driven models when quality data are available can significantly reduce the computational time without losing accuracy in the final results.

It is recommended to use such an approach whenever a minimum of 25,000 data samples are available, as they model the peak period better than the traditional approach. And it is highly recommended to use it in enclosed basins, governed by local climatic patterns.

To develop surrogate ANN models for wave downscaling some general rules can be followed. The optimal number of inputs is between 3 and 6. Hs, Tp and direction are the three main input variables that have to be used to develop wave downscaling data-driven models. If the tidal range is significant, the mean sea level should also be added as an input parameter. Finally, if the area of study has wind seas that are significant, the parameters related to wind (wind speed and wind direction) have also to be included in the model. The number of records available to train, validate and test the ANN has to be at least 25,000 and the optimum data-partitioning is around 70-80% for training and 30-20% for validating and testing. The recommended algorithm for this type of studies is the Levenberg-Marquardt, it is the algorithm that performed the best in both, model accuracy and computational cost. Referring to the model architecture, one hidden

layer is the optimum number of hidden layers to save computational time and avoid the overfitting of the neural network. The optimum number of hidden neurons has to be from one to 50 because above 50, model instabilities can occur (See Figures 20 and 21 from Appendix A). The transfer functions that perform the best are the sigmoid function between the input and the hidden layer and the linear function between the hidden layer and the output layer. Finally, to reduce model uncertainty, at least 40 iterations of the ANN have to be done when using trial and error approach. Hybrid models are also highly recommended to deal with model uncertainty (Portillo Juan et al., 2023).

In cases where data cannot be validated because of a lack of records, it is recommended to use a hybrid approach combining neural networks in those areas where records are available and numerical modelling for those areas where there are no records, in order to improve the accuracy in the calculation of the peak period.

Appendix A

In this section the process and results of the optimization of the ANN are presented.

To select the best architecture for the ANN the average ($\bar{E}$), maximum (E^{max}), and minimum (E^{min}) mean square error (MSE_{ANN}) (Equation 13) values were obtained with the training (E_{tr}), validation (E_{val}), and testing (E_{test}) data.

$$MSE_{ANN} = \frac{1}{n}\sum_{k=1}^{k_{max}}\sum_{j=1}^{j_{max}^{k}}\left(h_j^k - \hat{h}_j^k\right)^2 \quad (13)$$

where h_j^k and $\hat{h}_j^k$ denote the downscaled wave height after the jth test of the kth series and n the total number of tests of the dataset (Equation 14).

$$n = \sum_{k=1}^{k_{max}} j_{max}^{k} \quad (14)$$

The evolution of the loss function of the model (MSE_{ANN}) is presented in Figures 10 and 11.

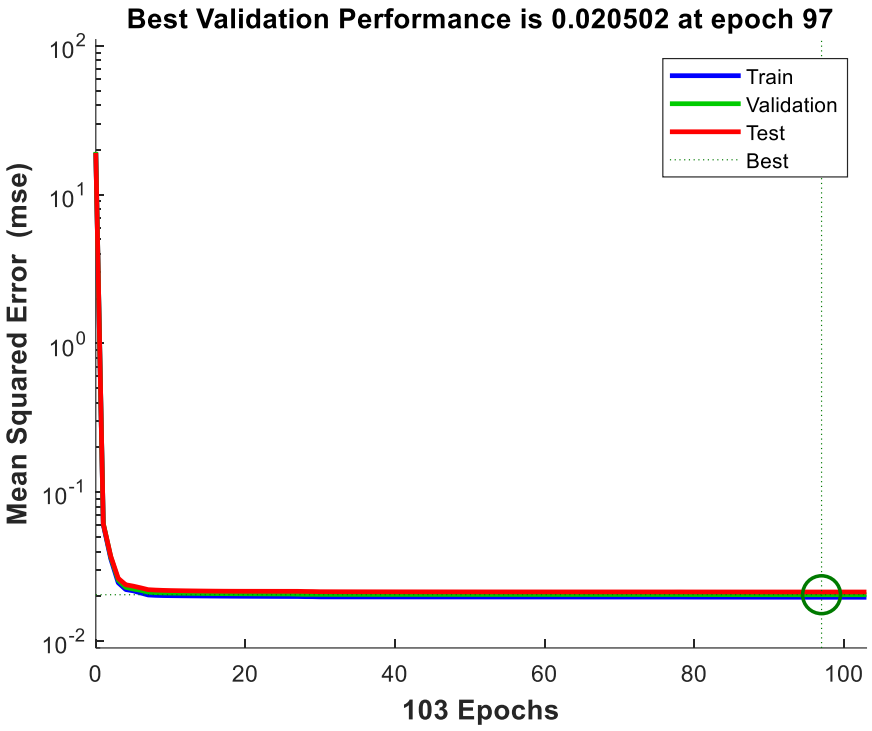


Figure 10. Evolution of loss function (mse) in Valencia.

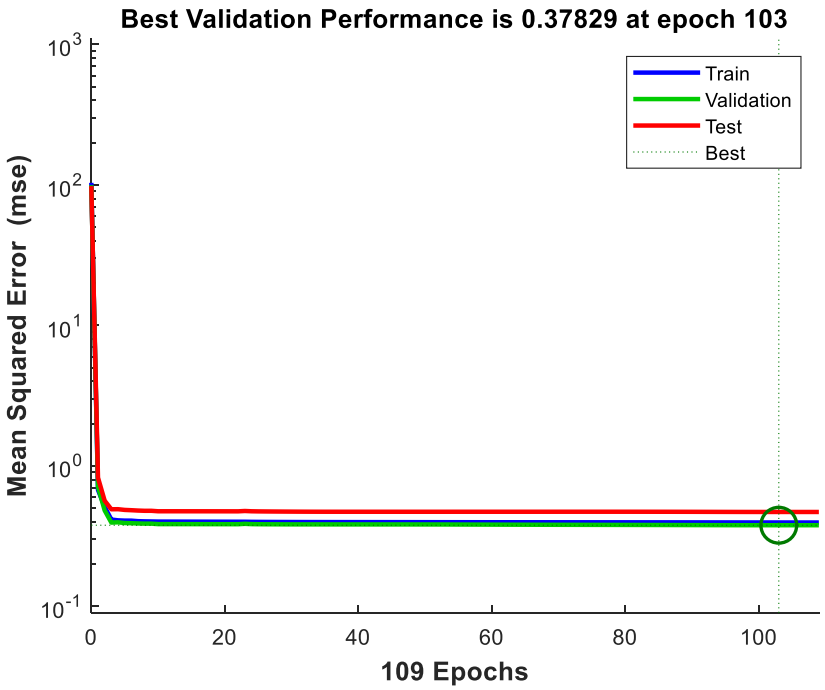


Figure 11. Evolution of loss function (mse) in Cabo de Palos.

In the case of Valencia, $E_{tr}, E_{val}, E_{test}$ decreased with the number of hidden neurons and the three variables studied ($E^{max}, \bar{E}, E^{min}$) reached their minima with 35 hidden neurons. Figure 12 shows the evolution of the different errors for all the ANN architectures and it can be seen that from 20 neurons onwards the errors in

the model are very small and similar, it is with 35 neurons where the minimum is reached in practically all the error measurements used.

The main bars represent $\bar{E}$, while their error bars represent E^{max} and E^{min}.

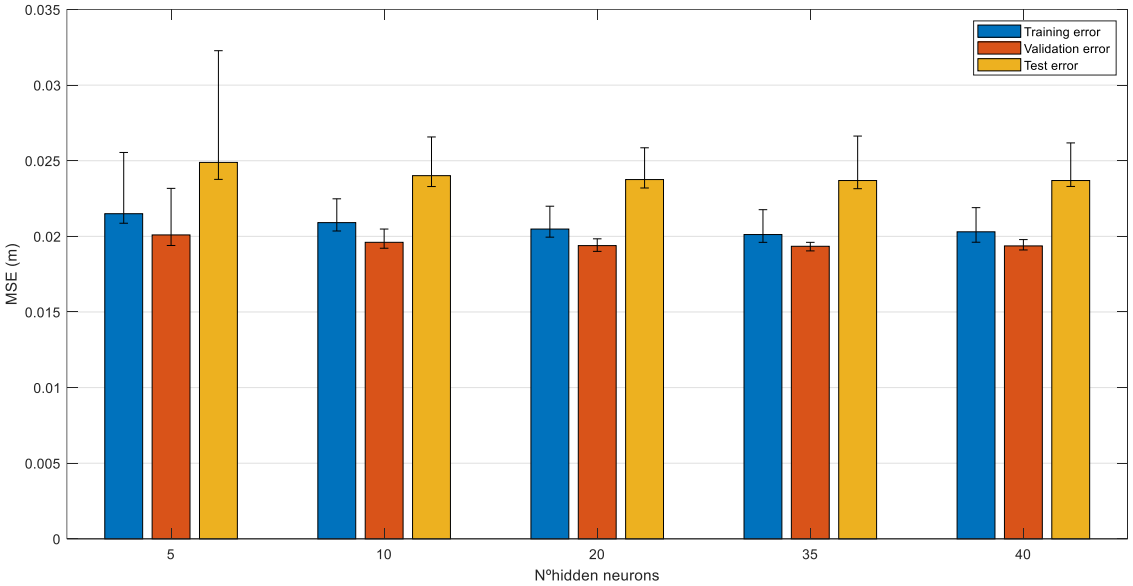


Figure 12. Training, validation and testing errors for each ANN architecture (Valencia case study).

The final errors are presented in Table 3.

Table 3. ANN errors in Valencia.

Training Error			**Validation Error**			**Test Error**		
$\boldsymbol{E_{tr}^{max}}$	$\overline{E_{tr}}$	E_{tr}^{min}	E_{val}^{max}	$\overline{E_{val}}$	E_{val}^{min}	E_{test}^{max}	$\overline{E_{test}}$	E_{test}^{min}
0.0218	0.0201	0.0196	0.0196	0.0193	0.0190	0.0266	0.0237	0.0231

In the case of Cabo de Palos, the results with different architectures present greater variability than in the Valencia case study. The mean validation and test error is lower for the architecture with 20 hidden neurons, but the mean training error is minimal for the case of 35 hidden neurons. To analyze better the data, all the errors have been plotted in Figure 13 following the same rules as in Figure 12. It is readily seen that, in general, the best performance is obtained by the ANN with 20 hidden neurons, which has therefore been retained for the analysis.

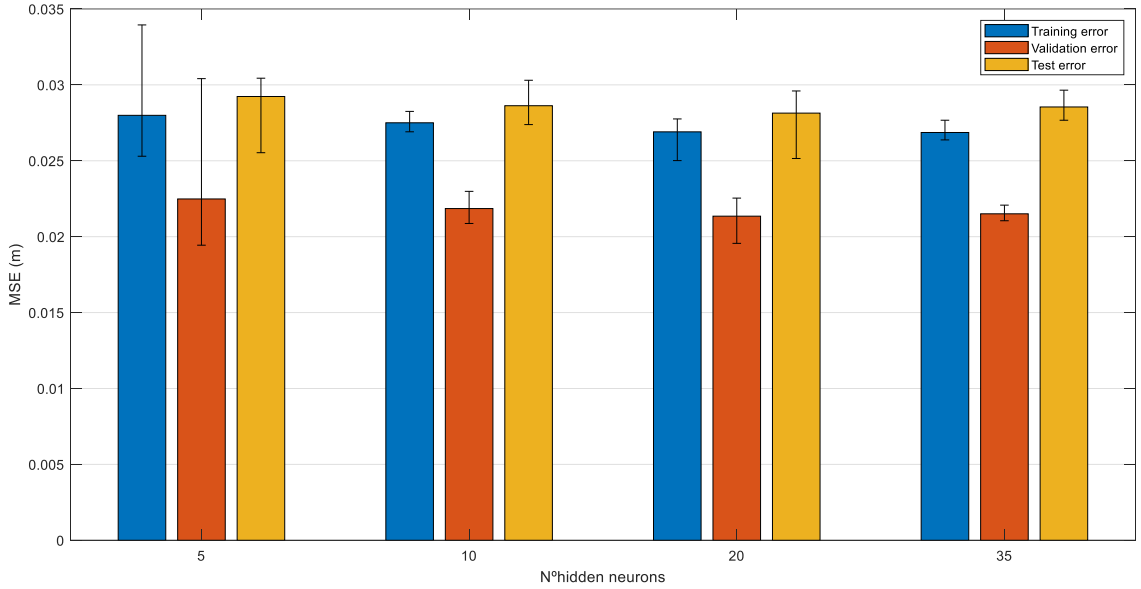


Figure 13. Training, validation and testing errors for each ANN architecture (Cabo de Palos case study).

Final error models are presented below (Table 4).

Table 4. ANN errors in Cabo de Palos.

Training Error			**Validation Error**			**Test Error**		
E_{tr}^{max}	$\overline{E_{tr}}$	E_{tr}^{min}	E_{val}^{max}	$\overline{E_{val}}$	E_{val}^{min}	E_{test}^{max}	$\overline{E_{test}}$	E_{test}^{min}
0.0278	0.0269	0.0250	0.0225	0.0213	0.0196	0.0296	0.0281	0.0251

After selecting the architecture for each buoy, [3-35-2] for Valencia buoy and [3-20-2] for Cabo de Palos buoy, the training algorithm was also optimized following the same methodology as with the ANN architecture optimization. In this case, the boxplot of the training, validation and testing error are shown for each buoy. It was considered that this type of graph was more appropriate to illustrate the average, maximum and minimum error and their deviation for each algorithm. Figures 14 to 16 show the boxplots of the error for the Valencia case study.

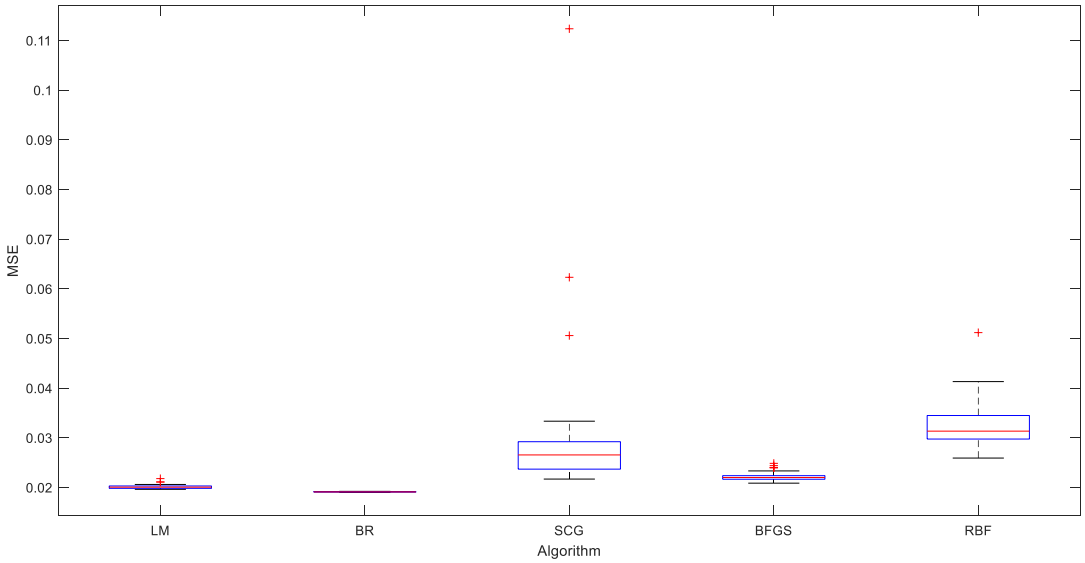


Figure 14. Training error boxplot for each algorithm (Valencia case study).

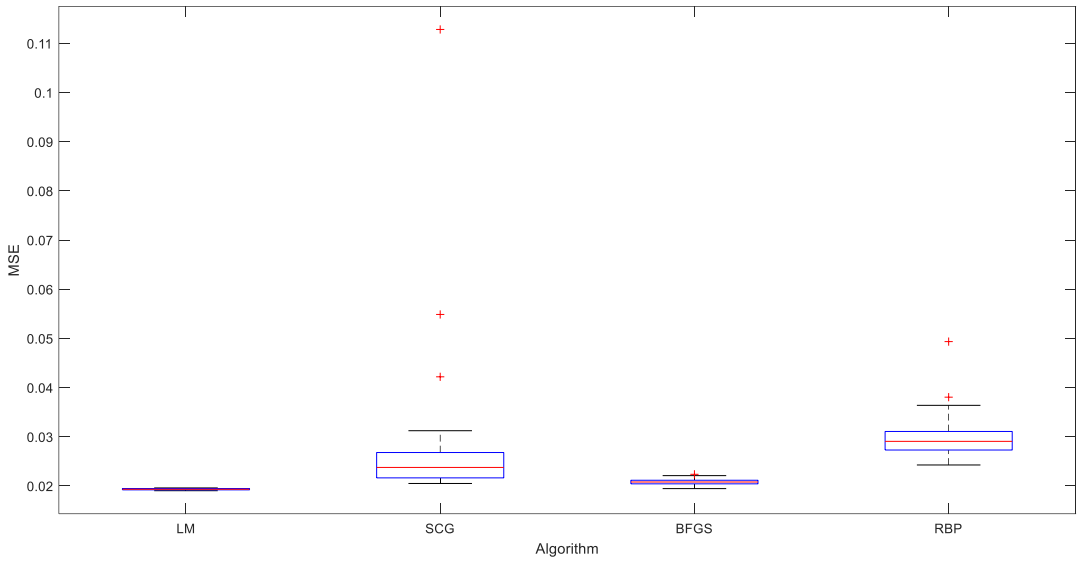


Figure 15. Validation error for each algorithm (Valencia case study).

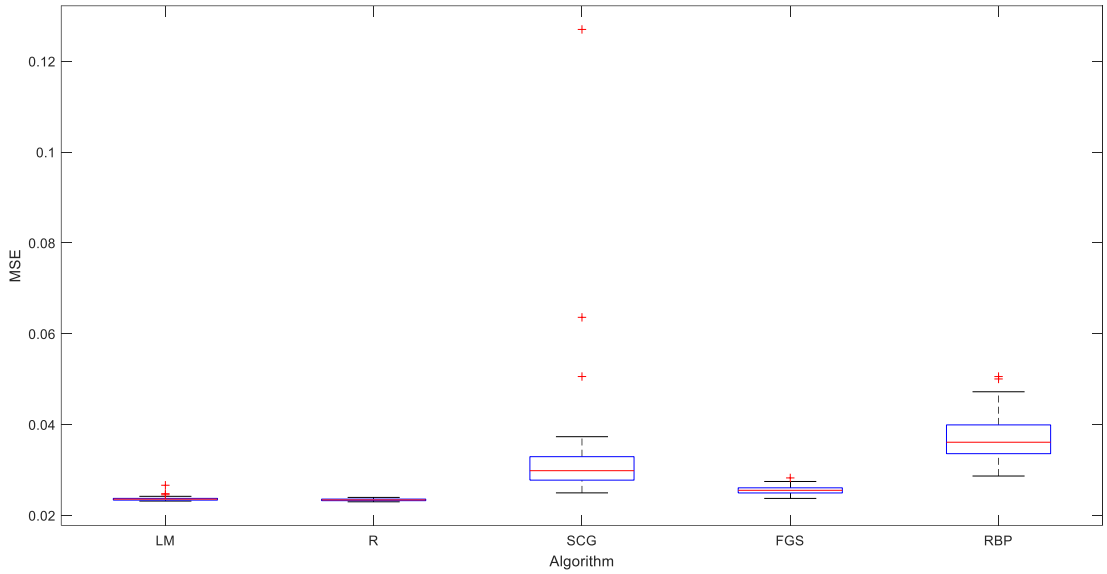


Figure 16. Test error for each algorithm (Valencia case study).

As can be seen from these Figures, RBP is the algorithm with the worst performance, followed by SCG. In addition, their performance is most variable, with large scatter in their results. More importantly, LM and BR are the algorithms that performed the best, with an average training error of 0.0201m and 0.01909 m, respectively, and an average testing error of 0.0237 m and 0.0234 m, respectively. In addition, the dispersion of errors of these two algorithms, unlike in the previous cases, is very small. Although the BR algorithm is slightly better than the LM, its computational cost is much higher. The average time for BR for training is around 410 s, vs. 5-20 s for the LM algorithm in a Microsoft Windows 10 Home v.10.0.19044 computer with a AMD Ryzen 7 3700U with Radeon Vega Mobile Gfx, 2300 Mhz CPU, with 16.0 GB of memory installed and 17.8 GB of memory virtual. Therefore, the LM algorithm has been chosen as the optimum algorithm for the Valencia ANN.

Figures 17 to 19 show the boxplots of the error for Cabo de Palos buoy. SCG is the algorithm that showed the worst performance, followed by RBP. Again, these algorithms showed the most variable performance, with scattered results. LM and BR are the algorithms that performed the best, with an average training error of 0.0269 m and 0.0220 m, respectively, and an average testing error of 0.0281 m and 0.0245 m, respectively. In addition, the error scatter of these two algorithms, unlike in the previous cases, is very small. Although the BR algorithm is slightly better than the LM, its computational cost is much higher. The average time for BR for training is around 220 s, while for the LM algorithm is around 5-20 s in a

Microsoft Windows 10 Home v.10.0.19044 computer with a AMD Ryzen 7 3700U with Radeon Vega Mobile Gfx, 2300 Mhz CPU, with 16.0 GB of memory installed and 17.8 GB of memory virtual. Therefore, the LM algorithm has also been chosen as the optimum algorithm for the Cabo de Palos ANN.

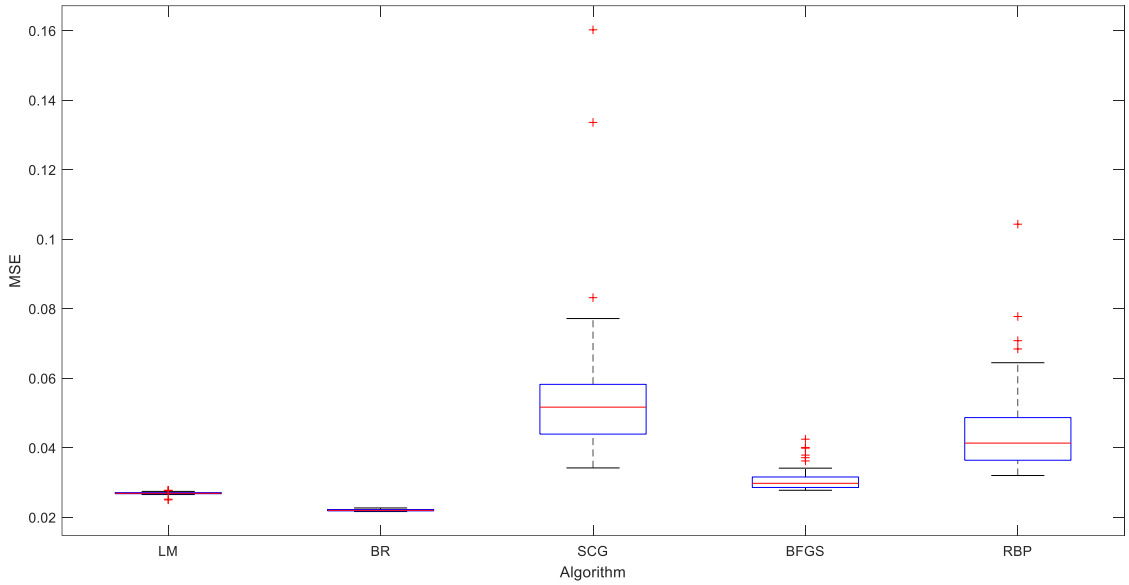


Figure 17. Training error boxplot for each algorithm (Cabo de Palos case study).

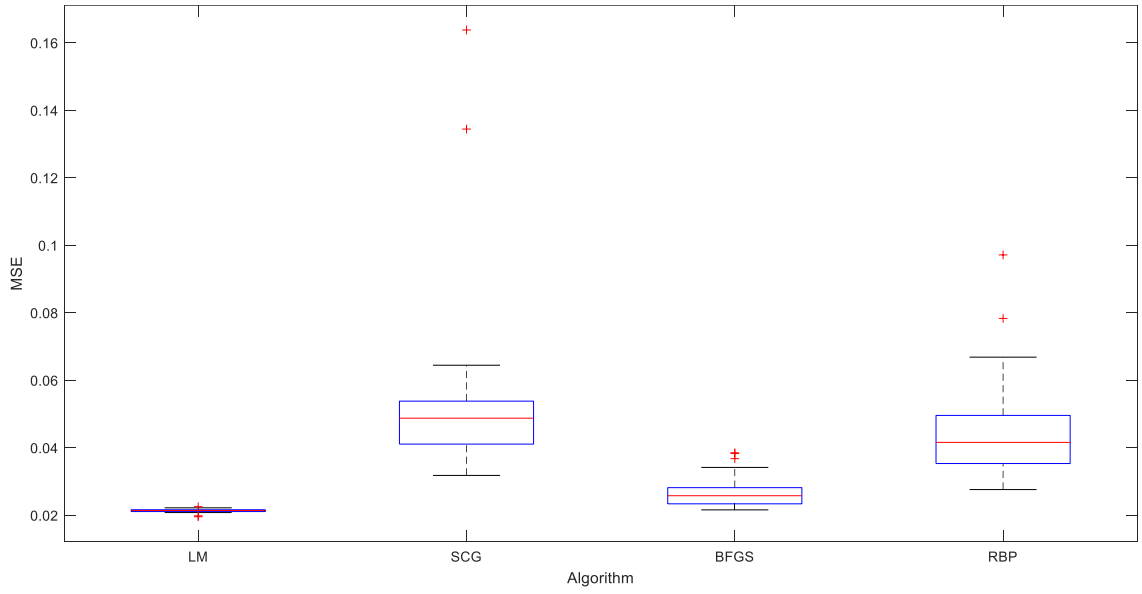


Figure 18. Validation error for each algorithm (Cabo de Palos case study).

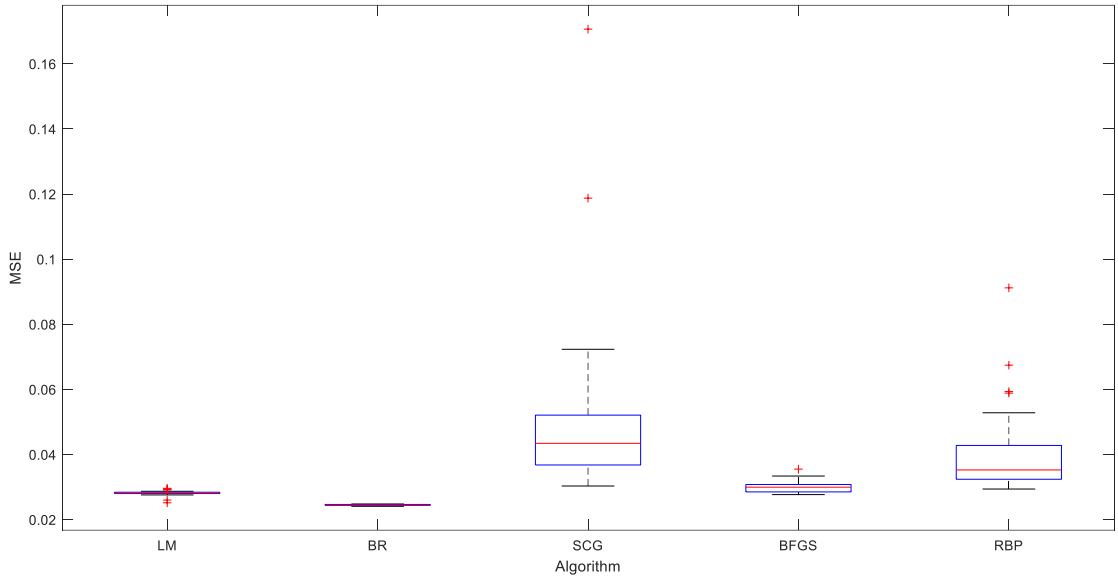


Figure 19. Test error for each algorithm (Cabo de Palos case study).

In the end, a quantification of the ANN uncertainty has been conducted. There are two types of uncertainty associated to ANN models, input uncertainty and model uncertainty. Input uncertainty is the uncertainty associated to those variables that are chosen as the inputs and the number of records that are used to train, validate and test the model. Model uncertainty is the uncertainty associated to the model itself, which means ANN architecture, algorithm, transfer functions, etc (Portillo Juan et al., 2023). A sensitivity analysis to quantify both uncertainties has been conducted. Different input variables have been tested and it has been concluded that the best performance of the model is given by between 3 and 6 input variables depending on the area of study. Referring to the uncertainty associated to the number of records, at least 25,000 records are needed to reduce input uncertainty significantly. In addition, a MATLAB code was developed to quantify model uncertainty and extract some general rules to develop data-driven models for wave downscaling (Figures 20 and 21). The RMSE was chosen as the loss function to evaluate the performance of the different models.

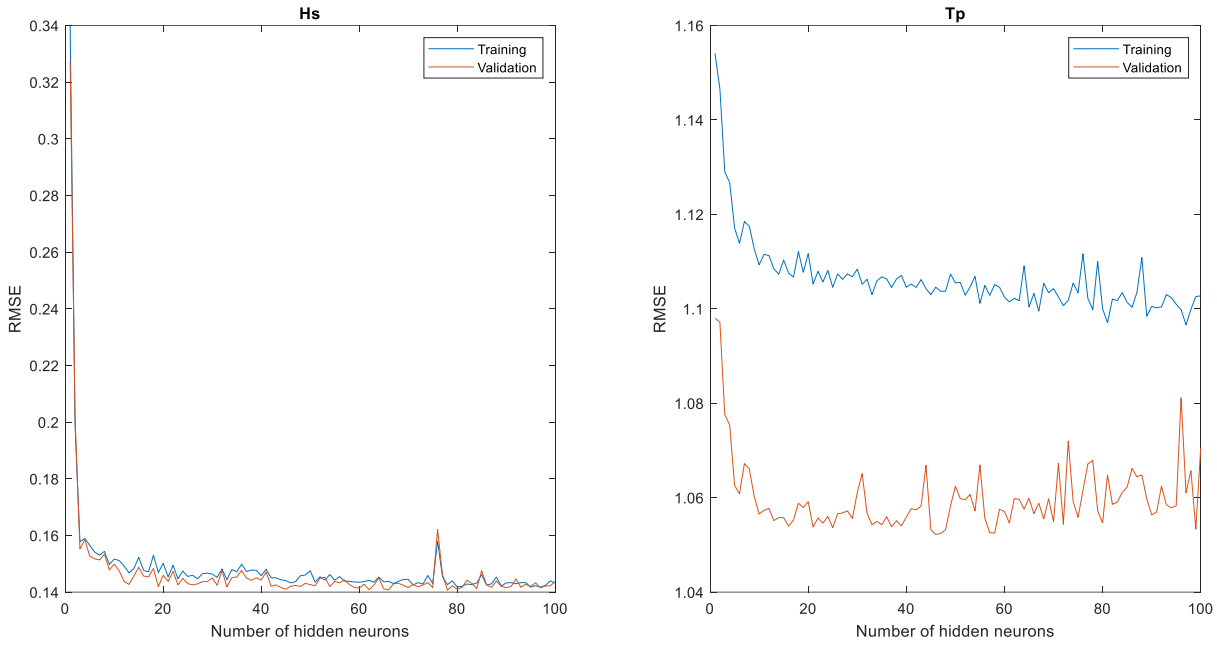


Figure 20. Evolution of RMSE for Hs and Tp in Valencia with different model architectures.

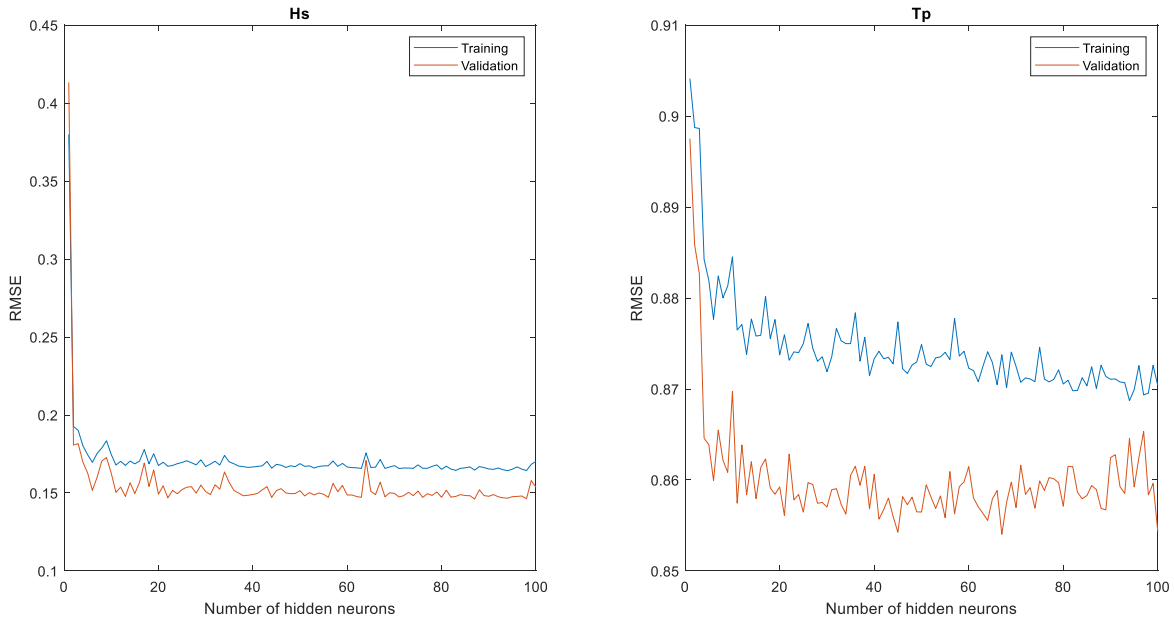


Figure 21. Evolution of RMSE for Hs and Tp in Valencia with different model architectures.

Finally, a trial and error methodology has been used to quantify initial weights, data-partitioning and algorithm uncertainty, resulting in an uncertainty up to 10%. A sensitivity analysis has also been made to analyse how to reduce model uncertainty and it has been concluded that at least 40 iterations are necessary to reduce data-driven models uncertainty.

References

Abhigna, P.;Jerritta, S.;Srinivasan, R.;Rajendran, V. y Ieee. (2017, 2017

Apr 06-08). *Analysis of Feed Forward and Recurrent Neural Networks in Predicting the Significant Wave Height at the Moored Buoys in Bay Of Bengal.* Paper presented at the IEEE International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, INDIA.

Agrawal, J. D. y Deo, M. C. (2004). Wave parameter estimation using neural networks. *Marine Structures, 17*(7), 536-550. doi:https://doi.org/10.1016/j.marstruc.2005.01.001

Allahdadi, M. N.;He, R. y Neary, V. S. (2019). Predicting ocean waves along the US east coast during energetic winter storms: sensitivity to whitecapping parameterizations. *Ocean Sci., 15*(3), 691-715. doi:10.5194/os-15-691-2019

Alomar, M.;Sánchez-Arcilla, A.;Bolaños, R. y Sairouni, A. (2014). Wave growth and forecasting in variable, semi-enclosed domains. *Continental Shelf Research, 87*, 28-40. doi:https://doi.org/10.1016/j.csr.2014.05.008

Amarouche, K.;Akpinar, A.;Bachari, N. E. I.;Cakmak, R. E. y Houma, F. (2019). Evaluation of a high-resolution wave hindcast model SWAN for the West Mediterranean basin. *Applied Ocean Research, 84*, 225-241. doi:10.1016/j.apor.2019.01.014

Amrutha, M. M.;Kumar, V. S.;Sandhya, K. G.;Nair, T. M. B. y Rathod, J. L. (2016). Wave hindcast studies using SWAN nested in WAVEWATCH III - comparison with measured nearshore buoy data off Karwar, eastern Arabian Sea. *Ocean Engineering, 119*, 114-124. doi:10.1016/j.oceaneng.2016.04.032

Aydogan, B. y Ayat, B. (2021). Performance evaluation of SWAN ST6 physics forced by ERA5 wind fields for wave prediction in an enclosed basin. *Ocean Engineering, 240*. doi:10.1016/j.oceaneng.2021.109936

Battjes, J. A. y Janssen, J. P. F. M. (1978). Energy Loss and Set-Up Due to Breaking of Random Waves. In *Coastal Engineering 1978* (pp. 569-587).

Bhowmick, S. A.;Kumar, R.;Chaudhuri, S. y Sarkar, A. (2011). Swell Propagation over Indian Ocean Region. *The International Journal of Ocean and Climate Systems, 2*(2), 87-99. doi:10.1260/1759-3131.2.2.87

Bingolbali, B.;Akpinar, A.;Jafali, H. y Van Vledder, G. P. (2019). Downscaling of wave climate in the western Black Sea. *Ocean Engineering, 172*, 31-45. doi:10.1016/j.oceaneng.2018.11.042

Booij, N.;Ris, R. C. y Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions 1. Model description and validation. *Journal of Geophysical Research: Oceans, 104*(C4), 7649-7666. doi:10.1029/98JC02622

Boukabara, S.-A. y Hoffmanb, R. N. (2022). Optimizing Observing Systems Using ASPEN: An Analysis Tool to Assess the Benefit and Cost Effectiveness of Observations to Earth System Applications. *Bulletin of the American Meteorological Society*.

Boukabara, S.-A.;Krasnopolsky, V.;Stewart, J. Q.;Maddy, E. S.;Shahroudi, N. y Hoffman, R. N. (2019). Leveraging Modern Artificial Intelligence for Remote Sensing and NWP: Benefits and Challenges. *Bulletin of the American Meteorological Society, 100*(12), ES473-ES491. doi:10.1175/bams-d-18-0324.1

Breunung, T. y Balachandran, B. (2023). Data-driven, high resolution ocean wave forecasting and extreme wave predictions. *Ocean Engineering, 268*, 113271. doi:https://doi.org/10.1016/j.oceaneng.2022.113271

Campos, R. M.;Krasnopolsky, V.;Alves, J.-H. y Penny, S. G. (2020). Improving NCEP's global-scale wave ensemble averages using neural networks. *Ocean Modelling, 149*, 101617. doi:https://doi.org/10.1016/j.ocemod.2020.101617

Camus, P.;Mendez, F. J. y Medina, R. (2011). A hybrid efficient method to downscale wave climate to coastal areas. *Coastal Engineering, 58*(9), 851-862. doi:https://doi.org/10.1016/j.coastaleng.2011.05.007

CAVALERI, L. y BERTOTTI, L. (2004). Accuracy of the modelled wind and wave fields in enclosed seas. *Tellus A, 56*(2), 167-175. doi:https://doi.org/10.1111/j.1600-0870.2004.00042.x

Chen, C.;Sasa, K.;Ohsawa, T.;Kashiwagi, M.;Prpić-Oršić, J. y Mizojiri, T. (2020). Comparative assessment of NCEP and ECMWF global datasets and numerical approaches on rough sea ship navigation based on numerical simulation and shipboard measurements. *Applied Ocean Research, 101*, 102219. doi:https://doi.org/10.1016/j.apor.2020.102219

Deo, M. C. y Sridhar Naidu, C. (1998). Real time wave forecasting using neural networks. *Ocean Engineering, 26*(3), 191-203. doi:https://doi.org/10.1016/S0029-8018(97)10025-7

Elkut, A. E.;Taha, M. T.;Abu Zed, A. B. E.;Eid, F. M. y Abdallah, M. A. (2021). Wind-wave hindcast using modified ECMWF ERA-Interim wind field in the Mediterranean Sea. *Estuarine Coastal and Shelf Science, 252*. doi:10.1016/j.ecss.2021.107267

Gao, Z.;Liu, X.;Yv, F.;Wang, J. y Xing, C. (2023). Learning wave fields evolution in North West Pacific with deep neural networks. *Applied Ocean Research, 130*, 103393. doi:https://doi.org/10.1016/j.apor.2022.103393

GEBCO. General Bathymetric Chart of the Oceans Retrieved from https://www.gebco.net/ Acessed on 07/01/2023

Gunaydin, K. (2008). The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. *Ocean Engineering, 35*(14-15), 1406-1415. doi:10.1016/j.oceaneng.2008.07.008

Hasselmann, K.;Barnett, T.;Bouws, E.;Carlson, H.;Cartwright, D.;Enke, K., . . . Walden, H. (1973). Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deut. Hydrogr. Z., 8*, 1-95.

Hasselmann, S. y Hasselmann, K. (1985). Computations and Parameterizations of the Nonlinear Energy Transfer in a Gravity-Wave Spectrum. Part I: A New Method for Efficient Computations of the Exact Nonlinear Transfer Integral. *Journal of Physical Oceanography, 15*(11), 1369-1377. doi:https://doi.org/10.1175/1520-0485(1985)015<1369:CAPOTN>2.0.CO;2

Hegermiller, C. A.;Antolinez, J. A. A.;Rueda, A.;Camus, P.;Perez, J.;Erikson, L. H., . . . Mendez, F. J. (2017). A Multimodal Wave Spectrum-Based Approach for Statistical Downscaling of Local Wave Climate. *Journal of Physical Oceanography, 47*(2), 375-386. doi:10.1175/jpo-d-16-0191.1

Hu, H.;van der Westhuysen, A. J.;Chu, P. y Fujisaki-Manome, A. (2021). Predicting Lake Erie wave heights and periods using XGBoost and LSTM. *Ocean Modelling, 164*, 101832. doi:https://doi.org/10.1016/j.ocemod.2021.101832

Huang, L.;Jing, Y.;Chen, H.;Zhang, L. y Liu, Y. (2022). A regional wind wave prediction surrogate model based on CNN deep learning network. *Applied Ocean Research, 126*, 103287. doi:https://doi.org/10.1016/j.apor.2022.103287

Iglesias, G.;Diz-Lois, G. y Pinto, F. T. (2010). Artificial Intelligence and headland-bay beaches. *Coastal Engineering, 57*(2), 176-183. doi:https://doi.org/10.1016/j.coastaleng.2009.10.004

Iglesias, G.;López, I.;Carballo, R. y Castro, A. (2009a). Headland-bay beach planform and tidal range: A neural network model. *Geomorphology, 112*(1), 135-143. doi:https://doi.org/10.1016/j.geomorph.2009.05.014

Iglesias, G.;López, I.;Castro, A. y Carballo, R. (2009b). Neural network modelling of planform geometry of headland-bay beaches. *Geomorphology, 103*(4), 577-587. doi:https://doi.org/10.1016/j.geomorph.2008.08.002

Iglesias, G.;Rabuñal, J.;Losada, M. A.;Pachón, H.;Castro, A. y Carballo, R. (2008). A virtual laboratory for stability tests of rubble-mound breakwaters. *Ocean Engineering, 35*(11), 1113-1120. doi:https://doi.org/10.1016/j.oceaneng.2008.04.014

IHC. (2013). Análisis Matemático y Estadístico de Variables Medioambientales (Version v.1.3.2). . Retrieved from https://ihcantabria.com/software-y-servicios-tic/software/

Jain, P. y Deo, M. (2008). Artificial Intelligence Tools to Forecast Ocean Waves in Real Time. *The Open Ocean Engineering Journal, 1*, 13-20.

Janssen, P. (2008). Progress in ocean wave forecasting. *Journal of Computational Physics, 227*, 3572-3594. doi:10.1016/j.jcp.2007.04.029

Joerges, C.;Berkenbrink, C. y Stumpe, B. (2021). Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks. *Ocean Engineering, 232*. doi:10.1016/j.oceaneng.2021.109046

Kazeminezhad, M. H. y Siadatmousavi, S. M. (2017). Performance evaluation of WAVEWATCH III model in the Persian Gulf using different wind resources. *Ocean Dynamics, 67*(7), 839-855. doi:10.1007/s10236-017-1063-2

Komen, G. J.;Cavaleri, L.;Donelan, M.;Hasselmann, K.;Hasselmann, S. y Janssen, P. (1996). *Dynamics and modelling of ocean waves*.

Komen, G. J.;Hasselmann, S. y Hasselmann, K. (1984). On the Existence of a Fully Developed Wind-Sea Spectrum. *Journal of Physical Oceanography, 14*(8), 1271-1285. doi:https://doi.org/10.1175/1520-0485(1984)014<1271:OTEOAF>2.0.CO;2

Liu , P. y Losada, I. J. (2002). Wave propagation modeling in coastal engieering. *Journal of Hydraulic Research - J HYDRAUL RES, 40*, 229-240. doi:10.1080/00221680209499939

Londhe, S. N.;Shah, S.;Dixit, P. R.;Nair, T. M. B.;Sirisha, P. y Jain, R. (2016). A Coupled Numerical and Artificial Neural Network Model for Improving Location Specific Wave Forecast. *Applied Ocean Research, 59*, 483-491. doi:10.1016/j.apor.2016.07.004

López y Iglesias, G. (2013). Artificial Intelligence for estimating infragravity energy in a harbour. *Ocean Engineering, 57*, 56-63. doi:https://doi.org/10.1016/j.oceaneng.2012.08.009

López y Iglesias, G. (2015). Artificial neural networks applied to port operability assessment. *Ocean Engineering, 109*, 298. doi:10.1016/j.oceaneng.2015.09.016

López, I. y Iglesias, G. (2014). Efficiency of OWC wave energy converters: A virtual laboratory. *Applied Ocean Research, 44*, 63-70. doi:https://doi.org/10.1016/j.apor.2013.11.001

MathWorks. (2021). MATLAB version: 9.11.0 (R2021b): Natick, Massachusetts: The MathWorks Inc. . Retrieved from https://www.mathworks.com

Medina, J. R.;Garrido, J. M.;Gómez-Martín, M. E. y Vidal, C. s. (2004). Armor Damage Analysis Using Neural Networks. *Coastal structures*, 236-248.

Minuzzi, F. C. y Farina, L. (2023). A deep learning approach to predict significant wave height using long short-term memory. *Ocean Modelling, 181*, 102151. doi:https://doi.org/10.1016/j.ocemod.2022.102151

Portillo Juan, N.;Matutano, C. y Negro Valdecantos, V. (2023). Uncertainties in the application of artificial neural networks in ocean engineering. *Ocean Engineering, 284*, 115193. doi:https://doi.org/10.1016/j.oceaneng.2023.115193

Portillo Juan, N. y Negro Valdecantos, V. (2022). Review of the application of Artificial Neural Networks in ocean engineering. *Ocean Engineering, 259*, 111947. doi:https://doi.org/10.1016/j.oceaneng.2022.111947

Portillo Juan, N.;Negro Valdecantos, V. y del Campo, J. M. (2022a). Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast. *Water, 14*(21), 3453. Retrieved from https://www.mdpi.com/2073-4441/14/21/3453

Portillo Juan, N.;Negro Valdecantos, V. y del Campo, J. M. (2022b). Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain. *Sustainability, 14*(12), 7507. Retrieved from https://www.mdpi.com/2071-1050/14/12/7507

Puertos del Estado. Oceanografía. Registros históricos de oleaje. Retrieved from https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx Acessed on 19/09/2022

Romano-Moreno, E.;Diaz-Hernandez, G.;Lara, J. L.;Tomas, A. y Jaime, F. F. (2022). Wave downscaling strategies for practical wave agitation studies in harbours. *Coastal Engineering, 175*. doi:10.1016/j.coastaleng.2022.104140

Rossum, G. V. (1989). Python (Version 3.10). Retrieved from https://www.python.org/downloads/windows/

Rusu, L.;Pilar, P. y Guedes Soares, C. (2008). Hindcast of the wave conditions along the west Iberian coast. *Coastal Engineering, 55*(11), 906-919. doi:https://doi.org/10.1016/j.coastaleng.2008.02.029

Rute Bento, A.;Martinho, P. y Guedes Soares, C. (2015). Numerical modelling of the wave energy in Galway Bay. *Renewable Energy, 78*, 457-466. doi:10.1016/j.renene.2015.01.024

Sadeghifar, T.;Lama, G. F. C.;Sihag, P.;Bayram, A. y Kisi, O. (2022). Wave height predictions in complex sea flows through soft-computing models: Case study of Persian Gulf. *Ocean Engineering, 245*, 110467. doi:https://doi.org/10.1016/j.oceaneng.2021.110467

Salah, P.;Reisi-Dehkordi, A. y Kamranzad, B. (2016). A hybrid approach to estimate the nearshore wave characteristics in the Persian Gulf. *Applied Ocean Research, 57*, 1-7. doi:10.1016/j.apor.2016.02.005

Standford University Machine Learning Course. Retrieved from https://www.coursera.org/learn/machine-learning/home Acessed on

Tolman, H. L. (1991). A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. *Journal of Physical Oceanography, 21*(6), 782-797.

Umesh, P. A. y Behera, M. R. (2020). Performance evaluation of input-dissipation parameterizations in WAVEWATCH III and comparison of wave hindcast with nested WAVEWATCH III-SWAN in the Indian Seas. *Ocean Engineering, 202*. doi:10.1016/j.oceaneng.2020.106959

Umesh, P. A. y Behera, M. R. (2021). On the improvements in nearshore wave height predictions using nested SWAN-SWASH modelling in the eastern coastal waters of India. *Ocean Engineering, 236*. doi:10.1016/j.oceaneng.2021.109550

Umesh, P. A.;Swain, J. y Balchand, A. N. (2018). Inter-comparison of WAM and WAVEWATCH-III in the North Indian Ocean using ERA-40 and QuikSCAT/NCEP blended winds. *Ocean Engineering, 164*, 298-321. doi:https://doi.org/10.1016/j.oceaneng.2018.06.053

Vanem, E.;Zhu, T. y Babanin, A. (2022). Statistical modelling of the ocean environment – A review of recent developments in theory and applications. *Marine Structures, 86*, 103297. doi:https://doi.org/10.1016/j.marstruc.2022.103297

Vieira, F.;Cavalcante, G.;Campos, E. y Taveira-Pinto, F. (2020). A methodology for data gap filling in wave records using Artificial Neural Networks. *Applied Ocean Research, 98*. doi:10.1016/j.apor.2020.102109

Wang, N.;Chen, Q.;Zhu, L. y Sun, H. (2022). Integration of data-driven and physics-based modeling of wind waves in a shallow estuary. *Ocean Modelling, 172*, 101978. doi:https://doi.org/10.1016/j.ocemod.2022.101978

Wang, R.;Kalnay, E. y Balachandran, B. (2019). Neural machine-based forecasting of chaotic dynamics. *Nonlinear Dynamics, 98*(4), 2903-2917. doi:10.1007/s11071-019-05127-x

Yang, C.-H.;Wu, C.-H.;Hsieh, C.-M.;Wang, Y.-C.;Tsen, I. F. y Tseng, S.-H. (2021). Deep Learning for Imputation and Forecasting Tidal Level. *Ieee Journal of Oceanic Engineering, 46*(4), 1261-1271. doi:10.1109/joe.2021.3073931

Zhang, R.;Corte-Real, J.;Moreira, M.;Kilsby, C.;Burton, A.;Fowler, H. J., . . . Sampaio, E. (2019). Downscaling climate change of mean climatology and extremes of precipitation and temperature: Application to a Mediterranean climate basin. *International Journal of Climatology, 39*(13), 4985-5005. doi:10.1002/joc.6122

3.9. A new methodology for the study of future maritime climate using Artificial Neural Networks and Monte Carlo simulations under the context of climate change

Abstract and Keywords

This paper proposes a new model to study future coastal maritime climate under climate change context. This new model combines statistical analysis, Monte Carlo simulations and Artificial Neural Networks (ANNs). Statistical analysis and Monte Carlo simulations are used to extrapolate future wave climate under climate change context at a regional level and ANNs are used to propagate these projected sea states obtained in deep water to the coast. The use of ANNs allows for the utilization of large amounts of data at a very low computational cost, and the use of Monte Carlo simulations enables the generation of future climate change projections at a regional level. The combination of the two methodologies results in a very accurate (MSE of 0.02 m and 1 s) and computationally inexpensive hybrid model that allows projections of coastal maritime climate considering climate change. This new methodology has been validated and applied in the Western Mediterranean for the long-term regime and for extreme events, obtaining increases in extreme events up to 1.5 m in wave height and up to 1.8 s in wave period by 2050.

Keywords: Artificial Neural Networks; Monte Carlo Simulations; Wave period; Hybrid model; Mediterranean Sea; Climate change

1. Introduction

Climate change is one of the foremost concerns today. The increase in greenhouse gas (GHG) emissions is causing global warming, which is affecting the ocean and cryosphere (Portillo Juan et al., 2022b). This increase of temperatures is provoking sea level rise (SLR) due to ocean thermal expansion and the thawing of polar regions and high mountain areas. Another consequence of the rising temperatures is the increment of both, the magnitude, and the frequency of extreme events (Amarouche & Akpinar, 2021; Vousdoukas et al., 2017; Vousdoukas et al., 2016). This is causing, at the same time, negative impacts in water quality, ecosystems, and cultures; great damages in human infrastructures, and social and economic losses (Abram et al., 2019; Bindoff et al., 2019; Collins et al., 2019; Hock et al.,

2019; Meredith et al., 2019; Oppenheimer et al., 2019). Therefore, given its damaging potential, scientific interest on climate change has grown exponentially over time since the 2000s (Hewawasam & Matsui, 2019; Klusak et al., 2023). Considerable scientific effort has been dedicated to study climate trends and make projections (Casas-Prat & Sierra, 2010, 2012; De Leo et al., 2021; Dieng et al., 2021; Yerdelen & Abdelkader, 2021). Interest on studying wave climate and making climate change projections is still growing and two main approaches can be distinguished in these studies.

The first approach to study wave climate and make future projections involves studying historical records or reconstructed wave data and analysing their trends to extrapolate them into the future. Hemer et al. (2007) were among the pioneers of this approach. Casas-Prat and Sierra (2010) applied it to analyse the wave storminess trends along the Catalan Coast and two years later investigated wave direction trends (Casas-Prat & Sierra, 2012). Wang et al. (2012) studied wave climate trends in the North Atlantic. Stefanakos and Vanem (2018) demonstrated the advantages of this approach and used nonstationary fuzzy modelling for wind and wave future projections. De Leo et al. (2020) also used this technique to detect and quantify trends in wave climate in the Mediterranean Sea. Amarouche and Akpinar (2021) used SWAN, Man-Kendall test, and Theil-Sen slope estimator to study the trends on storm wave intensity in the Western Mediterranean. This approach has been used in many studies due to its simplicity and low computational cost. However, it has also faced criticism for its perceived limitations, particularly in modelling extreme weather events. This criticism mainly comes from researchers who prefer the second approach, the emission scenarios approach (Camus et al., 2017; Casas-Prat & Sierra, 2013; Perez et al., 2015; Rueda et al., 2016; Vieira et al., 2020).

The emission scenarios approach is the traditional and most extended approach to study future wave climate. It is based on different emission scenarios that were defined by the Intergovernmental Panel on Climate Change (IPCC, 2000). The 5th assessment report of the IPCC (AR5) (A. IPCC, 2013) introduced four different emission scenarios referred to as Representative Concentration Pathways (RCPs) which became widely utilized in scientific studies. However, in their last report, the 6th assessment report (AR6), Masson-Delmotte et al. (2021) updated these scenarios and substituted them by Shared Socioeconomic Pathways (SSPs). Since the RCPs were introduced, the emission scenarios approach has been the most widespread method to project future wave climate among the scientific community.

Nonetheless, this method comes with a considerable computational burden due to the necessity of employing both, Global Circulation Models (GCMs) and Regional Circulation Models (RCMs). The utilization of diverse GCMs/RCMs yields distinct projections, making it necessary to consider various combinations of these models. Consequently, the requirement to evaluate multiple combinations substantially increases the computational cost associated with this technique. Lionello et al. (2008) were one of the first authors to use this approach. Charles et al. (2012) also applied different emission scenarios to forecast the evolution of wave heights and directions in the Bay of Biscay. Since then, different authors have preferred this approach.

The studies employing the emission scenarios approach utilize the concept of transfer learning, which involves deriving wave parameter values at one buoy from another buoy. This practice, also known as wave downscaling in ocean and coastal engineering, aims to extrapolate the wave climate observed at deep-water buoys to coastal buoys. Transfer learning between buoys is crucial for two main reasons: first, coastal buoy data can be influenced by refraction and shoaling effects that may distort results; and second, climate change projections are typically conducted at a global scale, necessitating the use of deep-water buoy data to infer conditions at coastal locations (Nerea Portillo Juan et al., 2023; Romano-Moreno et al., 2022). These transfer learning techniques can be implemented using either numerical methods (dynamical downscaling) or statistical models (statistical downscaling).

In the context of numerical approaches, Casas-Prat and Sierra (2013) explored five combinations of regional-global circulation models to forecast future wave climate in the Mediterranean coast. Vousdoukas et al. (2016) similarly used numerical models to predict extreme storm surge levels across European coastlines. Finally, Vieira et al. (2021) also utilized buoy transfer learning in the Iberian Peninsula.

On the other hand, in the statistical approach, Mori et al. (2013) utilized statistical methods to assess wave climate changes along the Japanese coast. Perez et al. (2015) proposed a statistical model based on Weather Types (WTs) for transfer learning (wave downscaling). Later, Rueda et al. (2016) extended this methodology to estimate maximum wave heights, while, Camus et al. (2017) applied statistical transfer learning to assess coastal impacts.

This approach based on emission scenarios to study future maritime climate involves significant computational costs and considerable uncertainty. This method relies on GCMs to obtain large-scale climate projections, which are

influenced by projected external forcings based on emission scenarios. In this approach there are many variables that add complexity and uncertainty to the model; including the choice of the emission scenario, the selection of GCMs, the definition of the initial conditions, or the parametrization of the climate models (Stefanakos & Vanem, 2018). To obtain wave climate projections from these atmospheric models, higher resolution regional climate projections of wind fields and various downscaling and transfer learning methods are required. These regional models introduce additional sources of uncertainty. Moreover, the need of using different combinations of GCM-RCMs results in a substantial computational cost. The alternative approach involves studying historical data series and trends and extrapolate these trends for future wave climate. However, as stated before, scientists typically reject this method due to its reliance on extensive historical data series for reliable future predictions. Furthermore, the reliability of extrapolations diminishes as they extend further into the future.

Nonetheless, in light of the significant expansion and advancement in data-driven models and Artificial Intelligence (AI) techniques (Pan & Xue, 2023; Peng et al., 2023; Slimani et al., 2024; Taha et al., 2022; Xu et al., 2022), there is now a promising avenue for addressing the challenges associated with the traditional emission scenarios approach by using trend analysis combined with AI.

Over the past few years, the evolution of AI has been remarkable, merging as a promising tool that reshapes traditional methodologies in many different fields. Ocean and maritime engineering are one of those fields in which AI has irrupted in the last years. However, compared to other scientific realms, such as medicine or analogous disciplines like water resources, it becomes evident that the level of AI development in our specific field lags behind. Nevertheless, the development of AI in our field is very promising, and the studies being conducted are yielding highly encouraging results (Nerea Portillo Juan et al., 2023).

Between 2010 and 2013, approximately 20 studies per year in the field of oceanography, utilizing AI techniques, were indexed in the Web of Science database. This figure increased until 50 studies per year from 2015 to 2017 and since then the number of studies has been doubling each year, with 179 studies in 2018, 395 in 2019, 535 in 2020, 672 in 2021 and staggering 1329 in 2022 (Figure 1). This data clearly reflects a significant and escalating trend, underscoring the growing inclination among scientists to embrace AI methodologies, and signifying a potential breakthrough in our field.

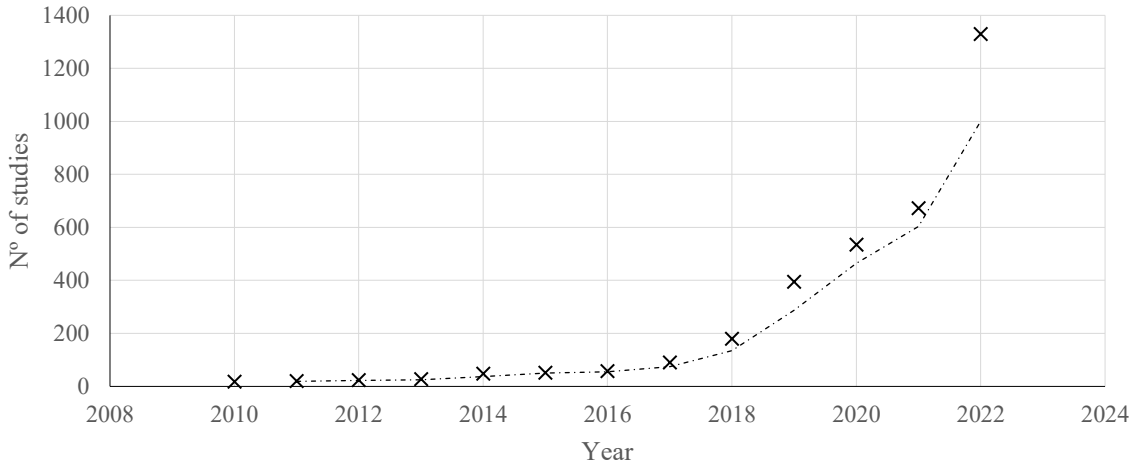


Figure 1. Evolution of the number of research studies that apply AI techniques in ocean and maritime engineering.

The significant advancements in AI capabilities enable more efficient data exploitation at a lower computational cost. It seems that these AI developments can address the limitations of the trend analysis approach. Now, with AI, we can apply AI techniques to extract insights from data and extrapolate trends. AI methodologies offer the potential to conduct in-depth studies on wave records and provide robust predictions of future wave climates, avoiding the computational costs of using emission scenarios (Figure 2).

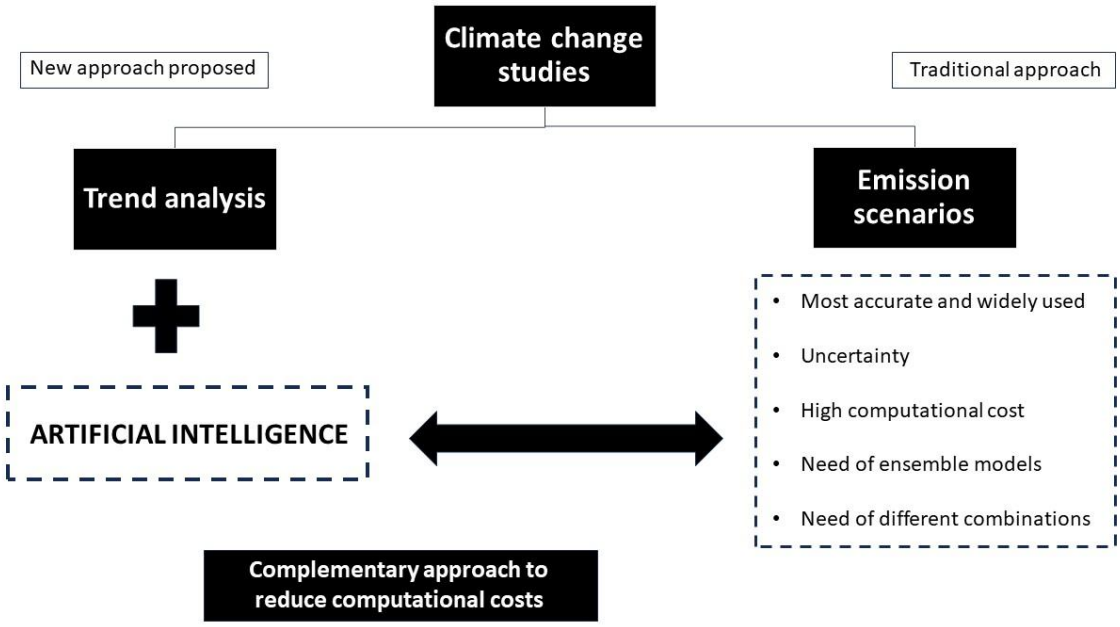


Figure 2. Comparison of the traditional approach and the new approach proposed for climate change studies.

Specifically talking about the different AI methodologies in ocean engineering, Fuzzy Inference Systems (FIS), Adaptive NeuroFuzzy Inference Systems (ANFIS), Artificial Neural Networks (ANNs) and Random Forests (RF) are probably the most widely spread AI-techniques.

FIS and ANFIS models are data-driven models that apply fuzzy logic and if-then rules, they were firstly applied by Kazeminezhad et al. (2005) and since then, different authors have used this methodology to predict wave climate (Kazeminezhad et al., 2005; Stefanakos, 2016; Stefanakos & Vanem, 2018; Zamani et al., 2008). RF is an algorithm in which trees are created, with the root node accessing a diverse bootstrap sample of the original dataset (Sadeghifar et al., 2022). Campos et al. (2021); Kim et al. (2019); Lecacheux et al. (2021); Sadeghifar et al. (2022); Su et al. (2019) are some of the examples that applied RF to ocean engineering studies.

ANNs are one the most popular soft-modelling techniques in ocean engineering, and therefore, the data-driven model that will be used in this research. Several authors have used ANNs to predict maritime climate: Agrawal and Deo (2004); Gunaydin (2008); López and Iglesias (2015); Londhe et al. (2016); Abhigna et al. (2017) and Vieira et al. (2020), among others. See Portillo Juan and Negro Valdecantos (2022) for more detail.

Recent studies in the field of artificial neural networks (ANNs) and ocean engineering have predominantly utilized deep hybrid networks to predict significant wave height, demonstrating promising outcomes. Duran-Rosal et al. (2017) were among the pioneering authors in this area. Atteia et al. (2022) developed a deep-learning-based hybrid model using satellite Synthetic Aperture Radar (SAR) to predict significant wave height. More recently, Gao et al. (2023) and Marangoni Gazineu Marinho Pinto et al. (2023) also introduced deep hybrid models for significant wave height estimation. In the current year, Ahmed et al. (2024) proposed a novel hybrid forecast system called CLSTM-BiGRU, integrating Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Bidirectional Gated Recurrent Unit (BiGRU). Patanè et al. (2024) used Convolutional Neural Networks (CNNs) along with short memory layers for their wave height predictions. Raj and Prakash (2024) designed a hybrid deep learning model combining CNNs and Bidirectional Long-Short Term Memory (BiLSTM) with Multivariate Variational Mode Decomposition (MVMD). Lastly, Saviz Naeini and Snaiki (2024) employed deep autoencoders and deep neural networks for significant wave height forecasting.

While ANNs have emerged as prominent data-driven models in ocean engineering research, particularly for wave height predictions in recent years, there is still considerable room for advancement. Although there are several studies that characterize wave climate using ANNs, very few incorporate the effect of climate change. Given the current circumstances, integrating climate change into wave characterization studies is imperative and failing to incorporate it risks hindering further progress in AI-based models. To the best of the author's knowledge, this research represents a pioneering effort proposing a methodology to integrate climate change into ANN models in ocean and coastal engineering. It introduces a novel methodology that applies Monte Carlo simulations to generate future climate scenarios that will feed the AI model, harnessing the strengths of both techniques. The Monte Carlo simulations facilitate data generation, while the ANN data-driven model effectively exploits this data. This approach allows for the study of future wave climate and climate prediction using AI models, which are far more computationally efficient than traditional models employed in our field.

Furthermore, by integrating these two techniques, significant differences were observed between the wave periods generated virtually and the actual wave periods, leading to decisive conclusions about the challenges of modelling wave periods with analytical equations. This finding is crucial for future studies, as it suggests that relationships used in previous studies may no longer hold under changing climate conditions.

The main objective of this paper is to present a complementary approach to the current established methodology, based on emission scenarios, that reduces computational costs by taking advantage of the great potential that AI is demonstrating. This alternative approach not only reduces computational costs, but also serves as a valuable tool for preliminary studies, offering a more efficient and less time-intensive complement to the existing methodology. This advancement will enable climate change assessments at local scales on a broader scale, rather than being limited to specific and critical cases, as has been the norm due to previous computational constraints. This shift is crucial, especially considering the worsening consequences of climate change each year.

The proposed model has been developed and validated in the Western Mediterranean Sea, demonstrating its applicability in obtaining maritime climate in shallow waters under climate change.

2. Methodology

The methodology outlined in this paper comprises three main steps. Firstly, an examination of historical maritime climate records and their statistical distributions is conducted. Using statistical and regression techniques, future projections are generated by determining the parameters of maritime climate distributions from the present until 2050. Secondly, a virtual database, basing on the trend analysis previously conducted, is constructed using Monte Carlo simulations. This database will we combined with the historical data records to feed the AI model designed in the final step of the methodology. This approach results in an AI model with robust data mining capabilities, fed by both, actual historical data and future projections, that can be effectively utilized to obtain maritime wave data in coastal locations across different time horizons (Figure 3). This section is structured in three main parts. Section 2.1. Study area and data, Section 2.2. Monte Carlo simulations and Section 2.3. ANN model.

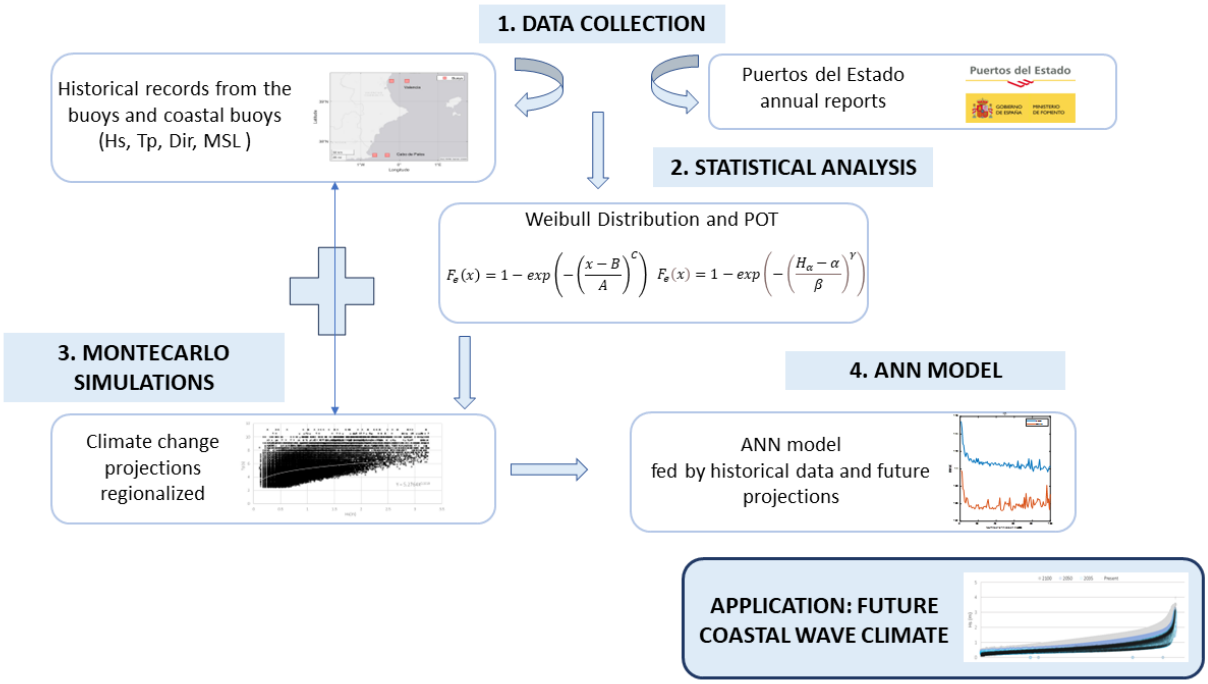


Figure 3. New methodology proposed.

b. Study area and data

The area of study is located in the Southeast (SE) of Spain. Due to its geography and orography, the wave and atmospheric patterns here are affected by regional and local phenomena (CAVALERI & BERTOTTI, 2004). Additionally, this area is projected to be significantly impacted by climate change; evidenced by increasing

occurrences of floods, more frequent heat waves, and intensifying extreme events that pose serious threats to coastal protection structures (Portillo Juan et al., 2022b). Most of the studies on this area have focused on specific parts, such as the Ebro Delta or the area of Catalonia, leaving many areas understudied (Portillo Juan et al., 2022a). Therefore, this research will concentrate on the least studied portion of the Spanish Mediterranean coast, which encompasses the coastal stretch from Valencia to Cabo de Palos (Figure 4). This area represents an important focus for understanding the local climate dynamics and assessing the potential impacts of climate change on coastal environments.

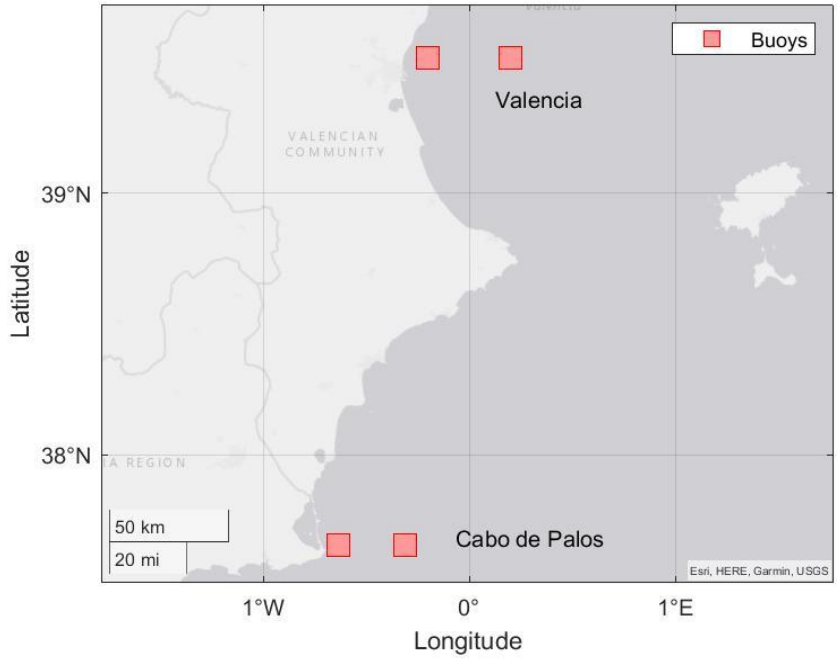


Figure 4. Location of the buoys used in the study.

In this study, buoy records were utilized as the primary data source. Although instrumental records are recognized as the most reliable data source, challenges often arise concerning their availability, spatial and temporal coverage, and inconsistencies over time due to updates in measuring devices, calibration techniques, or device failures. However, in this case, the instrumental records obtained from the buoys proved to be sufficiently accurate. Notably, there were abundant records, both temporally and spatially, as each buoy contributed more than 60,000 data samples (Table 1).

The deep-water and coastal buoys of Valencia and Cabo de Palos were used. Historical hourly wave height (Hs), peak period (Tp) and Direction (Dir) records were downloaded from the portal of Spain's State Ports (Puertos del Estado.

Oceanografía. Registros históricos de oleaje). While waves can be characterized using various parameters such as wave height, wave length, wave celerity, or wave frequency among others; ocean and coastal engineering designs typically focus on Hs, Tp and Dir to define maritime climate. Coastal and ocean engineers rely heavily on these parameters (Hs, Tp and Dir) when designing coastal protection structures and ports. Therefore, we chose to use these specific variables in our study to accurately characterize the wave conditions relevant to our engineering analysis.

Table 1. Available data in the buoys.

Buoy	H_s	T_p	Dir	Series availability	Temporal resolution
Valencia	X	X	X	15/09/2005 to 14/07/2022	Hourly
Coastal – Valencia	X	X	X	08/06/2005 to 30/10/2013	Hourly
Cabo de Palos	X	X	X	18/07/2006 to 14/07/2022	Hourly
Coastal – Cabo de Palos	X	X	-	14/11/1985 to 18/06/2012	Hourly

In addition, given the focus on studying and predicting maritime climate within the context of climate change, we also incorporated Mean Sea Level (MSL) to characterize future wave climate. The MSL records were also sourced from Spain's State Ports portal (*Puertos del Estado. Oceanografía. Registros históricos de oleaje*).

It is worth noting that another buoy, the Alicante buoy, is also available in this area. However, the Alicante buoy was excluded from the model development due to its shallow depth. Shallow depths can influence final results of wave measurements due to refraction and shoaling phenomena, potentially introducing inaccuracies into the analysis. Therefore, to ensure the accuracy and reliability of the model, data from the Alicante buoy were not included in this study.

To address the challenge posed by measurement inconsistencies and the absence of homogeneity in the data series, a data pre-processing step was undertaken. A specialized algorithm was crafted to detect and subsequently eliminate any temporal gaps in the dataset. Then, all the data were filtered to eliminate false values and errors. All wave height values under 0.1 m and periods under 1 s were removed. In cases where buoys fail to record data accurately, they often produce

very low values. Thus, these data points were deemed potential inaccuracies likely caused by this failure of the buoys and measurement noise.

After eliminating these time gaps and false values, the next step involved interpolating the missing values to achieve a consistently hourly homogeneous dataset. For shorter gaps, such as those spanning one day, interpolation occurs between adjacent data records. In the case of larger data gaps, SIMAR-WANA data points—virtual buoys provided by the Spanish Ports Authority—were employed to obtain the missing values.

It is important to explain that SIMAR-WANA data points are the official virtual buoys of the Spanish Government (*Puertos del Estado. Oceanografía. Puntos SIMAR-WANA*). The SIMAR-WANA dataset comprises two distinct subsets: SIMAR, covering the period from 1958 to 2005, and WANA, spanning from 2006 to the present day. The SIMAR data consist of simulations from atmospheric and wave models that cover the entire Spanish coast, using data from the AEMET (*AEMET. Agencia Estatal de Meteorología*). On the other hand, the WANA data originate from a sea state prediction system developed by Puertos del Estado itself in collaboration with AEMET (*AEMET. Agencia Estatal de Meteorología*). These data are diagnostic and analysis-driven, incorporating hourly assimilation, which means that the model provides wind and pressure fields consistent with the previous evolution of the modelled parameters. Both datasets use WaveWatch III model (Tolman et al., 2002) and give hourly data with a spatial resolution going from 0.25º to 0.05º. SIMAR-WANA virtual buoys represent an exceptional source of high-quality data, demonstrating superior performance for specific tasks compared to other virtual databases like Waverys or ERA5. (Juan Nerea et al., 2024).

In our data interpolation process, we opted to use linear regression for two primary reasons: it is straightforward and easy to implement, and we had access to virtual buoys (SIMAR-WANA data points), that facilitated interpolation using this simpler technique. However, it is important to highlight that ANNs are highly recommended for interpolation tasks. ANNs have shown excellence performance and, in some studies, they have even surpassed linear regression, particularly for addressing complex interpolation challenges (Ahmad, 2022; Alemam et al., 2023; Baek et al., 2020; Bagheri et al., 2014; Oda et al., 2012; Xie et al., 2018). Therefore, the authors of this study also recommend using ANNs for interpolation when virtual buoys are not available, especially in cases where significant time gaps

exist in the wave records. This approach can greatly enhance the accuracy and robustness of the interpolation model under challenging conditions.

After that, the matching of the buoys in deep water and coastal buoys was done using a time-pairing methodology, each observation of the buoy in deep water was paired with another observation of the coastal buoy by matching the hourly records of both buoys, obtaining the input-output pairs of the model. The record of the 1st of January of year 2000 at 00:00 of the deep-water buoy was paired with the record of the 1st of January of year 2000 at 00:00 of the buoy in shallow depth.

As a result of this first step, we obtained an hourly dataset from 2005 to 2022 for Valencia and from 2006 to 2022 for Cabo de Palos. With the dataset thoroughly pre-processed, the subsequent phase involved conducting a statistical analysis and executing Monte Carlo simulations.

c. Statistical analysis and Monte Carlo simulations

We compiled wave climate data spanning from 1985 to 2020 in Cabo de Palos (35 years) and from 2005 to 2020 in Valencia (15 years) from the reports of Puertos del Estado (Puertos del Estado. Banco de Datos Oceanográficos. Informes). We studied the parameters of wave distributions and their historical evolution to project values forward to 2050, examining both, long-term trends and extreme scenarios.

It is crucial to emphasize that longer time series data significantly enhance the accuracy and reliability of our projections, allowing for more robust long-term forecasts (IPCC, 2013; Meehl et al., 2007). We aim to project future wave climate up to the year 2050 – 30 years from 2020. However, in the case of Valencia, we only had 15 years of available records to study and analyse trends, which we considered was not enough to make accurate future projections (Amarouche & Akpinar, 2021; De Leo et al., 2021; NOAA. National Oceanic and Atmospheric Administration.). To address this limitation, we used Cabo de Palos buoy to aid in projecting the future wave distribution for Valencia. This approach was viable as the Cabo de Palos buoy provided a continuous 35-year time series and is geographically proximate to the Valencia buoy, enhancing the robustness of our projections.

Long-term wave regimes are typically described using a theoretical distribution function that fits wave data records corresponding to the middle or central zone of the histogram, reflecting average operating conditions. In Spain, the most commonly employed theoretical distributions for modelling long-term wave regimes include the Log-normal and Weibull distributions (Gómez et al., 2018).

The Log-normal distribution is often used for seas that exhibit right-skewness, whereas the Weibull distribution is preferred for seas with a diverse range of behaviours, such as the Mediterranean Sea (Informes anuales Oceanografía Puertos del Estado). Therefore, among these two distributions, the Spanish Government recommends using the Weibull distribution for modelling the long-term regime in the Mediterranean Sea. All statistics and data for the Spanish Mediterranean region are based on this distribution, which is why we have also chosen it for our work—to ensure consistency with existing records and to align with government recommendations (ROM03.91, 1992). Furthermore, the majority of researchers working in this field also utilize the Weibull distribution (Camus et al., 2019; Casas-Prat & Sierra, 2010; De Leo et al., 2021; Lionello et al., 2008; Portillo Juan et al., 2022a; Sánchez-Arcilla et al., 2008; Sierra et al., 2015).

Weibull distribution is characterized by 3 parameters (Equation 1): the scale parameter (A), the centring parameter(B) and the shape parameter (C) .

$$F_e(x) = 1 - exp(-((x - B)/A)^C) \quad [1]$$

Where the input x is the wave height and the output F_e (x) is the probability of no exceeding this value.

The values of A, B and C for the historical records were given by Puertos del Estado official reports (Informes anuales Oceanografía Puertos del Estado) and to obtain the values of A,B and C, from 2024 to 2050, we applied polynomial regression (Figure 5). The goodness of fit evaluation of the polynomial regression is presented in Table 2.

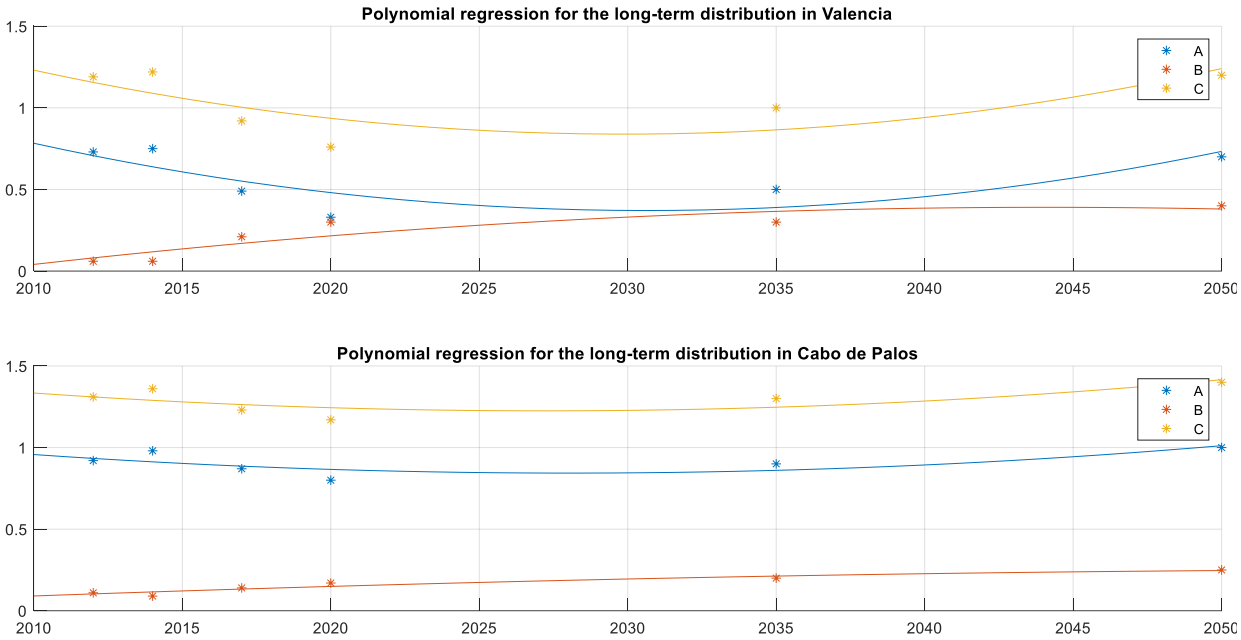


Figure 5. Statistical adjustment of parameters A, B and C for long-term distribution.

Table 2. Goodness of fit evaluation for the long-term distribution.

Parameters	R in Valencia	R in Cabo de Palos
A	0.8	0.77
B	0.91	0.96
C	0.75	0.77

For the extreme distribution we also chose a Weibull distribution to model sea states (Equation 2). This decision was based on Table 3 (Gómez et al., 2018), which outlines criteria for selecting distributions based on their domain of attraction for maximums and minimums.

Table 3. Selection criterion for extreme distributions. Taken from Castillo; and Pruneda (2001) and Gómez et al. (2018).

Distribution	**Domain of attraction**	
	For maximums	For minimums
Normal	Gumbel	Gumbel
Exponential	Gumbel	Weibull
Log-Normal	Gumbel	Gumbel
Gamma	Gumbel	Weibull

Maximum Gumbel	Gumbel	Gumbel
Minimum Gumbel	Gumbel	Gumbel
Rayleigh	Gumbel	Weibull
Uniform	Weibull	Weibull
Maximum Weibull	Weibull	Gumbel
Minimum Weibull	Gumbel	Weibull
Maximum Frechet	Frechet	Gumbel
Minimum Frechet	Gumbel	Frechet

Apart from these distributions, there are also other commonly employed in ocean and coastal engineering, such as the Peak Over the Threshold (POT) (Hosking & Wallis, 1987) or Pareto – Poisson (Pickands, 1975).

Based on the criteria outlined in Table 3, we selected the Weibull distribution to model extreme sea states, aligning with its use for long-term regime modelling and specifically for maximum values. This decision is reinforced by the recommendation of the Spanish Government, which also advocates for the use of the Weibull distribution to model extreme events in the Mediterranean Sea. Moreover, all official statistics and data for the Spanish Mediterranean region are based on this distribution (Informes anuales Oceanografía Puertos del Estado; ROM03.91, 1992). The Weibull distribution's suitability for extreme modelling is attributed to its flexibility, which allows it to accommodate a wide range of extreme behaviours, particularly relevant in the context of the Mediterranean Sea. Unlike the Gumbel distribution, which is often preferred for progressively behaving extremes, or the Frechet distribution, which is suitable for modelling rare large extreme events, the Weibull distribution offers versatility and is commonly employed in reliability assessments within the engineering field. Furthermore, the widespread adoption of the Weibull distribution in research is evidenced by numerous studies (Abadie et al., 2019; Avila et al., 2020; Casas-Prat & Sierra, 2013; Jebbad et al., 2022; Mase et al., 2013; Portillo Juan et al., 2022a; Sanchez-Arcilla et al., 2016; Vieira et al., 2020).

Extreme Weibull distribution is characterized by 3 parameters: the threshold (α), the scale parameter (β), and the shape parameter(γ).

$$F_e(x) = 1 - exp(-((H_\alpha - \alpha)/\beta)^\gamma) \quad [2]$$

Where the input H_α is the wave height and the output $F_e(x)$ is the probability of no exceeding this value.

It is important to note that although Equation 1 and 2 look very similar they refer to different concepts, and the parameters are different. While A,B , C and Equation 1 refers to the long-term distribution; α,ß, γ and Equation 2 refers to the extreme distribution. We chose these distributions and equations because they are the ones defined by the Spanish authority to characterize the maritime climate (Informes anuales Oceanografía Puertos del Estado).

To study and make projections for the extreme long-term distribution, the 95th percentile of the buoy records were taken as reference, that corresponds to a return period of 20 years according to the DNV standards (Hopstad et al., 2018) and Puertos del Estado reports (Informes anuales Oceanografía Puertos del Estado). This return period is of particular importance because is the return period that the consortium establishes to define extreme events and above which it covers the costs of the damage caused in maritime structures. We applied the same methodology used for the long-term distribution analysis, employing parameters specified by Puertos del Estado for the period spanning 2005 to the present. To project values for future years, we employed polynomial regression. The results for the extreme distribution are presented in Figure 6 and Table 4.

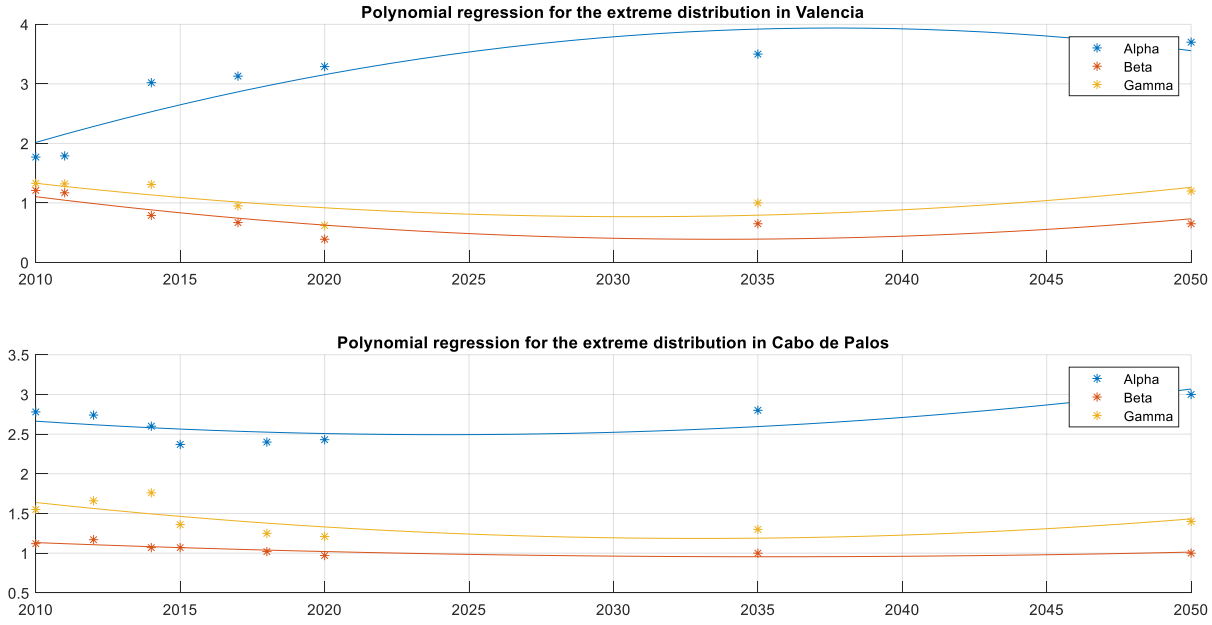


Figure 6. Statistical adjustment of parameters α, ß and γ for extreme distribution.

Table 4. Goodness of fit evaluation for the extreme distribution.

Parameters	R in Valencia	R in Cabo de Palos
α	0.90	0.79
β	0.83	0.84
γ	0.77	0.70

After conducting an analysis of the future wave climate and estimating the three parameters that characterize long-term and extreme wave distributions, Monte Carlo simulations were employed to generate the virtual dataset representing the future wave climate.

Various techniques are available to generate virtual datasets, with some of the most widely used in ocean and coastal engineering including bootstrapping sampling, Generative Adversarial Networks (GANs), Markov Chain Methods, Variational Auto-Encoders (VAEs), and Monte Carlo simulations.

Bootstrapping sampling use a dataset and resample this dataset with replacement from the own observed data (Bohlinger et al., 2019). This method allows to obtain the distribution of the dataset (Paulig & Okhrin, 2024). However, since our objective is to generate a dataset with a specified distribution rather than merely estimating an existing distribution, we opted not to use this technique.

GANs, introduced by Goodfellow et al. (2014), are deep learning models consisting of a generator and a discriminator. The generator produces virtual data while the discriminator distinguishes between real and virtual data. Although GANs aim to achieve optimal behaviour in both networks without distinguishing between real and virtual datasets (Li et al., 2024; Oh et al., 2021; Yi & Chen, 2023; Zhang et al., 2023), our objective is precisely to differentiate between the real dataset, that corresponds to the present wave climate, and the virtual dataset, that corresponds to the future wave climate; making GANs less suitable for our purpose.

Markov-Chain (MC) is a stochastic process used to model transitions between different states. It defines a set of probabilities for the next states (transitions) in a way that the probability of a given state depends only on the previous state (Grinstead & Snell, 1997; Kolios et al., 2023). Since MC methods rely on previous states for predictions, and we sought to generate data based on historical records, trend analysis, and statistical distribution with a broader approach, we discarded this methodology.

VAEs are also composed of two neural networks; the encoder, which reduces vector input dimensions and the decoder, which reconstructs the input. VAEs also implement probabilistic generation data, obtaining as an output a whole distribution of data instead of only one data sample (Bedja-Johnson et al., 2022; Xu et al., 2023; Yang et al., 2024). While powerful, VAEs are complex and involve computationally intensive training processes. Our objective is to reduce computational cost. Therefore, we favoured Monte Carlo simulations over VAEs for our objectives, which are presented in the next paragraph.

Monte Carlo simulations were originally proposed by Stanislaw Ulam and John von Neumann in the 1940s (Georgescu, 2023; Harrison, 2010; Ulam & Rabi, 2019) and they are a computational algorithm rooted in random sampling. This technique is widely employed to model the probability of diverse outcomes in processes that need precise definition due to the involvement of random variables. . A summary of this methodology can be found in Díaz et al. (2022). They outlined six key steps in its implementation:

1) Problem definition: Articulating the problem in terms of random variables, the main variables that define the process are identified and random values are assigned.

2) Probabilistic characterization: Quantifying the probabilistic characteristics of the identified random variables, this step entails characterizing the distribution of these variables.

3) Variable generation: The random variables that define the problem are transformed into real variables following the previously defined distribution.

4) Deterministic Evaluation: The problem is evaluated deterministically for each set, and a dataset is compiled.

5) Probabilistic Information Extraction: Extracting probabilistic information from the simulations and characterizing the new virtual dataset.

6) Process Accuracy Evaluation: Finally, the accuracy of the entire process is assessed through validation procedures.

In summary, Monte Carlo simulations were selected for generating the virtual dataset due to their suitability for modelling complex scenarios, efficient computation, and alignment with the objectives of our study regarding future wave climate analysis and scenario generation.

In this case, it was decided to run as many simulations as those corresponding to the time span between the last year of available records (2022) and the year 2050, which is the last year for which coastal maritime climate predictions are to be made, resulting in 227,760 simulations that correspond to the hourly records of the virtually generated maritime climate database. This criterion was decided to be consistent with the historical records, so that we have a real database taken by the buoys for the present maritime climate and the dataset generated by the Monte Carlo simulations for the future maritime climate, so that both datasets will later feed our AI model.

The simulations were conducted a total of 20 times to validate and assess result accuracy. To confirm the reliability of the simulations, a sensitivity analysis was executed. The findings revealed minimal variability in the key parameters defining the wave climate (Hs, Tp, and Dir). As a result, Monte Carlo simulations were deemed sufficiently valid and accurate for further analysis Table 5 shows the variation of the mean, maximum, minimum, standard deviation, and median across the 20 iterations tested for the entire virtual dataset.

Table 5. Variability of Hs, Tp and Dir trough the iterations of Monte Carlo simulations.

Variability	Hs (m)	Tp (s)	Dir (
Mean	± 0.005 m	± 0.005 s	± 0.1º
Maximum	± 0.6 m	± 0.2 s	± 15º s
Minimum	± 0.0002 m	± 0.001 s	± 0.03º s
Standard deviation	± 0.002 m	± 0.002 s	± 0.08º s
Median	± 0.002 m	± 0.004 s	± 0.08º s

The main variables chosen to model the process were Hs, Dir and MSL.

For Hs and Dir, we assigned them random values between 0 and 1. These values between 0 and 1 means the probability of occurrence. These probabilities were transformed into the variables Hs and Dir by applying the distribution function,

the Weibull distributions previously defined for Hs, and the uniform distribution tailored to the specific swell rose of each location for Dir (Figure 7).

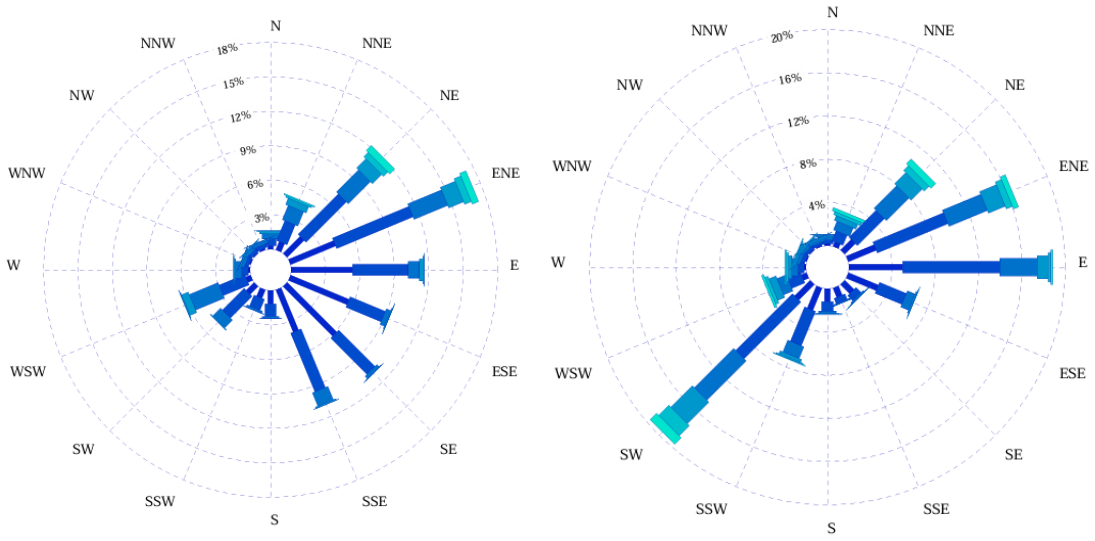


Figure 7. Swell roses in Valencia (left) and Cabo de Palos (right) deep buoys.

Once we obtained the series of values of the main variables, Hs and Dir, Tp was calculated by means of the relationship shown in Equation 3, which is the equation that the Spanish port authority proposes to model the relationship between Tp and Hs in this area (*Puertos del Estado. Oceanografía. Registros históricos de oleaje*).

$$T_p = a \times (H_s)^b \qquad [3]$$

The values of parameters a and b derived through an analysis of historical records from Valencia and Cabo de Palos, incorporating adjustments for both long-term and extreme distribution scenarios. These parameters were determined by fitting a trend line to the dataset of the pairs Hs-Tp using the formula $y = a \times (X)^2$, being Hs the X and Tp the Y.

Figure 8 shows the adjustment of the peak period and Table 6 the final values of a and b used for each case. The data points shown correspond to the Hs-Tp pairs from historical records of deep buoys in Valencia and Cabo de Palos.

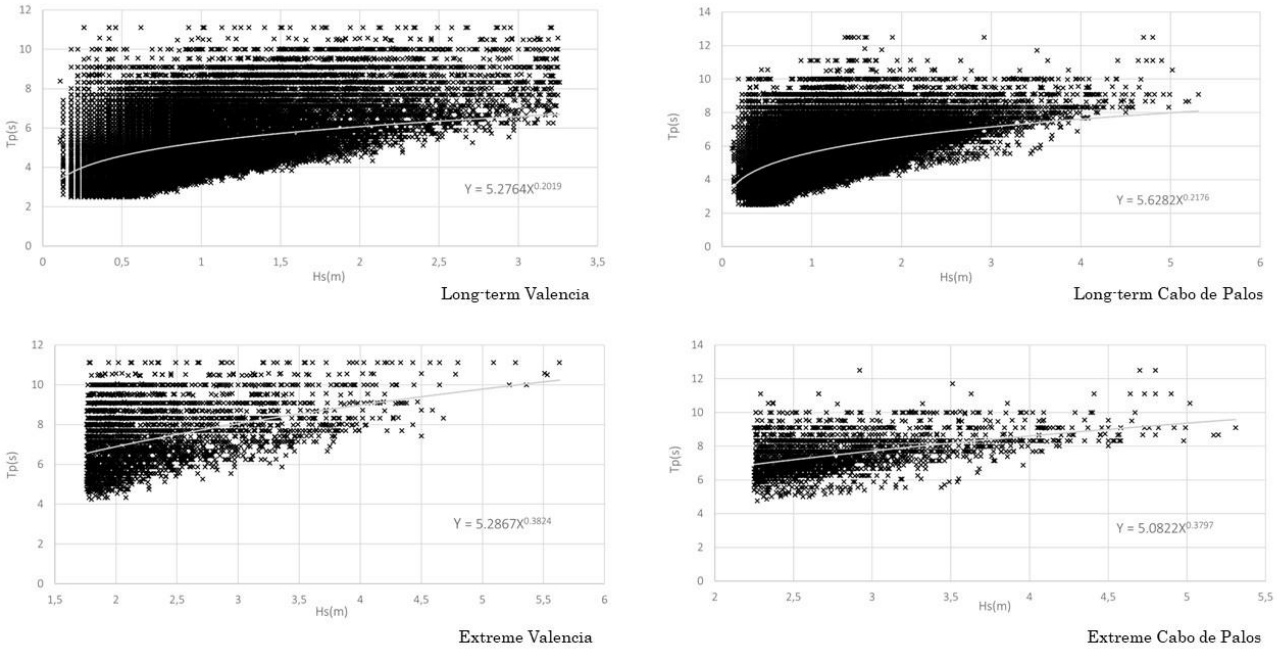


Figure 8. Hs-Tp relation.

We assumed that the relationship Hs-Tp remains constant in time and do not change.

Table 9. a and b parameters for Valencia and Cabo de Palos.

Parameters	Valencia		Cabo de Palos	
	Long-term	Extreme	Long-term	Extreme
a	5.28	5.29	5.63	5.08
b	0.2	0.38	0.22	0.38

For the MSL simulations, we considered an increase of Sea Level in 2050 of 0.25 m basing on previous studies (Luque et al., 2021; Portillo Juan et al., 2022b; Vitousek et al., 2017; Vousdoukas et al., 2017). Subsequently, we applied this increase progressively over time throughout the simulations, culminating in a rise of 0.25 meters by the target year of 2050.

Finally, the historical records and the virtual database were merged to create the dataset that would serve as input for the ANN model.

d. ANN model

Artificial Neural Networks (ANNs) are an AI-technique that try to mimic the human brain. They emulate the neurons by elements called nodes and the unions

between them by transfer functions. They are made of different layers: the input layer in which the known data is introduced into de model; the intermediate layers, known as hidden layers; and the output layer in which the final value searched is obtained. Each layer consists of different units (neurons or nodes) and each one is connected to the following layer by a transfer function (Basheer & Hajmeer, 2000; Krogh, 2008; Medina et al., 2004; Walczak, 2019; Yegnarayana, 2009).

Neural networks operate similarly to the human brain. They are trained using extensive datasets containing input-output pairs. By presenting specific inputs, the network generates corresponding outputs. Through this iterative process across numerous data samples, the model learns to discern relationships between variables and can subsequently predict new outcomes based on different inputs. In other words, we train the model to understand the correlations between the variables that define the problem. Once the ANN is trained, validated, and tested; the model can make new predictions (Nerea Portillo Juan et al., 2023). Figure 9 shows the scheme of an ANN applied to this specific study, where the superscript 0 means in deep water and the dw downscaled.

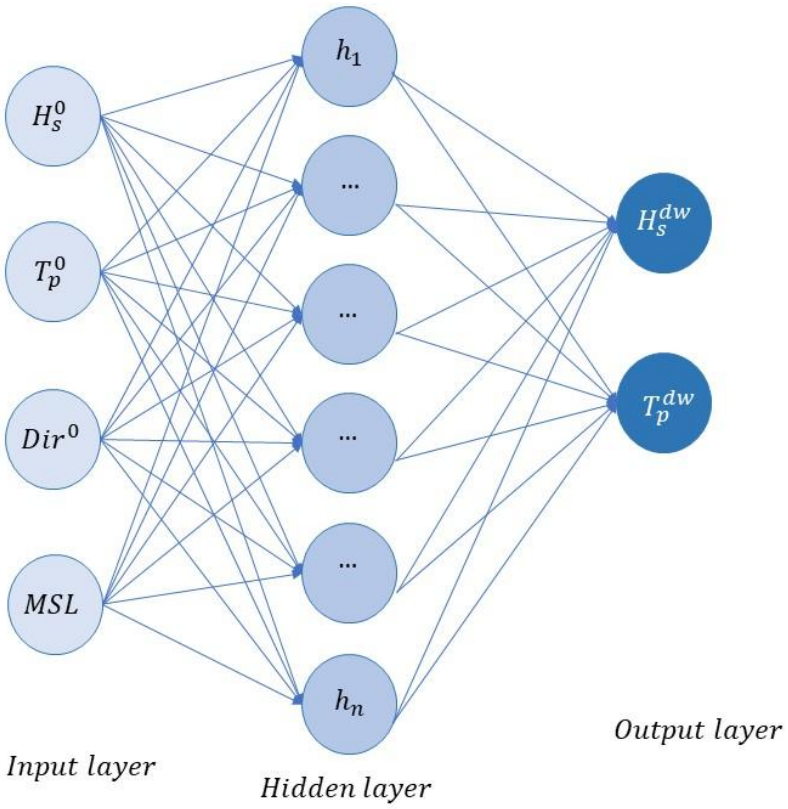


Figure 9. Scheme of the ANN.

The main steps to develop an ANN is to obtain the optimal architecture of the model, and the internal parameters and algorithm that perform best. However, one of the main problems of ANN models is the uncertainty inherent to this process, there is no clear and defined methodology to obtain the optimum parameters of an ANN and they are generally obtained by trial and error, which is a time consuming and precarious method (Nerea Portillo Juan et al., 2023). In the following paragraphs we will explained how we defined and optimized our ANN model.

Firstly, the type of neural network, the number of layers and the transfer functions of the model were selected based on scientific evidence (Portillo Juan & Negro Valdecantos, 2022). A 3-layer Feed Forward Multilayer Perceptron Neural Network (FFNNs) with the hyperbolic tangent sigmoid function between the input and the hidden layer and the linear transfer function between the hidden and output layer were used. Once these main parameters were selected, a trial-and-error approach was used to select the model architecture.

The number of input and output neurons is given by the physics of the problem, but the number of hidden neurons is not. The number of input neurons of the ANN was 4; wave height, peak period, direction and mean sea level (MSL) in the deep water buoys $\left(H_s^0, T_p^0, \theta^0, MSL\right)$ and the number of output neurons was 2: wave height and peak period in the coastal buoys $\left(H_s^{dw},\ T_p^{dw}\right)$, as they were the variables to be studied in this research paper.

The number of hidden neurons was optimized independently for each of the buoys. The two sub-models were trained a total of 40 times to remove the influence of the initialization weights and biases from the analysis.

To optimize the ANN architecture, we tested from 1 to 100 hidden neurons. The loss function to evaluate the performance of the models was the root mean square error (RMSE) (Equation 4).

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{k_{max}}\sum_{j=1}^{j_{max}^k}\left(h_j^k - \hat{h}_j^k\right)^2} \qquad [4]$$

where h_j^k and $\hat{h}_j^k$ denote the downscaled wave height after the jth test of the kth series and n the total number of tests of the dataset (Equation 5). h_j^k is the real output, which is the wave recorded by the coastal buoy, and $\hat{h}_j^k$ the wave predicted by the ANN.

$$n = \sum_{k=1}^{k_{max}} j_{max}^{k} \quad [5]$$

70% of the records were used to train the network and the remaining 30% to validate it and test it. The partition was done in contiguous blocks, taking the first records to train the ANN, and the subsequent 15% each allocated for model validation and testing. The Levenberg-Marquardt algorithm (LM) was used to optimize the architecture of the ANN. Table 7 shows the characteristics and hyperparameters selected to optimize the ANN.

Table 7. Setup values for the ANN optimization.

Characteristic	**Value**
Software	Matlab Deep Learning Toolbox (MathWorks, 2021)
Hardware	CORE i7 8th Gen Intel (R) UHD Graphics 620 CPU with a 16 GB RAM memory.
Average training time	1,929 s
Hyperparameters	**Value**
Nº epochs	1000
Batch size	1
Validation checks	6
Algorithm	LM
Learning rate	Adaptive based on the error

The results of the optimization of the ANN architecture are presented in Figure 10. To minimize model noise, we conducted the optimization process for the ANN architecture 10 times. The figure illustrates the average performance across these simulations, demonstrating the evolution of the RMSE for Hs and Tp in both the training and validation datasets across various architectures.

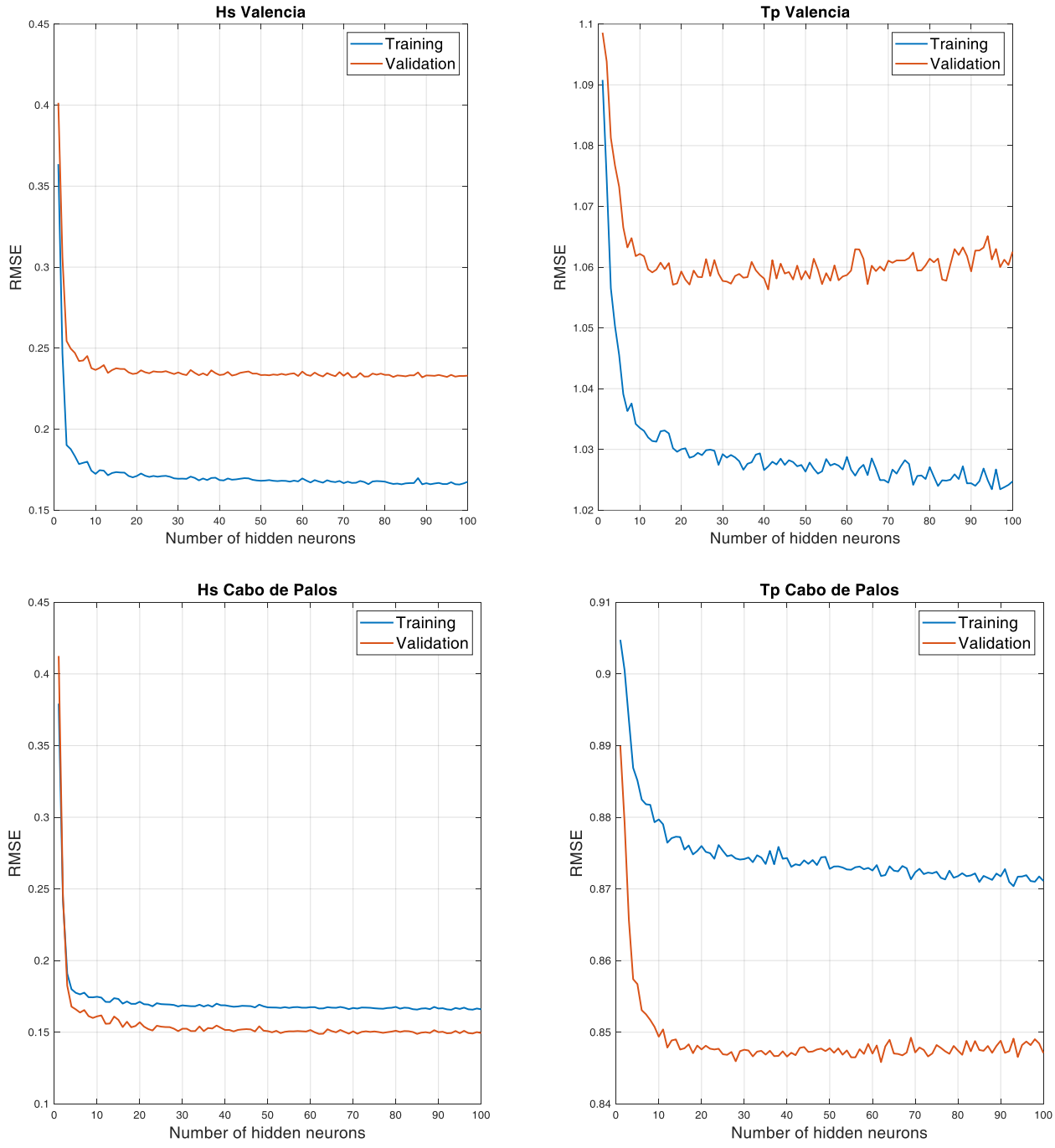


Figure 10. Evolution of RMSE for Hs and Tp in Valencia (up) and Cabo de Palos (down).

In Valencia the optimal behaviour of the network for both, the training and validation dataset, concerning the variables Hs and Tp, occurs between 20 and 50 hidden neurons. Acceptable training and validation errors are observed starting from 20 neurons. However, for Tp, a noticeable upward trend in validation error begins beyond 50 neurons, suggesting potential overfitting. Therefore, the range of 20 to 50 hidden neurons was selected to study its behaviour more deeply. Analysing

this stretch more specifically, the optimum behaviour was found to be at 33 hidden neurons. The selection of the optimal number of hidden neurons was based on obtaining the minimum of the mean of all errors, training and validation in the two study variables (Hs and Tp).

Following the same procedure as with Valencia buoy, in Cabo de Palos the optimum behaviour was reached at 22 hidden neurons. In the case of Cabo de Palos the increasing trend of the validation error in Tp is less pronounced, but it can be observed that beyond 30 neurons the Tp validation error stays the same and even slightly increases. While the Hs training and validation errors reach acceptable values beyond 20 neurons. Therefore, the range 20 – 30 neurons was studied with more detail, reaching the minimum errors at 22 hidden neurons.

After determining the architectural configurations for each buoy — [4-33-2] for the Valencia buoy and [4-22-2] for the Cabo de Palos buoy — the training algorithm underwent optimization. Various algorithms were considered for testing, including Resilient Back Propagation (RBP) (Riedmiller & Braun, 1993) as a gradient descent method; two conjugated gradient algorithms: Scaled Conjugated Gradient (SCG) (Møller, 1993) and Broyden-Fletcher-GoldFarb-Shanno (BFGS) (Fletcher, 1970; Goldfarb, 1970; Powell, 1994); Bayesian Regularization (BR) (Mackay, 1992) and the Quasi-Newton's algorithm Levenberg-Marquardt (LM) (Levenberg, 1944; Marquardt, 1963).

The selection of these algorithms was based on their documented effectiveness in scientific literature, specifically referencing the work of Portillo Juan and Negro Valdecantos (2022). To ensure robustness, each algorithm underwent training 40 times, mitigating the impact of weight and bias initialization on the analysis. In this case, the MSE (Equation 6) was used to evaluate the performance of the algorithms, following the same approach as in Nerea Portillo Juan et al. (2023). The boxplots of the results obtained are presented in Figure 11.

$$MSE = \frac{1}{n}\sum_{k=1}^{k_{max}}\sum_{j=1}^{j_{max}^{k}}\left(h_j^k - \hat{h}_j^k\right)^2 \qquad [6]$$

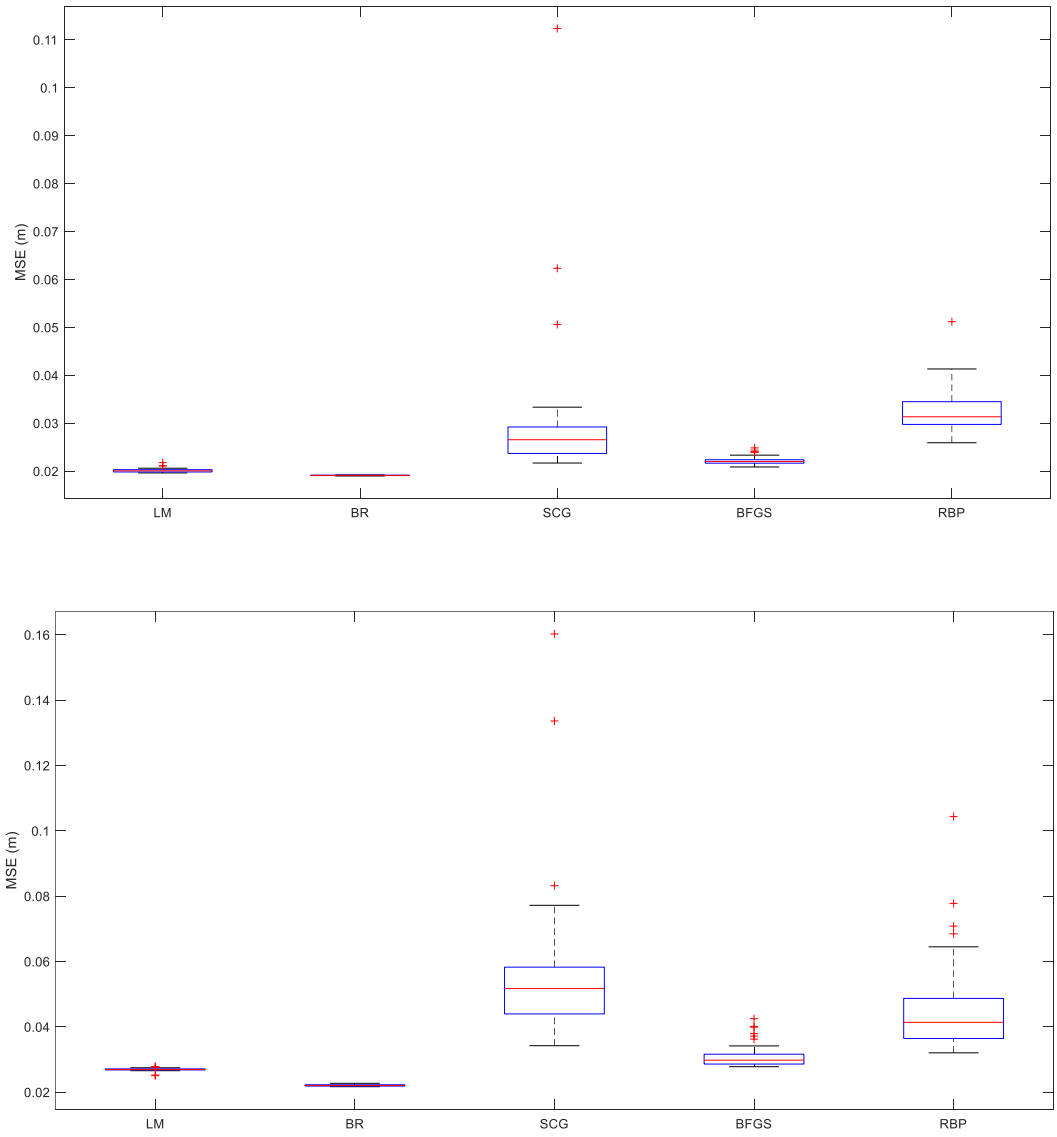


Figure 11. Boxplot of algorithm training error in Valencia (up)and Cabo de Palos (down)

In ANN for Valencia, the RBP was the algorithm with the worst performance, followed by SCG. In addition, their performance was the most scattered. LM and BR were the algorithms that performed the best and given that the LM algorithm had a lower computational cost than BR, LM algorithm was chosen as the optimum algorithm.

In Cabo de Palos, the performance of the algorithms was very similar to the performance in Valencia, with a higher error and dispersion associated with SCG algorithm. LM algorithm was also chosen as the optimum algorithm for the network.

The final characteristics of the model are presented in Table 8.

Table 8. ANN model

Location	Architecture	Algorithm	Transfer function
Valencia	4-33-2	Levenberg-Marquardt (LM)	Hyperbolic tangent sigmoid Linear
Cabo de Palos	4-22-2	Levenberg-Marquardt (LM)	Hyperbolic tangent sigmoid Linear

Finally, the overall performance of the ANN model has been summed-up in Table 9 and plotted in Figure 12.

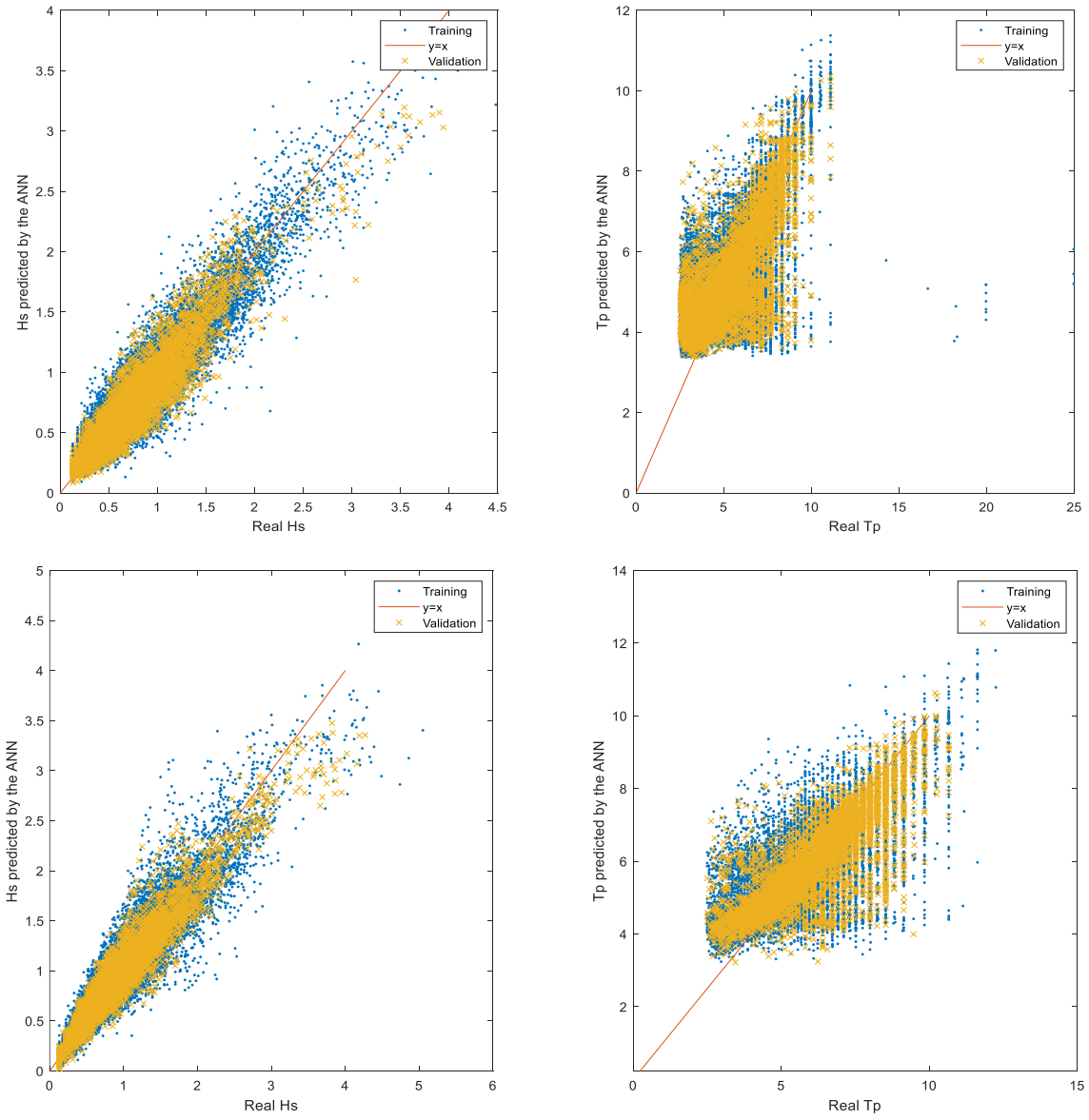


Figure 12. Performance of the model in Valencia (up) and Cabo de Palos (down).

As per the figure, the performance of the model is very accurate at both, Hs and Tp, with a maximum MSE of 0.021 m and 1.21 s. Table 9 shows these errors associated to the training and validation of the model. The dispersion associated with the variable Tp is higher due to the nature of this variable, but not due to the quality of the model. In fact, compared to numerical models, in this specific area neural networks have proved to perform better for Tp wave climate predictions (Nerea Portillo Juan et al., 2023).

Table 9. Training and validation error of the model

Variable	Error	Valencia	Cabo de Palos
Hs	RMSE Training	0.144	0.167
	RMSE Validation	0.142	0.147
Tp	RMSE Training	1.10	0.873
	RMSE Validation	1.05	0.863

3. Results

We applied the proposed methodology to derive present wave climate in shallow waters and forecasted wave climates for the years 2035 and 2050, also in coastal regions. The results obtained, along with comparisons, are presented and organized into two subsections. Subsection 3.1. dedicated to analyse long-term regime and subsection 3.2. focus on extreme regime.

3.1. Long-term regime

To evaluate the long-term regime, we used present data and generated a new dataset specifically for the deep-water buoys for the years 2035 and 2050 using again Monte Carlo simulations. Crucially, this dataset differed from those employed for model training. Subsequently, we applied this distinct dataset to the neural network model to derive coastal wave climate predictions for 2035 and 2050. The summarized results for the two locations under study are presented in Tables 10 and 11. Tables 10 and 11 provide the average values, standard deviations (SD), the mean increment of Hs and Tp, and the relative increment of Hs and Tp in shallow depth (coastal wave climate) for the years 2035 and 2050.

Table 10. Evolution of wave climate in the coast of Valencia under climate change

	Hs			Tp		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	0.66 m	0.64 m	0.82 m	5.25 s	4.81 s	5.02 s
SD	0.39 m	0.39 m	0.42 m	1.19 s	0.65 s	0.73 s
Δ	-	-0.03 m	0.17 m	-	-0.44 s	-0.23 s
%Δ	-	-4.55%	25.75%		-8.3%	-4.4%

Table 11. Evolution of wave climate in the coast of Cabo de Palos under climate change

	Hs			Tp		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	0.88 m	0.86 m	0.97 m	5.83 s	5.63 s	5.78 s
SD	0.47 m	0.50 m	0.51 m	1.15 s	0.71 s	0.71 s
Δ	-	-0.01 m	0.09 m	-	-0.20 s	-0.05 s
%Δ	-	-1.14%	10%		-3.43%	-0.9%

As can be seen from the tables, little changes in the maritime climate in the Western Mediterranean associated with climate change are expected.

By 2035, no significant increase in wave height is projected. In fact, our results show a minimal decrease of 0.03 m in Valencia and of 0.01 m in Cabo de Palos, which can be approximated to 0 considering that we depreciated all the values of wave heights that were below 0.1 m in our original databases (Section 2.1). These minor decreases can be attributed to the accuracy of our model, which excludes values below 0.1 m, as well as the inherent variability in wave characteristics and the dynamic nature of the wave climate. On the other hand, for the year 2050, we can expect wave increases. Valencia is projected to experience a more pronounced increase, with wave heights expected to rise by up to 0.2 m. In contrast, Cabo de Palos is forecasted to see a smaller increase, with wave heights expected to rise by up to 0.1 m. These projections highlight varying impacts on wave heights across different regions, underscoring the importance of conducting local studies to properly model localized effects of climate change.

Regarding the peak period, a reduction in values is anticipated for both 2035 and 2050. This is possibly because we assumed that relationship of Hs-Tp remained constant over time to not penalise extreme events too much. This approach is deliberately conservative to avoid disproportionately penalizing extreme events. Given the higher level of uncertainty associated with the peak period compared to wave height, we preferred not to overestimate its projections.

One striking observation is that the SD of Tp, instead of increasing over time, decreases. Initially, it seems logical that the standard deviations of the parameters defining the maritime climate would increase in the scenario of climate change. Indeed, this is the case with Hs, but the opposite occurs with Tp, with deviations nearly double in the present compared to the future.

This phenomenon arises from our methodology. We used real data when modelling wave states for the present period. However, for the 2035 and 2050 projections, we utilized a virtual database where we established a relationship between peak period (Tp) and significant wave height (Hs) based on the formula recommended by the Spanish authority responsible for climate characterization. The observation that present Tp values derived from actual records exhibit much greater dispersion compared to the future Tp values modelled using Equation 3 $\left(T_p = a \times (H_s)^b\right)$ underscores the inherent challenge in accurately modelling the relationship between Hs and Tp. While there exist analytical formulas linking Hs and Tp, the behaviour of the sea is far more complex and random than what can be captured by a simple formula. This complexity has been acknowledged previously by the authors in Portillo Juan et al. (2022a).

Figures 13 and 14 shows the evolution of Hs and Tp in time. The records were ordered from lowest to highest. The obtained Hs and Tp were sorted based on Hs in deep waters. This sorting method involved arranging the records from deep waters for the present, 2035, and 2050 in ascending order of Hs. Subsequently, this order was applied to the corresponding values of Hs and Tp in coastal waters, which were derived after applying the ANN model to waves in deep water.

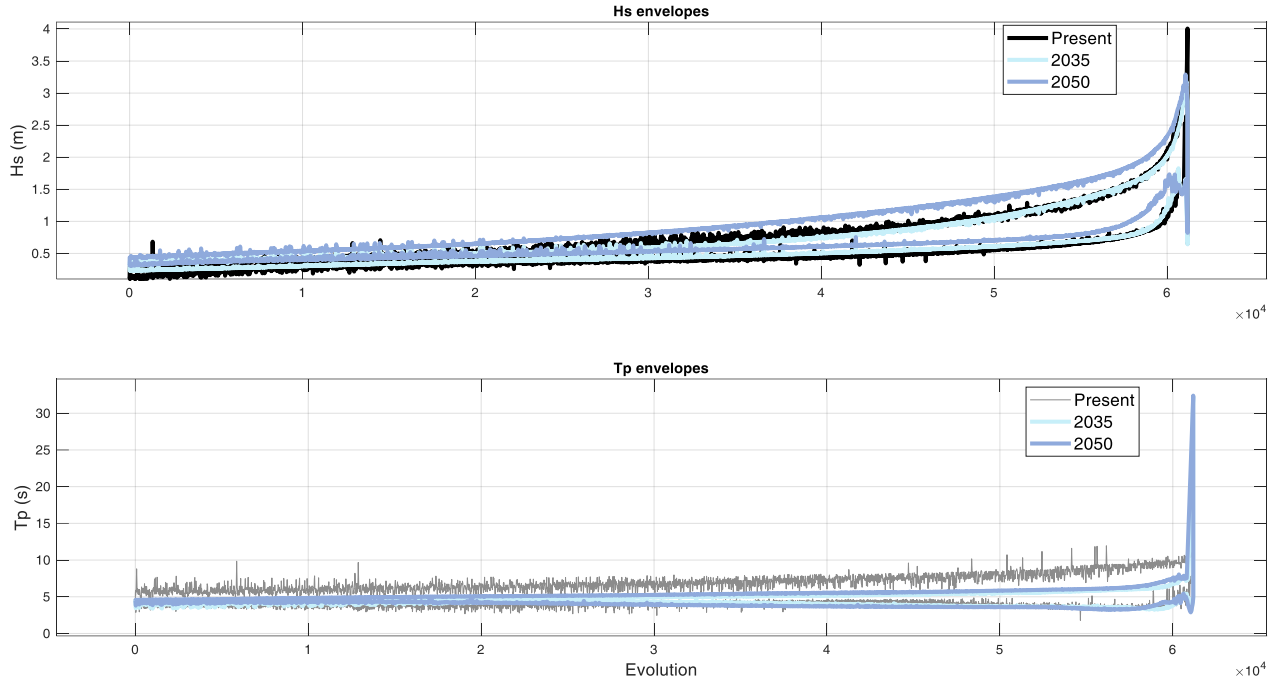


Figure 13. Hs (up) and Tp (down) evolution in the coast of Valencia under climate change scenario.

The black and grey lines represent the present wave climate, while the blue lines the wave climate in the future, as can be seen in the legend of the graph. As previously stated, from Figure 13, slightly shorter periods are observed for the future, especially those associated with small and medium waves. This is possibly due to the assumed conservative relationship between Hs and Tp previously mentioned. Even so, it can be seen in the graph how the periods associated with extreme events increase considerably with climate change.

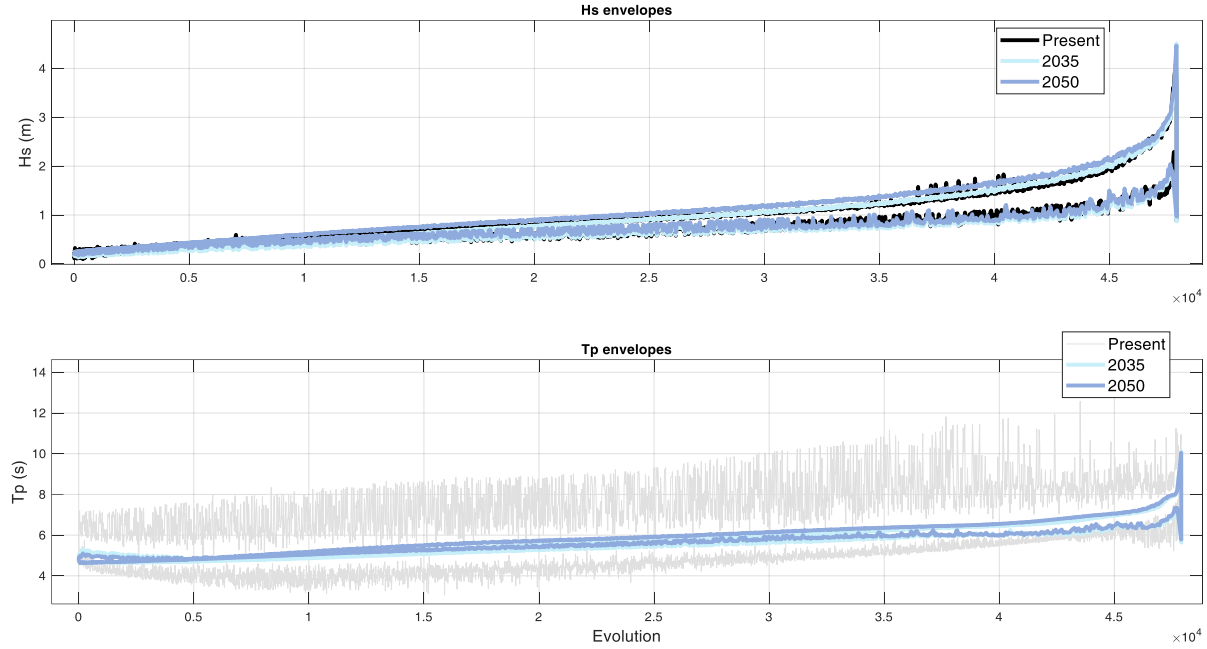


Figure 14. Hs (up) and Tp (down) evolution in the coast of Cabo de Palos under climate change scenario.

As already shown in Tables 10 and 11, it can be clearly seen that the future projections in the upper graph of Figure 14 (Hs in Cabo de Palos) are not as pronounced as in the upper graph of Figure 13 (Hs in Valencia), where the blue and grey stripes can be seen much more clearly.

With respect to the evolution of Tp in Cabo de Palos (lower graph in Figure 14), it can be clearly observed the phenomenon that we previously mentioned of the dispersed values of Tp in the present. As mentioned earlier, the primary cause of this significant dispersion lies in the use of two different sources of data: a real database for modelling the present wave climate and a virtual database for the future wave climate. The virtually generated data were created according to the relationship (Equation 3). However, as it can be observed, the wave behaviour in Cabo de Palos is far from conforming to this relationship. While areas like Valencia exhibit a relatively fulfilled relationship, $Tp = a\,Hs^b$ in regions like Cabo de Palos, this relationship does not hold true.

Furthermore, the method of sorting the values of Hs and Tp to create the graphs amplifies this appearance of scattered data. We sorted the data based on the value of Hs in deep waters, meaning that Tp is ordered based on Hs in deep waters rather than directly on Tp values. This ordering approach remains consistent for both the propagation of Hs and the virtually generated Tp values following Equation 3. However, in reality, the maximum values of Hs are not necessarily linked with the

maximum values of Tp; the behaviour of sea states is inherently more random. Therefore, the methodology used to create the graphs contributes to the perception that present Tp is much more dispersed than future Tp.

3.2. Extreme distribution

To study the extreme distribution, the 95th percentile of the records were considered, following the same procedure as in section 3.1. The results are presented in Tables 12 and 13.

Table 12. Evolution of wave climate in the coast of Valencia under climate change.

	Hs			Tp		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	1.69 m	3.10 m	3.23 m	7.12 s	9.00 s	8.80 s
SD	0.40 m	0.44 m	0.39 m	1.54 s	1.92 s	1.28 s
Δ	-	1.42 m	1.55 m	-	1.89 s	1.69 s
% Δ	-	84.02%	91.72%	-	26.55%	23.74%

Table 13. Evolution of wave climate in the coast of Cabo de Palos under climate change.

	Hs			Tp		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	2.09	2.76	2.89	7.47	8.18	8.29
SD	0.51 m	0.62 m	0.6 m	0.84 s	0.67 s	0.63 s
Δ	-	0.67	0.80	-	0.71	0.82
% Δ	-	32.06%	38.28%	-	9.50%	11.00%

In the case of extreme events, the increase in wave heights is much more pronounced than in the long-term regime. Increases of up to 1.5 m are reached at Valencia by the year 2050 and up to 0.8 m at Cabo de Palos. With respect to the peak period, it can be seen how it also increases up to 1.7 s, although in this case the increase is not as significant as for Hs.

Figures 15 and 16 show the same graphs presented for long-term distribution, but for extreme events.

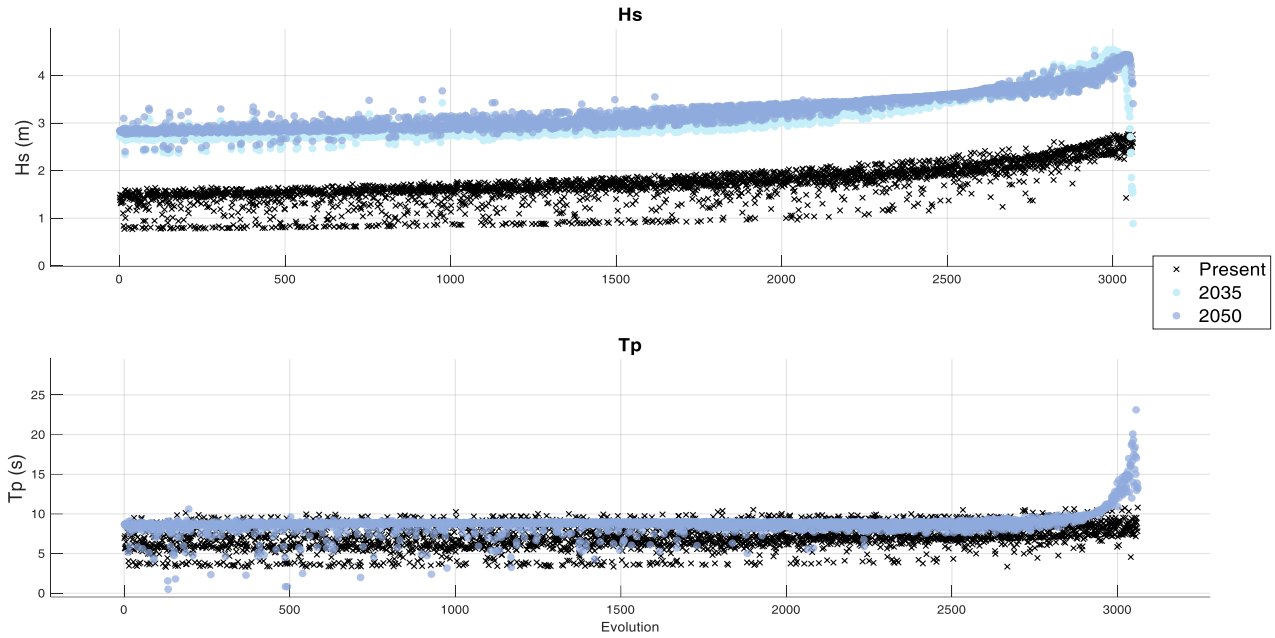


Figure 15. Extreme Hs (up) and Tp (down) evolution in Valencia under climate change scenario.

Figure 15 shows clearly how for the extreme regime there is a large difference between the current maritime climate and the future climate change scenarios. The wave heights associated with the 2035 and 2050 horizons for extreme events are clearly above those of today. In the case of the peak period, it can also be seen that current periods are below the peak periods associated with future extreme events.

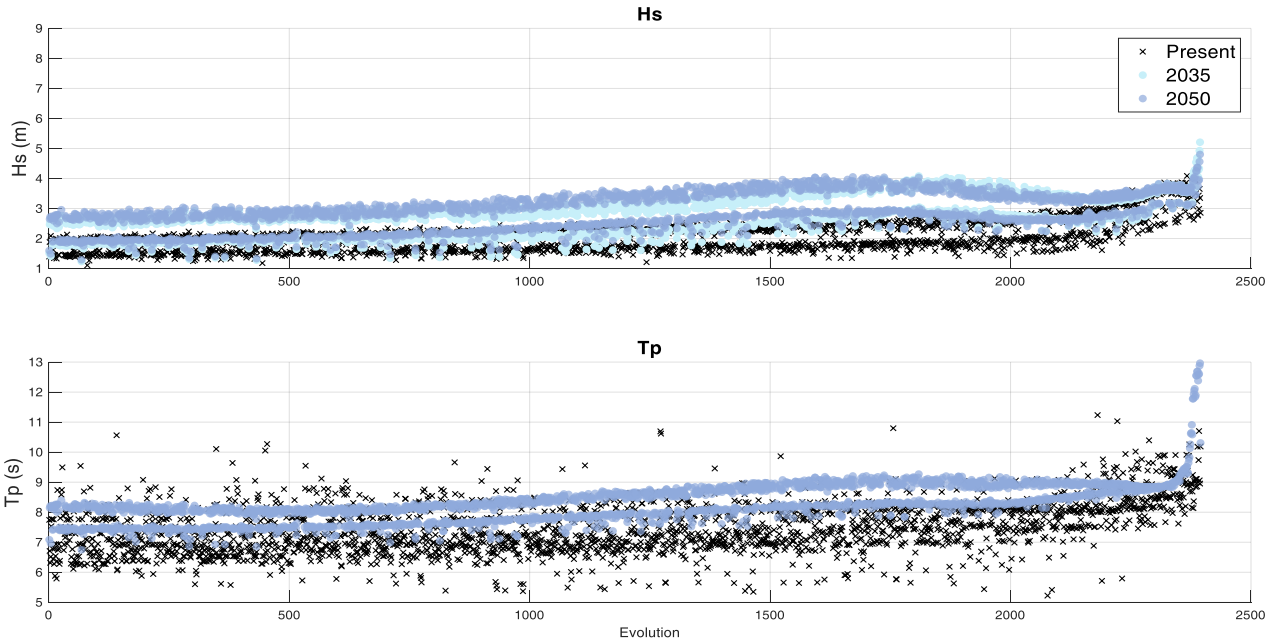


Figure 16. Extreme Hs (up) and Tp (down) evolution in Cabo de Palos under climate change scenario.

As happened with the long-term regime, the trends in Cabo de Palos are less pronounced than in Valencia, but it can be clearly appreciated how the values associated with present climate are far below the ones predicted for 2035 and 2050.

Finally, as a summary of the results obtained, Table 14 shows the changes in Hs and Tp expected for the Western Mediterranean in the long-term and extreme long-term distribution.

Table 14. Wave climate under climate change scenario in the Western Mediterranean

	ΔHs (m)		ΔTp (s)	
Year	**2035**	**2050**	**2035**	**2050**
Long-term	0.0 m	0.0 – 0.2 m	0.0 s	0.0 s
Extreme events	0.5 – 1.4 m	0.6 – 1.6 m	0.8 – 1.8 s	0.7 – 1.8 s

4. Discussion

The results obtained with the new methodology proposed after studying maritime climate in the Western Mediterranean show that there will be significative changes in the two frameworks studied, 2035 and 2050, especially for extreme events. Increments associated to the variable Hs appear to be more significant than those associated to Tp. However, as discussed in section 3, wave periods are already undergoing changes, and previous equations used to model Tp in relation to Hs may no longer accurately reflect reality. As demonstrated, the equation $T_p = a \times (H_s)^b$, commonly used to model Tp and also applied in this study, is not entirely representative of reality. This discrepancy may explain why changes in Tp do not appear as significant as changes in Hs. Tp is a wave climate variable with considerable uncertainty associated with it, and thus, it should be studied in much greater detail to draw conclusions with an acceptable level of certainty. Efforts should be directed towards finding an equation that better captures the behaviour of wave periods.

For the long-term regime significant wave height will increase up to 0.2 m in 2050 while the peak period will remain constant. However, for extreme climate events, the increments of wave height and peak period associated with climate change are much more significant, up to 1.5 m and 1.8 s. Although in the long-term regime the changes may seem insignificant, wave heights in the Mediterranean for extreme events can increase by up to 1.5 m. Extreme events are of the greatest importance

as they have the greatest capacity to cause structural, economic, and social damage. Wave height increases of up to 1.5 m would result in damage and even failure of existing coastal structures, as they are designed with lower parameters than those that would be under the climate change scenario. This would lead to great economic and social losses, so in this scenario it is necessary to develop an adaptation plan for these structures.

The results obtained in this study are in line with some of the previous studies (Table 15), while diverge from aspects of the existing literature. Table 15 presents the comparison of these results obtained in previous studies and the ones obtained in the present research.

Table 15. Results obtained compared to other studies.

Studies in line with the results obtained		
Authors	*Area of study*	*Conclusions*
Casas-Prat and Sierra (2013)	NorthWestern Mediterranean Sea	For the mean climate, relative changes of Hs up to 10% were obtained, whereas they were around 20% for the extreme climate.
Vousdoukas et al. (2016) Vousdoukas et al. (2017)	Europe	Significant increase in wave climate along the European coast, especially in extreme events.
Bricheno and Wolf (2018)	Europe	Modest changes in mean waves and strong changes in extreme waves. Larger waves in the area of Spain.
De Leo et al. (2020)	Mediterranean Sea	Trends in Hs are stronger for high percentiles compared to annual averages. Stronger trends appear in more energetic parts of the Mediterranean.
Wang et al. (2014)	Global study	Significant wave increases.
Amarouche and Akpinar (2021)	Western Mediterranean Sea	Increasing trend in the intensity and the duration of storms associated with higher extreme waves and sea levels.
Young and Ribal (2019)	Global study	Increase in wave heights, especially in extreme events.

Studies that contradict the results obtained		
Authors	*Area of study*	*Conclusions*
Casas-Prat and Sierra (2010)	NorthWestern Mediterranean Sea	Decreasing trend of significant wave height and mean period.
Perez et al. (2015)	Europe	Decreasing trend on wave-height, around 0.1 – 0.2 m.

As per the table, the results align with great part of the studies analysed in this research, there is a significant increasing trend in maritime climate in the Western Mediterranean Sea, especially in extreme events and energetic areas.

Many authors studied the effects of this changing climate, proving the importance of studying the changes that global warming is going to provoke in the wave climate. These changes in wave climate and extreme events have consequences in two main fields of coastal, ocean and ports engineering, the structural performance of coastal structures, and, the operability of ports. Some of the studies that proved the importance of correctly characterizing wave climate under climate change for coastal structures are the following: Chini and Stansby (2012) studied the increase of overtopping rates due to SLR and the increase of wave height, which affect the performance of coastal structures. Isobe (2013) proved that even if wave heights do not increase, the effect of SLR will have a big impact in overtopping rates and run ups, as well as in the armour stability of breakwaters, which would need an adaptation to these new climate conditions. Arns et al. (2017) concluded that the expected SLR will amplify the resulting design heights of coastal protection designs by an average of 48–56%. Almar et al. (2021) stated that under RCP 8.5, the globally aggregated annual overtopping hours by the end of the 21st-century is projected to be up to 50 times larger compared to present-day. Luo and Lin (2023) also studied the SLR hazard and its implications for flooding.

Regarding ports operations, Sanchez-Arcilla et al. (2016) highlighted all the impacts that climate change is going to have in ports. Camus et al. (2019) studied port operation downtimes due to climate change. Izaguirre et al. (2020) also studied how the operability of ports is going to change due to the increment of extreme events, SLR and wave heights and, Sierra et al. (2023) studied specifically the impacts on berthing areas. All of them and many more authors concluded that these changes in wave climate will provoke the decrease in the operating time of

the ports resulting in large economic losses. Therefore, future research lines should study how the changes obtained in wave height and wave period for 2035 and 2050 would affect the structural performance of coastal structures in the area of Valencia and Cabo de Palos and the operability of their ports.
In terms of accuracy and computational efficiency, our model achieved the same level of accuracy as the models presented in this section (MSE of 0.02 m for significant wave height and 1 s for peak wave period), but with significantly reduced computational time. The studies discussed predominantly utilized SWAN (Booij et al., 1999) for wave climate propagation and climate change projections, as well as Boussinesq and mild-slope models for harbour agitation studies, which are known for their high computational expense.

Focusing specifically on wave climate characterization and climate change projections—the primary areas of interest in our study—the proposed model offers substantial time savings. It reduces the development time by approximately 5 times, encompassing the training and optimization of ANN, and decreases the simulation runtime by around 20 times per simulation compared to the traditional models utilized in these studies. Moreover, ANNs and data-driven models, in general, have the added advantage of improving over time through use, as they continuously incorporate new data into their functioning. This ongoing integration of new data provides valuable feedback to the model itself.

The main limitation of this study concerns the availability of data series in Valencia. Our dataset only covered a 15-year time span. Longer data series generally provide more reliable predictions for trend analysis and extrapolation; the greater the length of the data series, the better the ability to predict future values with accuracy. In comparison to other studies employing trend analysis (Casas-Prat & Sierra, 2012; De Leo et al., 2021; Dieng et al., 2021), our sample size was smaller due to this limited timeframe. To mitigate this constraint, we employed statistical adjustments by incorporating data from Cabo de Palos buoy. In addition, historical records exhibited significant time gaps, which, as noted in previous sections, can affect the performance of the AI model. To solve this lack of data points and avoid model inconsistencies, we used interpolation for small data gaps and we supplement the data with data from virtual buoys, with the final aim of having homogeneous data series, strictly needed in AI-models.

Finally, future research lines should also study the problem of modelling Tp. As demonstrated, there exists a significant disparity between modelling Tp using the analytical equation that governs its behaviour and using real Tp values. While this

relationship may hold relatively true in certain locations like Valencia, there are other areas, such as Cabo de Palos, where it diverges significantly from reality. Therefore, future studies should aim to understand how Tp behaves across different locations, where the established relationship of $T_p = a \times (H_s)^b$ holds and where it does not, and how to properly model Tp to reflect its true behaviour. This understanding is crucial as it impacts all studies, regardless of the methodology used, whether AI or numerical models. It is of paramount importance for maritime and coastal engineers to accurately model Tp for effective planning and designs.

5. Conclusions

A new methodology to downscale wave climate under climate change scenario has been proposed and validated. This new methodology combines ANNs and Monte Carlo simulations to obtain future wave projections in the coast. Monte Carlo simulations are used to project future wave climate in deep waters and generate a virtual database that supplement historical records. With the AI-model fed by historical records and future projections of deep-water climate generated by Monte Carlo simulations, future sea states for different time frameworks can be downscaled to obtain future coastal maritime climate. This new methodology merges two of the most efficient tools used in ocean engineering. The use of Monte Carlo simulations allows making future predictions and generate huge amount of data — that is the fuel of AI-models — and the development of an ANN model allows the use of large amount of data that would not be possible with a numerical model due to its high computational cost. This results in a highly accurate model (MSE of 0.02 m for Hs and 1.1 s for Tp), thanks to the large amount of data used, with a very low computational cost, thanks to the use of AI tools (ANNs), reducing simulation time up to 20 times compared to numerical models.

This methodology has been applied to the Western Mediterranean and it has been concluded that increases in the wave height up to 0.2 m by 2050 in the long-term regime can be reached and up to 1.5 m in extreme events. Referring to the peak period, this increasing trend is less significant, with increases up to 1.8 s in extreme events.

Furthermore, after applying the methodology, it has been observed that there is a significant difference when modelling peak periods using real records versus following the equation proposed in maritime engineering. Therefore, the authors have called for a more thorough study on how to correctly model Tp.

References

Abadie, L., Galarraga, I., Markandya, A., & Sainz de Murieta, E. (2019). Risk measures and the distribution of damage curves for 600 European coastal cities. *Environmental Research Letters*, *14*. https://doi.org/10.1088/1748-9326/ab185c

Abhigna, P., Jerritta, S., Srinivasan, R., Rajendran, V., & Ieee. (2017, 2017 Apr 06-08). Analysis of Feed Forward and Recurrent Neural Networks in Predicting the Significant Wave Height at the Moored Buoys in Bay Of Bengal. [2017 international conference on communication and signal processing (iccsp)]. IEEE International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, INDIA.

Abram, N., Gattuso, J.-P., Prakash, A., Cheng, L., Chidichimo, M. P., Crate, S., Enomoto, H., Garschagen, M., Gruber, N., Harper, S., Holland, E., Kudela, R. M., Rice, J., Steffen, K., & von Schuckmann, K. (2019). *Framing and Context of the Report. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

AEMET. Agencia Estatal de Meteorología Retrieved 02/03/2023 from https://www.aemet.es/es/portada

Agrawal, J. D., & Deo, M. C. (2004). Wave parameter estimation using neural networks. *Marine Structures*, *17*(7), 536-550. https://doi.org/https://doi.org/10.1016/j.marstruc.2005.01.001

Ahmad, F. (2022). Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement [Article]. *CAAI Transactions on Intelligence Technology*, *7*(2), 200-218. https://doi.org/10.1049/cit2.12083

Ahmed, A. A. M., Jui, S. J. J., Al-Musaylh, M. S., Raj, N., Saha, R., Deo, R. C., & Saha, S. K. (2024). Hybrid deep learning model for wave height prediction in Australia's wave energy region [Article]. *Applied Soft Computing*, *150*, Article 111003. https://doi.org/10.1016/j.asoc.2023.111003

Alemam, M. K., Yong, B., & Sani-Mohammed, A. (2023). A Proposed Merging Methods of Digital Elevation Model Based on Artificial Neural Network and Interpolation Techniques for Improved Accuracy [Article]. *Artificial Satellites*, *58*(3), 122-170. https://doi.org/10.2478/arsa-2023-0009

Almar, R., Ranasinghe, R., Bergsma, E. W. J., Diaz, H., Melet, A., Papa, F., Vousdoukas, M., Athanasiou, P., Dada, O., Almeida, L. P., & Kestenare, E. (2021). A global analysis of extreme coastal water levels with implications for potential coastal overtopping. *Nature Communications*, *12*(1), 3775. https://doi.org/10.1038/s41467-021-24008-9

Amarouche, K., & Akpinar, A. (2021). Increasing Trend on Storm Wave Intensity in the Western Mediterranean. *Climate*, *9*(1), Article 11. https://doi.org/10.3390/cli9010011

Arns, A., Dangendorf , S., Jensen, J., Talke, S., Bender, J., & Pattiaratchi, C. (2017). Sea-level rise induced amplification of coastal protection design heights. *Scientific Reports*, *7*(1), 40171. https://doi.org/10.1038/srep40171

Atteia, G., Collins, M. J., Algarni, A. D., & Samee, N. A. (2022). Deep-Learning-Based Feature Extraction Approach for Significant Wave Height Prediction in SAR Mode Altimeter Data [Article]. *Remote Sensing*, *14*(21), Article 5569. https://doi.org/10.3390/rs14215569

Avila, D., Marichal, G. N., Padron, I., Quiza, R., & Hernandez, A. (2020). Forecasting of wave energy in Canary Islands based on Artificial Intelligence. *Applied Ocean Research*, *101*, Article 102189. https://doi.org/10.1016/j.apor.2020.102189

Baek, G., Kim, J., Ko, T. K., & Lee, S. (2020). Artificial neural network interpolation for magnetic field mapping in an air-core HTS quadruple magnet [Article]. *Cryogenics*, *107*, Article 103043. https://doi.org/10.1016/j.cryogenics.2020.103043

Bagheri, H., Sadjadi, S. Y., & Sadeghian, S. (2014). Exploring the role of genetic algorithms and artificial neural networks for interpolation of elevation in geoinformation models. Lecture Notes in Geoinformation and Cartography,

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, *43*(1), 3-31. https://doi.org/10.1016/s0167-7012(00)00201-3

Bedja-Johnson, Z., Wu, P., Grande, D., & Anderlini, E. (2022). Smart anomaly detection for Slocum underwater gliders with a variational autoencoder with long short-term memory networks. *Applied Ocean Research*, *120*, 103030. https://doi.org/https://doi.org/10.1016/j.apor.2021.103030

Bindoff, N. L., Cheung, W. W. L., Kairo, J. G., Arístegui, J., Guinder, V. A., Hallberg, R., Hilmi, N., Jiao, N., Karim, M. S., Levin, L., O'Donoghue, S., Purca Cuicapusa, S. R., Rinkevich, B., Suga, T., Tagliabue, A., & Williamson, P. (2019). *Changing Ocean, Marine Ecosystems, and Dependent Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Bohlinger, P., Breivik, Ø., Economou, T., & Müller, M. (2019). A novel approach to computing super observations for probabilistic wave model validation. *Ocean Modelling*, *139*, 101404. https://doi.org/https://doi.org/10.1016/j.ocemod.2019.101404

Booij, N., Ris, R. C., & Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research: Oceans*, *104*(C4), 7649-7666. https://doi.org/https://doi.org/10.1029/98JC02622

Bricheno, L. M., & Wolf, J. (2018). Future Wave Conditions of Europe, in Response to High-End Climate Change Scenarios. *Journal of Geophysical Research-Oceans*, *123*(12), 8762-8791. https://doi.org/10.1029/2018jc013866

Campos, R. M., Costa, M. O., Almeida, F., & Guedes Soares, C. (2021). Operational Wave Forecast Selection in the Atlantic Ocean Using Random Forests. *Journal of Marine Science and Engineering*, *9*(3), 298. https://www.mdpi.com/2077-1312/9/3/298

Camus, P., Losada, I. J., Izaguirre, C., Espejo, A., Menéndez, M., & Pérez, J. (2017). Statistical wave climate projections for coastal impact assessments. *Earth's Future*, *5*(9), 918-933. https://doi.org/https://doi.org/10.1002/2017EF000609

Camus, P., Tomás, A., Diaz-Hernandez, G., Rodriguez, B., Izaguirre, C., & Losada, I. J. (2019). Probabilistic assessment of port operation downtimes under climate change. *Coastal Engineering*.

Casas-Prat, M., & Sierra, J. P. (2010). Trend analysis of wave storminess: wave direction and its impact on harbour agitation. *Nat. Hazards Earth Syst. Sci.*, *10*(11), 2327-2340. https://doi.org/10.5194/nhess-10-2327-2010

Casas-Prat, M., & Sierra, J. P. (2012). Trend analysis of wave direction and associated impacts on the Catalan coast. *Climatic Change*, *115*(3), 667-691. https://doi.org/10.1007/s10584-012-0466-9

Casas-Prat, M., & Sierra, J. P. (2013). Projected future wave climate in the NW Mediterranean Sea. *Journal of Geophysical Research-Oceans*, *118*(7), 3548-3568. https://doi.org/10.1002/jgrc.20233

Castillo;, E., & Pruneda, R. E. (2001). *Estadística Aplicada* (Moralea, Ed.).

CAVALERI, L., & BERTOTTI, L. (2004). Accuracy of the modelled wind and wave fields in enclosed seas. *Tellus A*, *56*(2), 167-175. https://doi.org/https://doi.org/10.1111/j.1600-0870.2004.00042.x

Charles, E., Idier, D., Delecluse, P., Déqué, M., & Le Cozannet, G. (2012). Climate change impact on waves in the Bay of Biscay, France. *Ocean Dynamics*, *62*(6), 831-848. https://doi.org/10.1007/s10236-012-0534-8

Chini, N., & Stansby, P. K. (2012). Extreme values of coastal wave overtopping accounting for climate change and sea level rise. *Coastal Engineering*, *65*, 27-37. https://doi.org/https://doi.org/10.1016/j.coastaleng.2012.02.009

Collins, M., Sutherland, M., Bouwer, L., Cheong, S.-M., Frölicher, T., Jacot Des Combes, H., Koll Roxy, M., Losada, I., McInnes, K., Ratter, B., Rivera-Arriaga, E., Susanto, R. D., Swingedouw, D., & Tibig, L. (2019). *Extremes, Abrupt Changes and Managing Risk. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

De Leo, F., Besio, G., & Mentaschi, L. (2021). Trends and variability of ocean waves under RCP8.5 emission scenario in the Mediterranean Sea. *Ocean Dynamics*, *71*(1), 97-117. https://doi.org/10.1007/s10236-020-01419-8

De Leo, F., De Leo, A., Besio, G., & Briganti, R. (2020). Detection and quantification of trends in time series of significant wave heights: An application in the Mediterranean Sea. *Ocean Engineering*, *202*, 107155. https://doi.org/https://doi.org/10.1016/j.oceaneng.2020.107155

Díaz, H., Teixeira, A. P., & Guedes Soares, C. (2022). Application of Monte Carlo and Fuzzy Analytic Hierarchy Processes for ranking floating wind farm locations. *Ocean Engineering*, *245*, 110453. https://doi.org/https://doi.org/10.1016/j.oceaneng.2021.110453

Dieng, H. B., Cazenave, A., Gouzenes, Y., & Sow, B. A. (2021). Trends and inter-annual variability of altimetry-based coastal sea level in the Mediterranean Sea: Comparison with tide gauges and models. *Advances in Space Research*, *68*(8), 3279-3290. https://doi.org/10.1016/j.asr.2021.06.022

Duran-Rosal, A. M., Fernandez, J. C., Gutierrez, P. A., & Hervas-Martinez, C. (2017). Hybridization of neural network models for the prediction of Extreme Significant Wave Height segments. 2016 IEEE Symposium Series on Computational Intelligence, SSCI 2016,

Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, *13*(3), 317-322. https://doi.org/10.1093/comjnl/13.3.317

Gao, R., Li, R., Hu, M., Suganthan, P. N., & Yuen, K. F. (2023). Significant wave height forecasting using hybrid ensemble deep randomized networks with neurons pruning [Article]. *Engineering Applications of Artificial Intelligence*, *117*, Article 105535. https://doi.org/10.1016/j.engappai.2022.105535

Georgescu, I. (2023). The early days of Monte Carlo methods. *Nature Reviews Physics*, *5*(7), 372-372. https://doi.org/10.1038/s42254-023-00608-w

Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, *24*(109), 23-26. https://doi.org/10.2307/2004873

Gómez, R., Molina, R., Castillo, C., Rodríguez, I., & López, J. D. (2018). *CONCEPTOS Y HERRAMIENTAS PROBABILÍSTICAS PARA EL CÁLCULO DEL RIESGO EN EL ÁMBITO PORTUARIO*. Gobierno de España. Organismo Público Puertos del Estado.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., & Bengio, Y. (2014). Generative adversarial networks. *Communications of the ACM*, *63*, 139 - 144.

Grinstead, C. M., & Snell, J. L. (1997). *Introduction to probability*. American Mathematical Soc.

Gunaydin, K. (2008). The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. *Ocean Engineering*, *35*(14-15), 1406-1415. https://doi.org/10.1016/j.oceaneng.2008.07.008

Harrison, R. L. (2010). Introduction To Monte Carlo Simulation. *AIP Conf Proc*, *1204*, 17-21. https://doi.org/10.1063/1.3295638

Hemer, M. A., Church, J. A., & Hunter, J. R. (2007). Waves and Climate Change on the Australian Coast. *Journal of Coastal Research*, 432-437. http://www.jstor.org/stable/26481627

Hewawasam, V., & Matsui, K. (2019). Historical development of climate change policies and the Climate Change Secretariat in Sri Lanka. *Environmental Science & Policy*, *101*, 255-261. https://doi.org/https://doi.org/10.1016/j.envsci.2019.09.001

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., & Steltzer, H. (2019). *High Mountain Areas. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Hopstad, A. L. H., Argyriadis, K., Manjock, A., Goldsmith, J., & Ronold, K. O. (2018). DNV GL Standard for Floating Wind Turbines. ASME 2018 1st International Offshore Wind Technical Conference,

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, *29*(3), 339-349. https://doi.org/10.1080/00401706.1987.10488243

Informes anuales Oceanografía Puertos del Estado. (01/09/2022). https://www.puertos.es/es-es

IPCC. (2000). IPCC Special report. Emissions scenarios. Technical Report.

IPCC. (2013). *Climate Change 2013: The Physical Science Basis.Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*

(T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley, Eds.). Cambridge University Press.

IPCC, A. (2013). Climate change 2013: the physical science basis. *Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*, *1535*.

Isobe, M. (2013). Impact of global warming on coastal structures in shallow water. *Ocean Engineering*, *71*, 51-57. https://doi.org/https://doi.org/10.1016/j.oceaneng.2012.12.032

Izaguirre, C., Losada, I. J., Camus, P., González-Lamuño, P., & Stenek, V. (2020). Seaport climate change impact assessment using a multi-level methodology. *Maritime Policy & Management*, *47*(4), 544-557. https://doi.org/10.1080/03088839.2020.1725673

Jebbad, R., Sierra, J. P., Mosso, C., Mestres, M., & Sanchez-Arcilla, A. (2022). Assessment of harbour inoperability and adaptation cost due to sea level rise. Application to the port of Tangier-Med (Morocco) [Article]. *Applied Geography*, *138*, Article 102623. https://doi.org/10.1016/j.apgeog.2021.102623

Juan Nerea, P., Rodríguez Javier, O., & Valdecantos Vicente, N. (2024). Comparison of the SIMAR-WANA, ERA-5, and Waverys Databases for Maritime Climate Estimations and the Implications of Coastal Protection Structures. *Journal of Waterway, Port,*

Coastal, and Ocean Engineering, *150*(1), 04023021. https://doi.org/10.1061/JWPED5.WWENG-2017

Kazeminezhad, M. H., Etemad-Shahidi, A., & Mousavi, S. J. (2005). Application of fuzzy inference system in the prediction of wave parameters. *Ocean Engineering*, *32*(14), 1709-1725. https://doi.org/https://doi.org/10.1016/j.oceaneng.2005.02.001

Kim, G.-U., Seo, K.-H., & Chen, D. (2019). Climate change over the Mediterranean and current destruction of marine ecosystem. *Scientific Reports*, *9*, Article 18813. https://doi.org/10.1038/s41598-019-55303-7

Klusak, P., Agarwala, M., Burke, M., Kraemer, M., & Mohaddes, K. (2023). Rising Temperatures, Falling Ratings: The Effect of Climate Change on Sovereign Creditworthiness [Article]. *Management Science*, *69*(12), 7468-7491. https://doi.org/10.1287/mnsc.2023.4869

Kolios, A., Richmond, M., Koukoura, S., & Yeter, B. (2023). Effect of weather forecast uncertainty on offshore wind farm availability assessment. *Ocean Engineering*, *285*, 115265. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115265

Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, *26*(2), 195-197. https://doi.org/10.1038/nbt1386

Lecacheux, S., Rohmer, J., Paris, F., Pedreros, R., Quetelard, H., & Bonnardot, F. (2021). Toward the probabilistic forecasting of cyclone-induced marine flooding by overtopping at Reunion Island aided by a time-varying random-forest classification approach. *Natural Hazards*, *105*(1), 227-251. https://doi.org/10.1007/s11069-020-04307-y

Levenberg, K. (1944). A METHOD FOR THE SOLUTION OF CERTAIN NON – LINEAR PROBLEMS IN LEAST SQUARES. *Quarterly of Applied Mathematics*, *2*, 164-168.

Li, R., Song, B., Chen, Y., Jin, X., Zhou, D., Han, Z., Chen, W.-L., & Cao, Y. (2024). Deep learning reconstruction of high-Reynolds-number turbulent flow field around a cylinder based on limited sensors. *Ocean Engineering*, *304*, 117857. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.117857

Lionello, P., Cogo, S., Galati, M. B., & Sanna, A. (2008). The Mediterranean surface wave climate inferred from future scenario simulations. *Global and Planetary Change*, *63*(2-3), 152-162. https://doi.org/10.1016/j.gloplacha.2008.03.004

Londhe, S. N., Shah, S., Dixit, P. R., Nair, T. M. B., Sirisha, P., & Jain, R. (2016). A Coupled Numerical and Artificial Neural Network Model for Improving Location Specific Wave Forecast. *Applied Ocean Research*, *59*, 483-491. https://doi.org/10.1016/j.apor.2016.07.004

López, M., & Iglesias, G. (2015). Artificial neural networks applied to port operability assessment. *Ocean Engineering*, *109*, 298. https://doi.org/10.1016/j.oceaneng.2015.09.016

Luo, X., & Lin, T. (2023). Probabilistic Sea Level Rise Hazard Analysis Based on the Current Generation of Data and Protocols. *Journal of Structural Engineering*, *149*(3), 04022252. https://doi.org/10.1061/JSENDH.STENG-11413

Luque, P., Gomez-Pujol, L., Marcos, M., & Orfila, A. (2021). Coastal Flooding in the Balearic Islands During the Twenty-First Century Caused by Sea-Level Rise and Extreme Events. *Frontiers in Marine Science*, *8*, Article 676452. https://doi.org/10.3389/fmars.2021.676452

Mackay, D. J. C. (1992). Bayesian Interpolation. *Neural Computation*, *4*, 415-447.

Marangoni Gazineu Marinho Pinto, P., Martins Campos, R., Gallo, M. N., & Parente Ribeiro, C. E. (2023). Predicting significant wave height with artificial neural networks in the South Atlantic Ocean: a hybrid approach [Article]. *Ocean Dynamics*, *73*(6), 303-315. https://doi.org/10.1007/s10236-023-01546-y

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431-441. https://doi.org/10.1137/0111030

Mase, H., Tsujio, D., Yasuda, T., & Mori, N. (2013). Stability analysis of composite breakwater with wave-dissipating blocks considering increase in sea levels, surges and waves due to climate change. *Ocean Engineering*, *71*, 58-65. https://doi.org/10.1016/j.oceaneng.2012.12.037

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., & Gomis, M. (2021). Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, *2*.

MathWorks. (2021). *MATLAB version: 9.11.0 (R2021b)*. In Natick, Massachusetts: The MathWorks Inc. . https://www.mathworks.com

Medina, J. R., Garrido, J. M., Gómez-Martín, M. E., & Vidal, C. s. (2004). Armor Damage Analysis Using Neural Networks. *Coastal structures*, 236-248.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., Stouffer, R. J., & Taylor, K. E. (2007). THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research. *Bulletin of the American Meteorological Society*, *88*(9), 1383-1394. https://doi.org/https://doi.org/10.1175/BAMS-88-9-1383

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M. M. C., Ottersen, G., Pritchard, H., & Schuur, E. A. G. (2019). *Polar Regions. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, *6*(4), 525-533. https://doi.org/https://doi.org/10.1016/S0893-6080(05)80056-5

Mori, N., Shimura, T., Yasuda, T., & Mase, H. (2013). Multi-model climate projections of ocean surface variables under different climate scenarios—Future change of waves, sea level and wind. *Ocean Engineering*, *71*, 122-129. https://doi.org/https://doi.org/10.1016/j.oceaneng.2013.02.016

NOAA. National Oceanic and Atmospheric Administration. Retrieved 13/04/2024 from https://www.ncei.noaa.gov/

Oda, K., Kitamura, S., & Lee, M. (2012). Applicability of artificial neural network to spatial interpolation of soil properties in kansai international airport. Proceedings of the International Offshore and Polar Engineering Conference,

Oh, S., Jin, H. K., Joe, S. J., & Ki, H. (2021). Prediction of structural deformation of a deck plate using a GAN-based deep learning method. *Ocean Engineering*, *239*, 109835. https://doi.org/https://doi.org/10.1016/j.oceaneng.2021.109835

Oppenheimer, M., Glavovic, B. C., Hinkel, J., van de Wal, R., Magnan, A. K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R. M., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., & Sebesvari, Z. (2019). *Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Pan, X., & Xue, Y. (2023). Advancements of Artificial Intelligence Techniques in the Realm About Library and Information Subject - A Case Survey of Latent Dirichlet Allocation Method [Article]. *Ieee Access*, *11*, 132627-132640. https://doi.org/10.1109/ACCESS.2023.3334619

Patanè, L., Iuppa, C., Faraci, C., & Xibilia, M. G. (2024). A deep hybrid network for significant wave height estimation [Article]. *Ocean Modelling*, *189*, Article 102363. https://doi.org/10.1016/j.ocemod.2024.102363

Paulig, N., & Okhrin, O. (2024). Robust path following on rivers using bootstrapped reinforcement learning. *Ocean Engineering*, *298*, 117207. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.117207

Peng, H., Wu, C., Nguyen, D., Schuemann, J., Mairani, A., Pu, Y., & Jiang, S. (2023). Recent Advancements of Artificial Intelligence in Particle Therapy [Article]. *IEEE transactions on radiation and plasma medical sciences*, *7*(3), 213-224. https://doi.org/10.1109/TRPMS.2023.3241102

Perez, J., Menendez, M., Camus, P., Mendez, F. J., & Losada, I. J. (2015). Statistical multi-model climate projections of surface ocean waves in Europe. *Ocean Modelling*, *96*, 161-170. https://doi.org/https://doi.org/10.1016/j.ocemod.2015.06.001

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, *3*(1), 119-131, 113. https://doi.org/10.1214/aos/1176343003

Portillo Juan, N., & Negro Valdecantos, V. (2022). Review of the application of Artificial Neural Networks in ocean engineering. *Ocean Engineering*, *259*, 111947. https://doi.org/https://doi.org/10.1016/j.oceaneng.2022.111947

Portillo Juan, N., Negro Valdecantos, V., & del Campo, J. M. (2022a). Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast. *Water*, *14*(21), 3453. https://www.mdpi.com/2073-4441/14/21/3453

Portillo Juan, N., Negro Valdecantos, V., & del Campo, J. M. (2022b). Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain. *Sustainability*, *14*(12), 7507. https://www.mdpi.com/2071-1050/14/12/7507

Portillo Juan, N., Olalde Rodríguez, J., Negro Valdecantos, V., & Iglesias, G. (2023). Data-driven and physics-based approach for wave downscaling: A comparative study. *Ocean Engineering*, *285*, 115380. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115380

Portillo Juan, N., Olalde Rodríguez, J., Negro Valdecantos, V., & Iglesias, G. (2023). *Data-driven and physics-based approach for wave downscaling: A comparative study*.

Powell, M. J. D. (1994). A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation. In S. Gomez & J.-P. Hennart (Eds.), *Advances in Optimization and Numerical Analysis* (pp. 51-67). Springer Netherlands. https://doi.org/10.1007/978-94-015-8330-5_4

Puertos del Estado. Banco de Datos Oceanográficos. Informes. (01/01/2023). https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx

Puertos del Estado. Oceanografía. Puntos SIMAR-WANA. Retrieved 04/02/2023 from https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx

Puertos del Estado. Oceanografía. Registros históricos de oleaje. (19/09/2022). https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx

Raj, N., & Prakash, R. (2024). Assessment and prediction of significant wave height using hybrid CNN-BiLSTM deep learning model for sustainable wave energy in Australia [Article]. *Sustainable Horizons*, *11*, Article 100098. https://doi.org/10.1016/j.horiz.2024.100098

Riedmiller, M., & Braun, H. (1993, 28 March-1 April 1993). A direct adaptive method for faster backpropagation learning: the RPROP algorithm. IEEE International Conference on Neural Networks,

ROM03.91. (1992). *Wave Climate on the Spanish Coast*. Spanish Government.

Romano-Moreno, E., Diaz-Hernandez, G., L. Lara, J., Tomás, A., & F. Jaime, F. (2022). Wave downscaling strategies for practical wave agitation studies in harbours.

Coastal Engineering, *175*, 104140. https://doi.org/https://doi.org/10.1016/j.coastaleng.2022.104140

Rueda, A., Camus, P., Mendez, F. J., Tomas, A., & Luceno, A. (2016). An extreme value model for maximum wave heights based on weather types. *Journal of Geophysical Research-Oceans*, *121*(2), 1262-1273. https://doi.org/10.1002/2015jc010952

Sadeghifar, T., Lama, G. F. C., Sihag, P., Bayram, A., & Kisi, O. (2022). Wave height predictions in complex sea flows through soft-computing models: Case study of Persian Gulf. *Ocean Engineering*, *245*, 110467. https://doi.org/https://doi.org/10.1016/j.oceaneng.2021.110467

Sánchez-Arcilla, A., González-Marco, D., & Bolaños, R. (2008). A review of wave climate and prediction along the Spanish Mediterranean coast. *Natural Hazards and Earth System Sciences*, *8*, 1217-1228.

Sanchez-Arcilla, A., Pau Sierra, J., Brown, S., Casas-Prat, M., Nicholls, R. J., Lionello, P., & Conte, D. (2016). A review of potential physical impacts on harbours in the Mediterranean Sea under climate change. *Regional Environmental Change*, *16*(8), 2471-2484. https://doi.org/10.1007/s10113-016-0972-9

Saviz Naeini, S., & Snaiki, R. (2024). A novel hybrid machine learning model for rapid assessment of wave and storm surge responses over an extended coastal region [Article]. *Coastal Engineering*, *190*, Article 104503. https://doi.org/10.1016/j.coastaleng.2024.104503

Sierra, J., Sánchez-Arcilla, A., Gironella, X., Garcia, V., Altomare, C., Mösso, C., González-Marco, D., Gomez, J., Barceló, M., & Barahona, C. (2023). Impact of climate change on berthing areas in ports of the Balearic Islands: adaptation measures. *Frontiers in Marine Science*, *10*. https://doi.org/10.3389/fmars.2023.1124763

Sierra, J. P., Casas-Prat, M., Virgili, M., Moesso, C., & Sanchez-Arcilla, A. (2015). Impacts on wave-driven harbour agitation due to climate change in Catalan ports. *Natural Hazards and Earth System Sciences*, *15*(8), 1695-1709. https://doi.org/10.5194/nhess-15-1695-2015

Slimani, H., El Mhamdi, J., & Jilbab, A. (2024). Assessing the advancement of artificial intelligence and drones' integration in agriculture through a bibliometric study [Article]. *International Journal of Electrical and Computer Engineering*, *14*(1), 878-890. https://doi.org/10.11591/ijece.v14i1.pp878-890

Stefanakos, C. (2016). Fuzzy time series forecasting of nonstationary wind and wave data. *Ocean Engineering*, *121*, 1-12. https://doi.org/https://doi.org/10.1016/j.oceaneng.2016.05.018

Stefanakos, C. N., & Vanem, E. (2018). Nonstationary fuzzy forecasting of wind and wave climate in very long-term scales. *Journal of Ocean Engineering and Science*, *3*(2), 144-155. https://doi.org/https://doi.org/10.1016/j.joes.2018.04.001

Su, H., Yang, X., & Yan, X. H. (2019, 28 July-2 Aug. 2019). Estimating Ocean Subsurface Salinity from Remote Sensing Data by Machine Learning. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium,

Taha, A., Ochs, V., Kayhan, L. N., Enodien, B., Frey, D. M., Krähenbühl, L., & Taha-Mehlitz, S. (2022). Advancements of Artificial Intelligence in Liver-Associated Diseases and Surgery [Article]. *Medicina (Lithuania)*, *58*(4), Article 459. https://doi.org/10.3390/medicina58040459

Tolman, H. L., Balasubramaniyan, B., Burroughs, L. D., Chalikov, D., Chao, Y. Y., Chen, H. S., & Gerald, V. M. (2002). Development and Implementation of Wind-Generated Ocean Surface Wave Modelsat NCEP. *Weather and Forecasting*, *17*, 311-333.

Ulam, S. M., & Rabi, I. I. (2019). Adventures of a Mathematician. *Mathematics: People · Problems · Results*.

Vieira, B. F. V., Pinho, J. L. S., & Barros, J. A. O. (2021). Extreme wave value analysis under uncertainty of climate change scenarios off Iberian Peninsula coast [Article]. *Ocean Engineering*, *229*, Article 109018. https://doi.org/10.1016/j.oceaneng.2021.109018

Vieira, F., Cavalcante, G., Campos, E., & Taveira-Pinto, F. (2020). A methodology for data gap filling in wave records using Artificial Neural Networks. *Applied Ocean Research*, *98*, Article 102109. https://doi.org/10.1016/j.apor.2020.102109

Vitousek, S., Barnard, P. L., Limber, P., Erikson, L., & Cole, B. (2017). A model integrating longshore and cross-shore processes for predicting long-term shoreline response to climate change. *Journal of Geophysical Research: Earth Surface*, *122*(4), 782-806. https://doi.org/https://doi.org/10.1002/2016JF004065

Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Verlaan, M., & Feyen, L. (2017). Extreme sea levels on the rise along Europe's coasts. *Earths Future*, *5*(3), 304-323. https://doi.org/10.1002/2016ef000505

Vousdoukas, M. I., Voukouvalas, E., Annunziato, A., Giardino, A., & Feyen, L. (2016). Projections of extreme storm surge levels along Europe. *Climate Dynamics*, *47*(9-10), 3171-3190. https://doi.org/10.1007/s00382-016-3019-5

Walczak, S. (2019). Artificial Neural Networks. In *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. IGI Global. https://doi.org/10.4018/978-1-5225-7368-5.ch004

Wang, X. L., Feng, Y., & Swail, V. R. (2012). North Atlantic wave height trends as reconstructed from the 20th century reanalysis. *Geophysical Research Letters*, *39*(18). https://doi.org/https://doi.org/10.1029/2012GL053381

Wang, X. L., Feng, Y., & Swail, V. R. (2014). Changes in global ocean wave heights as projected using multimodel CMIP5 simulations. *Geophysical Research Letters*, *41*(3), 1026-1034. https://doi.org/https://doi.org/10.1002/2013GL058650

Xie, M., Huang, M., Bai, Y., Hu, Z., & Deng, Y. (2018). A robust data interpolation based on a back propagation artificial neural network operator for incomplete acquisition in wireless sensor networks [Article]. *Journal of Sensors*, *2018*, Article 7853695. https://doi.org/10.1155/2018/7853695

Xu, G., Ji, C., Xu, Y., Yu, E., Cao, Z., Wu, Q., Lin, P., & Wang, J. (2023). Machine learning in coastal bridge hydrodynamics: A state-of-the-art review. *Applied Ocean Research*, *134*, 103511. https://doi.org/https://doi.org/10.1016/j.apor.2023.103511

Xu, W., Tian, X., Anwaier, A., Qu, Y., Shi, G., Zhang, H., & Ye, D. (2022). A systematic review of current advancements of artificial intelligence in genitourinary cancers [Article]. *China Oncology*, *32*(1), 68-74. https://doi.org/10.19401/j.cnki.1007-3639.2022.01.009

Yang, Z., Xu, M., Wang, S., Li, J., Peng, Z., Jin, F., & Yang, Y. (2024). Detection of wind turbine blade abnormalities through a deep learning model integrating VAE and neural ODE. *Ocean Engineering*, *302*, 117689. https://doi.org/https://doi.org/10.1016/j.oceaneng.2024.117689

Yegnarayana, B. (2009). *ARTIFICIAL NEURAL NETWORKS*. PHI.

Yerdelen, C., & Abdelkader, M. (2021). HYDROLOGICAL DATA TREND ANALYSIS WITH WAVELET TRANSFORM. *Comptes Rendus De L Academie Bulgare Des Sciences*, *74*(8), 1194-1202. https://doi.org/10.7546/crabs.2021.08.11

Yi, J., & Chen, Z. (2023). Prediction of deck grillages lifting deformation using a data-driven Def-GAN network. *Ocean Engineering*, *287*, 115788. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115788

Young, I. R., & Ribal, A. (2019). Multiplatform evaluation of global trends in wind speed and wave height. *Science*, *364*(6440), 548-552. https://doi.org/doi:10.1126/science.aav9527

Zamani, A., Solomatine, D., Azimian, A., & Heemink, A. (2008). Learning from data for wind–wave forecasting. *Ocean Engineering*, *35*(10), 953-962. https://doi.org/https://doi.org/10.1016/j.oceaneng.2008.03.007

Zhang, J., Zhang, X., Wang, X., Ning, P., & Zhang, A. (2023). Reconstructing 3D ocean subsurface salinity (OSS) from T–S mapping via a data-driven deep learning model. *Ocean Modelling*, *184*, 102232. https://doi.org/https://doi.org/10.1016/j.ocemod.2023.102232

4. Discussion

This section presents a detailed analysis of the results obtained from the application of the developed AI models. An examination is conducted to identify model errors and compare performance with traditional methodologies. Critical aspects of the models are highlighted, and a set of general rules and guidelines is provided to support future work in this area. Additionally, a dedicated subsection outlines the specific contributions of each article included in the thesis.

4.1. Wave downscaling

4.1.1. Results

4.1.1.1. Long-term regime

The summarized results for long-term regime for the two locations under study are presented in Tables 4-1 and 4-2. They provide the average values, standard deviations (SD), the mean increment (Δ), and the relative increment (%Δ) of Hs and Tp in shallow depth (coastal wave climate) for the years 2035 and 2050.

Table 4-1. Evolution of wave climate in Valencia under climate change.

	Hs			**Tp**		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	0.66 m	0.64 m	0.82 m	5.25 s	4.81 s	5.02 s
SD	0.39 m	0.39 m	0.42 m	1.19 s	0.65 s	0.73 s
Δ	-	-0.03 m	0.17 m	-	-0.44 s	-0.23 s
%Δ	-	-4.55%	25.75%		-8.3%	-4.4%

Table 4-2. Evolution of wave climate in Cabo de Palos under climate change.

	Hs			**Tp**		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	0.88 m	0.86 m	0.97 m	5.83 s	5.63 s	5.78 s
SD	0.47 m	0.50 m	0.51 m	1.15 s	0.71 s	0.71 s
Δ	-	-0.01 m	0.09 m	-	-0.20 s	-0.05 s
%Δ	-	-1.14%	10%		-3.43%	-0.9%

As can be seen from the tables, little changes in the long-term maritime climate in the Western Mediterranean associated with climate change are expected. One striking observation is that the SD of Tp, instead of increasing over time, decreases. Initially, it seems logical that the standard deviations of the parameters defining the maritime climate would increase in the scenario of climate change. Indeed, this is the case with Hs, but the opposite occurs with Tp.

This phenomenon results from the applied methodology. Real data were used to model wave states for the present period. In contrast, the 2035 and 2050 projections were based on a virtual database, in which a relationship between peak period (Tp) and significant wave height (Hs) was established using the formula recommended by the Spanish authority responsible for climate characterization. The fact that present Tp values derived from actual records exhibit much greater dispersion compared to the future Tp values modelled using Equation 7 $\left(T_p = a \times (H_s)^b\right)$ underscores the inherent challenge in accurately modelling the relationship between Hs and Tp. While there exist analytical formulas linking Hs and Tp, the behaviour of the sea is far more complex and random than what can be captured by a simple formula. This complexity has been acknowledged previously by the authors in Portillo Juan et al. (2022a) (Paper 3.4.).

Figures 4-1 and 4-2 show the evolution of Hs and Tp in time. The records were ordered from lowest to highest. The obtained Hs and Tp were sorted based on Hs in deep waters. This sorting method involved arranging the records from deep waters for the present, 2035, and 2050, in ascending order of Hs. Subsequently, this order was applied to the corresponding values of Hs and Tp in coastal waters, which were derived after applying the ANN model to waves in deep water. Furthermore, the method of sorting the values of Hs and Tp to create the graphs amplifies the appearance of scattered data. The data was sorted based on the value of Hs, meaning that Tp is ordered based on Hs in deep waters rather than directly on Tp values. This ordering approach remains consistent for both the propagation of Hs and the virtually generated Tp values following Equation 7 However, the maximum values of Hs are not necessarily linked with the maximum values of Tp; the behaviour of sea states is inherently more random.

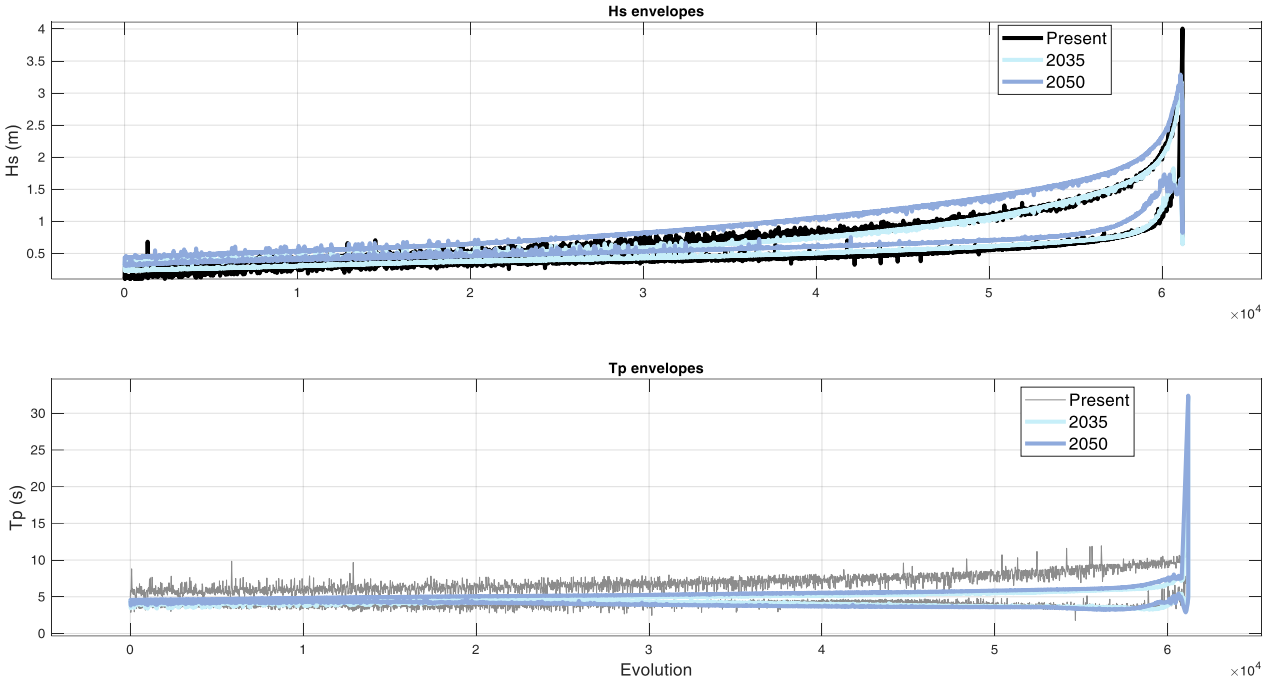


Figure 4-1. Hs (up) and Tp (down) evolution in Valencia under climate change.

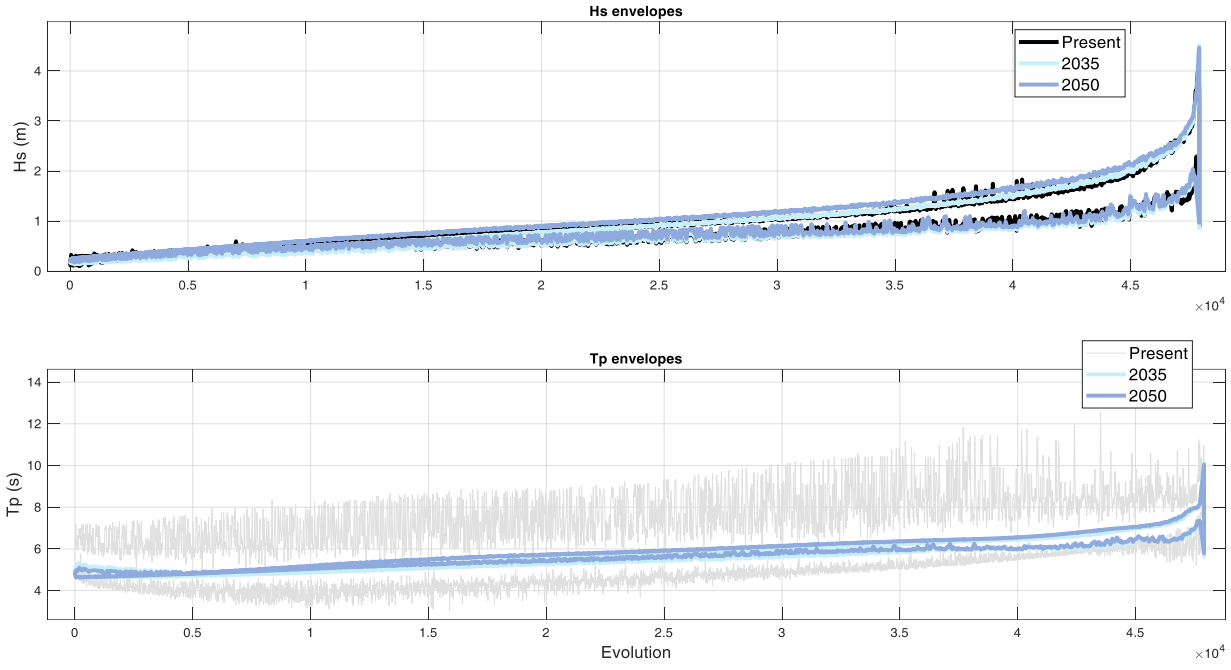


Figure 4-2. Hs (up) and Tp (down) evolution in Cabo de Palos under climate change.

The black and grey lines represent the present wave climate, while the blue lines the wave climate in the future, as can be seen in the legend of the graph.

Future projections in the upper graph of Figure 4-2 (Hs in Cabo de Palos) are not as pronounced as in the upper graph of Figure 4-1 (Hs in Valencia), where the blue and grey stripes can be seen much more clearly.

With respect to the evolution of Tp in Cabo de Palos (lower graph in Figure 4-2), it can be clearly observed the phenomenon that it was previously mentioned of the dispersed values of Tp in the present. The virtually generated data were created according to Equation 7. However, as it can be observed, the wave behaviour in Cabo de Palos is far from conforming to this relationship. While areas like Valencia exhibit a relatively fulfilled relationship, in regions like Cabo de Palos, this relationship does not hold true.

4.1.1.2. Extreme regime

To study the extreme distribution, the 95th percentile of the records were considered, following the same procedure as in section 4.1.1.1. The results are presented in Tables 4-3 and 4-4.

Table 4-3. Evolution of extreme wave climate in Valencia under climate change.

	Hs			**Tp**		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	1.69 m	3.10 m	3.23 m	7.12 s	9.00 s	8.80 s
SD	0.40 m	0.44 m	0.39 m	1.54 s	1.92 s	1.28 s
Δ	-	1.42 m	1.55 m	-	1.89 s	1.69 s
% Δ	-	84.02%	91.72%	-	26.55%	23.74%

Table 4-4. Evolution of extreme wave climate in Cabo de Palos under climate change.

	Hs			**Tp**		
Year	**2020**	**2035**	**2050**	**2020**	**2035**	**2050**
Mean	2.09	2.76	2.89	7.47	8.18	8.29
SD	0.51 m	0.62 m	0.6 m	0.84 s	0.67 s	0.63 s
Δ	-	0.67	0.80	-	0.71	0.82
% Δ	-	32.06%	38.28%	-	9.50%	11.00%

In the case of extreme events, the increase in wave heights is significant and is much more pronounced than in the long-term regime. Increases of up to 1.5 m are reached at Valencia by the year 2050 and up to 0.8 m at Cabo de Palos. With respect to the peak period, it can be seen how it also increases up to 1.7 s.

In the case of Tp, the increase is not as pronounced as it is for Hs. This can once again be explained by the greater difficulty in accurately modelling Tp and by the

fact that the theoretical relationship between Hs and Tp does not always hold perfectly in reality.

Figures 4-3 and 4-4 show the same graphs presented for long-term distribution, but for extreme events.

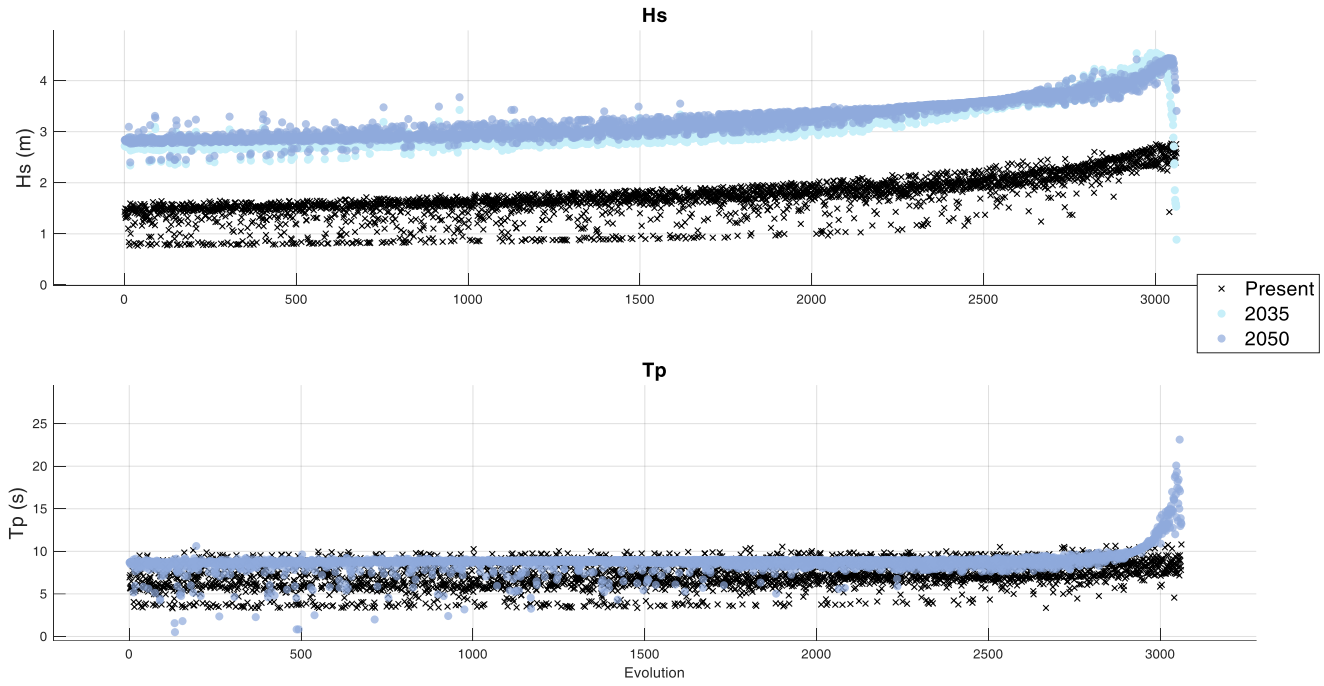


Figure 4-3. Extreme Hs (up) and Tp (down) evolution in Valencia under climate change.

Figure 4-3 shows how for the extreme regime there is a large difference between the current maritime climate and the future climate change scenarios. The wave heights associated with 2035 and 2050 for extreme events are above those of today. In the case of the peak period, it can also be seen that current periods are below the peak periods associated with future extreme events.

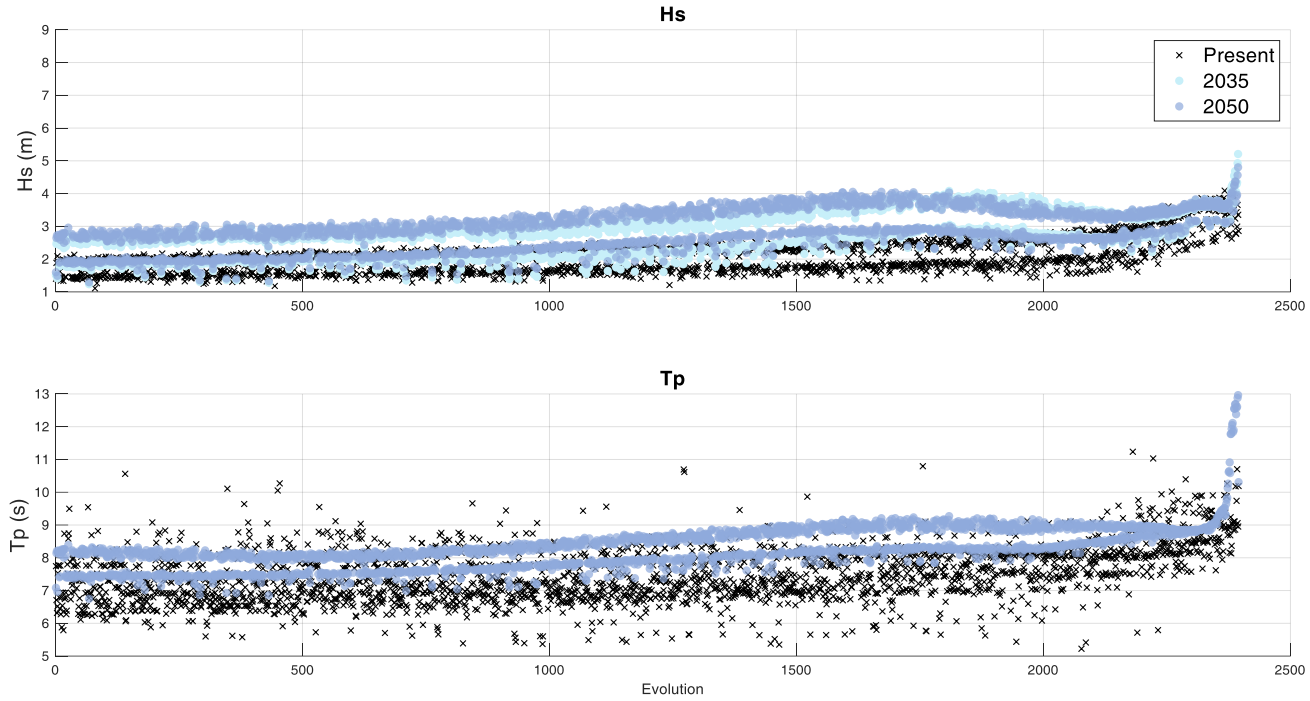


Figure 4-4. Extreme Hs (up) and Tp (down) evolution in Cabo de Palos under climate change.

As happened with the long-term regime, the trends in Cabo de Palos are less pronounced than in Valencia, but it can be clearly appreciated how the values associated with present climate are far below the ones predicted for 2035 and 2050.

As a summary of the results, Table 4-5 shows the expected changes in Hs and Tp for the Western Mediterranean. Extreme events are going to increase significantly, which is particularly relevant for our ports and structures.

Table 4-5. Wave climate under climate change scenario in the Western Mediterranean.

	ΔHs (m)		**ΔTp (s)**	
Year	**2035**	**2050**	**2035**	**2050**
Long-term	0.0 m	0.0 – 0.2 m	0.0 s	0.0 s
Extreme events	0.5 – 1.4 m	0.6 – 1.6 m	0.8 – 1.8 s	0.7 – 1.8 s

4.1.2. Comparison and guide for AI model development

After the ANN model for wave downscaling was designed, trained, and tested, its performance was compared with that of SWAN, the most widely used numerical model for wave downscaling among coastal and maritime engineers.

SWAN simulations were run using the user interface developed by the IHC (IHC, 2013). A DTM (Digital Terrain Model) of the coastline with the bathymetries from GEBCO was developed, generating a point cloud with XYZ information. The WGS84/UTM Zone 30N geodesic system was used.

Apart from the wave records used in the ANN model wind parameters were also considered as secondary parameters for the SWAN model. These values were obtained from the same source (*Puertos del Estado. Oceanografía. Registros históricos de oleaje*): the wind velocity (W_v) and wind direction (W_{dir}) were used.

Two model grids were used for the present study, an outer and a nested grid. The use of more than one grid has the aim of improving the resolution and accuracy of the model. The outer grid, which includes the deep-water wave data, has a resolution of 200 m with square elements; dimensions of 46400 × 39800 m and 232 X-nodes and 199 Y-nodes in Valencia, and dimensions of 51600 × 42200 m and 258 X-nodes and 211 Y-nodes in Cabo de Palos. The nested grid, which has a resolution of 40 m with square elements, has dimensions of 11000 × 9000 m and 225 X-nodes and 275 Y-nodes in Valencia and 8000 × 7600 m and 200 X-nodes and 190 Y-nodes in Cabo de Palos.

The SWAN model version was 41.31AB and operated using an implicit integration scheme. During simulations, the model was set to run in stationary model with the third-generation option (GEN3). The physics included whitecapping parametrization with $C_{ds} = 2.36 \cdot 10^{-5}$ (Komen et al., 1996); depth limited wave breaking with $\gamma = 0.73$ and $\alpha = 1$ (Battjes & Janssen, 1978); bottom friction with Cb = 0.038 m2·s− 3 using JONSWAP spectrum (Hasselmann et al., 1973); quadruplet interactions with $\lambda = 0.25$ and $C_{nl45} = 3 \cdot 10^7$ using the Discrete Interaction Approximation (DIA) (Hasselmann & Hasselmann, 1985); and refraction as default values. The spectrum of the model considered 36 frequencies (0.0356 Hz–1.00 Hz) and 72 directions. The shape of the spectra was defined with JONSWAP spectrum ($\gamma = 10, \sigma = 10/\gamma = 3.3, \sigma = 20$) depending on the energy of the incident waves (in function of the wave height).

The forcing functions of the wave data were applied uniformly over the entire model domain. In all cases a sea level of +0.00 m (above LAT) and no currents were considered. Finally, the model output (in this case only Hs, Tp and Dir) values were extracted and processed.

To compare the performance of SWAN and the AI-based models, the R, MSE, RMSE, MAE, MAPE and Bias – previously defined in Section 2. Methodology – were used.

Figure 4-5 shows the regression plot of the ANN models in Valencia and Cabo de Palos. The performance of the proposed model is very accurate, with a correlation coefficient of approx. 0.95. The Cabo de Palos models show a slightly better performance than the Valencia model, which may be related to the quality of the data available in this location.

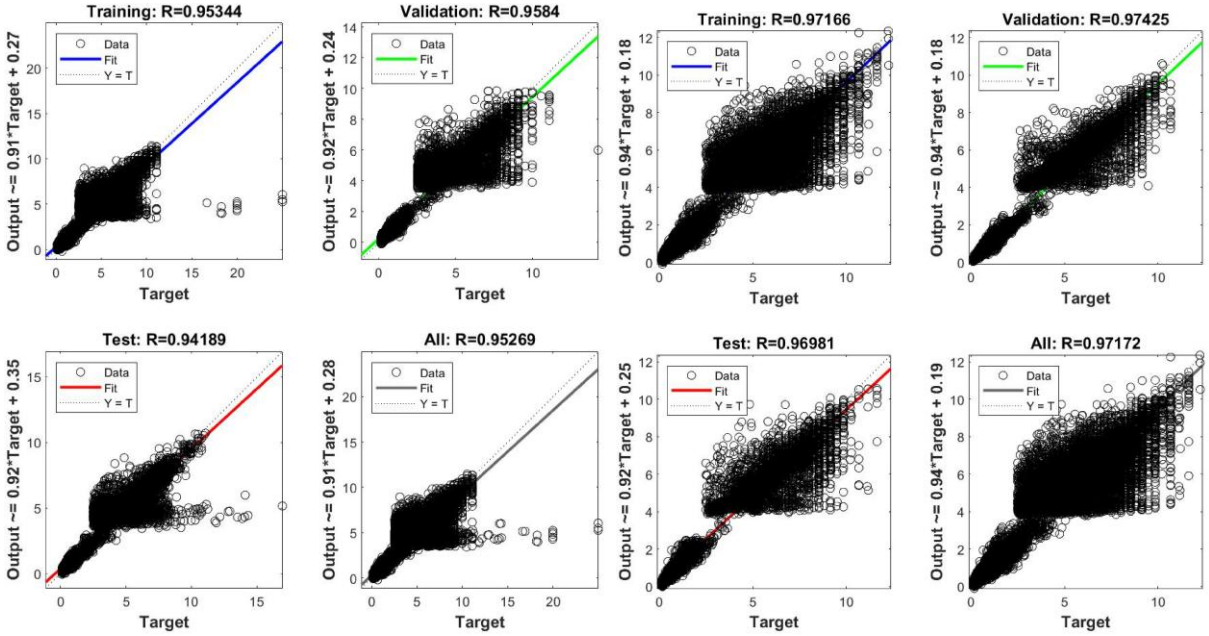


Figure 4-5. Regression plots in Valencia (left) and Cabo de Palos (right).

Figure 4-6 presents the error histogram of the ANN in Valencia and Cabo de Palos, respectively. Error histograms represent in the X-axis the difference between the target values and the results obtained, and in the Y-axis the number of data samples that have these errors. In the case of Valencia around 85,000 data samples obtained with the ANN model have an overprediction of 0.07303 (main bar in the graph on the left), while in Cabo de Palos around 60,000 data samples obtained with the ANN model have an overprediction of 0.1558 (main bar in the graph of the right). The remaining bars of both graphs show the number of samples that present higher errors.

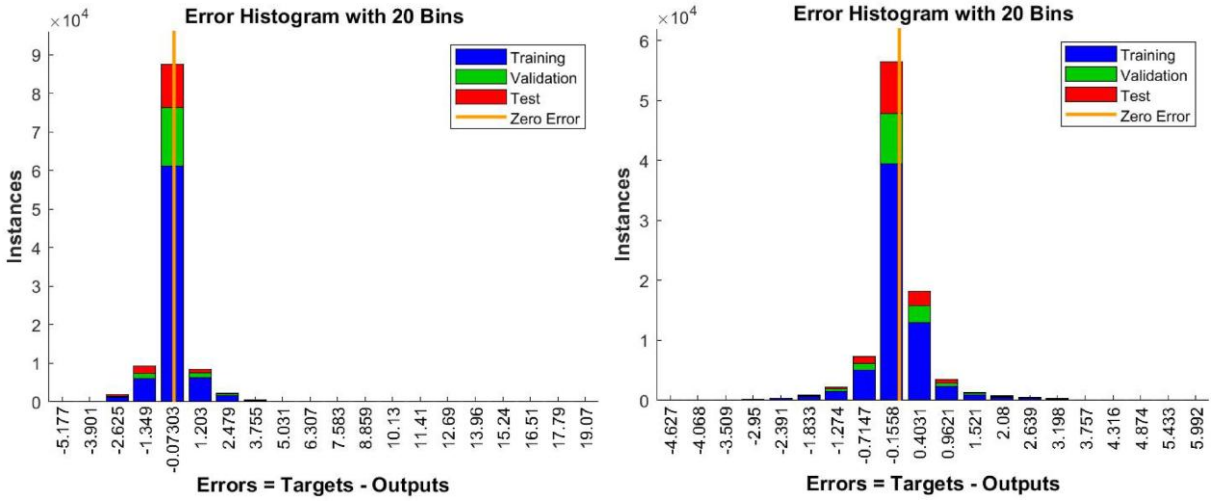


Figure 4-6. Error histogram in Valencia (left) and Cabo de Palos (right).

Figure 4-7 presents the results for the SWAN model. From these figures it can already be seen that the accuracy in the prediction of periods in SWAN is slightly lower.

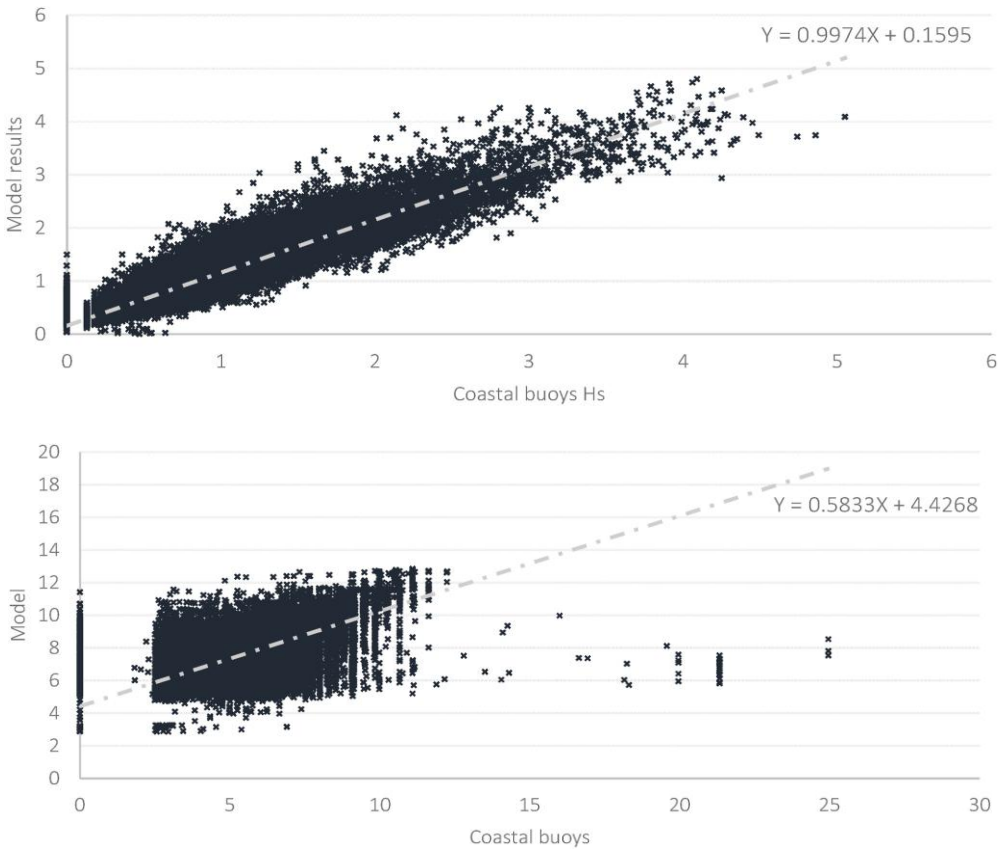


Figure 4-7. Regression plot for SWAN model. Hs up and Tp down.

Table 4-6 presents the values of the statistical analysis for both variables analysed, Hs and Tp, and both models. The ANN model performs slightly better than SWAN.

Table 4-6. Statistical analysis of the models for wave downscaling.

	Hs (m)		Tp (s)	
	SWAN	**ANN**	**SWAN**	**ANN**
R	0.94	0.96	0.72	0.96
MSE	0.06	0.023	6.2	1.00
RMSE	0.24	0.15	2.49	1.00
MAE	0.18	0.11	2.31	0.68
MAPE	0.33	0.05	0.49	0.05
Bias	0.15	0.002	2.26	0.03

After implementing both approaches for wave downscaling, the physics-based approach with SWAN model and the data-driven approach with Feed Forward Multilayer Perceptron Neural Networks, it has been proved that both approaches perform similarly for Hs wave downscaling, but that the data-driven approach even slightly outperforms the SWAN model for the Tp variable. Models with a physical approach lose accuracy in modelling the peak period of the waves due to the greater randomness that characterises this variable. This randomness makes Tp prediction much more suitable for data-driven models, such as neural networks, which are adept at capturing random patterns.

Moreover, in the case of areas such as the Spanish Mediterranean, enclosed basins, this difference between models is even greater, as these areas are governed by local wind and wave patterns which global or larger-scale models struggle to capture. These global models are subsequently used to feed the propagation models. Their errors are thus carried over to the propagation models, causing them to lose accuracy. On the other hand, data-driven models take the data directly from instrumental records, thus avoiding these possible initial errors.

The only thing that AI models require is precisely large amounts of data and consistent series. These models base their performance on the data that feed them, so that the more information available, the better they tend to perform. For this study there were hourly records of 6 and 8 years available, i.e., around 60,000 data samples. 25,000 samples are usually sufficient for a neural network to predict accurately in the long term, and 6,000 for the short term (Portillo Juan & Negro

Valdecantos, 2022). In this case, the number of data samples available was double than the minimum required for a neural network to perform well. The availability of enough quality data added to the fact that the Spanish Mediterranean is a closed sea governed by local phenomena that are difficult to model with a physical-approach model resulted in the ANNs model slightly outperforming the SWAN model for Tp wave downscaling.

As for the computational cost, it is widely known that models with a data-driven approach require much less time and computational resources than physical models. This is, in fact, one of the great advantages of these models and one of the main reasons why they are selected over traditional models. The only drawback is that although this data processing is much faster, it is also much less intuitive. Neural networks are "black-box" models whose results are difficult to explain and analyse. Furthermore, although there are different ways of optimizing these models, the most common, trial and error, was used in this study. This is a tedious process, which could be accused of being untechnical, as it may not ensure that the optimal model is reached. In spite of these drawbacks, in the present study neural networks have shown a better performance than the SWAN model with much less computational cost.

Regarding the input parameters used in the models, wave and wind parameters were used for SWAN, while only wave parameters were used for the ANN because maritime climate data alone were sufficient to achieve the desired results. Even using fewer data than SWAN, neural networks have shown a greater ability to model the Tp.

Regarding potential lines of research in the specific field of wave downscaling and maritime climate studies, future studies should focus on how to better leverage trend analysis and extrapolation of historical records. These aspects have been addressed at a basic level in this work; however, the field holds significantly more potential that remains unexplored due to time constraints.

4.2. Structural assessment

4.2.1. Results

The outcomes derived from applying the presented model at the port of Valencia are detailed in this subsection.

When evaluating the safety factors for sliding, both sections met the criteria with a significant margin of safety, rendering it unnecessary to consider analysing these results. Accordingly, the analysis will proceed to examine the remaining two variables: the required increase in height for the section to remain within the maximum allowable overtopping limits, and the weight increment of the slope in the case of the composite breakwater to satisfy safety requirements.

Figure 4-8 illustrates the increase in the vertical breakwater height necessary to satisfy design conditions, depicted as a function of projected wave height and period.

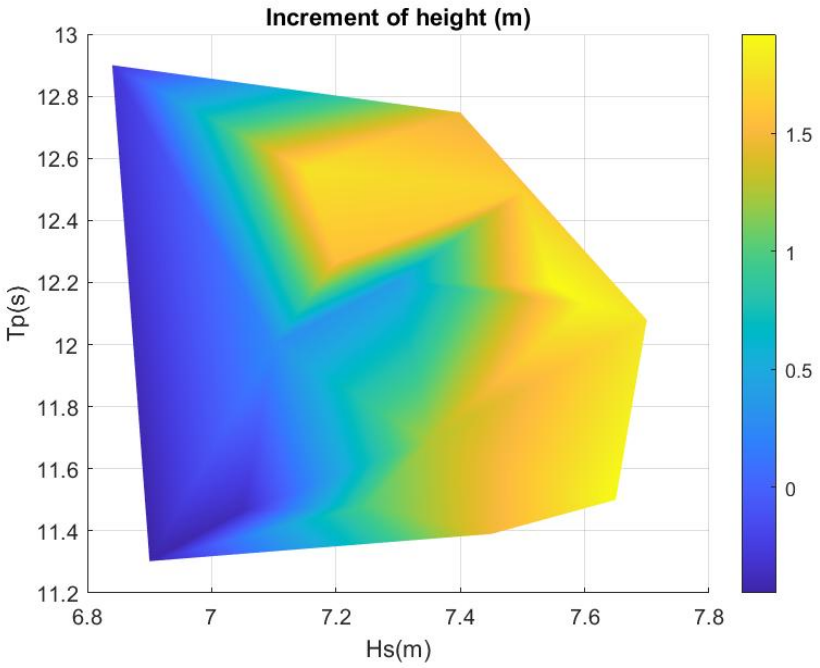


Figure 4-8. Increment of height needed for the vertical breakwater under different wave scenarios for 2035 and 2050.

This height augmentation spans from 0 m for the initial design conditions to 1.2 m for scenarios characterized by higher wave height and longer periods. As can be seen in the figure, the design of a vertical breakwater is very sensitive to the wave period. Pressures, forces, transmission, and other phenomena that affects its structural and hydraulic behaviour depends on the wave period. Notably, higher wave heights necessitate more substantial adaptation measures, as evidenced by the needed increased of the breakwater. Additionally, the figure reveals that even for less extreme wave heights but coupled with relatively high periods, considerable increases in seawall crest height are imperative for optimal design performance.

Figure 4-9 illustrates the analogous scenario for the compound breakwater. In this instance, the transmitted wave height of the breakwater is depicted on the X-axis. Per the figure, increase of height of the compound breakwater varies much less with wave conditions maintaining a consistently lower value (approximately 0.4 m in all cases). Similar to the vertical breakwater, the wave period plays a pivotal role in shaping the design of a compound breakwater. Careful consideration of wave period is imperative for their adaptation too, particularly when dealing with longer periods coupled with medium waves, which typically necessitate more robust adaptation measures, as can be observed in the presented figures.

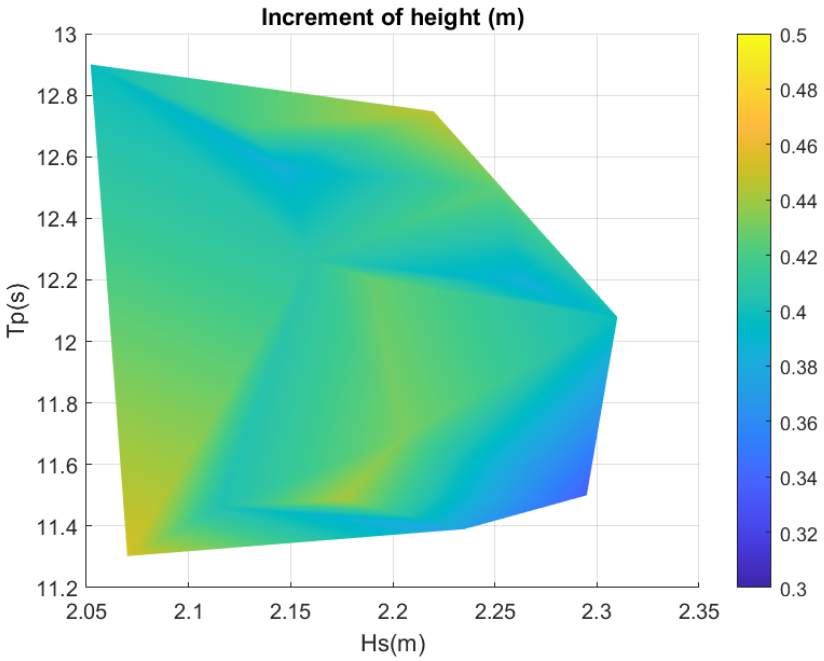


Figure 4-9. Increment of height needed for the compound breakwater under different wave scenarios for 2035 and 2050.

In Figure 4-10, the incremental weight required to meet safety standards is depicted, determined through the computation of the variable block damage N_{od}. Depending on the prevailing wave conditions, adjustments to the composite breakwater may necessitate weight increases of up to more than 1 t. Again, especially critical are large wave heights and high periods as illustrated by the yellow area in the upper right corner of the figure, indicative of the highest wave heights and periods.

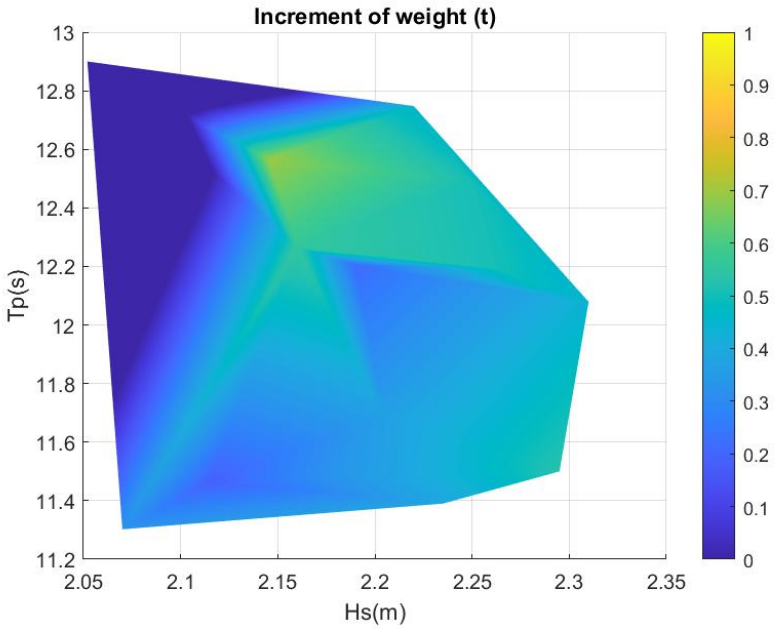


Figure 4-10. Increment of weight needed for the compound breakwater under different wave scenarios for 2035 and 2050.

The outcomes derived from the application of the two models to the Port of Valencia underscore the need for developing adaptation plans of coastal structures to climate change. Noteworthy events such as Storm Emma in 2018, Gloria in 2020, and Filomena in 2021, or the recent Storm Nelson in 2024 emphasize the importance of aligning our structures with evolving climate conditions. With extreme storms growing in frequency and intensity, the associated economic and social repercussions are escalating. To avert more severe consequences, it is essential to conduct comprehensive studies of structures situated along our coast.

This study reveals the need of increasing the crown height by up to 1.2 m in the North breakwater (head section) of the Port of Valencia and augmenting the weight of the compound section. Implementing these adaptation measures requires an examination of various options, necessitating a detailed study of the Valencia area. The different alternatives spans increasing the breakwater's crown height and adding more blocks to the armour to constructing a reef-breakwater or exploring nature-based solutions such as the utilization of seagrass meadows, which yield equivalent effects.

Future lines of research should investigate the best adaptation alternatives for the North Breakwater of Valencia, considering aspects such as safety, costs, and level of invasiveness. Additionally, the most resilient and sustainable options should be identified to achieve an optimal adaptation solution.

4.2.2. Comparison and guide for AI model development

Following the development of two new AI models for coastal protection structural adaptation—one focused on vertical breakwaters and the other on compound breakwaters—several conclusions have been reached regarding the application of artificial intelligence techniques to climate change and structural adaptation studies, from which a series of general rules and insights can be derived.

Both ANNs models have demonstrated commendable performance, showing low error rates and high correlation coefficients. This is particularly noteworthy as AI models significantly reduce the computational cost compared to the established numerical models used for coastal structures studies. Nonetheless, certain considerations warrant attention.

Foremost, the need for more extensive and higher-quality data as the complexity of the modelled process increases. Predicting or modelling maritime climate is simple because it involves few variables, allowing accurate models with small datasets. However, things get more complicated when dealing with processes like structural studies. These models, which encompass a multitude of variables, render the modelling process more intricate and less accurate, even with the incorporation of advanced AI techniques.

A comparative analysis between this second phase of the study (structural analysis) and the previous phase, focused on the propagation and projection of wave conditions, reveals two points of particular interest:

1. A significant increase in computational cost: structural adaptation models are approximately 100 times more expensive than wave propagation models.
2. A marginal decrease in accuracy: a loss of correlation coefficients of around 0.1 is observed in the structural adaptation studies compared to the maritime climate studies, which is likely related to the greater complexity of the process being modelled.

Similar contrasts emerge when comparing the models developed for the vertical breakwater and the composite breakwater. Notably, the composite breakwater model exhibits inferior results – r of 0.87 compared to a correlation coefficient of 0.98 for the vertical breakwater – despite utilizing a larger dataset (221,257 data samples compared to 141,660 for the vertical breakwater) Furthermore, the composite breakwater model necessitated more time for training, validation, and testing. This also may be due to the complexity of the modelled process, because

the structural and hydraulic behaviour of a composite breakwater is usually more complex than that of a vertical breakwater and, consequently, extracting the relationships between the variables that are used as input and the final output of the ANN is more difficult and requires more information and data to arrive at a consistent model.

The computational cost of the model concerning the quantity of input parameters employed has also been studied. The increase of computational cost of the ANN with the rising complexity of the modelled process and the associated variables involved highlights the importance of correctly select only the essential input variables. It has been shown that incorporating only one input more can lead to a substantial increase in computational time, ranging from 15% to 50%, contingent on modelling factors such as the number of epochs and the specific characteristics of the ANN (Figure 4-11). Therefore, to mitigate unnecessary computational expenses, it is imperative to conduct a sensitivity and correlation analysis before developing any ANN model to discern which variables are necessary to model the process and which can be ignored.

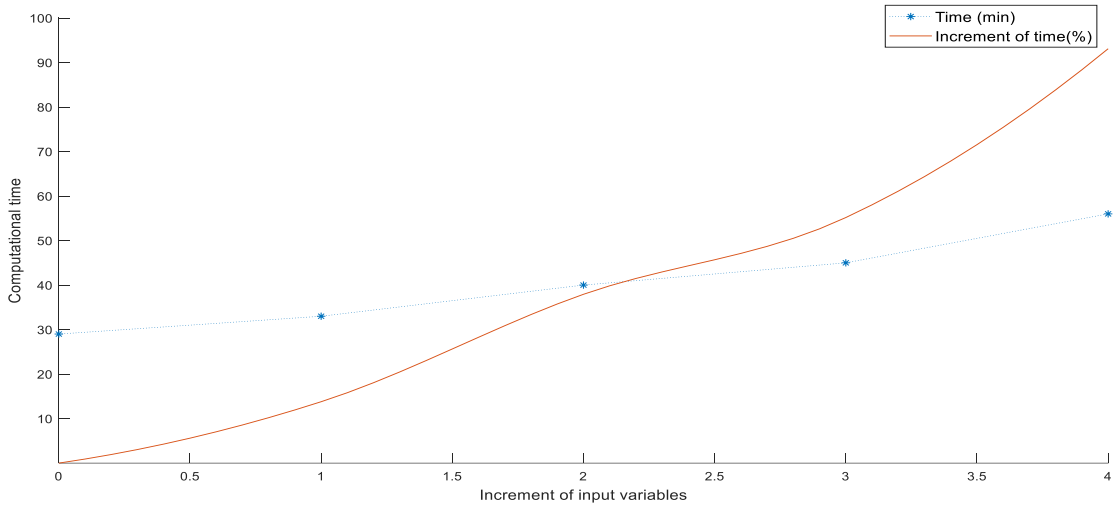


Figure 4-11. Increment of computational time with the increment of inputs (100 epochs).

Referring to hybrid models, incorporating AI techniques and optimization algorithms, such as Genetic Algorithms (GAs), is highly advantageous for minimizing computational time and improving the model performance. The implementation of a hybrid model, combining ANNs with GAs, resulted in a remarkable reduction in computational time up to 30-40 times, contingent upon the number of iterations — compared to the conventional trial-and-error approach. This synergy of AI and optimization techniques not only expedites the modelling

process but also enhances the overall efficiency of the model, showcasing a significant leap forward in computational effectiveness.

Finally, when analysing the errors of the model, it has been seen that the great majority of the errors are concentrated in the extreme values. Figure 4-12 shows the increase of the percentage of the RMSE of the total model with the percentile of the variable. The cumulative RMSE of the model and the proportionate distribution within each percentile was computed in intervals of 10 by 10 and graphically represented. Notably, in the context of composite breakwaters, nearly 60% of the accumulated model error is discerned in the 90th percentile. Both graphical representations illustrate a consistent elevation in accumulated error with percentile increments, which means that the model presents more errors the more extreme the values to be predict are.

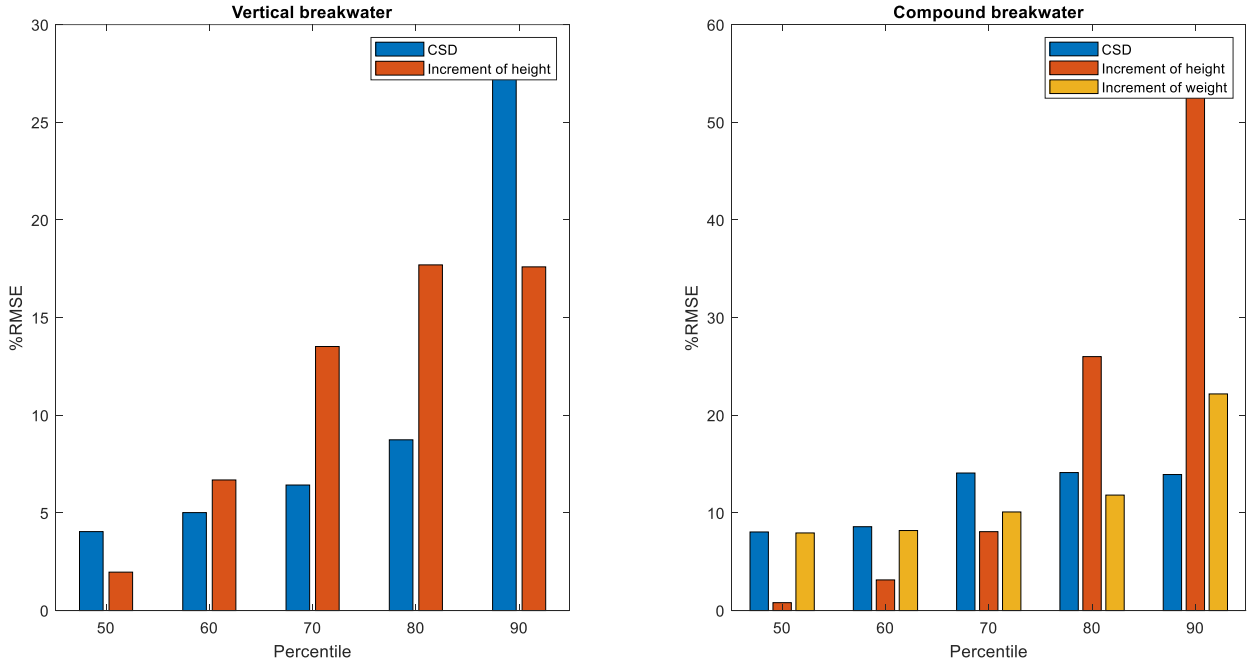


Figure 4-12. Distribution of the errors of the ANN-GA model.

This is something very common in AI models. Such models learn trends from existing data and may struggle when extrapolating to cases characterized by extreme values due to their underrepresentation, leading to diminished predictive accuracy in these scenarios. This is one of the main drawbacks of AI models and the reason they have been most criticized. In fact, the main limitation of the model proposed is this, that it can overestimate or underestimate adaptation measures for very extreme events, and therefore, future research lines should focus on addressing this issue and seek solutions to improve the prediction of extremes.

Lastly, it is crucial to underscore the significance of adopting a probabilistic approach and incorporating the peak period consideration when assessing the safety of structures. In certain cases, the sea state associated with the highest wave height may not necessarily be the most destructive. The peak period assumes a pivotal role in the design of vertical and composite breakwaters, exerting considerable influence on their safety. Consequently, these variables demand important attention. Therefore, when scrutinizing the adaptation of coastal protection structures, it is imperative to embrace a probabilistic and non-deterministic approach, evaluating the impact of various sea states on the structure, with particular emphasis on diverse peak periods.

4.3. Study of port agitation

4.3.1. Results

The deep learning model was successfully implemented and validated at the port of Valencia. Once the model was validated, it was used to study port agitation in the port itself under different wave climate conditions, based on present wave climate records and future projections. Periods from 5 s to 9 s and wave directions in sectors of 22.5º were studied (Figure 4-13).

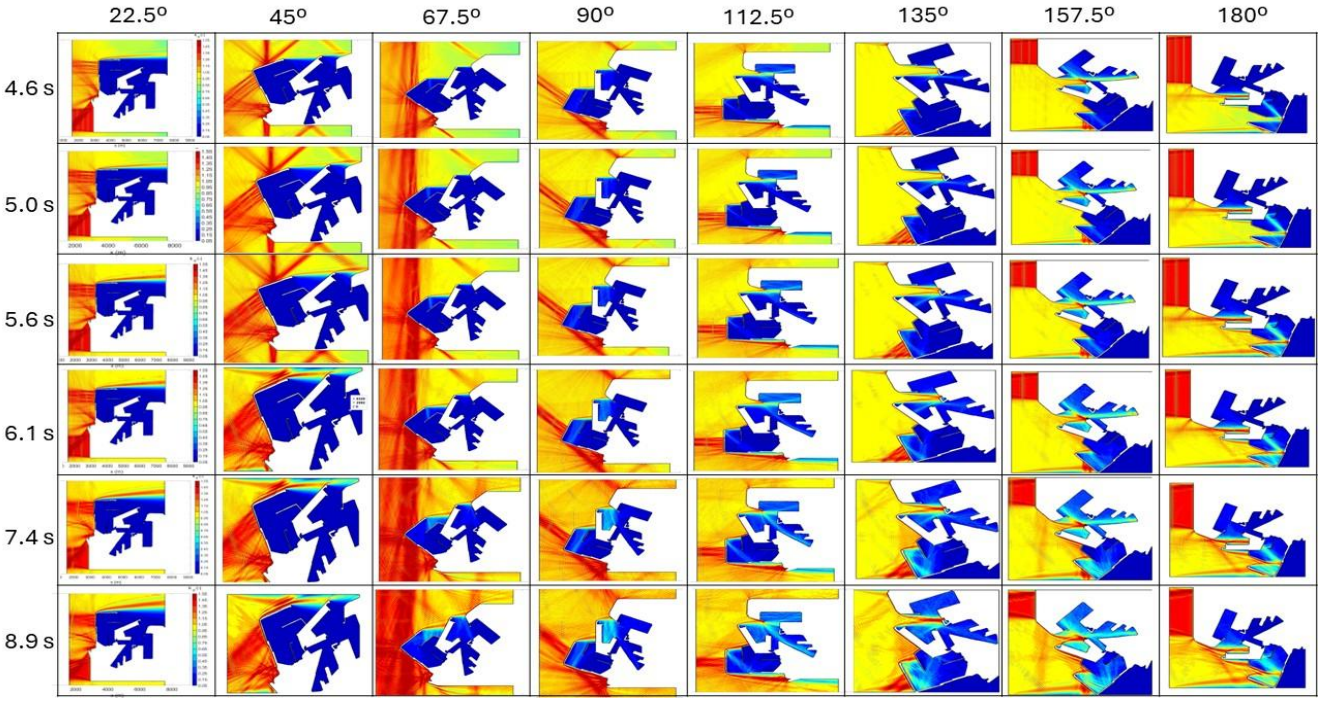


Figure 4-13. Port agitation in Valencia under different wave climate scenarios.

Although the peak period currently in the Valencia area is more around 5-6 s, there are increasingly more weather events in the Mediterranean associated with much longer peak periods, around 10 s (Portillo Juan et al., 2022a). That is why it is important to analyse port agitation under these periods. In fact, as seen in Figure 4-13, port agitation situations under longer periods are more critical than in the case of shorter periods. A comparison between the last row of the image matrix in Figure 4-13, corresponding to a wave period of 8.9 seconds, and the first row, corresponding to a period of 4.6 seconds, reveals a marked predominance of red and yellow hues. These colours indicate Kd values greater than one, signifying wave height amplification within the port. Regarding wave direction, the most problematic conditions are observed in the sector around 90º, indicating that storms originating from the East to the South (up to 180º) pose the greatest concern.

The worst port agitation conditions would occur for periods of 8.9 s and directions of 90º, 112.5º, 135º, 157.5º, and 180º, which are the 5 images on the right side of the last row of Figure 4-13 (Cases 4.1, 5.1, 6.1, 7.1 and 8.1). Since these are the worst cases, the analysis will focus on them (Figure 4-14).

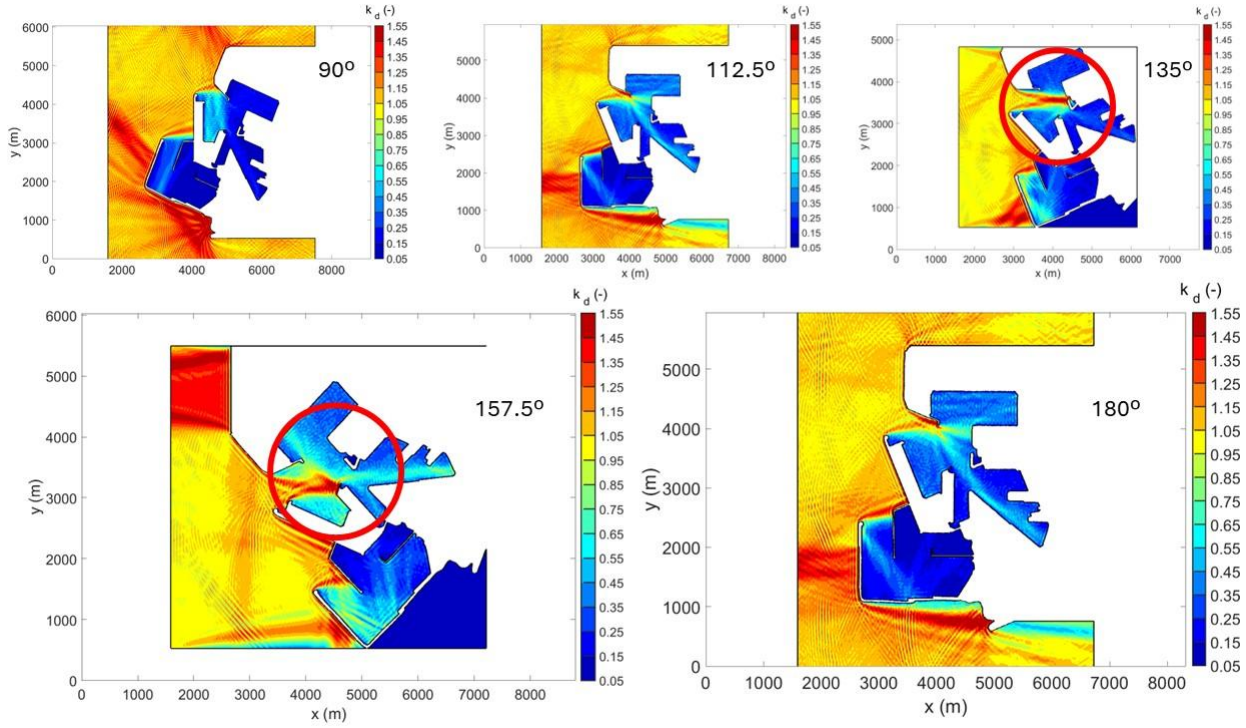


Figure 4-14. Worst cases of port agitation in Valencia (Tp= 8.9 s).

In general, the port of Valencia is a sheltered and protected port, even for longer periods that are predicted to occur more frequently in the future. The situations of

greater vulnerability occur for waves with directions of 135 and 157.5º, which are highlighted in red in Figure 4-14. For those conditions in the areas highlighted in red in the image, waves can amplify and cause problems in the port.

4.3.2. Comparison and guide for AI model development

The results obtained by the proposed AI model were compared to those generated by traditional numerical methods and it was verified that the results are practically the same. Figure 4-15 demonstrates the remarkable consistency between the outcomes obtained from both approaches.

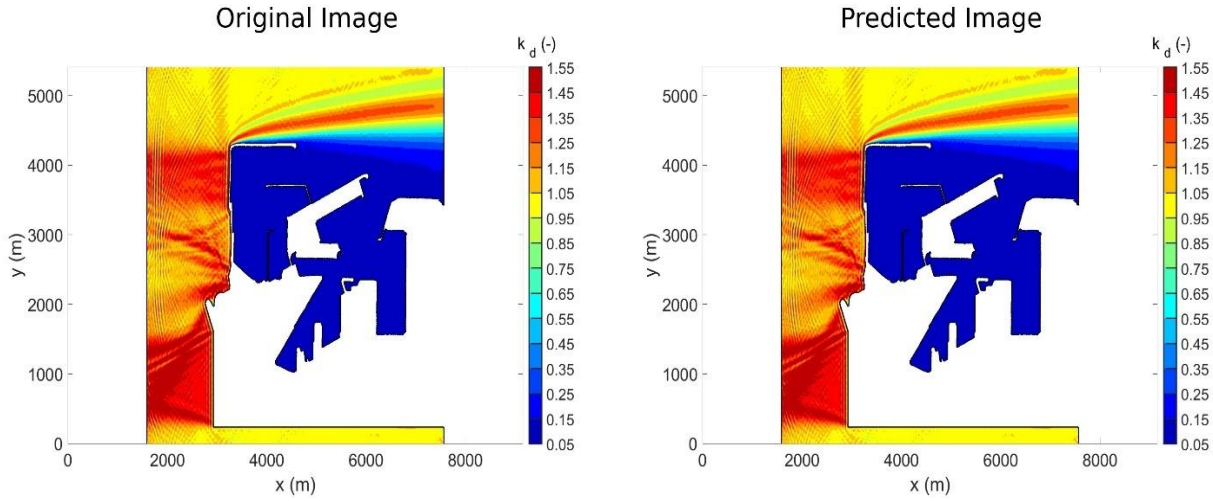


Figure 4-15. Comparison of the numerical model output (left) ant the AI model output (right).

The new proposed model not only demonstrates excellent results, but also significantly reduces all computational costs associated with numerical models of port agitation.

The trained AI model eliminates the need for extensive preprocessing tasks associated to numerical modelling like calibration, bathymetry development, and port layout adjustments. Once the AI model is trained, studying a new case only requires inputting the new conditions, whereas with numerical models, the entire model must be redeveloped. Additionally, as neural networks thrive on data, continuous usage enhances their precision, emphasizing the iterative improvement inherent to AI models when employed correctly.

In the event that other cases are to be studied, it would be as simple as entering the peak period, direction, and port layout into the Python console. The primary time and computational resource consumption occur during model optimization,

but once optimized, training takes only half an hour – on the specified CPU – and obtaining new cases simply involves entering the conditions into the already trained model.

Implementing these types of models would allow for significant savings in resources and computational time, which is of particular interest in the field. Given the current situation of climate change, where climate and maritime patterns are shifting, it is necessary to study how these changes will affect our ports. While existing studies predominantly focus on large ports due to their economic significance, small ports, most vulnerable to climate change, have been overlooked due to resource constraints. Implementing these resource-saving methodologies facilitates vital research in underserved areas, crucially contributing to our understanding and mitigation of port-related challenges in the face of climate change.

For the development of these models, it is highly recommended incorporating more complex data, such as bathymetry and port layout, directly from images themselves. This approach condenses essential information effectively and integrates these data into the model using CNNs, which have demonstrated an excellent performance in image processing tasks, instead of the more commonly used FFNNs in ocean engineering, that have much more limitations.

4.4. Contributions of the Publications

The contribution of each scientific publication derived from this thesis has been explained throughout the document. However, this section is dedicated to explaining in detail how the thesis was structured based on the scientific output.

This thesis is organized around two fundamental blocks: climate change and artificial intelligence. Accordingly, three blocks of publications can be distinguished (Table 4-7).

Table 4-7. Organization of scientific output

Block	Topic	Papers
Block1	Climate change	3.2., 3.3., 3.4. and 3.7.
Block2	Artificial Intelligence	3.5., 3.6. and 1 of Appendix B
Block3	Application of AI to Climate Change	3.8, 3.9. and 2 to 4 of Appendix B

The thesis began with a review of the state of the art in climate change, structured into two main phases:

1. First Phase: Climate change was addressed as a global phenomenon in its broadest sense, analysing its political, social, and economic implications. The objective of this phase was to achieve a deep understanding of what is truly happening. This included a technical and scientific analysis of the phenomenon, as well as a historical exploration of both the Earth and humanity. The aim was to place climate change in its temporal and dimensional context, assessing its implications, the responsibility of nations, potential solutions, and the magnitude of the problem. Many studies analyse the implications of climate change on different systems without first analysing what actually is. The abundance of available information and discussions about phenomena like global warming today can create the false impression that we truly understand them, which is often not the case. This first phase of study was intended to ensure that we genuinely understood what is happening. A specific analysis of the consequences of climate change makes much more sense when framed within a historical, social, and economic context. As a result of this first phase of study, the article titled - *A New Climate Change Analysis Parameter: A Global or a National Approach Dilemma* was published (DOI: doi.org/10.3390/en15041522)

2. Second Phase: After analysing climate change in its broadest sense, a specific review of the state of the art on its impact as a global phenomenon on the port and coastal system was conducted. In this second phase, the consequences of climate change in a specific context, namely our coasts and ports, were analysed. This analysis focused on studying how it will influence maritime climate, waves, sea level, and other key variables. The analysis focused on understanding how these different climate variables will be affected by the new situation we are facing. Based on this, the impact of these changes on our ports and coasts was studied, evaluating social and economic consequences as well as potential solutions and adaptation pathways. As a result of this second phase of research, the article titled *Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain* was published (DOI: doi.org/10.3390/su14127507)

After an exhaustive review of the state of the art on climate change and its study at both global (Phase I) and specific (Phase II) levels, an analysis of the evolution

of historical maritime climate records in Spain was conducted. This study focused on reviewing historical records and identifying trends, analysing the behaviour of maritime climate, identifying extreme events, and projecting future trends. Statistical studies and trend extrapolations were carried out using various tools to define Spain's specific situation in the face of climate change. As a result of this analysis, the article titled *Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast* was published (DOI: doi.org/10.3390/w14213453)

After studying climate change at global, regional, and local levels, a specific analysis of climate change in the Spanish Mediterranean was conducted. The main objective of this thesis is to develop a comprehensive study of climate change applied to ports, defining a complete methodology that allows these studies to be addressed at both ELU (structural level) and ELS (operational level). The ultimate goal is to propose an alternative and complementary methodology to the numerical models currently established in port and coastal engineering.

To achieve this, it was decided to use methodologies based on AI, developing data models that could overcome the limitations of traditional methodologies. Following the same approach applied in the climate change study, the first step in any research is always to conduct an exhaustive analysis of the state of the art. Therefore, the process began by investigating the various AI techniques and evaluating the level of development of these methodologies both globally and specifically in port engineering. This phase corresponds to Block 2 of Table 4-7.

The first milestone of this second phase was the completion of the AI course from Coursera (*Standford University Machine Learning Course*). Based on the material from this course, an exhaustive review of the most appropriate methodology for port and coastal engineering was conducted: Artificial Neural Networks (ANNs). As a result of this work, the article titled *Review of the application of Artificial Neural Networks in ocean engineering* was published (DOI: doi.org/10.1016/j.oceaneng.2022.111947).

During the study of this methodology, the high uncertainty associated with AI techniques was identified as the main drawback. This uncertainty affects both model parameters and their optimization and overall performance. In response to this issue, a more detailed study of these uncertainties was conducted, leading to the publication of a second article titled *Uncertainties in the application of Artificial Neural Networks in ocean engineering* (DOI:

doi.org/10.1016/j.oceaneng.2023.115193) which delves into the analysis of these uncertainties with the goal of better understanding the functioning of AI models.

Both articles focus on the application of ANNs in port and coastal engineering. To gain a more extensive view of AI in maritime and ocean engineering and following the same research line as the first climate change article, the field of study was expanded. An analysis of the state of the art of all AI techniques applied in maritime engineering was conducted, comparing their differences and applications, evaluating which models are the most appropriate, and extrapolating the applicability of certain techniques to other disciplines. As a result of this phase, the article titled *Advancements in Artificial Intelligence-Based Models in Maritime and Ocean Engineering* (Article 1 of Annex B) is currently in the process of publication.

After analysing the state of the art in both conceptual blocks of the thesis — climate change and Artificial Intelligence — the development of AI models in the three defined phases was carried out:

1. Study of maritime climate and propagation from deep to coastal waters.
2. Study of ELU (structural) of coastal protection structures.
3. Study of port operability.

In the development of the maritime climate propagation model, the question arose as to which of the available databases would be most suitable for characterizing the maritime climate in the Mediterranean. To resolve this uncertainty, an article was published comparing the different available data sources: *Comparison of SIMAR-WANA, ERA-5 and Waverys databases for maritime climate estimations and implications for coastal protection structures* (DOI: doi.org/10.1061/JWPED5.WWENG-2017).

Once this study was completed, the design of the AI model for maritime climate propagation began. As a first step, it was considered prudent to demonstrate the applicability of neural networks in maritime climate studies. This effort led to the publication of the article titled *Data-driven and physics-based approach for wave downscaling: A comparative study* (DOI: doi.org/10.1016/j.oceaneng.2023.115380).

After demonstrating the viability of neural networks in climate propagation studies, AI models specifically oriented towards maritime climate propagation in the context of climate change were designed. This part of the study culminated in the publication of the article titled *A new methodology for the study of future*

maritime climate using Artificial Neural Networks and Monte Carlo simulations under the context of climate change (DOI: doi.org/10.1016/j.ocemod.2024.102384).

Additionally, further research was conducted on how to make climate change projections using neural networks, specifically LSTMs. As a result of this work, the article *Improving Multi-Variable Wave Forecasting with AI: Integrating LSTM and Random Forest Using a Window-and-Flatten Technique* (fourth article in Annex B) is currently in the process of being published.

Following the completion of the first phase of the climate change study, the models for the second and third phases — the study of coastal protection structures and the study of port operability — were designed.

As a result of this final phase of the doctorate, the articles titled *An Artificial Intelligence approach for coastal adaptation studies: Insights from a case study in the Mediterranean Sea* (Second article of Annex B) and *Implementing Deep Learning in port agitation studies under the context of climate change* (Third article of Annex B) are in the process of publication.

5. Conclusions

Over three years ago, I delved into the world of Artificial Intelligence. At that time, the only thing I knew about AI was absolutely nothing. So, I started from scratch with the goal of applying these techniques to our professional field: port and coastal engineering.

This document aims to reflect the entire learning process I went through while attempting to implement these models. That is why the thesis follows a gradual approach, where the complexity of the models increases over time, mirroring my own learning journey.

When I first began studying AI and reviewing scientific papers, I found a world that was both unfamiliar and extremely complex. It was frustrating to grapple with documents that assumed a level of knowledge that neither I nor many other professionals unfamiliar with these models had. Therefore, I decided then that my thesis should make this knowledge accessible to the general public and not limit it to an elite group. There is no point in developing highly complex models that can only be implemented by specialized scientists; we must make them accessible to professionals and continue breaking down the barrier between science and industry. This is the goal I have personally pursued with this document.

The key to survival, both for humans and animals, is adaptation. Those who best adapt to new situations are the ones who survive. As a society, we face the challenge of adapting to climate change; as engineers, we have another challenge: to avoid falling behind and to adapt to a society immersed in change, new technologies, and a continuous dynamism.

AI appears to be the future of our society. Although port and coastal engineers currently have some of the best numerical models, we cannot ignore the development of AI. In a few years, these models could become obsolete if we do not incorporate new technologies. If our models fall behind, so will we and our field.

The potential of AI is enormous, and the sooner we start leveraging it, the better our models will become. Moreover, it is essential that these models are understandable and that we facilitate their development for technicians by providing clear guidelines and rules to assist in their implementation.

We cannot rely on our current models to remain the best; in a few years, they could become outdated. We need to explore new methodologies like AI, even if this involves unexpected or unsuccessful outcomes.

As researchers, we must learn to "waste time" exploring new methodologies and investigating new paths. Most of these paths will lead nowhere, but there will be one that allows us to progress. We must be willing to pay the price of investigating paths that may not lead to success.

Research is precisely that: testing, trying, and learning from each attempt. This thesis is just the beginning of what AI can contribute to our field, the result of the work of a small team over three years that has encountered numerous obstacles. However, if we continue to research and learn from our failures, we will keep advancing. With cooperation, interdisciplinary teams, and technicians, we can develop very powerful models, but this requires time, resources, and people.

We must prepare our field for the development of AI models and adapt our systems accordingly. AI depends on data, so it is urgent to improve our data collection systems and develop new monitoring tools so that in the future, we have the necessary tools to implement AI in our ports and coasts.

We need support, resources, time, cooperation, the willingness to fail and try again, and, as always, we also need funding.

We must continue seeking new solutions, striving for progress and advancement.

We have to adapt, and we must do so now, both as engineers and as a society.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., . . . Zheng., X. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. In.

Ahmed, F., Xiang, X. B., Jiang, C. C., Xiang, G., & Yang, S. L. (2023). Survey on traditional and AI based estimation techniques for hydrodynamic coefficients of autonomous underwater vehicle. *Ocean Engineering*, *268*, Article 113300. https://doi.org/10.1016/j.oceaneng.2022.113300

Albert, J. I. G., Burgos, M. A., Barquín, G. G., Carvajal, A. L., & Albiach, J. C. C. (2012). *Diques de Abrigo en los puertos de interés general del Estado* (P. d. Estado, Ed.).

Alemam, M., Yong, B., & Sani-Mohammed, A. (2023). A Proposed Merging Methods of Digital Elevation Model Based on Artificial Neural Network and Interpolation Techniques for Improved Accuracy. *Artificial Satellites*, *58*, 122-170. https://doi.org/10.2478/arsa-2023-0009

Allsop, W., & Hettiarachchi, S. (1989). *Wave reflections in harbours: design, construction and performance of wave absorbing structures*.

Battjes, J. A., & Janssen, J. P. F. M. (1978). Energy Loss and Set-Up Due to Breaking of Random Waves. In *Coastal Engineering 1978* (pp. 569-587). https://doi.org/doi:10.1061/9780872621909.034

Bento, P. M. R., Pombo, J. A. N., Mendes, R. P. G., Calado, M. R. A., & Mariano, S. J. P. S. (2021). Ocean wave energy forecasting using optimised deep learning neural networks. *Ocean Engineering*, *219*, Article 108372. https://doi.org/10.1016/j.oceaneng.2020.108372

Bindoff, N. L., Cheung, W. W. L., Kairo, J. G., Arístegui, J., Guinder, V. A., Hallberg, R., Hilmi, N., Jiao, N., Karim, M. S., Levin, L., O'Donoghue, S., Purca Cuicapusa, S. R., Rinkevich, B., Suga, T., Tagliabue, A., & Williamson, P. (2019). *Changing Ocean, Marine Ecosystems, and Dependent Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Boehringer, C. (2003). The Kyoto Protocol: A Review and Perspectives. *Oxford Review of Economic Policy*, *19*.

Booij, N., Ris, R. C., & Holthuijsen, L. H. (1999). A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research: Oceans*, *104*(C4), 7649-7666. https://doi.org/https://doi.org/10.1029/98JC02622

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M., & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*.

Casas-Prat, M., & Sierra, J. P. (2010). Trend analysis of wave storminess: wave direction and its impact on harbour agitation. *Nat. Hazards Earth Syst. Sci.*, *10*(11), 2327-2340. https://doi.org/10.5194/nhess-10-2327-2010

Casas-Prat, M., & Sierra, J. P. (2012). Trend analysis of wave direction and associated impacts on the Catalan coast. *Climatic Change*, *115*(3), 667-691. https://doi.org/10.1007/s10584-012-0466-9

Castillo;, E., & Pruneda, R. E. (2001). *Estadística Aplicada* (Moralea, Ed.).

Collins, M., Sutherland, M., Bouwer, L., Cheong, S.-M., Frölicher, T., Jacot Des Combes, H., Koll Roxy, M., Losada, I., McInnes, K., Ratter, B., Rivera-Arriaga, E., Susanto, R. D., Swingedouw, D., & Tibig, L. (2019). *Extremes, Abrupt Changes and Managing Risk. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Cramer, W., Guiot, J., Fader, M., Garrabou, J., Gattuso, J.-P., Iglesias, A., Lange, M. A., Lionello, P., Llasat, M. C., Paz, S., Peñuelas, J., Snoussi, M., Toreti, A., Tsimplis, M. N., & Xoplaki, E. (2018). Climate change and interconnected risks to sustainable development in the Mediterranean. *Nature Climate Change*, *8*(11), 972-980. https://doi.org/10.1038/s41558-018-0299-2

Demuth, H., & Beale, M. (1992). *Neural Network Toolbox* (M. Edition, Ed.).

Díaz, H., Teixeira, A. P., & Guedes Soares, C. (2022). Application of Monte Carlo and Fuzzy Analytic Hierarchy Processes for ranking floating wind farm locations. *Ocean Engineering*, *245*, 110453. https://doi.org/https://doi.org/10.1016/j.oceaneng.2021.110453

Dixit, P., & Londhe, S. (2016). Prediction of extreme wave heights using neuro wavelet technique. *Applied Ocean Research*, *58*, 241-252. https://doi.org/https://doi.org/10.1016/j.apor.2016.04.011

Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. *Comput. J.*, *7*, 149-154.

Franco, L., Gerloni, M. d., & Meer, J. v. d. (2010). WAVE OVERTOPPING ON VERTICAL AND COMPOSITE BREAKWATERS.

Francois, C. (2015). KERAS. In.

Gao, S., Huang, J., Li, Y., Liu, G., Bi, F., & Bai, Z. (2021). A forecasting model for wave heights based on a long short-term memory neural network. *Acta Oceanologica Sinica*, *40*(1), 62-69. https://doi.org/10.1007/s13131-020-1680-3

Garcia-Alonso, L., Moura, T. G. Z., & Roibas, D. (2020). The effect of weather conditions on port technical efficiency. *Marine Policy*, *113*, 103816. https://doi.org/https://doi.org/10.1016/j.marpol.2020.103816

GEBCO. (07/01/2023). *General Bathymetric Chart of the Oceans* https://www.gebco.net/

Ghasri, M. (2023). Hybrid Artificial Neural Network with Genetic Algorithm. In: MATLAB Central File Exchange.

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, *33*(8). https://doi.org/https://doi.org/10.1029/2006GL025734

Giorgi, F., & Lionello, P. (2008). Climate change projections for the Mediterranean region. *Global and Planetary Change*, *63*(2), 90-104. https://doi.org/https://doi.org/10.1016/j.gloplacha.2007.09.005

Goda, Y. (1970). A SYNTHESIS OF BREAKER INDICES.

Goda, Y. (1975). Irregular Wave Deformation in the Surf Zone. *Coastal Engineering in Japan*, *18*(1), 13-26. https://doi.org/10.1080/05785634.1975.11924196

Goda, Y. (2001). Performance-Based Design of Caisson Breakwaters with New Approach to Extreme Wave Statistics. *Coastal Engineering Journal*, *43*(4), 289-316. https://doi.org/10.1142/S0578563401000384

Goda, Y. (2010). *Random seas and design of maritime structures* (Vol. 33). World Scientific Publishing Company.

Gómez, R., Molina, R., Castillo, C., Rodríguez, I., & López, J. (2018). Conceptos y Herramientas Probabilísticas Para El Cálculo Del Riesgo En El. In: Ámbito.

Google Earth 7.3.6.9796 64-bit. (2016). http://www.earth.google.com

Gopinath, D. I., & Dwarakish, G. S. (2015). Wave Prediction Using Neural Networks at New Mangalore Port along West Coast of India. *Aquatic Procedia*, *4*, 143-150. https://doi.org/https://doi.org/10.1016/j.aqpro.2015.02.020

Graves, A. (2012). Long Short-Term Memory. In A. Graves (Ed.), *Supervised Sequence Labelling with Recurrent Neural Networks* (pp. 37-45). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24797-2_4

Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural Network Design* PWS Publishing.

Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Kruseman, P., Meerburg, A., Muller, P., Olbers, D., Richter, K., Sell, W., & Walden, H. (1973). Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Deut. Hydrogr. Z.*, *8*, 1-95.

Hasselmann, S., & Hasselmann, K. (1985). Computations and Parameterizations of the Nonlinear Energy Transfer in a Gravity-Wave Spectrum. Part I: A New Method for Efficient Computations of the Exact Nonlinear Transfer Integral. *Journal of Physical Oceanography*, *15*(11), 1369-1377. https://doi.org/https://doi.org/10.1175/1520-0485(1985)015<1369:CAPOTN>2.0.CO;2

Hinkel, J., Lincke, D., Vafeidis, A. T., Perrette, M., Nicholls, R. J., Tol, R. S. J., Marzeion, B., Fettweis, X., Ionescu, C., & Levermann, A. (2014). Coastal flood damage and adaptation costs under 21st century sea-level rise. *Proceedings of the National Academy of Sciences*, *111*(9), 3292-3297. https://doi.org/10.1073/pnas.1222469111

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., & Steltzer, H. (2019). *High Mountain Areas. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Holland, J. H. (1992). Genetic Algorithms. *Scientific American*, *267*(1), 66-73. http://www.jstor.org/stable/24939139

Hopstad, A. L. H., Argyriadis, K., Manjock, A., Goldsmith, J., & Ronold, K. O. (2018). DNV GL Standard for Floating Wind Turbines. ASME 2018 1st International Offshore Wind Technical Conference,

IHC. (2013). *Análisis Matemático y Estadístico de Variables Medioambientales (Version v.1.3.2).* . In https://ihcantabria.com/software-y-servicios-tic/software/

IMO. (2012). *International Shipping Facts and Figures – Information Resources on Trade, Safety, Security, Environment.* Maritime Knowledge Centre. International Maritime Organization. .

Informes anuales Oceanografía Puertos del Estado. (01/09/2022). https://www.puertos.es/es-es

IPCC. (2000). IPCC Special report. Emissions scenarios. Technical Report. . In.

Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface Water Mapping by Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(11), 4909-4918. https://doi.org/10.1109/JSTARS.2017.2735443

Janssen, P. A. E. M. (2008). Progress in ocean wave forecasting. *Journal of Computational Physics*, *227*(7), 3572-3594. https://doi.org/https://doi.org/10.1016/j.jcp.2007.04.029

Jcntsje, & Meer, V. d. (1988). Rock slopes and gravel beaches under wave attack.

Juan Nerea, P., Rodríguez Javier, O., & Valdecantos Vicente, N. (2024). Comparison of the SIMAR-WANA, ERA-5, and Waverys Databases for Maritime Climate Estimations and the Implications of Coastal Protection Structures. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *150*(1), 04023021. https://doi.org/10.1061/JWPED5.WWENG-2017

Kalra, R., Deo, M. C., Kumar, R., & Agarwal, V. K. (2005). RBF network for spatial mapping of wave heights. *Marine Structures*, *18*(3), 289-300. https://doi.org/https://doi.org/10.1016/j.marstruc.2005.09.003

Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, *80*(5), 8091-8126. https://doi.org/10.1007/s11042-020-10139-6

Komen, G. J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., & Janssen, P. (1996). *Dynamics and modelling of ocean waves*.

Lara, J. L., Lucio, D., Tomas, A., Di Paolo, B., & Losada, I. J. (2019). High-resolution time-dependent probabilistic assessment of the hydraulic performance for historic coastal structures: application to Luarca Breakwater. *Philos Trans A Math Phys Eng Sci*, *377*(2155), 20190016. https://doi.org/10.1098/rsta.2019.0016

Losada, M. A., Baquerizo, A., Ortega-Sanchez, M., & Avila, A. (2011). Coastal Evolution, Sea Level, and Assessment of Intrinsic Uncertainty. *Journal of Coastal Research*, 218-228. https://doi.org/10.2112/si59-023.1

Lucio, D., Lara, J., Tomas, A., & Losada, I. J. (2019). *Probabilistic hydraulic design of non-conventional breakwaters in shallow water conditions*. https://doi.org/10.18451/978-3-939230-64-9_027

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G., & Aarninkhof, S. (2018). The State of the World's Beaches. *Scientific Reports*, *8*(1), 6641. https://doi.org/10.1038/s41598-018-24630-6

Mandal, S., & Prabaharan, N. (2006). Ocean wave forecasting using recurrent neural networks. *Ocean Engineering*, *33*(10), 1401-1410. https://doi.org/10.1016/j.oceaneng.2005.08.007

Marhon, S. A., Cameron, C. J. F., & Kremer, S. C. (2013). Recurrent Neural Networks. In M. Bianchini, M. Maggini, & L. C. Jain (Eds.), *Handbook on Neural Information Processing* (pp. 29-65). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36657-4_2

Masselink, G., & Hughes, M. (2003). *An Introduction to Coastal Processes and Geomorphology*. https://doi.org/10.4324/9780203783740

MathWorks. (2021). *MATLAB version: 9.11.0 (R2021b)*. In Natick, Massachusetts: The MathWorks Inc. . https://www.mathworks.com

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1956). *A Proposal for the Darmouth Research Project on Artificial Intelligence* Darmouth

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. *Bulletin and Mathematical Biophysics* 115-133.

Medina, J. R., Garrido, J. M., Gómez-Martín, M. E., & Vidal, C. s. (2004). Armor Damage Analysis Using Neural Networks. *Coastal structures*, 236-248.

Mentaschi, L., Vousdoukas, M. I., Pekel, J.-F., Voukouvalas, E., & Feyen, L. (2018). Global long-term observations of coastal erosion and accretion. *Scientific Reports*, *8*(1), 12876. https://doi.org/10.1038/s41598-018-30904-w

Meredith, M., Sommerkorn, M., Cassotta, S., Derksen, C., Ekaykin, A., Hollowed, A., Kofinas, G., Mackintosh, A., Melbourne-Thomas, J., Muelbert, M. M. C., Ottersen, G., Pritchard, H., & Schuur, E. A. G. (2019). *Polar Regions. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Møller, M. (1993). Moller, M.F.: A Scaled Conjugate Gradient Algorithm For Fast Supervised Learning. Neural Networks 6, 525-533. *Neural Networks*, *6*, 525-533. https://doi.org/10.1016/S0893-6080(05)80056-5

Oppenheimer, M., Glavovic, B. C., Hinkel, J., van de Wal, R., Magnan, A. K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R. M., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., & Sebesvari, Z. (2019). *Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* (H.-O. Pörtner, D. C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, & N. M. Weyer, Eds.).

Portillo Juan, N., Matutano, C., & Negro Valdecantos, V. (2023). Uncertainties in the application of artificial neural networks in ocean engineering. *Ocean Engineering*, *284*, 115193. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115193

Portillo Juan, N., & Negro Valdecantos, V. (2022). Review of the application of Artificial Neural Networks in ocean engineering. *Ocean Engineering*, *259*, 111947. https://doi.org/https://doi.org/10.1016/j.oceaneng.2022.111947

Portillo Juan, N., & Negro Valdecantos, V. (2024). A novel model for the study of future maritime climate using artificial neural networks and Monte Carlo simulations under the context of climate change. *Ocean Modelling*, *190*, 102384. https://doi.org/https://doi.org/10.1016/j.ocemod.2024.102384

Portillo Juan, N., Negro Valdecantos, V., & del Campo, J. M. (2022a). Analysis of Monthly Recorded Climate Extreme Events and Their Implications on the Spanish Mediterranean Coast. *Water*, *14*(21), 3453. https://www.mdpi.com/2073-4441/14/21/3453

Portillo Juan, N., Negro Valdecantos, V., & del Campo, J. M. (2022b). A New Climate Change Analysis Parameter: A Global or a National Approach Dilemma. *Energies*, *15*(4), 1522. https://www.mdpi.com/1996-1073/15/4/1522

Portillo Juan, N., Negro Valdecantos, V., & del Campo, J. M. (2022c). Review of the Impacts of Climate Change on Ports and Harbours and Their Adaptation in Spain. *Sustainability*, *14*(12), 7507. https://www.mdpi.com/2071-1050/14/12/7507

Portillo Juan, N., Negro Valdecantos, V., & Troch, P. (2024). *Advancements in Artificial Intelligence -Based Models in Maritime and Ocean Engineering* [Journal Article].

Portillo Juan, N., Olalde Rodríguez, J., Negro Valdecantos, V., & Iglesias, G. (2023). Data-driven and physics-based approach for wave downscaling: A comparative study. *Ocean Engineering*, *285*, 115380. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.115380

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, *12*, 241-254.

Puertos del Estado. Oceanografía. Registros históricos de oleaje. (19/09/2022). https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx

Radder, A. C., & Dingemans, M. W. (1985). Canonical equations for almost periodic, weakly nonlinear gravity waves. *Wave Motion*, *7*(5), 473-485. https://doi.org/https://doi.org/10.1016/0165-2125(85)90021-6

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning internal representations by error propagation.* (J. L. McClelland, Ed.). MIT Press, Cambridge.

Ruscio, F., Costanzi, R., Gracias, N., Quintana, J., & Garcia, R. (2023). Autonomous boundary inspection of Posidonia oceanica meadows using an underwater robot. *Ocean Engineering*, *283*, 114988. https://doi.org/https://doi.org/10.1016/j.oceaneng.2023.114988

Saghiri, A. M., Vahidipour, S. M., Jabbarpour, M. R., Sookhak, M., & Forestiero, A. (2022). A Survey of Artificial Intelligence Challenges: Analyzing the Definitions, Relationships, and Evolutions. *Applied Sciences*, *12*(8), 4054. https://www.mdpi.com/2076-3417/12/8/4054

Schleussner, C.-F., Rogelj, J., Schaeffer, M., Lissner, T., Licker, R., Fischer, E. M., Knutti, R., Levermann, A., Frieler, K., & Hare, W. (2016). Science and policy

characteristics of the Paris Agreement temperature goal. *Nature Climate Change*, *6*(9), 827-835. https://doi.org/10.1038/nclimate3096

Standford University Machine Learning Course. (01/01/2022). Retrieved 01/01/2022 from https://www.coursera.org/learn/machine-learning/home

Takagi, H., Kashihara, H., Esteban, M., & Shibayama, T. (2011). ASSESSMENT OF FUTURE STABILITY OF BREAKWATERS UNDER CLIMATE CHANGE. *Coastal Engineering Journal*, *Vol.53*, 21p.

Teslyuk, V., Kazarian, A., Kryvinska, N., & Tsmots, I. (2021). Optimal Artificial Neural Network Type Selection Method for Usage in Smart House Systems. *Sensors*, *21*(1), 47. https://www.mdpi.com/1424-8220/21/1/47

Tolman, H. L. (1991). A third-generation model for wind waves on slowly varying, unsteady, and inhomogeneous depths and currents. *Journal of Physical Oceanography*, *21*(6), 782-797.

Troch, P. (1998). *A numerical model for propagation and transformation of linear water waves.*

Troch, P. (2018). Mildwave reference manual. Technical Report v4.1. In. Ghent University, Department of Civil Engineering.

Umesh, P. A., & Behera, M. R. (2020). Performance evaluation of input-dissipation parameterizations in WAVEWATCH III and comparison of wave hindcast with nested WAVEWATCH III-SWAN in the Indian Seas. *Ocean Engineering*, *202*, 106959. https://doi.org/https://doi.org/10.1016/j.oceaneng.2020.106959

UNCTAD. (2019). *Review of maritime transport*. United Nations Publication.

UNFCC. (1992). *Convención Marco de las Naciones Unidas Sobre el Cambio Climático* New York, USA.

van der Meer, J., Allsop, W., Bruce, T., Rouck, J., Kortenhaus, A., Pullen, T., Schüttrumpf, H., Troch, P., & Zanuttigh, B. (2016). *EurOtop: Manual on wave overtopping of sea defences and related sturctures - An overtopping manual largely based on European research, but for worlwide application (2nd edition)*.

van der Meer, J., Regeling, H., & Waal, J. (2001). Wave Transmission: Spectral Changes and Its Effects on Run-Up and Overtopping. *Coastal Engineering 2000 - Proceedings of the 27th International Conference on Coastal Engineering, ICCE 2000*, *276*. https://doi.org/10.1061/40549(276)168

Wang, H., Yin, J., Wang, X., & Ieee. (2020, 2020

Aug 22-24). Optimization of Wavelet Neural Network Model for Tide Prediction Based on Genetic Algorithm.*Chinese Control and Decision Conference* [Proceedings of the 32nd 2020 chinese control and decision conference (ccdc 2020)]. 32nd Chinese Control And Decision Conference (CCDC), Hefei, PEOPLES R CHINA.

Wang, N., Chen, Q., Zhu, L., & Sun, H. (2022). Integration of data-driven and physics-based modeling of wind waves in a shallow estuary. *Ocean Modelling*, *172*, 101978. https://doi.org/https://doi.org/10.1016/j.ocemod.2022.101978

Yao, Y., Yang, Y., Wang, Y., & Zhao, X. (2019). Artificial intelligence-based hull structural plate corrosion damage detection and recognition using

convolutional neural network. *Applied Ocean Research*, *90*, 101823. https://doi.org/https://doi.org/10.1016/j.apor.2019.05.008

Yedjour, D., & Benyettou, A. (2018). Symbolic interpretation of artificial neural networks based on multiobjective genetic algorithms and association rules mining. *Applied Soft Computing*, *72*, 177-188. https://doi.org/https://doi.org/10.1016/j.asoc.2018.08.007

Yogitha, R., & Mathivanan, G. (2018). *Performance Analysis of Transfer Functions in an Artificial Neural Network*. https://doi.org/10.1109/ICCSP.2018.8524387

Appendices

Appendix A

Wave propagation model codes

Example of ANN architecture optimization

```
clear
datainput = readmatrix("CP_inputs_conSLR.xlsx");
dataoutput = readmatrix("CP_outputs_HsyTp.xlsx");
datainput=transpose(datainput);
dataoutput=transpose(dataoutput);
NANI=isnan(datainput);
NANO=isnan(dataoutput);
NANO = [zeros(2,size(NANO,2)); NANO];
NAN=NANI|NANO;
valoresNAN=sum(NAN);
datainput = datainput(:,~valoresNAN)
dataoutput = dataoutput(:,~valoresNAN)
for i=1:100
    hidden=i;
    net= feedforwardnet(i,'trainlm');
    net=init(net);
    net.divideFcn = 'divideblock';
    [net, tr] = train(net, datainput, dataoutput);
    y_train=net(datainput(:,tr.trainInd));
    y_val=net(datainput(:,tr.valInd));
    y_trainTrue=dataoutput(:,tr.trainInd);
    y_valTrue=dataoutput(:,tr.valInd);
    error_train=(y_train-y_trainTrue).^2;
    error_val=(y_val-y_valTrue).^2;
    mse_train_Hs(:,i)=mean(error_train(1,:));
    mse_train_Tp(:,i)=mean(error_train(2,:));
    mse_val_Hs(:,i)=mean(error_val(1,:));
    mse_val_Tp(:,i)=mean(error_val(2,:));
end
rmse_train_Hs=sqrt(mse_train_Hs);
rmse_val_Hs=sqrt(mse_val_Hs);
rmse_train_Tp=sqrt(mse_train_Tp);
rmse_val_Tp=sqrt(mse_val_Tp);
subplot(1,2,1)
plot(1:100,rmse_train_Hs)
xlabel("Number of hidden neurons");
ylabel("RMSE");
title("Hs");
hold on
```

```
plot(1:100,rmse_val_Hs)
legend("Training","Validation");
hold off
subplot(1,2,2)
plot(1:100,rmse_train_Tp)
xlabel("Number of hidden neurons");
ylabel("RMSE");
title("Tp");
hold on
plot(1:100,rmse_val_Tp)
legend("Training","Validation");
hold off
```

Example of algorithm optimization

```
datainput = readmatrix("Valencia_Inputs.xlsx");
dataoutput = readmatrix("Valencia_Outputs.xlsx");
datainput=transpose(datainput);
dataoutput=transpose(dataoutput);
net = feedforwardnet(20,'trainscg');
[net, tr] = train(net, datainput, dataoutput);
```

Example of ANNs model evaluation

```
clear
datainput = readmatrix("CP_inputs_conSLR.xlsx");
dataoutput = readmatrix("CP_outputs_HsyTp.xlsx");
datainput=transpose(datainput);
dataoutput=transpose(dataoutput);
NANI=isnan(datainput);
NANO=isnan(dataoutput);
NANO = [zeros(2,size(NANO,2)); NANO];
NAN=NANI|NANO;
valoresNAN=sum(NAN);
datainput = datainput(:,~valoresNAN)
dataoutput = dataoutput(:,~valoresNAN)
net= feedforwardnet(22,'trainlm');
net=init(net);
net.divideFcn = 'divideblock';
[net, tr] = train(net, datainput, dataoutput);
    y_train=net(datainput(:,tr.trainInd));
    y_val=net(datainput(:,tr.valInd));
    y_trainTrue=dataoutput(:,tr.trainInd);
    y_valTrue=dataoutput(:,tr.valInd);
subplot(1,2,1)
plot(y_trainTrue(1,:),y_train(1,:),'.')
xlabel("Real Hs");
ylabel ("Hs predicted by the ANN");hold on
plot(0:4,0:4);
plot(y_valTrue(1,:),y_val(1,:),'x')
legend("Training", "y=x","Validation");
```

```
hold off
subplot(1,2,2)
plot(y_trainTrue(2,:),y_train(2,:),'.')
xlabel("Real Tp");
ylabel ("Tp predicted by the ANN");
hold on
plot(0:10,0:10);
plot(y_valTrue(2,:),y_val(2,:),'x')
legend("Training", "y=x","Validation");
hold off
```

Structural assessment model codes

Data featuring (Matlab)

```
%Lectura datos
clear
datosi=readmatrix("DiqueVertical2035.xlsx");
```

```
%Parametros basicos
Hd=2.5*datosi(:,1);
L0=9.81*datosi(:,2).^2/(2*pi());
```

```
%Resolucion L
for i=1:999000
    fun= @(X)(X-9.81/(2*pi())*datosi(i,2)^2*tanh(2*pi()*datosi(i,5)/X));
    L(i)=fzero(fun,9.81*datosi(i,2)^2/(2*pi()));
end
L=L';
```

```
%Parametros alfa auxiliares
alfa1 =
0.60+1/2*((4*pi().*datosi(:,5)./L)./(sinh(4*pi().*datosi(:,5)./L))).^2;
alfa21 = (datosi(:,5)- datosi(:,7))./(3*datosi(:,5)).*(Hd./datosi(:,7)).^2;
alfa22 = 2*datosi(:,7)./Hd;
alfa2 = min(alfa21,alfa22);
alfa3 = 1 - datosi(:,6)./datosi(:,5).*(1-1./cosh(2*pi().*datosi(:,5)./L));
```

```
%Presiones
tetha = 1.5*Hd;
p1 = (alfa1+alfa2)*1.025*9.81.*Hd;
p2 = p1./cosh(2*pi().*datosi(:,5)./L);
p3 = alfa1.*alfa3*1.025*9.81.*Hd;
```

```
p4 = (tetha-datosi(:,8))./tetha.*p1;
```

```
%Fuerzas y momentos
f1 = p4.*datosi(:,8);
cdg1 = datosi(:,6)+datosi(:,8)/2;
m1 = f1.*cdg1;
f2 = 1/2*(p1-p4).*datosi(:,8);
cdg2 = datosi(:,6)+datosi(:,8)/3;
m2 = f2.*cdg2;
f3=p3.*datosi(:,6);
cdg3 = datosi(:,6)/2;
m3 = f3.*cdg3;
f4 = 1/2*(p1-p3).*datosi(:,6);
cdg4 = 2/3*datosi(:,6);
m4 = f4.*cdg4;
```

```
%Coeficientes de seguridad
csd = 0.6*(datosi(:,10)*9.81-p3.*datosi(:,11))./(f1+f2+f3+f4);
csv = (datosi(:,10)*9.81.*datosi(:,11)/2+m1+m2+m3+m4-
p3.*datosi(:,11)/3)./(p3.*datosi(:,11)/3);
```

```
%overtopping
q = sqrt(9.81*datosi(:,1).^3).*0.082.*exp(-3*(datosi(:,9))./datosi(:,1));
INCRq = q*1000 - datosi(:,12);
NewRc = datosi(:,9);
INCRh = zeros(999000,1);
```

```
for i=1:999000
    if INCRq(i)>0
        NewRc(i)= (-
1/3*log(datosi(i,12)/(1000*0.082*sqrt(9.81*datosi(i,1)^3)))*(datosi(i,1)));
        INCRh(i) = NewRc(i) - datosi(i,9);
    end
end
```

```
output = [csd INCRh];
```

```
valores = find(INCRh);
outputf = [csd(valores) INCRh(valores)];
inputf = datosi(valores, :);
outputf2 = [outputf; output(1:20:end,:)];
```

```
inputf2 = [inputf; datosi(1:20:end,:)];
save outputf2;
save inputf2;
```

```
load inputf2050.mat
load outputf2050.mat
inputDV =[inputf2; inputf2050];
outputDV = [outputf2; outputf2050];
writematrix(inputDV, "I_DV_f2.xlsx");
writematrix(outputDV, "O_DV_f2.xlsx");
%Lectura datos
clear
datosi=readmatrix("DiqueCompuesto2050.xlsx");
```

```
%Parametros basicos
Hd=2.5*datosi(:,1);
L0=9.81*datosi(:,2).^2/(2*pi());
```

```
%Resolucion L
for i=1:999000
    fun= @(X)(X-9.81/(2*pi())*datosi(i,2)^2*tanh(2*pi()*datosi(i,5)/X));
    L(i)=fzero(fun,9.81*datosi(i,2)^2/(2*pi()));
end
L=L';
```

```
%Parametros alfa auxiliares
alfa1 =
0.60+1/2*((4*pi().*datosi(:,5)./L)./(sinh(4*pi().*datosi(:,5)./L))).^2;
alfa21 = (datosi(:,5)- datosi(:,7))./(3*datosi(:,5)).*(Hd./datosi(:,7)).^2;
alfa22 = 2*datosi(:,7)./Hd;
alfa2 = min(alfa21,alfa22);
alfa3 = 1 - datosi(:,6)./datosi(:,5).*(1-1./cosh(2*pi().*datosi(:,5)./L));
```

```
%Presiones
tetha = 1.5*Hd;
p1 = (alfa1+alfa2)*1.025*9.81.*Hd;
p2 = p1./cosh(2*pi().*datosi(:,5)./L);
p3 = alfa1.*alfa3*1.025*9.81.*Hd;
p4 = (tetha-datosi(:,8))./tetha.*p1;
```

```
%Fuerzas y momentos
f1 = p4.*datosi(:,8);
```

```
f2 = 1/2*(p1-p4).*datosi(:,8);
f3=p3.*datosi(:,6);
f4 = 1/2*(p1-p3).*datosi(:,6);
```

```
%Coeficientes de seguridad
csd = 0.6*(datosi(:,10)*9.81-p3.*datosi(:,11))./(f1+f2+f3+f4);
```

```
%overtopping
tan = rand([999000,1])*0.3 + 0.5;
```

```
iribarren = 1.25*tan.*datosi(:,2)./sqrt(datosi(:,1));
Ru2=zeros(999000,1);
for i=1:999000
    if iribarren(i) >1.5
        Ru2(i)=0.96*iribarren(i);
    else
        Ru2(i)=1.17*iribarren(i)^0.46;
    end
end
```

```
q = sqrt(9.81*datosi(:,1).^3).*8e-5.*exp(3.10*(Ru2-datosi(:,9))./datosi(:,1));
INCRq = q*1000 - datosi(:,12);
NewRc = datosi(:,9);
INCRh = zeros(999000,1);
```

```
for i=1:999000
    if INCRq(i)>0
        NewRc(i)= Ru2(i)-
(1/3.10*log(datosi(i,12)/(0.08*sqrt(9.81*datosi(i,1)^3)))*(datosi(i,1)));
        INCRh(i) = NewRc(i) - datosi(i,9);
    end
end
```

```
%Daño piezas
Tm = datosi(:,2)/1.2;
N=1000;
delta=(2.4/1.025-1);
sm=2*pi().*datosi(:,1)./(9.81*Tm.^2);
Nod=((2.5*datosi(:,1)./(delta*sm.^-0.1.*(datosi(:,13)./2.4).^(1/3))-
1)*1000^0.13/6.7).^(1/0.4);
```

```
NewW=datosi(:,13);
```

```
for i=1:999000
    if Nod(i) > 0.5
        NewW(i)= 2.4*(2.5*datosi(i,1)/(delta*sm(i)^-
0.1*6.7*0.5^0.4/1000^0.13+1))^3;
    end
end
incrW=NewW-datosi(:,13);
```

```
output = [csd INCRh incrW];
valores2 = find (incrW);
valores = find(INCRh);
outputf = [csd(valores) INCRh(valores) incrW(valores); csd(valores2)
INCRh(valores2) incrW(valores2)];
inputf = [datosi(valores, :);datosi(valores2, :)];
outputf2DC50 = [outputf; output(1:40:end,:)];
inputf2DC50 = [inputf; datosi(1:40:end,:)];
save outputf2DC50;
save inputf2DC50;
```

Model generation (Python)

VERTICAL BREAKWATER

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from deap import base, creator, tools, algorithms
import random
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split

# Load data
dfX = pd.read_excel('I_DV2050_f2.xlsx')
dfY = pd.read_excel('O_DV2050_f2.xlsx')

# Data preprocessing, se  dividen en train y validation y los escalamos (SOLO X)
scaler = MinMaxScaler()
x_train, x_test, Y_train, Y_test = train_test_split(dfX, dfY, test_size=0.3,
random_state=1234)
X_train = scaler.fit_transform(x_train)
X_test = scaler.transform(x_test)

# Define ANN model function
def create_model(input_shape, weights=None):
    model = keras.Sequential([
        layers.InputLayer(input_shape=input_shape),
```

```
        layers.Dense(47, activation='relu'),
        layers.Dense(1, activation='linear')
    ])

    if weights is not None:
        model.set_weights(weights)  # Ensure the weights are structured correctly

    model.compile(optimizer='adam', loss='mean_squared_error')
    return model

# Define fitness function
def fitness(individual):
    num_layers = 2  # Adjust based on model structure
    model = create_model(input_shape=X_train.shape[1:]) #Crea un modello nuevo

    weights_shapes = [w.shape for w in model.get_weights()] #Extrae la forma de
los pesos del mdelo
    split_indices = np.cumsum([np.prod(shape) for shape in weights_shapes]) #Lo
divide en arrays para cada capa
    weight_arrays = np.split(np.array(individual), split_indices[:-1])
    reshaped_weights = [w.reshape(shape) for w, shape in zip(weight_arrays,
weights_shapes)] #Les da forma para ajustarlos al modelo

    model.set_weights(reshaped_weights) #Reasigna los pesos
    loss = model.evaluate(X_test, Y_test, verbose=0) #Evalua el modelo
    return (loss,)

# Genetic Algorithm setup
creator.create("FitnessMin", base.Fitness, weights=(-1.0,)) #define la funcion
objetivo a minimizar (Fitness, mse) weights=(-1.0,)=minimizar
creator.create("Individual", list, fitness=creator.FitnessMin) #Set de pesos de la
red

toolbox = base.Toolbox()
toolbox.register("attribute", random.uniform, -1, 1) #Genera pesos aleatorios
entre -1 y 1
model = create_model(input_shape=X_train.shape[1:]) #Crea el modelo con los
pesos
n_weights = sum([np.prod(w.shape) for w in model.get_weights()])
toolbox.register("individual", tools.initRepeat, creator.Individual,
toolbox.attribute, n=n_weights) #Genera un individuo
toolbox.register("population", tools.initRepeat, list, toolbox.individual) #Genera
una población

toolbox.register("mate", tools.cxTwoPoint) #Utiliza two points crossover
toolbox.register("mutate", tools.mutGaussian, mu=0, sigma=1, indpb=0.1)
#Utiliza mutación Gaussiana
```

```
toolbox.register("select", tools.selTournament, tournsize=3) #Torunament
selection of 3. Toma 3 individuos de la población y se queda con el mejor
toolbox.register("evaluate", fitness) #Llama a la función objetivo, fitness

# Run Genetic Algorithm
population_size = 200 #En cada iteracion se cogen 200 individuos
num_generations = 100 #Iteraciones
crossover_prob = 0.8
mutation_prob = 0.2

population = toolbox.population(n=population_size)
algorithms.eaSimple(population, toolbox, cxpb=crossover_prob,
mutpb=mutation_prob, ngen=num_generations, verbose=True)

# Retrieve best individual
best_individual = tools.selBest(population, k=1)[0]
best_model = create_model(input_shape=X_train.shape[1:]) #Se queda con el
mejor modelo

# Reshape weights to match the model
weights_shapes = [w.shape for w in best_model.get_weights()]
split_indices = np.cumsum([np.prod(shape) for shape in weights_shapes])
weight_arrays = np.split(np.array(best_individual), split_indices[:-1])
reshaped_weights = [w.reshape(shape) for w, shape in zip(weight_arrays,
weights_shapes)]
best_model.set_weights(reshaped_weights)

# Evaluate best model
loss = best_model.evaluate(X_test, Y_test)
print(f'Final loss: {loss}')

# Save best model
best_model.save('best_model.h5')
```

- **TEST**

```
import tensorflow as tf
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler

# Load the model
model = keras.models.load_model("DVmodel.h5")
input_data = pd.read_excel('Input DV.xlsx')

scaler = MinMaxScaler()
input_data_norm = scaler.fit_transform(input_data)
```

```
predictions = model.predict(input_data_norm)
```

COMPOUND BREAKWATER HEIGHT

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from deap import base, creator, tools, algorithms
import random
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split

# Load data
dfX = pd.read_excel('I_DC2050_f2.xlsx')
dfY = pd.read_excel('O_DC2050_f2.xlsx')

# Data preprocessing, se  dividen en train y validation y los escalamos (SOLO X)
scaler = MinMaxScaler()
x_train, x_test, Y_train, Y_test = train_test_split(dfX, dfY, test_size=0.3,
random_state=1234)
X_train = scaler.fit_transform(x_train)
X_test = scaler.transform(x_test)

# Define ANN model function
def create_model(input_shape, weights=None):
    model = keras.Sequential([
        layers.InputLayer(input_shape=input_shape),
        layers.Dense(39, activation='relu'),
        layers.Dense(1, activation='linear')
    ])

    if weights is not None:
        model.set_weights(weights)  # Ensure the weights are structured correctly

    model.compile(optimizer='adam', loss='mean_squared_error')
    return model

# Define fitness function
def fitness(individual):
    num_layers = 2  # Adjust based on model structure
    model = create_model(input_shape=X_train.shape[1:]) #Crea un modello nuevo

    weights_shapes = [w.shape for w in model.get_weights()] #Extrae la forma de
los pesos del mdelo
```

```
    split_indices = np.cumsum([np.prod(shape) for shape in weights_shapes]) #Lo
divide en arrays para cada capa
    weight_arrays = np.split(np.array(individual), split_indices[:-1])
    reshaped_weights = [w.reshape(shape) for w, shape in zip(weight_arrays,
weights_shapes)] #Les da forma para ajustarlos al modelo

    model.set_weights(reshaped_weights) #Reasigna los pesos
    loss = model.evaluate(X_test, Y_test, verbose=0) #Evalua el modelo
    return (loss,)

# Genetic Algorithm setup
creator.create("FitnessMin", base.Fitness, weights=(-1.0,)) #define la funcion
objetivo a minimizar (Fitness, mse) weights=(-1.0,)=minimizar
creator.create("Individual", list, fitness=creator.FitnessMin) #Set de pesos de la
red

toolbox = base.Toolbox()

# Initialize weights based on Glorot initialization
def glorot_initializer(layer_shape):
    # Fan-in and fan-out for the layer
    fan_in = layer_shape[0]  # Number of input units
    fan_out = layer_shape[1] if len(layer_shape) == 2 else np.prod(layer_shape[1:])
# Number of output units
    limit = np.sqrt(6 / (fan_in + fan_out))  # Xavier/Glorot initialization
    return random.uniform(-limit, limit)

# Custom function to generate weights using Glorot initialization for the entire
model
def generate_glorot_weights(model):
    weight_shapes = [w.shape for w in model.get_weights()]
    weights = []
    for shape in weight_shapes:
        # Generate weights for each layer using Glorot initialization
        layer_weights = np.array([glorot_initializer(shape) for _ in
range(np.prod(shape))])
        weights.append(layer_weights)
    return np.concatenate(weights)

# Initialize the population using the Glorot initialization
model = create_model(input_shape=X_train.shape[1:])  # Create model to
calculate number of weights
n_weights = sum([np.prod(w.shape) for w in model.get_weights()])  # Calculate
total number of weights

# Register the individual creation function
def create_individual():
```

```
    return generate_glorot_weights(model)

toolbox.register("individual", tools.initIterate, creator.Individual,
create_individual)  # Create an individual using Glorot weights
toolbox.register("population", tools.initRepeat, list, toolbox.individual)  #
Generates a population

toolbox.register("mate", tools.cxTwoPoint)  # Use two-point crossover
toolbox.register("mutate", tools.mutGaussian, mu=0, sigma=1, indpb=0.1)  #
Gaussian mutation
toolbox.register("select", tools.selTournament, tournsize=3)  # Tournament
selection of 3
toolbox.register("evaluate", fitness)  # Call fitness function

# Run Genetic Algorithm
population_size = 200 #En cada iteracion se cogen 200 individuos
num_generations = 100 #Iteraciones
crossover_prob = 0.8
mutation_prob = 0.2

population = toolbox.population(n=population_size)
algorithms.eaSimple(population, toolbox, cxpb=crossover_prob,
mutpb=mutation_prob, ngen=num_generations, verbose=True)

# Retrieve best individual
best_individual = tools.selBest(population, k=1)[0]
best_model = create_model(input_shape=X_train.shape[1:]) #Se queda con el
mejor modelo

# Reshape weights to match the model
weights_shapes = [w.shape for w in best_model.get_weights()]
split_indices = np.cumsum([np.prod(shape) for shape in weights_shapes])
weight_arrays = np.split(np.array(best_individual), split_indices[:-1])
reshaped_weights = [w.reshape(shape) for w, shape in zip(weight_arrays,
weights_shapes)]
best_model.set_weights(reshaped_weights)

# Evaluate best model
loss = best_model.evaluate(X_test, Y_test)
print(f'Final loss: {loss}')

# Save best model
best_model.save('model DCgLOROT.h5')
```

- **TEST**

```
import tensorflow as tf
import pandas as pd
```

```
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler

# Load the model
model = keras.models.load_model("model DCgLOROT.h5")
input_data = pd.read_excel('Input DC.xlsx')

scaler = MinMaxScaler()
input_data_norm = scaler.fit_transform(input_data)

predictions = model.predict(input_data_norm)
```

COMPOUND BREAKWATER HEIGHT

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from deap import base, creator, tools, algorithms
import random
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split

# Load data
dfX = pd.read_excel('I_DC2050_f2.xlsx')
dfY = pd.read_excel('O_DC2050_f2w.xlsx')

# Data preprocessing, se  dividen en train y validation y los escalamos (SOLO X)
scaler = MinMaxScaler()
x_train, x_test, Y_train, Y_test = train_test_split(dfX, dfY, test_size=0.3,
random_state=1234)
X_train = scaler.fit_transform(x_train)
X_test = scaler.transform(x_test)

# Define ANN model function
def create_model(input_shape, weights=None):
    model = keras.Sequential([
        layers.InputLayer(input_shape=input_shape),
        layers.Dense(39, activation='relu'),
        layers.Dense(1, activation='linear')
    ])

    if weights is not None:
        model.set_weights(weights)  # Ensure the weights are structured correctly
```

```
    model.compile(optimizer='adam', loss='mean_squared_error')
    return model

# Define fitness function
def fitness(individual):
    num_layers = 2  # Adjust based on model structure
    model = create_model(input_shape=X_train.shape[1:]) #Crea un modello nuevo

    weights_shapes = [w.shape for w in model.get_weights()] #Extrae la forma de
los pesos del mdelo
    split_indices = np.cumsum([np.prod(shape) for shape in weights_shapes]) #Lo
divide en arrays para cada capa
    weight_arrays = np.split(np.array(individual), split_indices[:-1])
    reshaped_weights = [w.reshape(shape) for w, shape in zip(weight_arrays,
weights_shapes)] #Les da forma para ajustarlos al modelo

    model.set_weights(reshaped_weights) #Reasigna los pesos
    loss = model.evaluate(X_test, Y_test, verbose=0) #Evalua el modelo
    return (loss,)

# Genetic Algorithm setup
creator.create("FitnessMin", base.Fitness, weights=(-1.0,)) #define la funcion
objetivo a minimizar (Fitness, mse) weights=(-1.0,)=minimizar
creator.create("Individual", list, fitness=creator.FitnessMin) #Set de pesos de la
red

toolbox = base.Toolbox()

# Initialize weights based on Glorot initialization
def glorot_initializer(layer_shape):
    # Fan-in and fan-out for the layer
    fan_in = layer_shape[0]  # Number of input units
    fan_out = layer_shape[1] if len(layer_shape) == 2 else np.prod(layer_shape[1:])
# Number of output units
    limit = np.sqrt(6 / (fan_in + fan_out))  # Xavier/Glorot initialization
    return random.uniform(-limit, limit)

# Custom function to generate weights using Glorot initialization for the entire
model
def generate_glorot_weights(model):
    weight_shapes = [w.shape for w in model.get_weights()]
    weights = []
    for shape in weight_shapes:
        # Generate weights for each layer using Glorot initialization
        layer_weights = np.array([glorot_initializer(shape) for _ in
range(np.prod(shape))])
        weights.append(layer_weights)
```

```
    return np.concatenate(weights)

# Initialize the population using the Glorot initialization
model = create_model(input_shape=X_train.shape[1:])  # Create model to
calculate number of weights
n_weights = sum([np.prod(w.shape) for w in model.get_weights()])  # Calculate
total number of weights

# Register the individual creation function
def create_individual():
    return generate_glorot_weights(model)

toolbox.register("individual", tools.initIterate, creator.Individual,
create_individual)  # Create an individual using Glorot weights
toolbox.register("population", tools.initRepeat, list, toolbox.individual)  #
Generates a population

toolbox.register("mate", tools.cxTwoPoint)  # Use two-point crossover
toolbox.register("mutate", tools.mutGaussian, mu=0, sigma=1, indpb=0.1)  #
Gaussian mutation
toolbox.register("select", tools.selTournament, tournsize=3)  # Tournament
selection of 3
toolbox.register("evaluate", fitness)  # Call fitness function

# Run Genetic Algorithm
population_size = 200 #En cada iteracion se cogen 200 individuos
num_generations = 100 #Iteraciones
crossover_prob = 0.8
mutation_prob = 0.2

population = toolbox.population(n=population_size)
algorithms.eaSimple(population, toolbox, cxpb=crossover_prob,
mutpb=mutation_prob, ngen=num_generations, verbose=True)

# Retrieve best individual
best_individual = tools.selBest(population, k=1)[0]
best_model = create_model(input_shape=X_train.shape[1:]) #Se queda con el
mejor modelo

# Reshape weights to match the model
weights_shapes = [w.shape for w in best_model.get_weights()]
split_indices = np.cumsum([np.prod(shape) for shape in weights_shapes])
weight_arrays = np.split(np.array(best_individual), split_indices[:-1])
reshaped_weights = [w.reshape(shape) for w, shape in zip(weight_arrays,
weights_shapes)]
best_model.set_weights(reshaped_weights)
```

```
# Evaluate best model
loss = best_model.evaluate(X_test, Y_test)
print(f'Final loss: {loss}')

# Save best model
best_model.save('model DCwgLOROT.h5')
```

- **TEST**

```
import tensorflow as tf
import pandas as pd
import numpy as np
from tensorflow import keras
from sklearn.preprocessing import MinMaxScaler

# Load the model
model = keras.models.load_model("model DCwgLOROT.h5")
input_data = pd.read_excel('Input DC.xlsx')

scaler = MinMaxScaler()
input_data_norm = scaler.fit_transform(input_data)

predictions = model.predict(input_data_norm)
```

Output graphs (Matlab)

```
load inputf2DC.mat
load outputf2DC.mat
inputDC =[inputf2DC; inputf2DC50];
outputDC = [outputf2DC; outputf2DC50];
writematrix(inputDC, "I_DC_f2.xlsx");
writematrix(outputDC, "O_DC_f2.xlsx");
DC = readmatrix("ERRORES_DC_junto.xlsx");
DV = readmatrix("errores_DV_junto.xlsx");



csd_av_dc=74;
incrh_av_dc=0.07;
incrw_av_dc=5.85;
sum_rmse_csd_dc=52871;
sum_rmse_incrh_dc=4459;
sum_rmse_incrw_dc=52147;
csd_av_dv=225;
incrh_av_dv=0.35;
sum_rmse_csd_dv=24048;
sum_rmse_incrh_dv=4209;
```

```
rmse_DC_max=0;
rmse_DV_max=0;
rmse_DC_maxh=0;
rmse_DV_maxh=0;
rmse_DC_maxw=0;
%percentil 90
for i=1:size(DC,1)
    if DC(i,1)>160
        rmse_DC_max=rmse_DC_max+DC(i,11);
    end
    if DC(i,2)>0.27
        rmse_DC_maxh=rmse_DC_maxh+DC(i,12);
    end
    if DC(i,3)>14
        rmse_DC_maxw=rmse_DC_maxw+DC(i,13);
    end
end
for j=1:size(DV,1)
    if DV(j,1)>500
        rmse_DV_max=rmse_DV_max+DV(j,8);
    end
    if DV(j,2)>1.2
        rmse_DV_maxh=rmse_DV_maxh+DV(j,9);
    end
end
p_DC_CSD=[0 0 0 0 0];
p_DC_incrh=[0 0 0 0 0];
p_DC_incrw=[0 0 0 0 0];
p_DV_CSD=[0 0 0 0 0];
p_DV_incrh=[0 0 0 0 0];


p_DC_CSD(1,5)=rmse_DC_max/sum_rmse_csd_dc*100;
p_DC_incrh(1,5)=rmse_DC_maxh/sum_rmse_incrh_dc*100;
p_DC_incrw(1,5)=rmse_DC_maxw/sum_rmse_incrw_dc*100;
p_DV_CSD(1,5)=rmse_DV_max/sum_rmse_csd_dv*100;
p_DV_incrh(1,5)=rmse_DV_maxh/sum_rmse_incrh_dv*100;
rmse_DC_max=0;
rmse_DV_max=0;
rmse_DC_maxh=0;
rmse_DV_maxh=0;
rmse_DC_maxw=0;
%percentil 80
for i=1:size(DC,1)
    if DC(i,1)>115 %percentil80
        rmse_DC_max=rmse_DC_max+DC(i,11);
    end
    if DC(i,2)>0.15 %percentil80
        rmse_DC_maxh=rmse_DC_maxh+DC(i,12);
    end
    if DC(i,3)>9 %percentil80
```

```
        rmse_DC_maxw=rmse_DC_maxw+DC(i,13);
    end
end
for j=1:size(DV,1)
    if DV(j,1)>350 %percentil80
        rmse_DV_max=rmse_DV_max+DV(j,8);
    end
    if DV(j,2)>0.6 %percentil80
        rmse_DV_maxh=rmse_DV_maxh+DV(j,9);
    end
end


p_DC_CSD(1,4)=rmse_DC_max/sum_rmse_csd_dc*100 -p_DC_CSD(1,5);
p_DC_incrh(1,4)=rmse_DC_maxh/sum_rmse_incrh_dc*100 - p_DC_incrh(1,5);
p_DC_incrw(1,4)=rmse_DC_maxw/sum_rmse_incrw_dc*100 - p_DC_incrw(1,5);
p_DV_CSD(1,4)=rmse_DV_max/sum_rmse_csd_dv*100 - p_DV_CSD(1,5);
p_DV_incrh(1,4)=rmse_DV_maxh/sum_rmse_incrh_dv*100 - p_DV_incrh(1,5);


rmse_DC_max=0;
rmse_DV_max=0;
rmse_DC_maxh=0;
rmse_DV_maxh=0;
rmse_DC_maxw=0;
%percentil70
for k=1:size(DC,1)
    if DC(k,1)>90
        rmse_DC_max=rmse_DC_max+DC(k,11);
    end
    if DC(k,2)>0.09
        rmse_DC_maxh=rmse_DC_maxh+DC(k,12);
    end
    if DC(k,3)>6.5
        rmse_DC_maxw=rmse_DC_maxw+DC(k,13);
    end
end
for l=1:size(DV,1)
    if DV(l,1)>276
        rmse_DV_max=rmse_DV_max+DV(l,8);
    end
    if DV(l,2)>0.25
        rmse_DV_maxh=rmse_DV_maxh+DV(l,9);
    end
end


p_DC_CSD(1,3)=rmse_DC_max/sum_rmse_csd_dc*100 - p_DC_CSD(1,4)- p_DC_CSD(1,5);
p_DC_incrh(1,3)=rmse_DC_maxh/sum_rmse_incrh_dc*100 - p_DC_incrh(1,4)-
p_DC_incrh(1,5);
```

```
p_DC_incrw(1,3)=rmse_DC_maxw/sum_rmse_incrw_dc*100 - p_DC_incrw(1,4)-
p_DC_incrw(1,5);
p_DV_CSD(1,3)=rmse_DV_max/sum_rmse_csd_dv*100 - p_DV_CSD(1,4)- p_DV_CSD(1,5);
p_DV_incrh(1,3)=rmse_DV_maxh/sum_rmse_incrh_dv*100 - p_DV_incrh(1,4)-
p_DV_incrh(1,5);


rmse_DC_max=0;
rmse_DV_max=0;
rmse_DC_maxh=0;
rmse_DV_maxh=0;
rmse_DC_maxw=0;


%percentil60
for i=1:size(DC,1)
    if DC(i,1)>72
        rmse_DC_max=rmse_DC_max+DC(i,11);
    end
    if DC(i,2)>0.05
        rmse_DC_maxh=rmse_DC_maxh+DC(i,12);
    end
    if DC(i,3)>5.1
        rmse_DC_maxw=rmse_DC_maxw+DC(i,13);
    end
end
for n=1:size(DV,1)
    if DV(n,1)>212
        rmse_DV_max=rmse_DV_max+DV(n,8);
    end
    if DV(n,2)>0.12
        rmse_DV_maxh=rmse_DV_maxh+DV(n,9);
    end
end


p_DC_CSD(1,2)=rmse_DC_max/sum_rmse_csd_dc*100 - p_DC_CSD(1,4) - p_DC_CSD(1,3)
- p_DC_CSD(1,5);
p_DC_incrh(1,2)=rmse_DC_maxh/sum_rmse_incrh_dc*100 - p_DC_incrh(1,4) -
p_DC_incrh(1,3)- p_DC_incrh(1,5);
p_DC_incrw(1,2)=rmse_DC_maxw/sum_rmse_incrw_dc*100 - p_DC_incrw(1,4) -
p_DC_incrw(1,3)- p_DC_incrw(1,5);
p_DV_CSD(1,2)=rmse_DV_max/sum_rmse_csd_dv*100 - p_DV_CSD(1,4)- p_DV_CSD(1,3)-
p_DV_CSD(1,5);
p_DV_incrh(1,2)=rmse_DV_maxh/sum_rmse_incrh_dv*100 - p_DV_incrh(1,4) -
p_DV_incrh(1,3)- p_DV_incrh(1,5);


rmse_DC_max=0;
rmse_DV_max=0;
rmse_DC_maxh=0;
```

```
rmse_DV_maxh=0;
rmse_DC_maxw=0;


%percentil50
for o=1:size(DC,1)
    if DC(o,1)>57
        rmse_DC_max=rmse_DC_max+DC(o,11);
    end
    if DC(o,2)>0.028
        rmse_DC_maxh=rmse_DC_maxh+DC(o,12);
    end
    if DC(o,3)>4.05
        rmse_DC_maxw=rmse_DC_maxw+DC(o,13);
    end
end
for p=1:size(DV,1)
    if DV(p,1)>170
        rmse_DV_max=rmse_DV_max+DV(p,8);
    end
    if DV(p,2)>0.065
        rmse_DV_maxh=rmse_DV_maxh+DV(p,9);
    end
end


p_DC_CSD(1,1)=rmse_DC_max/sum_rmse_csd_dc*100 - p_DC_CSD(1,4) - p_DC_CSD(1,3)
- p_DC_CSD(1,2)- p_DC_CSD(1,5);
p_DC_incrh(1,1)=rmse_DC_maxh/sum_rmse_incrh_dc*100 - p_DC_incrh(1,4) -
p_DC_incrh(1,3) - p_DC_incrh(1,2)- p_DC_incrh(1,5);
p_DC_incrw(1,1)=rmse_DC_maxw/sum_rmse_incrw_dc*100 - p_DC_incrw(1,4) -
p_DC_incrw(1,3)- p_DC_incrw(1,2)- p_DC_incrw(1,5);
p_DV_CSD(1,1)=rmse_DV_max/sum_rmse_csd_dv*100 - p_DV_CSD(1,4)- p_DV_CSD(1,3)-
p_DV_CSD(1,2)- p_DV_CSD(1,5);
p_DV_incrh(1,1)=rmse_DV_maxh/sum_rmse_incrh_dv*100 - p_DV_incrh(1,4) -
p_DV_incrh(1,3)- p_DV_incrh(1,2)- p_DV_incrh(1,5);
```

```
hold off
subplot(1,2,1)
x=[50,60,70,80, 90];
p_DV=[p_DV_CSD;p_DV_incrh];
bar(x,p_DV);
xlabel("Percentile");
ylabel("%RMSE");
legend('CSD','Increment of height');
title('Vertical breakwater');
subplot(1,2,2)
x=[50,60,70,80,90];
p_DC=[p_DC_CSD;p_DC_incrh;p_DC_incrw];
bar(x,p_DC);
```

```
xlabel("Percentile");
ylabel("%RMSE");
legend('CSD','Increment of height', 'Increment of weight');
title('Compound breakwater');
```

```
clear
load NN_DV.mat
I_DV=readmatrix("Input DV.xlsx");
O_DV=zeros(size(I_DV,1),2);
for i=1:size(I_DV,1)
    O_DV(i,:)=net(I_DV(i,:)')
end
x=I_DV(:,1);
y=I_DV(:,2);
z=O_DV(:,2);
dt = delaunayTriangulation(x,y) ;
tri = dt.ConnectivityList ;
xi = dt.Points(:,1) ;
yi = dt.Points(:,2) ;
F = scatteredInterpolant(x,y,z);
zi = F(xi,yi) ;
trisurf(tri,xi,yi,zi)
view(2)
colorbar
xlabel('Hs(m)');
ylabel('Tp(s)');
title('Increment of height (m)')
shading interp
```

```
clear
load NN_DC.mat
I_DC=readmatrix("Input DC.xlsx");
O_DC=zeros(size(I_DC,1),3);
for i=1:size(I_DC,1)
    O_DC(i,:)=net(I_DC(i,:)')
end
x=I_DC(:,1);
y=I_DC(:,2);
z=O_DC(:,3);
dt = delaunayTriangulation(x,y) ;
tri = dt.ConnectivityList ;
xi = dt.Points(:,1) ;
yi = dt.Points(:,2) ;
F = scatteredInterpolant(x,y,z);
zi = F(xi,yi) ;
trisurf(tri,xi,yi,zi)
view(2)
limits = [61,63];
c = colorbar;
set(gca,'clim',limits([1,end]))
xlabel('Hs(m)');
ylabel('Tp(s)');
```

```
title('Increment of weight (t)')
shading interp
```

Port operability assessment model codes

import numpy as np

import cv2

import tensorflow as tf

from tensorflow import keras

from tensorflow.keras import layers, models

import pandas as pd

from skimage.metrics import structural_similarity as ssim

import keras.backend as K

from skimage.transform import resize

import matplotlib.pyplot as plt

Assuming X_wave_period, X_wave_direction, and X_port_image are your input variables,

and Y_images is your output images

Load data from Excel file

excel_file = r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\wave_data.xlsx'

wave_data = pd.read_excel(excel_file)

Assuming your Excel file has columns named 'wave_period' and 'wave_direction'

X_wave_period = wave_data['wave_period'].values

X_wave_direction = wave_data['wave_direction'].values

```
# Load and preprocess X_port_images

X_port_images = []

# image_paths is the list of directories of the input images

image_paths = [r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia4colores_DIR_22.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia4colores_DIR_22.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia4colores_DIR_22.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia4colores_DIR_22.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia4colores_DIR_22.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia4colores_DIR_22.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia_DIR_45.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia_DIR_45.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia_DIR_45.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia_DIR_45.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia_DIR_45.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia_DIR_45.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_067.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_067.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_067.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_067.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_067.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_067.5.bmp',
```

r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_90.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_90.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_90.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_90.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_90.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_90.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_112.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_112.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_112.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_112.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_112.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5colores_DIR_112.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5coloresretocado2020_dir_135.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5coloresretocado2020_dir_135.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5coloresretocado2020_dir_135.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5coloresretocado2020_dir_135.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Inputs\PuertoValencia5coloresretocado2020_dir_135.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes

```
agitacion\Inputs\PuertoValencia5colores_DIR_157.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_157.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_157.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_157.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_157.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_157.5.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_180.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_180.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_180.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_180.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_180.bmp',
r'C:\Users\Puertos\.spyder-py3\01_Modelo                    CNNs_redes
agitacion\Inputs\PuertoValencia5colores_DIR_180.bmp']
for image_path in image_paths:
    image = cv2.imread(image_path)
    if image is None: # Check if there are any failures in reading the images
        print(f"Failed to load image from: {image_path}")
    else:
        X_port_images.append(image)
X_port_images = np.array(X_port_images)


# Load Y_images
Y_images = []
```

label_paths = [r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\001.1_22.5_8.9s.bmp' , r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\001.2_22.5_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\001.3_22.5_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\001.4_22.5_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\001.5_22.5_5.0s.bmp',r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\001.6_22.5_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\002.1_45_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\002.2_45_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\002.3_45_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\002.4_45_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\002.5_45_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\002.6_45_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\003.1_67.5_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\003.2_067.5_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\003.3_67.5_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\003.4_67.5_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\003.5_67.5_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\003.6_67.5_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\004.1_90_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\004.2_90_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\004.3_90_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\004.4_90_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\004.5_90_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\004.6_90_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes

agitacion\Outputs\005.1_112.5_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\005.2_112.5_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\005.3_112.5_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\005.4_112.5_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\005.5_112.5_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\005.6_112.5_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\006.1_135_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\006.2_135_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\006.3_135_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\006.4_135_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\006.5_135_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\006.6_135_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\007.1_157.5_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\007.2_157.5_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\007.3_157.5_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\007.4_157.5_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\007.5_157.5_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\007.6_157.5_4.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\008.1_180_8.9s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\008.2_180_7.4s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\008.3_180_6.1s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\008.4_180_5.6s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\008.5_180_5.0s.bmp', r'C:\Users\Puertos\.spyder-py3\01_Modelo CNNs_redes agitacion\Outputs\008.6_180_4.6s.bmp']

for label_path in label_paths:

```
        label_image = cv2.imread(label_path)
        if label_image is None: # Check if there are any failures in reading the images
            print(f"Failed to load image from: {image_path}")
        else:
            Y_images.append(label_image)
Y_images = np.array(Y_images)

# Assuming image_paths and label_paths are lists containing paths to image files
# Ensure X_wave_period, X_wave_direction, X_port_images, and Y_images have compatible shapes and formats
max_wave_period_value = 8.9
max_wave_direction_value = 180

# Normalize input variables if necessary
X_wave_period = X_wave_period / max_wave_period_value  # Normalize wave period
X_wave_direction = X_wave_direction / max_wave_direction_value  # Normalize wave direction
X_port_images = X_port_images / 255.0  # Normalize port image (assuming pixel values are in range [0, 255])
Y_images = Y_images/255.0

#Añadimos todo en una matriz
X_wave = np.column_stack((X_wave_period, X_wave_direction))


# Define input shapes
wave_shape = X_wave.shape
port_image_input_shape = X_port_images.shape
```

```
def create_model(wave_shape, port_image_input_shape, output_shape):

    # Wave input branch
    wave_input = layers.Input(shape=wave_shape, name='wave_input')
    wave_branch = layers.Dense(4, activation='relu')(wave_input)
    wave_branch = layers.Dense(8, activation='relu')(wave_branch)

    # Port image input branch
    port_image_input = layers.Input(shape=port_image_input_shape, name='port_image_input')
    port_image_branch = layers.Conv2D(8, (3, 3), padding='same', activation='relu')(port_image_input)
    port_image_branch = layers.MaxPooling2D((2, 2))(port_image_branch)
    port_image_branch = layers.Conv2D(16, (3, 3), activation='relu')(port_image_branch)
    port_image_branch = layers.MaxPooling2D((2, 2))(port_image_branch)
    port_image_branch = layers.Flatten()(port_image_branch)
    port_image_branch = layers.Dense(32, activation='relu')(port_image_branch)

    # Concatenate the branches
    concatenated = layers.concatenate([wave_branch, port_image_branch])

    # Calculate the total number of units
    output_units = np.prod(output_shape[1:])

    # Define the output layer
    output_layer = layers.Dense(output_units, activation='sigmoid')(concatenated)
```

```
    output_layer = layers.Reshape(output_shape[1:])(output_layer)

    # Define the model
    model = models.Model(inputs=[wave_input, port_image_input], outputs=output_layer)

    return model

# Define the number of output neurons
num_output_neurons = 1728 * 1152 * 3  # Number of pixels * number of channels

# Create the model
model = create_model(wave_shape[1:], port_image_input_shape[1:], Y_images.shape)

# Compile the model
model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=0.0005), loss='mse')


# Train the model

history = model.fit([X_wave, X_port_images], Y_images, epochs=100, batch_size=12, validation_split=0.3)

# Get the minimum loss
#MSE

min_mse = min(history.history['loss'])
```

```
print("MSE Loss min:", min_mse)


# Reverse normalization on Y_images
original_max_value = 255.0  # Assuming original pixel values were in range [0, 255]
normalized_max_value = 1.0  # Assuming normalization scaled pixel values to range [0, 1]


# Reverse normalization
Y_images = Y_images * original_max_value / normalized_max_value



## Validation set and reverse normalization
validation_images = model.predict([X_wave, X_port_images])
validation_images = validation_images * original_max_value / normalized_max_value


# Define the number of images you want to display
num_images_to_display = 1  # Change this number as needed


# Convert images to the correct data type (assuming Y_images is a list of images)
Y_images = [np.uint8(image) for image in Y_images]
validation_images = [np.uint8(image) for image in Y_images]


# Display the validation images
for i in range(num_images_to_display):
    plt.figure(figsize=(50, 50))


    # Display original image
```

```
    plt.subplot(1, 2, 1)
    plt.imshow(cv2.cvtColor(Y_images[i], cv2.COLOR_BGR2RGB), cmap='viridis')
    plt.title("Original Image", fontsize=70)
    plt.axis('off')  # Turn off axis labels

    # Display predicted image
    plt.subplot(1, 2, 2)
    plt.imshow(cv2.cvtColor(validation_images[i], cv2.COLOR_BGR2RGB), cmap='viridis')
    plt.title("Predicted Image", fontsize=70)
    plt.axis('off')

    plt.tight_layout()
    plt.show()


# Save the model
model.save('agitation_model.h5')
```

Appendix B

This appendix includes the three articles that are currently under review. It has been decided not to include the content of these articles in the thesis to avoid potential issues with their publication.

1. Advancements in Artificial Intelligence -Based Models in Maritime and Ocean Engineering
2. An Artificial Intelligence approach for coastal adaptation studies: Insights from a case study in the Mediterranean Sea
3. Implementing Deep Learning in port agitation studies under the context of climate change
4. Improving Multi-Variable Wave Forecasting with AI: Integrating LSTM and Random Forest Using a Window-and-Flatten Technique

Appendix C

The content of this appendix corresponds to research paper 4 in Appendix B.

This appendix explains the process of developing a LSTM model to predict future wave states. In addition, a Random Forest model (RF) is also developed and compared to the LSTM.

The first step in developing any AI model is to collect and preprocess the relevant data. Historical data from the Valencia buoy sourced from Puertos del Estado database were used in this study. The dataset consists of hourly records spanning from 2005 to 2024, totalling 167,970 data samples. Of these, 70% were used for model training and 30% for validation. To assess the model's performance on unseen data, the final 120 hours of records were reserved for testing. The dataset includes measurements of wave characteristics, temperature, pressure, sea level, and wind parameters.

All the data was preprocessed. Wave height measurements below 0.05 m were considered false or unreliable and replaced with NaN values, while wave periods shorter than 0.1 s were also deemed invalid and substituted with NaN. Missing data due to time gaps in the time series were handled with interpolation. For gaps smaller than or equal to 4 hours, linear interpolation was applied to ensure the continuity of the dataset. However, for gaps exceeding 4 hours, interpolation alone was considered unreliable. In these cases, data from the two closest buoys to Valencia — Tarragona and Cabo de Palos — were used. The missing values for Valencia were extrapolated as the average of the corresponding measurements from these nearby buoys, incorporating a spatially informed estimation process (Figure A-1).

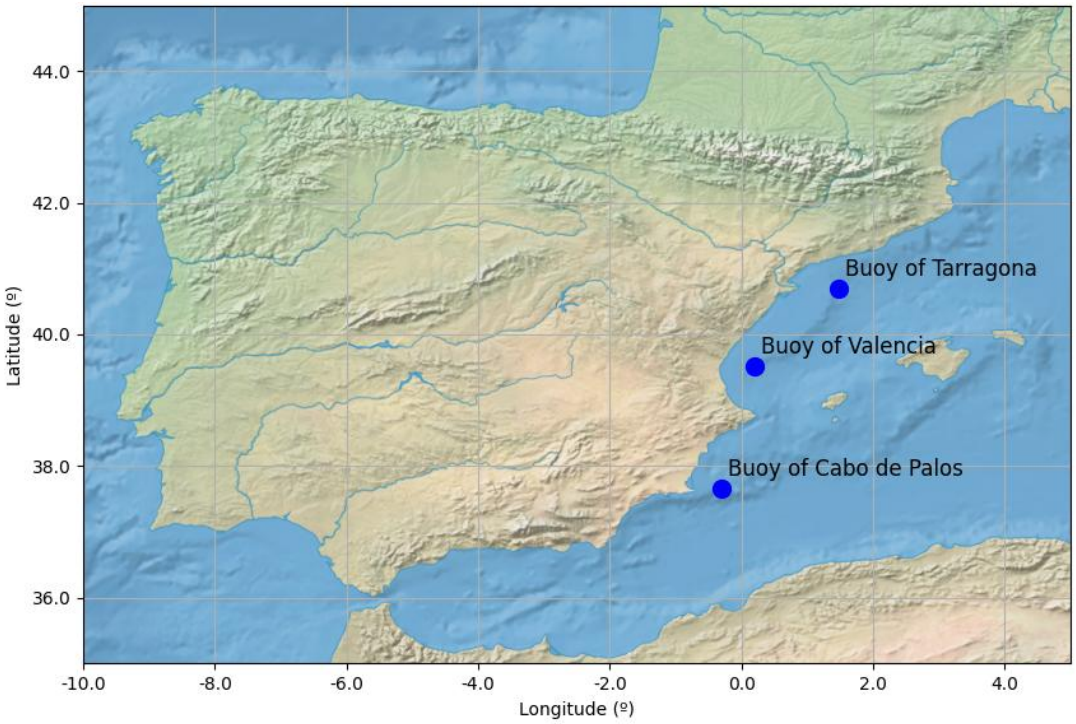


Figure A-1. Study area.

After collecting and preprocessing all the data, feature engineering was performed. A sensitivity analysis was conducted to identify the most relevant variables for each model. This analysis followed a trial-and-error methodology, where various input variables were tested to assess their influence on the model's performance. Specifically, a stepwise backward selection approach was employed, which involved initially testing all models with all variables included. Subsequently, variables that did not contribute meaningfully to improving the model or that introduced noise were systematically eliminated. To enhance the robustness of the model, cross-validation techniques were also applied to minimize the impact of the initial weights and ensure a more generalized and reliable outcome.

The goal of this sensitivity analysis was to evaluate the significance of each variable in predicting the desired outcomes and to select those that would provide the most accurate and insightful results. Out of the 19 raw variables initially downloaded from the dataset, 11 variables were deemed most appropriate for inclusion in the model. These variables were the date (t), significant wave height (Hs), peak wave period (Tp), wave direction (Dir), wind speed (Ws), wind direction (Wd), air temperature (aT), air pressure (aP), sea level (SL), sea surface temperature (SST) and meteorological tide (MT).

To better capture the interactions between the different physical phenomena involved in wave generation, several derived variables were defined. These derived

variables were designed to model complex relationships and interactions that raw variables alone could not fully represent. To define these derived variables, the same process used for selecting the raw variables was followed: a trial-and-error approach with a stepwise backward selection methodology and cross-validation techniques. The variables designed to model the interactions between physical processes are presented in Table A-1 (Equations 54 to 58).

Table A-1. Definition of derived variables modelling interaction between physical processes.

Variable	Definition	
Wave Power (P)	$\frac{\rho g H_s^2 T_p}{32\pi}$	[54]
SL change rate (SLC)	$SL(t+1) - SL(t)$	[55]
Wind –Wave (WW)	$H_s * W_s$	[56]
Air – Sea Temperature (AST)	$SST - aT$	[57]
Wind – Temperature (WT)	$\sqrt{(SST - aT)^2 + {W_s}^2}$	[58]

In addition to these derived variables, a window-and-flatten approach was applied with the objective of optimizing how temporal patterns in the data are captured and utilized in prediction models. To implement this, two additional variables based on the window-and-flatten methodology were defined, both of which are presented in Table A-2, bringing the total number of input variables to 18.

Table A-2. Definition of rolling variables.

Variable	Definition
Rolling SL (RSL)	Moving average of SL with a widow size = 3
Rolling Deviation SST (RSST)	Moving deviation of SST with a widow size = 3

Finally, the window-and-flatten methodology was implemented, in which all previous data records within a defined window size were consolidated into a single feature vector for future wave prediction (Figure A-2). This approach ensures that the model does not rely solely on the most recent time step but instead considers the entire sequence of observations within the window, preserving valuable

temporal dependencies. The resulting vector serves as the model's input, enabling it to predict the next wave height more effectively. The window size was optimized following a trial-and-error approach to find the most effective size for capturing temporal dependencies and ensuring the model's accuracy.

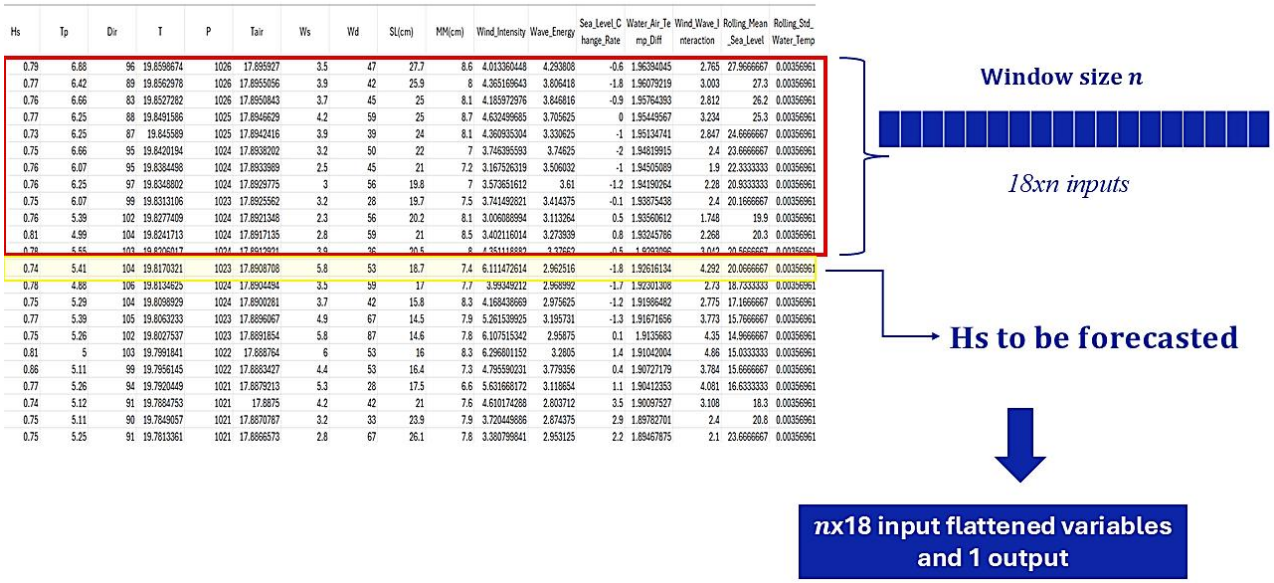


Figure A-2. Window and flatten process.

Like other types of ANNs, the architecture and hyperparameters of the LSTM model require careful optimization to achieve the best performance. To determine the optimal configuration, the same trial-and-error methodology and cross-validation techniques used for input data selection were employed. The characteristics of the optimized LSTM model are detailed in Table A-3.

Table A-3. LSTM model characteristics.

Hyperparameter	**Value**
N° epochs	20
Batch size	32
Architecture	50 memory cells
Activation function	Tanh
Optimizer	Adam
Data Partition	80 % training 20% validation
Scaler	MaxMin Scaler

The results of the validation of the LSTM are presented in Figure A-3.

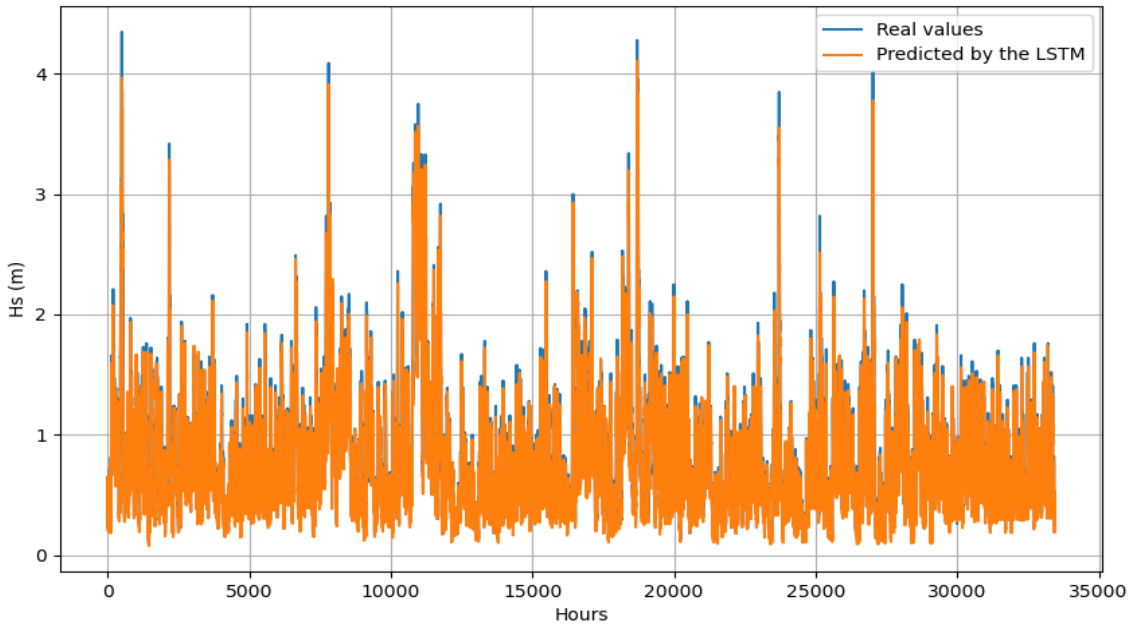


Figure A-3. Validation of the LSTM model.

The LSTM model was designed to predict Hs, Tp and wave direction, developing three different models for each variable. It was tested on three datasets (K1, K2 and K3). These test cases were designed to evaluate the model under distinct wave conditions: K1 represents a scenario with low initial wave energy, K2 begins at a high-energy point exceeding 80% of observed wave heights, and K3 features an abrupt transition in wave conditions.

The test phase was designed to simulate a real-time forecasting scenario using a recursive, multi-step approach in which each prediction depends on previous model outputs rather than actual values. This method, while more challenging, reflects realistic operational conditions and leads to the accumulation of errors over time. Consequently, the test results represent a conservative estimate, providing the lower bound of model performance. Results are presented in Table A-4 to Table A-6 and Figure A-4 to Figure A-6.

Table A-4. Performance of the LSTM model for Hs in the different datasets.

Metrics	**K1 (55 hours)**	**K2 (88 hours)**	**K3 (43 hours)**
MSE	0.0136 m^2	0.0370 m^2	0.0138 m^2
RMSE	0.1170 m	0.1930 m	0.1170 m
MAE	0.0849 m	0.1260 m	0.0975 m
MAPE	10.17%	13.53%	12.40%
R^2	0.77	0.72	0.91

The same procedure as with the wave height was followed for Tp. Figure A-5 shows the performance of the LSTM model. Given the poor results under our conservative testing framework, we also evaluated the model's performance when the real sea state conditions were updated every 6 hours, to assess whether periodic correction could improve accuracy. These 6 hours update highly improved the performance of the model, as can be seen in Figure A-5 and Table A-5.

Table A-5. Performance of the LSTM model for Tp in the different datasets.

Performance Metrics	**K1 (48 hours)**		**K2 (65 hours)**		**K3 (32 hours)**	
	Normal	6h update	Normal	6h update	Normal	6h update
MSE	0.298 s^2	0.220 s^2	1.510 s^2	0.407 s^2	0.375 s^2	0.170 s^2
RMSE	0.546 s	0.468 s	1.230 s	0.638 s	0.613 s	0.412 s
MAE	0.423 s	0.354 s	0.915 s	0.438 s	0.445 s	0.277 s
MAPE	9.48%	8.10%	16.26%	8.57%	8.94%	5.89%
R^2	0.54	0.66	-0.97	0.454	0.68	0.86

Table A-6. Performance of the LSTM model for wave direction in the different datasets.

Performance Metrics	**K1 (112 hours)**	**K2 (5 hours)**	**K3 (30 hours)**
MSE	130.18 ^{o2}	243.15 ^{o2}	982.61 ^{o2}
RMSE	11.41^{o}	15.60 ^{o}	31.34 ^{o}
MAE	7.05 ^{o}	13.50 ^{o}	25.68 ^{o}
R^2	0.98	0.14	0.92

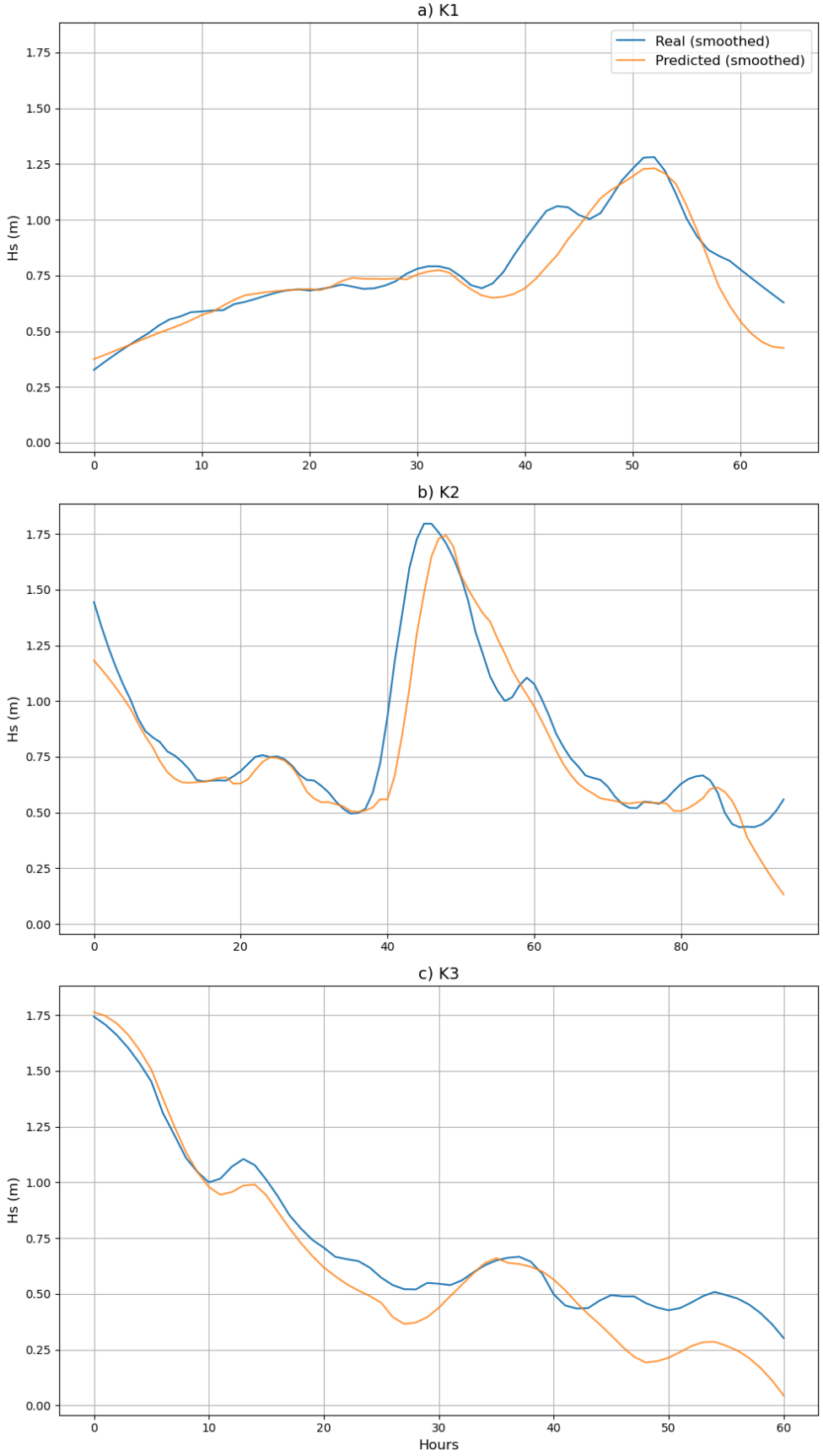


Figure A-4. Hs prediction for each of the subsets by the LSTM model.

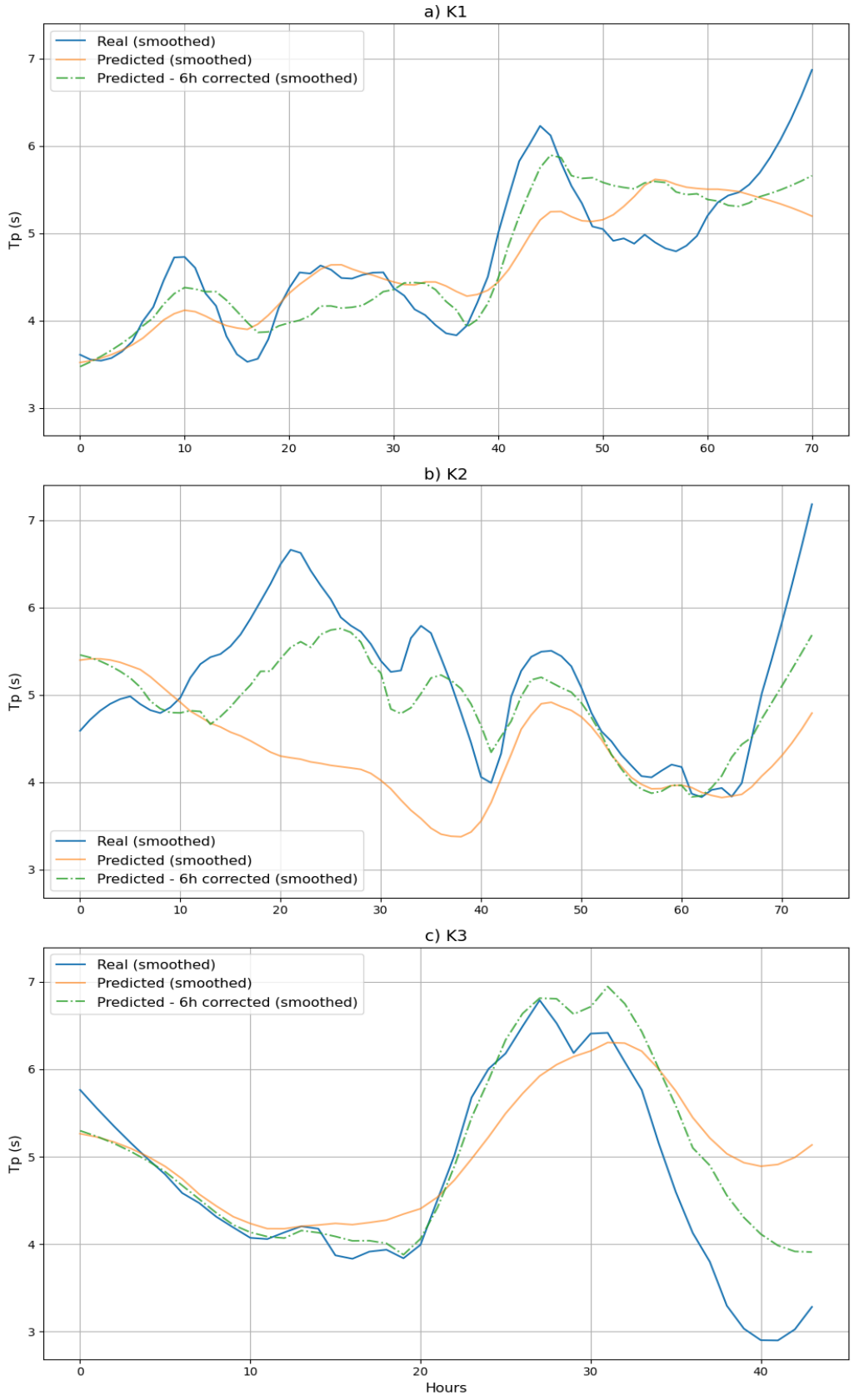


Figure A-5. Tp prediction for each of the subsets by the LSTM model.

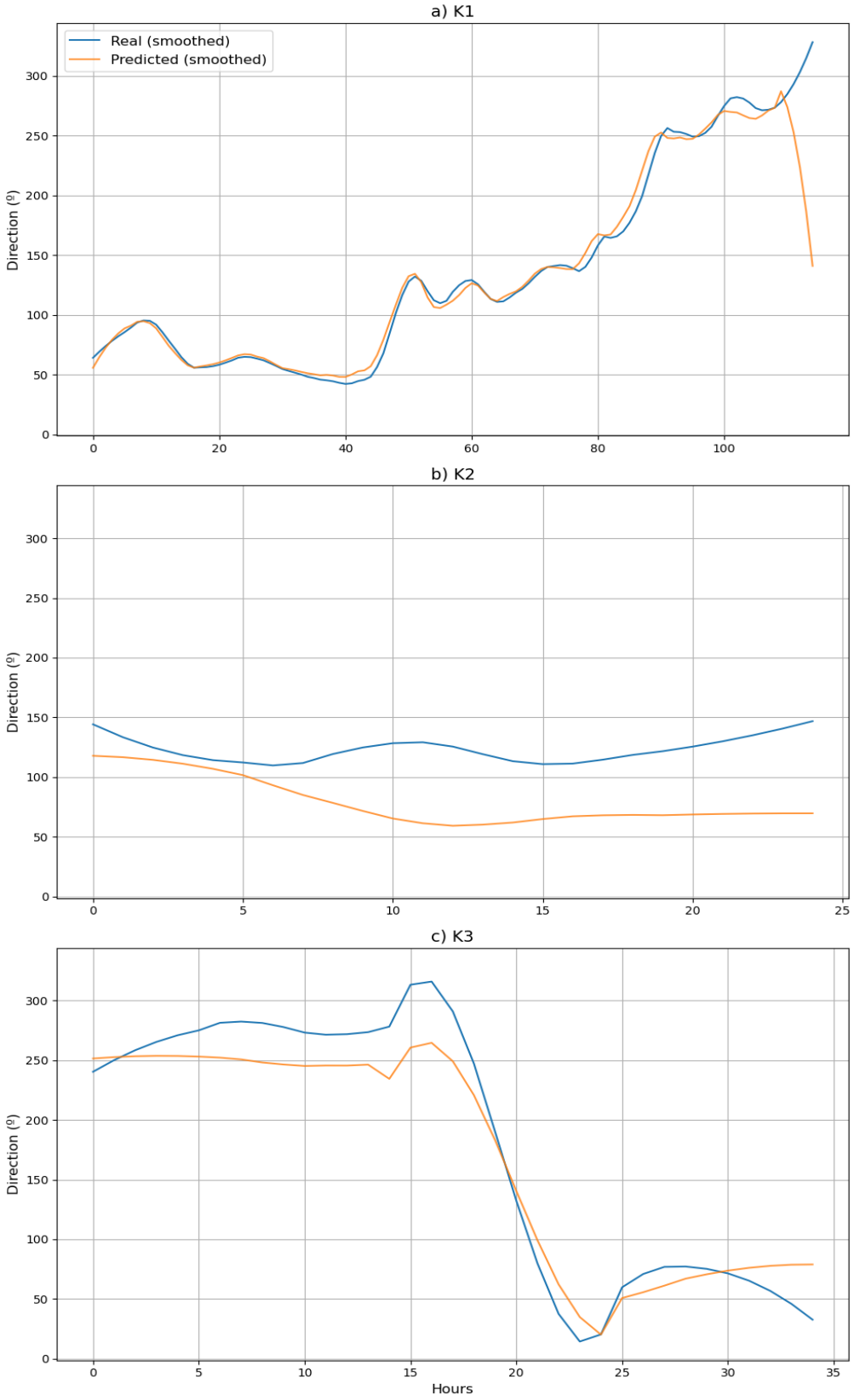


Figure A-6. Wave direction prediction for each of the subsets by the LSTM model.

Apart from LSTMs, Random Forest models (RFs) were also tested to predict future wave states. RFs are a machine learning technique introduced by Breiman (2001) that build upon the concept of decision trees (DTs). RFs are a type of ensemble method, which enhances the predictive performance of models by combining multiple simpler learners—DTs—into a more robust model. The final prediction in a RF is obtained by aggregating the outputs of individual DTs. For classification tasks, RFs use majority voting, while for regression tasks, they compute the average of the predictions. Each DT in the RF is constructed using a different bootstrap sample, a technique where subsets of the dataset are randomly drawn with replacement. This method, known as bootstrapping, ensures that each tree is trained on a slightly different dataset, which introduces variability and reduces the likelihood of overfitting to the training data technique (Harris & Grzes, 2019). To further promote independence and decorrelation among the DTs, RFs apply feature sampling, which randomly selects a subset of features at each split in a DT (Campos et al., 2021). This random feature selection prevents certain dominant features from overly influencing the model. By reducing correlations between trees, this approach enhances the overall accuracy and robustness of the ensemble model. RFs are an example of the bagging (Bootstrap Aggregating) method, which combines predictions from multiple base models to reduce variance and improve generalization. The combination of bootstrapping for data sampling and feature selection at split points ensures that RF models are less prone to overfitting than individual DTs, while still capturing complex relationships in the data.

The same process used in optimizing the LSTM models was employed to determine the optimal hyperparameters for the RF model. The specific characteristics of the optimized RF model are summarized in Table A-7.

Table A-7. RF model characteristics.

Hyperparameter	Value
Nº estimators	100
Maximum depth	10
Bootstrap	True
Criterion	MSE
Data Partition	80 % training 20% validation
Scaler	MaxMin Scaler

The results of the validation of the RF are presented in Figure A-7.

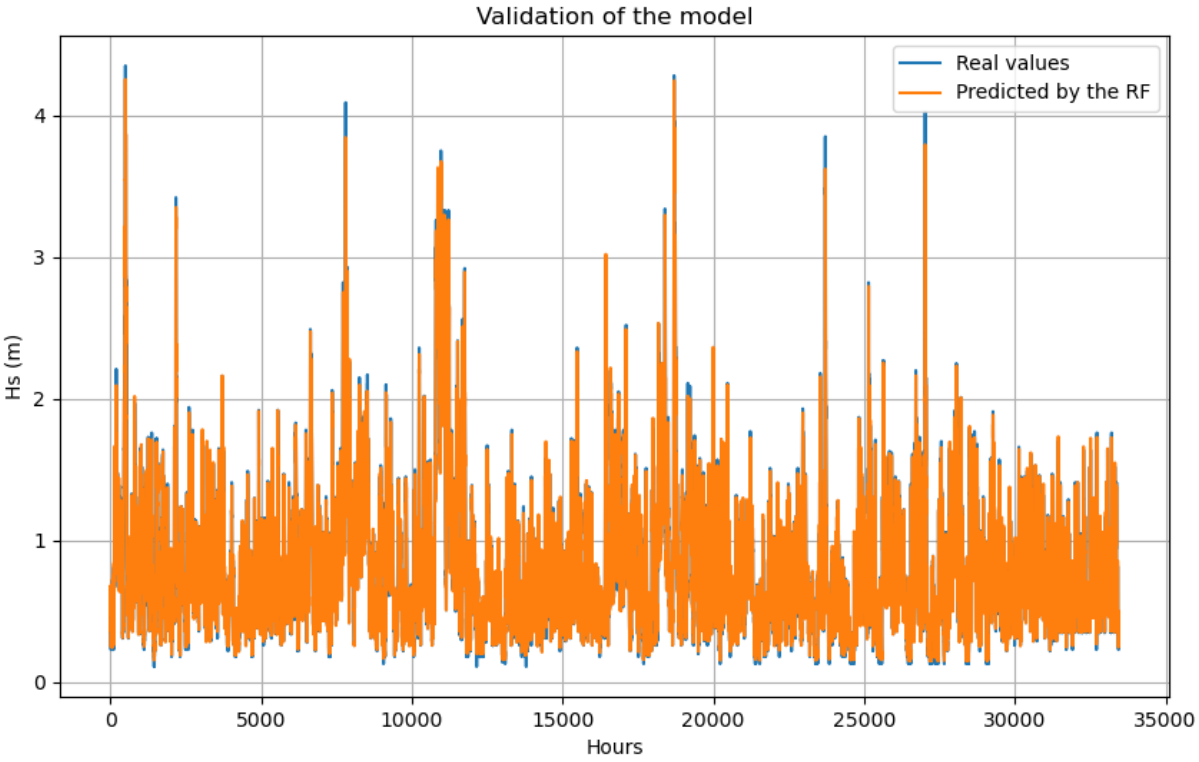


Figure A-7. Validation of the RF model.

Table A-8 to A-10 and Figure A-8 to A-10 show the performance of the RF model for Hs, Tp and Dir prediction.

Table A-8. Performance of the RF model for Hs in the different datasets.

Metrics	**K1 (31 hours)**	**K2 (53 hours)**	**K3 (30 hours)**
MSE	0.0017 m^2	0.0575 m^2	0.1170 m^2
RMSE	0.0409 m	0.2399 m	0.3420 m
MAE	0.0327 m	0.1874 m	0.3169 m
MAPE	5.42 %	21.40 %	41.50 %
R^2	0.89	0.67	0.25

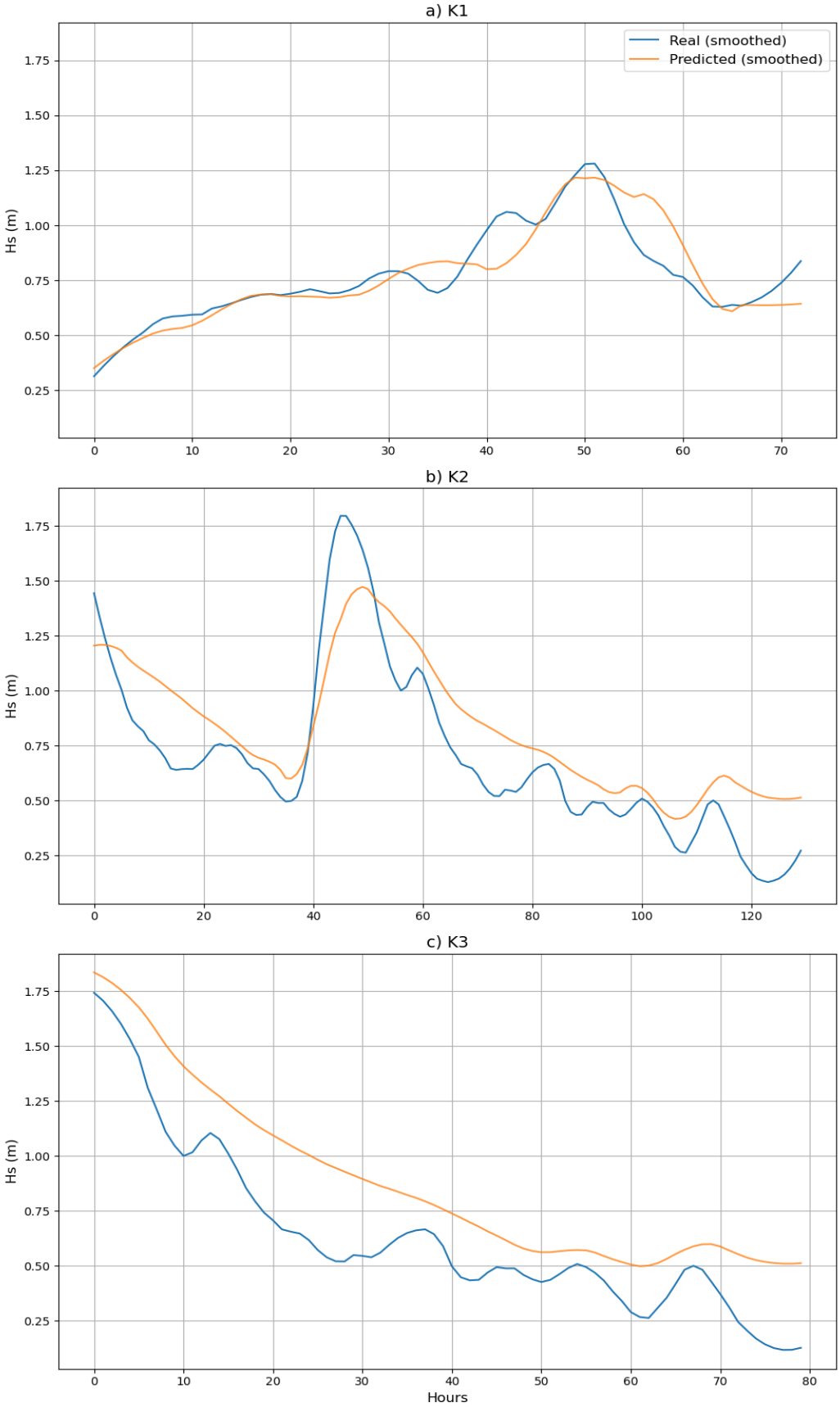


Figure A-8. Hs prediction for each of the subsets by the RF model.

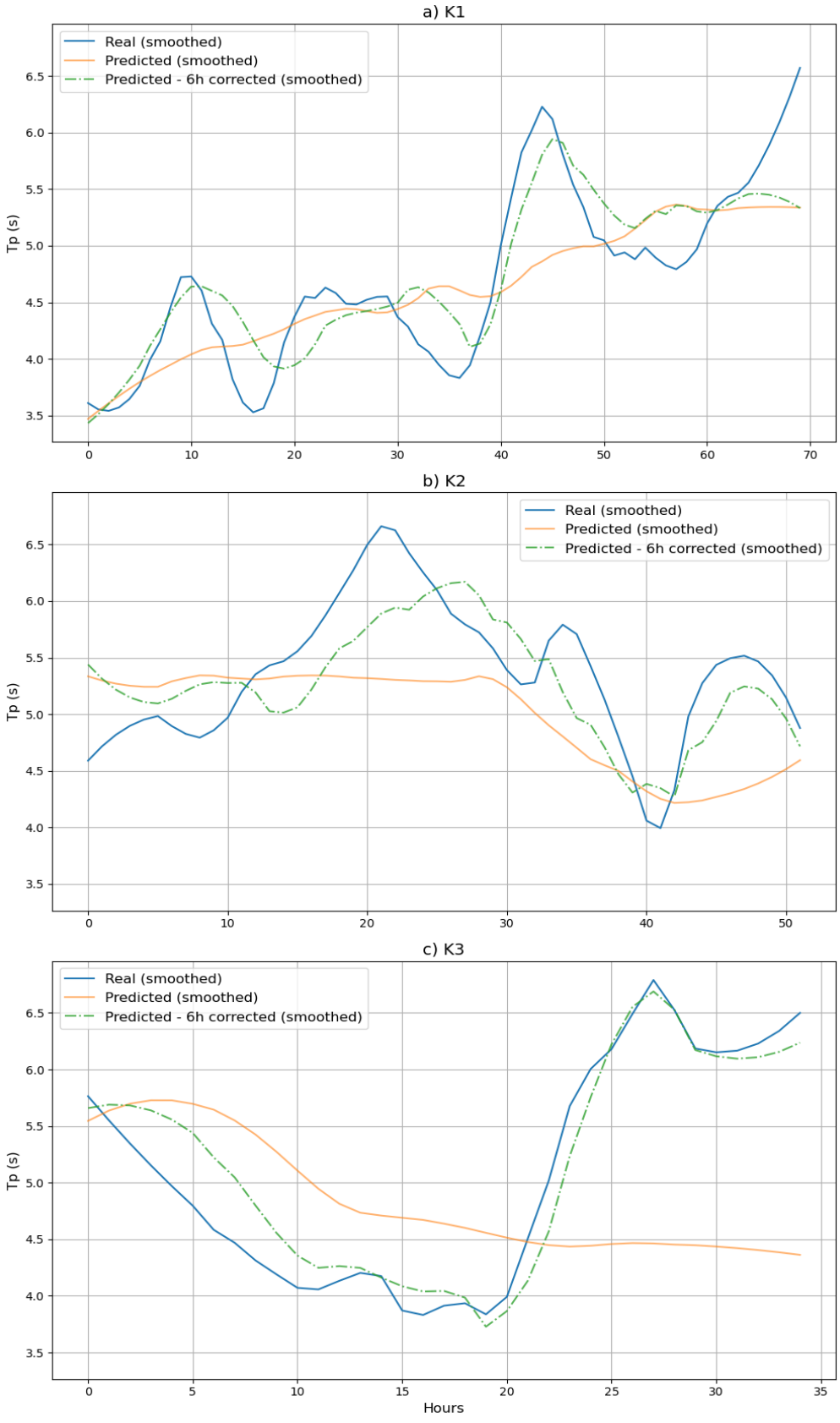


Figure A-9. Tp prediction for each of the subsets by the RF model.

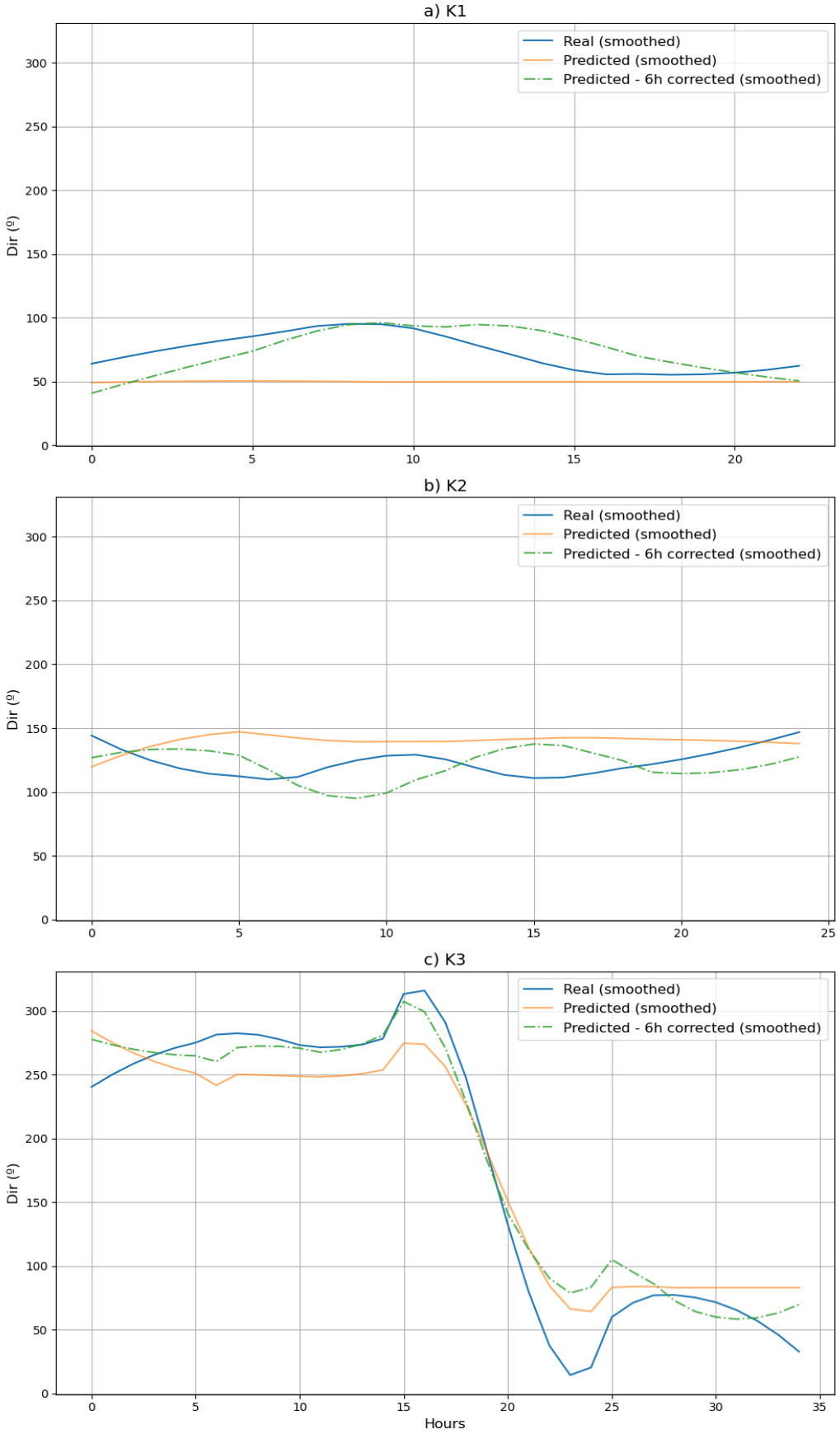


Figure A-10. Wave direction prediction for each of the subsets by the RF model

Similar to the LSTM results, the initial performance of the RF model was notably poor. To address this, we applied a correction by updating the real values every 6 hours. This adjustment led to a significant improvement in the model's accuracy, demonstrating the benefits of periodic correction in mitigating error accumulation.

Table A-9. Performance of the RF model for Tp in the different datasets.

Performance Metrics	**K1 (30 hours)**		**K2 (31 hours)**		**K3 (21 hours)**	
	Normal	6h update	Normal	6h update	Normal	6h update
MSE	0.2729 s^2	0.260 s^2	0.5442 s^2	0.2949 s^2	0.7289 s^2	0.2478 s^2
RMSE	0.5224 s	0.510 s	0.7377 s	0.5430 s	0.8537 s	0.4978 s
MAE	0.4107 s	0.386 s	0.5942 s	0.4245 s	0.7443 s	0.3624 s
MAPE	10.35 %	9.86 %	10.56 %	7.76 %	18.24 %	8.37 %
R^2	0.22	0.24	-0.21	0.34	-0.57	0.56

For wave direction prediction, the RF model performed significantly worse than the LSTM. Consequently, we applied the same strategy used for Tp by introducing a 6-hour correction interval. This allowed us to assess whether periodic updates could enhance the model's performance. Despite this adjustment, the RF model continued to perform poorly. Even with periodic corrections, it quickly converged to a nearly constant value and failed to capture directional changes. This behaviour suggests that the RF model lacks the sensitivity required to effectively model the variability and dynamics of wave direction.

Table A-10. Performance of the RF model for wave direction in the different datasets.

Performance Metrics	**K1 (24 hours)**		**K2 (24 hours)**		**K2 (30 hours)**	
	Normal	6h update	Normal	6h update	Normal	6h update
MSE	735.13 ^{o2}	290.05 ^{o2}	554.33 ^{o2}	560.88 ^{o2}	1454.70 ^{o2}	1349.39 ^{o2}
RMSE	27.11^{o}	17.03 ^{o}	23.54 ^{o}	23.68 ^{o}	38.14 ^{o}	36.73 ^{o}
MAE	22.96 ^{o}	11.8 ^{o}	19.33 ^{o}	20.15 ^{o}	30.85 ^{o}	22.24 ^{o}
R^2	-2.49	-0.38	-2.23	-2.27	0.88	0.89

The new approach proposed in this study, which emphasizes the modelling of wave variables using a window-and-flatten technique, demonstrated promising performance, even when applied to relatively simple models. Notably, the RF models, when combined with this technique, achieved predictive accuracy for significant Hs comparable to that of LSTM networks, despite the latter being specifically designed for sequence-based tasks. While LSTMs continued to outperform RF in capturing overall temporal trends and longer-term dependencies, RF models excelled in predicting individual values with high precision. However, a critical limitation of the RF approach was its tendency to get stuck at a constant value after a certain prediction horizon, which severely hindered its effectiveness for long-term forecasting.

This window-and-flatten method shows considerable potential beyond the scope of this study. By transforming sequential data into vectorized input windows, the approach allows non-sequential algorithms like RF to extract meaningful temporal features. This not only expands the range of machine learning models applicable to time series forecasting, but also opens the door to developing hybrid modelling strategies. In such frameworks, traditional time series models, like LSTM or other recurrent architectures, can be combined with non-sequential algorithms to leverage the strengths of both paradigms, potentially enhancing predictive performance, robustness, and interpretability in complex forecasting tasks.

Given these findings, future research should aim to conduct comparative evaluations of different modelling strategies, including LSTM, RF, and hybrid approaches, across various marine environments. Model performance is highly dependent on regional characteristics. A model that performs well in the Mediterranean Sea, a low-energy and locally forced basin, may not be equally effective in higher-energy systems like the Atlantic Ocean or Cantabrian Sea, where wave dynamics are more complex and heavily influenced by distant swell.

Through the evaluation of different wave parameters, it becomes evident that modelling peak period remains a significant challenge, even for AI-based approaches known for handling nonlinear and noisy data. This reinforces the need for continued innovation in model design and feature engineering, particularly when dealing with highly variable or poorly represented parameters.

Another critical factor that emerged is the sensitivity of model performance to the testing dataset. As demonstrated, results varied significantly depending on whether the model was tested during low-energy conditions (K1), high-energy

events (K2), or abrupt transitions (K3). This underscores the importance of evaluating models across diverse sea states to avoid misleading conclusions based on narrow or unrepresentative datasets.

In terms of limitations, the stagnation behaviour observed in RF models warrants further investigation. Hybrid models that integrate the temporal learning strengths of LSTM with the feature-level precision of RF could provide a path forward. Additionally, extending these methodologies to support long-term forecasting would be critical for their application in climate change studies. Developing robust, reliable long-range models could significantly enhance our ability to perform risk assessments, plan coastal infrastructure, and support decision-making processes in the context of a changing climate.