Methodology of air traffic flow clustering and 3-D prediction of air traffic density in ATC sectors based on machine learning models

Francisco Pérez Moreno a,*, Víctor Fernando Gómez Comendador a, Raquel Delgado-Aguilera Jurado a, María Zamarreño Suárez a, Dominik Janisch b, Rosa María Arnaldo Valdés a

a Department of Aerospace Systems, Air Transport and Airports, Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain
b ATM Research and Development Reference Centre (CRIDA), 28022 Madrid, Spain

1. Introduction

The demands of air traffic operations are increasing significantly over the last decade (de Arruda Jr, Weigang, & Milea, 2015), and are expected to still increase in the coming years. This increase in demand leads to a disruption of the balance between the demand for air traffic operations and the capacity of the ATM system. The development of new technologies and procedures is a key issue to help the ATC service increase their capacity and match the ever-increasing demand (Gómez, Goron, Groza, & Letia, 2016). In order to develop new technologies, numerous international programmes and projects are being developed, such as the Single European Sky ATM Research (SESAR) in Europe, the Next Generation Air Transportation System (NextGen) in the United States, and the Aviation System Block Upgrades (ASBU) framework from International Civil Aviation Organization (ICAO) (Zhang, Liu, Hu, & Zhu, 2018).

At present, there is no universal indicator of complexity (Gianazza, 2007, Airspace configuration using air traffic complexity metrics, 2007). There are currently many industrial and research projects whose main objective is to achieve a global definition of complexity. This topic is of major interest to the Air Traffic Management industry. But airspace complexity is a very challenging concept which will depend on many factors such as the distribution of aircraft in the airspace, occupancy, air traffic flows distribution or weather conditions (Gómez Comendador, Arnaldo Valdés, Vidosavljevic, Sánchez Cidoncha, & Zheng, 2019). Within these projects’ framework, the definition and estimation of airspace future complexity is a widely studied topic. Airspace complexity is a concept related to ATCO workload. Thus, research will seek to reduce airspace complexity as it is expected that this will reduce the workload of the air traffic controllers (Xie, Zhang, Ge, Dong, & Chen, 2021). Besides helping to define the airspace complexity, research projects are trying to develop new technologies to assess and estimate this airspace complexity, such as the use of trajectory-based operations.

Keywords:
Air Traffic Flows
Machine Learning
Trajectory clustering
Flight Level
ATM System Capacity

ABSTRACT

The increase in the demand for aircraft operations has caused the ATM system to become overloaded as it no longer has sufficient capacity to respond to this increase in demand. For this reason, many projects have emerged with the aim of increasing the capacity of the ATM system through the development of new technologies. This paper proposes a solution that would allow predicting and evaluating the traffic density in a three-dimensional basis in one or several ATC sectors. The final goal of this methodology is to analyse the complexity of these ATC sectors.

This paper proposes, first, the two-dimensional structuring of traffic density in a set of air traffic flows identified from historical operational data in the sector of analysis. In the vertical dimension, the traffic will be still structured in flight levels. As a subsequent step, a prediction of this structured traffic density is attempted by means of machine learning models.

This proposed methodology will try to facilitate the work of the ATC service by allowing them to have a picture of the air traffic organisation of the ATC sector before the real operation occurs. The application of this methodology will allow the adjustment of the ATC service resources. In addition, it will allow the complexity of the sectors to be assessed, as this complexity will strongly depend on how the traffic is structured.
that aim to estimate airspace complexity (Radisic, Novak, & Juricic, 2017).

Prior knowledge of complexity is of great interest to the ATC service. The big problem at present is the presence of areas of high complexity, and the workload of ATCOs who operate there will be very high. While there are areas that are very low in complexity and where ATCOs will have very little workload. A robust definition of complexity, and the ATC service’s knowledge of the complexity in the airspace, will enable them to divide this airspace into different sectors to distribute this complexity evenly. This will have a direct effect on the ATCOs, as the workload of the most saturated ATCOs will be reduced.

The role played by controllers in ATM is fundamental, and their impact on aviation safety is very high. ATCOs experience high mental limitations and workload due to the complexity of the airspace. This complexity is the result of various factors such as traffic density, weather conditions, and aircraft operations. The ATCOs must be able to effectively manage this complexity to ensure the safety and efficiency of air traffic operations.

The most prominent tool included in research projects which aim to estimate complexity is machine learning. Many of these projects have developed models in which the complexity of air traffic is assessed based on the characteristics of the air traffic itself (Xiao, Zhang, Cai, & Cao, 2016). In assessing traffic in terms of its complexity, it is not necessary to have such an exhaustive analysis.

Prior knowledge of complexity is of great interest to the ATC service. The big problem at present is the presence of areas of high complexity, and the workload of ATCOs who operate there will be very high. While there are areas that are very low in complexity and where ATCOs will have very little workload. A robust definition of complexity, and the ATC service’s knowledge of the complexity in the airspace, will enable them to divide this airspace into different sectors to distribute this complexity evenly. This will have a direct effect on the ATCOs, as the workload of the most saturated ATCOs will be reduced.

The role played by controllers in ATM is fundamental, and their impact on aviation safety is very high. ATCOs experience high mental limitations and workload due to the complexity of the airspace. This complexity is the result of various factors such as traffic density, weather conditions, and aircraft operations. The ATCOs must be able to effectively manage this complexity to ensure the safety and efficiency of air traffic operations.

The most prominent tool included in research projects which aim to estimate complexity is machine learning. Many of these projects have developed models in which the complexity of air traffic is assessed based on the characteristics of the air traffic itself (Xiao, Zhang, Cai, & Cao, 2016). In assessing traffic in terms of its complexity, it is not necessary to have such an exhaustive analysis.

Prior knowledge of complexity is of great interest to the ATC service. The big problem at present is the presence of areas of high complexity, and the workload of ATCOs who operate there will be very high. While there are areas that are very low in complexity and where ATCOs will have very little workload. A robust definition of complexity, and the ATC service’s knowledge of the complexity in the airspace, will enable them to divide this airspace into different sectors to distribute this complexity evenly. This will have a direct effect on the ATCOs, as the workload of the most saturated ATCOs will be reduced.
depends upon the availability of an ADS receiver nearby. In this paper OpenSky’s data are not used because in the areas near the Pyrenees and the Atlantic Ocean, where the sectors are located, there are some lack of ADS receiver’s coverage and information for certain aircraft may be lost. This lack of coverage can be checked on OpenSky’s own website (OpenSky, 2022) and has been also stated by previous research works (Martínez-Prieto, et al., 2017).

To overcome these limitations, this paper will take a different approach from the general 4-D trajectory prediction literature. The aim of the paper is to forecast aircraft organisation based on their timetable. To this end, the methodology followed will be to organise the airspace into air traffic flows, according to the main traffic patterns of the sector. These air traffic flows will be identified according to a customised methodology based on the characteristics of historical traffic in the sector. Once these flows have been defined, the traffic in the sector will be classified into the air traffic flow that best suits the trajectory of the aircraft, using machine learning algorithms. This approach will attempt to obtain a 2-D image of the traffic in the sector.

In order to know the traffic density completely, it will also be necessary to know how aircraft are distributed vertically. This methodology proposes to make a prediction of the FLs of the aircraft flying in the sector, also using machine learning models. Adding this to the previous analysis, the traffic in the sector will be known in 3-D.

The advantages of this comprehensive methodology will be:

- The information needed to define the flows and to classify aircraft within the flows will be much reduced compared to the information needed to estimate the full trajectory. This results in an easier collection of information and a lower computational cost of the methodology.
- Estimating air traffic in the sector can be achieved by using a simpler model. By not incurring a lot of detail, it will not require an accurate and sophisticated model that is more complex to understand and develop.
- The results will be easily interpretable. The purpose of this air traffic analysis is to help to understand the complexity of airspace. The simple interpretation of the results fits very well with this objective.
- An advantage of the methodology is that it can be applied to different time horizons. In this way, traffic density can be predicted in the short term, but can also be extended to the long term. In this paper, the analysis is focused on a daily time horizon.

Beyond the actual methodology of classifying air traffic into air traffic flows and flight levels, an estimation of this will be made with the help of machine learning. AI-based 4-D trajectory prediction has numerous benefits for the ATM system, such as improved safety and a decrease in mid-air collisions (Escamilla Núnez, Mora Camino, & Bouadi, 2017). But by predicting traffic density, these advantages are indirectly maintained. Traffic density is related to the workload of ATCOs directly and indirectly, being the indirect relationship through air traffic conflicts (Corver, Unger, & Grote, 2016). Hence, if predictions of air traffic density can reduce the workload of the ATCOs, similar benefits to those obtained with 4-D trajectory prediction will be achieved. This makes estimation through machine also learning a fundamental part of the work carried out in this paper.

Along these lines, this section will be divided into subsections in which the main points of the analysis in this paper will be explained. First, the process of creating flows from the main traffic patterns of the ATC sector will be discussed. Then, the process of distribution of the vertical density of aircraft by flight levels will be described. Finally, the machine learning models developed to obtain the desired traffic density prediction will be described.

### 2.1. Air traffic flows identification and clustering methodology

The first step to know the air traffic density is to understand how it will be organised in the horizontal plane. In this paper, it has been decided to design a three-step process for the creation of air traffic flows:

- Definition of flows based on geographical characteristics of historical traffic within the sector.
- Clustering of flows based on proximity of trajectories.
- Elimination of uncrowded routes.

The first step in this process will be the definition of operational traffic flows. Following the methodology for the creation of air traffic flows proposed in (Verdonk Gallego, et al., 2019), the creation of flows through the longitude and latitude of aircraft entering and leaving the sector is proposed. In addition, in order to differentiate climbing/descending traffic flows that may overlap their 2-D trajectory, the flows shall be differentiated by the attitude of aircraft entering and exiting the ATC sector. The attitude of an aircraft is the orientation of an aircraft with respect to the horizon (Zolotukhin & Nesterov, 2015). In this paper, the attitude describes the type of flight the aircraft is operating at a given time. The attitude will determine whether a flight is climbing, descending, or cruising. It will also help to know if the aircraft is changing it flight level (if in climb attitude it increases the flight level, if in descent attitude it decreases the flight level) or if the flight level remains stable cruising:

\[
\text{flw} : \mathbb{R}^4 \rightarrow \text{flw}(lng_{in}, lat_{in}, lng_{out}, lat_{out}, Att_{in}, Att_{out})
\]  

(1)

The definition of the variable \(\text{flw}\) is done by defining a straight line from the entry point of the ATC sector of each operation to its exit point. Therefore, the slope will be:

\[
m_{\text{flw}} = \frac{lat_{out} - lat_{in}}{lng_{out} - lng_{in}}
\]  

(2)

According to this definition, an air traffic flow will be created for each aircraft. In real operation, traffic flows are defined by a number of
aircraft with similar air routes. To account for this fact, and following the proposal of (Gui, Zhou, Wang, Liu, & Sun, 2020), a tolerance has been defined for the entry and exit points of the individual flights. This tolerance shall be established as a radius around the reference point of the flow. Therefore, flights which have entry points within this circle of influence should be considered as part of the same flow. An example of how this tolerance is defined is shown in Fig. 1.

Flights of which entry and exit points are in proximity, and which have the same evolution on entry to and exit from the ATC sector, shall be considered to be within the same elementary air traffic flow. The tolerance must be sufficient to be able to accommodate the distance between flights with close entry and exit points, but not large enough to cause aircraft with different characteristics to be within the same flow.

The flows will thus be redefined as follows:

$$\text{flw} : \mathbb{R}^4 \rightarrow \text{flw}(\text{lng}_{\text{in}} \pm \tau_{\text{flw}}, \text{lat}_{\text{in}} \pm \tau_{\text{flw}}, \text{lng}_{\text{out}} \pm \tau_{\text{flw}}, \text{lat}_{\text{out}} \pm \tau_{\text{flw}}, \text{Att}_{\text{in}}, \text{Att}_{\text{out}})$$  \hspace{1cm} (3)

This being an iterative process. Flights are ordered in chronological order, meaning that they will be ordered according to the time and day they enter the sector. The first flight shall be the one that first enters the sector. The first flight will be the reference of the first elementary air traffic flow, flw. When a flight does not comply with the assigned tolerance, or has different attitudes from the previous flights, flw will be defined and this will be its reference point. The rest of the flights will be compared with both flows. This will form an iterative process in which more elementary flows will be continuously creating.

Fig. 1 shows how flows are formed from the entry points of

---

**Fig. 2.** Identification of clustered air traffic flows.

**Fig. 3.** Comparative between Elementary air traffic flows (a) and clustered air traffic flows (b) in LECMPAU.
individual operations and defining a tolerance around them.

But the real aim of this clustering algorithm is to study the main traffic patterns in the ATC sector. Therefore, the solution is to cluster the elementary flows. Just as an air traffic flow is defined by aircraft with similar characteristics, in this paper a clustered flow will be defined from elementary flows with similar characteristics. To keep the model simple, the clustered flows will consist of elementary flows with similar entry and exit points, and with similar entry and exit attitudes.

\[ FLW : \mathbb{R}^{N+2} \rightarrow FLW(\text{flw}_i, \text{flw}_j, \ldots, \text{flw}_N, \text{Att}_{in}, \text{Att}_{out}) \] (4)

Fig. 2 shows the process of identifying clustered air traffic flows based on previously defined elementary air traffic flows.

The process of creating clustered air traffic flows will be an iterative process based on a tolerance. The process shall be analogous to the process of creating elementary flows, but with a higher tolerance. This can be expressed as:

\[ FLW : \mathbb{R}^{N} \rightarrow FLW(\text{flw}_i \pm t_{FLW}, \text{flw}_j \pm t_{FLW}, \ldots, \text{flw}_N \pm t_{FLW}, \text{Att}_{in}, \text{Att}_{out}) \] (5)

Fig. 2 shows the process of identifying clustered air traffic flows based on previously defined elementary air traffic flows.

When identifying clustered air traffic flows, the variability of the flows is much lower than the variability in the identification of elementary air traffic flows. This means that the methodology actually captures the behaviour of aircraft operating similarly within the ATC sector.

In Fig. 3 a comparison is made between the elementary flows of LECMPAU in Fig. 3a and the clustered flows of the same sector in Fig. 3b. The tolerances \( t_{flw} = 5\text{NM} \) and \( t_{FLW} = 15\text{NM} \) have been used in the identification process as an example.

In en-route ATC sectors, traffic is usually structured, so this methodology should correctly capture the traffic flows in the sector. But departure/arrival ATC will be more complex, as there will be highly variable traffic departing from or arriving at one or more airports. This variety in routes will have to be considered (Murça, Hansman, Li, & Ren, 2018). In such cases, two new parameters will be included in the formation of air traffic flows: the origin of the flight for departing flights, and the destination of the flight for arriving flights. If the origins or destinations are the airports of influence within the ATC sector studied, the flows will be clustered taking into account this variable. By aggregating this, it is possible to further reduce the clustered flows in departure/arrival sectors, and to have a more realistic picture of how the traffic is structured. In these cases, clustered flows are defined:

\[ FLW_{or} : \mathbb{R}^{N+3} \rightarrow FLW_{or}(\text{flw}_i \pm t_{FLW}, \text{flw}_j \pm t_{FLW}, \ldots, \text{flw}_N \pm t_{FLW}, \text{Att}_{in}, \text{Att}_{out}, \text{or}) \] (6)

\[ FLW_{dest} : \mathbb{R}^{N+3} \rightarrow FLW_{dest}(\text{flw}_i \pm t_{FLW}, \text{flw}_j \pm t_{FLW}, \ldots, \text{flw}_N \pm t_{FLW}, \text{Att}_{in}, \text{Att}_{out}, \text{dest}) \] (7)

Fig. 4a represents the clustered flows of the GCCCRNE sector without taking into account the influential airports within the sector. Fig. 4b shows the flows of the same sector taking into account origins and destinations and keeping in the identification process the tolerances \( t_{flw} = 5\text{NM} \) \( t_{FLW} = 15\text{NM} \).

Fig. 5. Analysis of aircraft vertical density.
Eventually, after the formation of the clustered air traffic flows, both en-route and departure/arrival ATC sectors, will contain air traffic flows with very few operations. This is due to the operation of certain aircraft outside the norm. Regulations in the analysed sector or in adjacent sectors may mean that, in extraordinary cases, aircraft may have to change their trajectory. These instances are outside the standard operation of the sector, so they should be identified and cleared, as long as this volume of traffic does not affect the operation within the ATC sector, as the objective of this algorithm is to identify the main traffic flows. In order to eliminate these flows, it has been decided to take a criterion of the number of aircraft passing through the flow in the period of one year. Specifically, it has been decided to eliminate flows

Fig. 6. Traffic density estimation scheme using machine learning.

Fig. 7. Number of flows with respect to the tolerance $t_{flw}$.
containing less than 100 aircraft in a year. This limit represents an average of approximately 2 aircraft per week.

This decision has been taken since, if the flow were to be formed by aircraft with regular operation, it would not affect the operation of the sector at all to remove a flow with two flights per week. And if this flow is formed through concentrated in time operations, they will not have a great influence either, and furthermore they will not be representative of common traffic flows, but will be a temporary flow caused by a special event.

This is expected to greatly reduce the number of flows identified, without incurring the elimination of many operations within the sector. Thus, it will still be possible to study the traffic distribution of the sector in an accurate and optimised way by focusing the study only on the routes that are actually used.

With this algorithm, the traffic density in the different flows of the sectors will be studied. This proposed algorithm has certain advantages that make its application interesting.

- Conceptual and application simplicity.
- Algorithm based on real operation. This means that it shows the real traffic patterns in the ATC sector, and that the results can vary with the historical sample presented.
- Adaptability. The algorithm will adapt to the nature of the sector. If it is an en-route ATC sector, the origin/destination dependency shall

Fig. 8. Number of flows with respect to the tolerance $t_{FLW}$ within LECMPAU.

Fig. 9. Number of flows with respect to the tolerance $t_{FLW}$ within GCCCRNE.
not be applied. If it is a departure/arrival ATC sector, the origin/destination dependency shall be applied. In addition, the tolerances in the definition of $f_{l_w_i}$ and $L_{W_i}$ are variable to allow the best fit to different use cases.

- Low computational cost. By reducing the number of flows, and disregarding exceptional traffic flows, the computational cost will be reduced without affecting the accuracy of the results, or the analysis of the results.

Once this methodology has been applied, it will be possible to study the traffic in the horizontal plane. This will provide a picture of how aircraft will be distributed in the airspace. But the information will not be complete if one wants to know how the traffic is totally organised and

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Quality parameters for LECMPAU.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max distance between ENAIRE/algorithm</td>
<td>0.54NM</td>
</tr>
<tr>
<td>Clustered Traffic/ Total Traffic</td>
<td>0.977</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Quality parameters for GCCCRNE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max distance between ENAIRE/algorithm</td>
<td>0.38NM</td>
</tr>
<tr>
<td>Clustered Traffic/ Total Traffic</td>
<td>0.963</td>
</tr>
</tbody>
</table>

Fig. 10. Clustered air traffic flows in LECMPAU compared with air traffic routes defined in INSIGNIA (ENAIRE, 2021).

Fig. 11. Clustered air traffic flows in GCCCRNE compared with air traffic routes defined in INSIGNIA.
what its density is. For this, it will be necessary to know how the aircraft are distributed vertically.

2.2. Vertical distribution organisation per flight levels

The analysis and prediction of vertical trajectories is a problem that has also been studied by many researchers (Verdgon Gallego, Gómez Comendador, Sáez Nieto, Orenge Imaz, & Arnaldo Valdés, 2018). The vertical distribution of aircraft is very important in determining airspace complexity. The vertical separation between aircraft plays a key role in potential collisions and their safe resolution (Radanovic, Piera Eroles, Boca, & Ramos Gonzalez, 2018).

The vertical distribution of aircraft is more straightforward to perform. In the historical data of the operations, the attitudes of the aircraft and the flight levels that the different aircraft will occupy at the entry and exit of each ATC sector are available. In particular, $FL_{in}$ is used to refer to the flight level at the entry to the ATC sector, and $FL_{out}$ to refer to the flight level at the exit of the ATC sector. Moreover, it is known that aircraft will try not to change their flight level too much in order to optimise their vertical trajectory. Related to this, concepts such as CCO or CDA, which allow fuel or emissions savings (Pérez Castan, et al., 2018), are of great interest to the industry.

Despite not having an exhaustive knowledge of the vertical profile of aircraft over the entire trajectory, the literature specifies that in order to have knowledge of the complexity of the airspace, it is sufficient to be aware of the aircraft that are on a climbing and descending trajectory (Gianazza & Guittet, 2006, Selection and evaluation of air traffic complexity metrics, 2006). This can be assessed with $Att_{in}, Att_{out}, FL_{in}$ and $FL_{out}$. Fig. 5 shows a comparison of the different types of flights according to the combination of these variables.

Hence, it shall be this information the used for the complexity assessment. In addition, this information can also be used to obtain additional parameters if needed. These parameters can be, for example, the flight levels occupied in the air traffic flows, or the aircraft per flight level at the entry or exit of the flow on a given day or at a given time, since the times of entry and exit of the aircraft sector will also be available.

Therefore, with the four key parameters $Att_{in}, Att_{out}, FL_{in}$ and $FL_{out}$, the vertical density of aircraft can be examined in a simple way that fits in with the overall objective of the paper.

2.3. Machine learning prediction models

With the aforementioned organisation, a structure that allows the assessment of the traffic density in the horizontal plane and also in the vertical dimension is obtained. The real problem is the large increase in demand, which means that in the future the ATM system will not be able to handle it. For this reason, the main objective of this paper is to estimate the aircraft behaviour in the future. This topic is being studied by many researchers (Li, et al., 2021; Zhang, Liu, Wang, Song, & Liu, 2020). They all agree that new technologies developed to estimate the behaviour of aircraft in airspace will be done through machine learning models.

In this paper, a machine learning model will be proposed to estimate the traffic density in both horizontal plane and vertical dimension. This machine learning model will use as input general flight characteristics presented in timetables. The objective of this model is: “to be able to estimate the horizontal and vertical behaviour of an aircraft within an ATC sector from the predicted timetable”. If the ATC service has the flight schedule available the day before its operation, it will be capable of estimating the traffic density in different sectors by means of this forecast. This will have the great benefit of allowing the ATC service to foresee the possible saturation areas, and which ATC sectors will have the most complex operation, and to address these in advance of this operation. Knowing where the complexity will be higher, and being able to reduce this complexity, will lead to a reduction in the workload of the ATCOs (Cao, et al., 2018).

To predict traffic density, the machine learning model proposed in this paper is schematically shown in Fig. 6.

Fig. 6 indicates that this machine learning model will be divided into three successive steps. Short-term air traffic flow prediction is essential in the field of air traffic flow management. Many researchers have focused on traffic flow prediction to aid in management and planning (Zhang, et al., 2021). Therefore, firstly, based on the flight information and using the previously developed classification methodology, the air traffic flow to which each flight will belong within the analysed ATC sector will be estimated.

This air traffic flow shall be used as additional input to the second model, which will estimate $FL_{in}$ from the general flight information and the air traffic flow to which it belongs ($Att_{in}$ and $Att_{out}$ is intrinsic to the prediction of air traffic flows). Lastly, the estimation of $FL_{out}$ will be carried out. For this, the general flight information, the air traffic flow and the entry flight level shall be used as input variables.

This successive machine learning approach will make it possible to estimate air traffic density over any time horizon, although this paper is limited to its application in the short term. Specifically, estimations will be made from data from the day before the operation. The organisation of the machine learning model in three steps has certain advantages that will help to make the estimations more reliable.

- The independence of the algorithms means that each algorithm focuses on finding parameters for the estimation of the three target variables separately. Related to this independence between algorithms, the hyperparameters of each algorithm can be customised. Hyperparameters will have a great influence on the results provided by a machine learning algorithm (Probst, Boulesteix, & Bischl, 2019), so this variety can be very beneficial for the overall model.
- Although there is independence between the machine learning algorithms, the input variables will be related. For algorithms 2 and 3, the target variables of the previous models will be added to the general flight information. This makes the input information more and more detailed, if the algorithms provide good results. Thanks to it, the algorithms will be able to learn clearer patterns and give better results. An example of a very clear correlation is $FL_{in}$ and $FL_{out}$. The exit flight level will largely depend on the flight level at which the aircraft enters the sector. Therefore, adding $FL_{in}$ to the $FL_{out}$ estimation is expected to improve the results of this algorithm.

The organisation of Fig. 6 makes the explainability of the complete model clearer. By being able to separate the models, but linking and relating them to each other, a general model is obtained that is very easy to explain.

- If a more complex and accurate model is desired, this can be done easily. The machine learning model is divided into modules. This organisation makes it easy to add new modules or to modify existing ones. This makes the machine learning model very flexible.

Once the methodology for describing the traffic density of the sector has been established, and the prediction process using machine learning has been described, a test will be carried out in section 3 as an example. The flows will be created, and then the performance of the three different machine learning models will be evaluated.

3. Results

The methodology for identifying air traffic flows in one ATC sector and estimating its traffic density has been presented. The next step is the validation and testing of this methodology. For this purpose, an example of application will be carried out. This example is based on the operation in the LECMPAU and GCCCRNE ATC sectors. These sectors have been chosen since LECMPAU represents a typical route sector, with very structured operation and little variation throughout the year. And GCCCRNE represents a departure/arrival sector with operation highly dependent on departures and arrivals at the airports within the sector, which operation is very variable due to being in the Canary Islands.

In order to produce these results, data from the flight plans and the actual operation in these sectors during 2019 has been used. The data has been obtained by the Spanish service provider ENAIRE, and processed and validated by the company C.R.I.D.A. As all the operational data for 2019 is available, it has been decided that both the flow clustering algorithm and the machine learning models will use all the data. Therefore, the intervals of the different simulations will be 2019.

This section will be divided by obtaining results from:

1. Identification of clustered flows in both ATC sectors: This is the first step of the proposed methodology. The objective is to identify the main air traffic flows in the sectors and to classify the operation within these flows in order to subsequently estimate the traffic density.

2. Estimation of air traffic flow through which aircraft will operate: This estimation is the first step in the estimation of traffic density in the sectors by machine learning models.

3. Estimation of entry and exit flight levels of aircraft in the sector: With this estimation, the traffic density estimation is completed.

<table>
<thead>
<tr>
<th>Table 6</th>
<th>Structure of $FL_{in}$ prediction model.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input variable</td>
<td>Output variable</td>
</tr>
<tr>
<td>Origin-Destination</td>
<td>$FL_{in}$</td>
</tr>
<tr>
<td>Airline</td>
<td></td>
</tr>
<tr>
<td>Aircraft type</td>
<td></td>
</tr>
<tr>
<td>Hour of the flight (based on IOBT)</td>
<td></td>
</tr>
<tr>
<td>Day of the week of the flight (based on IOBT)</td>
<td></td>
</tr>
<tr>
<td>Month of the flight (based on IOBT)</td>
<td></td>
</tr>
<tr>
<td>Clustered air traffic flow</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 12. Relative importance of clustered air traffic flows prediction model.
3.1. Identification of clustered air traffic flows

In obtaining results, the ultimate goal is to make a prediction of how aircraft will be organised in the sector. The first step is to create air traffic flows in which to subsequently classify aircraft. This process will be implemented by means of the traffic flow identification and clustering algorithm.

The process will depend on the aircraft entry and exit data of the sector, the two variable tolerances for the formation of elementary and clustered flows, and the nearby airports that may influence the operation in the case of the approach sectors.

∑

The longitudes and latitudes of entry and exit of the ATC sector will be part of the database used for the development of this paper. The origin and destination airports used will also be well-known data, as they are intrinsic parameters of the geography of the ATC sectors analysed. Therefore, the only variability in the model will be the two tolerances.

The identification of elementary air traffic flows will depend on \( t_{\text{flw}} \).

To see which tolerance best adjusts to the analysed sectors. A sensitivity analysis is performed, shown in Fig. 8 for LECMPAU and in Fig. 9 for GCCCRNE. For the choice of the tolerance, a second sensitivity analysis is performed, shown in Fig. 8 for LECMPAU and in Fig. 9 for GCCCRNE. The number of flows identified is after the removal of flows with less than 100 flights in the year. In this approach, it is planned to study the performance of the algorithm by means of the final results of the algorithm.

In the LECMPAU sector, it will not be necessary to consider any nearby airports, as it is an en-route sector. In the case of GCCCRNE, it will be necessary to take into account the airports GCLP, GCRR and GCFV which are located within this sector, and which will influence the operations of the sector.

In LECMPAU, the number of clustered flows decreases as \( t_{\text{flw}} \) increases. But when reaching a tolerance of 20NM, the number of clustered flows becomes very steady. This makes the decision range between these 20NM and 28NM. Within these 20NM and 28NM, the selection criterion is the elimination of as few aircraft as possible belonging to flows with less than 100 flights. Therefore, the final selection will be \( t_{\text{flw}} = 25 \text{NM} \), where 0.45% of the total number of aircraft is eliminated for being in non-operational flows.

Table 7: Structure of \( FL_{\text{in}} \) prediction model.

<table>
<thead>
<tr>
<th>Input variable</th>
<th>Output variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin-Destination</td>
<td>( FL_{\text{in}} )</td>
</tr>
<tr>
<td>Airline</td>
<td></td>
</tr>
<tr>
<td>Hour of the flight (based on I0BT)</td>
<td></td>
</tr>
<tr>
<td>Day of the week of the flight (based on I0BT)</td>
<td></td>
</tr>
<tr>
<td>Month of the flight (based on I0BT)</td>
<td></td>
</tr>
<tr>
<td>Clustered air traffic flow</td>
<td>( FL_{\text{in}} )</td>
</tr>
</tbody>
</table>

Table 8: Summary of the testing indicators for the \( FL_{\text{in}} \) prediction model.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Scenario</th>
<th>Precision</th>
<th>Recall</th>
<th>( F1 )-score</th>
<th>( FL_{\text{in}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECMPAU</td>
<td>Mean</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.10</td>
<td>0.12</td>
<td>0.11</td>
<td>–</td>
</tr>
<tr>
<td>GCCCRNE</td>
<td>Mean</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 9: Summary of the testing indicators for the \( FL_{\text{in}} \) prediction model.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Scenario</th>
<th>Precision</th>
<th>Recall</th>
<th>( F1 )-score</th>
<th>( FL_{\text{in}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECMPAU</td>
<td>Mean</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>–</td>
</tr>
<tr>
<td>GCCCRNE</td>
<td>Mean</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 10: K-fold cross validation of \( FL_{\text{in}} \) prediction model.

<table>
<thead>
<tr>
<th>Sector</th>
<th>K-fold 1</th>
<th>K-fold 2</th>
<th>K-fold 3</th>
<th>K-fold 4</th>
<th>K-fold 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECMPAU</td>
<td>0.84</td>
<td>0.85</td>
<td>0.87</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>GCCCRNE</td>
<td>0.90</td>
<td>0.92</td>
<td>0.93</td>
<td>0.94</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Therefore, after looking at Fig. 8 and Fig. 9, it is concluded that the chosen tolerances are 25NM for LECMPAU, and 19NM for GCCCRNE. This will result in 19 and 36 clustered air traffic flows, respectively. These results will give a picture of how the sector is organised and the final flows for both sectors. These tolerances have been chosen according to two fundamental criteria:

- $t_{flw}$ has been chosen with the objective of being a tolerance that is capable of grouping aircraft that operate in a similar pattern, but without being high enough to bring together aircraft with different routes at an operational level. For this reason, it has been considered that the optimum tolerance is the one that produces the change in slope of Fig. 7.
- $t_{FLW}$ has been chosen to eliminate the fewest number of aircraft from the analysis due to the fact that they belong to flows with less than 100 operations per year.

Different combinations of tolerances have been tested. But the main difference in the final result is the classification of aircraft into different flows and the number of aircraft eliminated, with the number of final flows remaining almost constant. For this reason, the selection criteria have been established to be correct. Additional tolerance selection criteria will be studied in the future. Work is also underway on an algorithm with automatic selection of tolerances that are better adapted to the nature of the different sectors.

It is important to verify that the identified operational flows are correct. To this end, these flows will be compared with the traffic routes identified by ENAIRE. To test the algorithm in a case of moderate complexity, the flows identified in LECMPAU are first compared. Being an en-route ATC sector, with more structured traffic, the algorithm is expected to perform better than in cases with more irregular traffic. The results are presented in Fig. 10.

The clustered flows in the LECMPAU sector look similar to the routes identified by the Spanish service provider ENAIRE, through its INSIGNIA tool:

- Flows identified with blue colour correspond to flights crossing the sector from south to north or vice versa. At INSIGNIA, they correspond to the route departing from the BANEV notification point and leaving the sector via the LUSEM or BEGUY points, among others, to the north of the sector.
- Green flows correspond to traffic flows in the northwest of the sector. Compared to INSIGNIA, these are routes such as UM190 or UP181.
- Yellow flows are operations from the east to the west of the sector. These routes are very well identified by departing from points such as LEGAM or BAGAS and leaving the sector via the notification point GOSVI or LATEK.
- The red flows correspond to aircraft crossing the sector from east to west, but along a route further south of the sector. Specifically, these correspond to route UN725.
- The flows identified in black are two traffic flows crossing the LECMPAU sector diagonally. Specifically, they correspond to route UN175 passing through LETU, and to a crossing between the west of the sector and the north, passing through the PPN navaid.

The operation in the LECMPAU sector is concentrated on the PPN navaid. The clustering algorithm created, despite not taking this information into account, is able to identify that many of the flows will converge on this navaid. The blue, yellow and black flows will pass through the area where the PPN navaid is located.

In order to have a greater validation of the algorithm, it has been decided to validate it on two additional parameters. The first parameter is the distance between the entry and exit points of the routes defined by ENAIRE and the points of the flows defined by the algorithm. The second parameter is the percentage of traffic that has been clustered by the algorithm. Two limits have been set to guarantee a correct functioning of the clustering algorithm.

<table>
<thead>
<tr>
<th>Sector</th>
<th>K-fold 1</th>
<th>K-fold 2</th>
<th>K-fold 3</th>
<th>K-fold 4</th>
<th>K-fold 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LECMPAU</td>
<td>0.98</td>
<td>0.98</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>GCCCRNE</td>
<td>0.97</td>
<td>0.98</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 11: K-fold cross validation of $FL_{out}$ prediction model.

![Fig. 13. Relative importance of $FL_{in}$ prediction model.](image-url)
distance < 1NM; \( P_{in}, P_{out} \)  

\[
\text{ClusteredTraffic} \geq 0.95
\]

These two parameters are intended to validate the correct identification of flows and the correct clustering of aircraft into flows, respectively. These two quality requirements of the algorithm have been validated by ENAIRE’s ATCOs, giving validity to the selected parameters and margins.

In the case of LECMPAU, Table 1 below identifies the maximum distance and traffic that has been finally clustered by the algorithm.

In this case, both quality parameters are well below the preset margins. In addition, ENAIRE’s controllers have also validated the flows identified by the algorithm, and considered that the flows are tailored to the main traffic flows in the sector, according to their experience.

For these reasons, it is concluded in this first test of the algorithm that in an en-route sector such as LECMPAU, traffic flows can be captured accurately when compared with the operating procedures defined by the service provider ENAIRE. The next step is to test the air traffic flow clustering algorithm in an ATC sector with a high percentage of climbing/descending flights. For this purpose, the performance of this algorithm will be tested in the GCCCRNE ATC sector. This sector was chosen because, being a sector within the airspace of the Canary Islands, it will have a significant amount of interisland traffic, which will mean that the percentage of climbing/descending aircraft will be important. GCCCRNE will have a much more variable behaviour than LECMPAU, having also within the sector the airports GCFV, GCRR and GCLP. This will also test whether the algorithm is able to add the influence of the airports in the sectors where it is necessary.

In this case, the departure and arrival procedures at the airports within the sector, as well as the traffic routes operated, have been obtained from INSIGNIA. The air traffic flow clustering algorithm is able to identify flows of different characteristics. These are represented in different colours in Fig. 11

Table 12
Summary of the testing indicators for the prediction models from all 2019.

<table>
<thead>
<tr>
<th>Model</th>
<th>Scenario</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air traffic flow</td>
<td>Mean</td>
<td>0.90</td>
<td>0.94</td>
<td>0.92</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.14</td>
<td>0.08</td>
<td>0.11</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Worst-case</td>
<td>0.38</td>
<td>0.65</td>
<td>0.48</td>
<td>17, LECMPAU_CL</td>
</tr>
<tr>
<td></td>
<td>Best-case</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>30, LECMPAU_CL</td>
</tr>
<tr>
<td>FL_{in}</td>
<td>Mean</td>
<td>0.85</td>
<td>0.89</td>
<td>0.87</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.11</td>
<td>0.07</td>
<td>0.09</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Worst-case</td>
<td>0.67</td>
<td>0.82</td>
<td>0.73</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>Best-case</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>420</td>
</tr>
<tr>
<td>FL_{out}</td>
<td>Mean</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Standard deviation</td>
<td>0.15</td>
<td>0.06</td>
<td>0.12</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Worst-case</td>
<td>0.49</td>
<td>0.77</td>
<td>0.60</td>
<td>330</td>
</tr>
<tr>
<td></td>
<td>Best-case</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>360, 370, 380, 390, 400, 410, 430, 450</td>
</tr>
</tbody>
</table>

Fig. 14. Relative importance of \( FL_{out} \) prediction model.
algorithm. In order to be able to identify in detail the operation in the vicinity of the airport, it would be necessary to apply this algorithm to sectors of the airport’s TMA (Terminal Manoeuvring Area).

For further validation of the performance of the algorithm, the two quality parameters defined in Equations (11) and (12) are depicted in Table 2.

The parameters are again above the established margins. In addition, the results have been re-validated with ENAIRE’s ATCOs. These controllers have again determined that the identified flows actually coincide with real traffic flows in the GCCCRNE sector.

With this second test, the flow identification and clustering algorithm has been validated in two operating situations of different complexity. In both cases, flows have been identified corresponding to the real operation, and the algorithm has been validated.

3.2. Clustered air traffic flow prediction

Once the clustered air traffic flows have been identified, the future operation has to be classified into these different flows. To this end, the machine learning model to predict the flow that each aircraft will occupy will be developed in the following section. Based on the general characteristics of a flight, set out in the timetable ATC service has, this model will predict the flow through which it will operate in each of the sectors in which the model is to be applied. To be able to apply this algorithm in an ATC sector, it will be necessary to apply the flow characterisation algorithm. This is why this prediction will be made on the LECMPAU and GCCCRNE sectors. As the flow clustering algorithm has been validated in these two sectors, the results of the machine learning model are expected to be correct.

According to the results of projects with a similar topic to the one in this paper (Speiser, Miller, Tooze, & Ip, 2019), it is therefore decided to use a Random Forest algorithm to develop the flow prediction model. Specifically, the parameters that define this Random Forest model are:

- Number of trees: 200
- Function to measure the quality of a split: Gini
- Maximum depth of the trees: 10
- Minimum samples required to split a node: 2
- Maximum samples in a leaf: 1

In addition to the parameters of the algorithm itself, it is very important to know which will be the input and output variables in the model. Specifically, the model will be structured as is presented in Table 3.

With this input data distribution, the machine learning model is trained and tested. From the presented sample of flights operating in LECMPAU and GCCCRNE during 2019, it is decided to use the 80% of the sample for training and 20% for testing the model. Two types of indicators will be used to test the model (Géron, 2017):

- Accuracy: This indicator calculates the percentage of correctly classified data. Accuracy gives a preliminary idea of the model, and in order to have a correct model it should be above 0.85.
- Precision, Recall and F1-Score: These indicators will be calculated for each of the classes, and will allow a more comprehensive assessment of the accuracy of the model. These are calculated as:

\[
\text{precision} = \frac{TP}{TP + FP} \quad (13) \\
\text{recall} = \frac{TP}{TP + FN} \quad (14) \\
F1 - score = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{1}{2}FP + FN} \quad (15)
\]

These values will not be as important as Accuracy, and may sometimes be below 85%. However, the higher these indicators are, the better the performance of the model will be. Additionally, the models will be evaluated by means of K-fold cross Validation. In k-fold cross-validation a dataset X is randomly split into k exclusive subsets X1, . . . , Xk of approximately equal size and the holdout method is repeated k times. Each time, one of the k subsets is used as the test set and the other k – 1 subsets are put together to form a training set (Sengur, 2009). With these different splits of the dataset, Accuracy can be calculated in the different cases and the model can be evaluated with a larger sample and of a different nature. This indicator is more robust than the simple Accuracy, as it allows a more exhaustive analysis of the sample. Thus, if the results are positive, it strengthens the possibility of real application of this tool.

In this case, and in order to be consistent with the Accuracy evaluation, the dataset has been split into 5 parts. 4 of the parts will be used for training and the other for testing. In this way, we will have in each case 80% training data, and 20% testing data. This will allow a solid comparison with the previous indicators. Once the evaluation indicators are determined, the model is trained and tested. First, the Accuracy for each of the sectors is shown, giving an overview of the model developed.

\[
\text{Accuracy}_{\text{LECMPAU}} = 0.97 \quad (16)
\]
For further detail, Table 2 shows a summary of the results, as 19 clustered flows have been identified in LECMPAU and 36 in GCCCRNE. This is a large number of flows to be analysed, so Table 4 presents the mean of the different cases, the standard deviation and the worst-case and best-case scenario and which flow they correspond to, in each of the ATC sectors.

In both cases, the prediction is remarkably accurate. All mean values are above 0.9. In the case of LECMPAU, even the worst-case scenario has
performance parameters around 0.9. This can also be seen from the fact that the standard deviation is 0.03. This value means that all parameter values will be close to the mean values, although with slight differences. However, these differences will be well above the established minimum value of 0.85.

In GCCCRNE, as there is a larger number of flows, in some cases the performance is worse, as in flow 3 with indicators' values just above 0.7. But this occurs only in one flow within the sector. On the other hand, in both sectors the model is able to get it right on all occasions when a flight will pass through certain flows, which is very beneficial for

Fig. 17. Comparison between FLout density in real (blue) and predicted (red) scenarios. Three different days are selected, 3rd April (a), 25th July (b) and 15th October (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
the model. The standard deviation in this case is higher. It happens that in some cases the indicators do fall below the established limit, but these are a minority. In the majority of cases, the indicators are above 0.9.

In this case, the K-fold cross Validation indicators are presented in Table 5 for each of the sectors analysed. This table presents the Accuracy for each of the divisions made during the K-fold cross Validation. In LECPMUA, the scores vary from 0.96 to 0.97. This corresponds to the Accuracy of 0.97, in some cases being 1% lower. This variability is small and is considered to be within valid operating ranges. In GCCCRNE, the scores vary from 0.88 to 0.92. The Accuracy in this sector is 0.92, so the decrease is slightly higher, at 4%. Even so, the results are correct in all cases and are above the established limit of 0.85.

From an operational point of view, LECPMUA is a more structured sector, so the operation will be more similar between different days than in GCCCRNE. In GCCCRNE there is more variability in traffic. This can be seen in the K-fold cross Validation scores.

To get a better understanding of the model and to be able to fully validate it, an analysis of the explainability will be conducted, using the relative importance of the model to know if the training is correct from sector, so the operation will be more similar between different days than in GCCCRNE. In GCCCRNE, the scores vary from 0.88 to 0.92. The Accuracy in this sector is 0.92, so the decrease is slightly higher, at 4%. Even so, the results are correct in all cases and are above the established limit of 0.85.

To complete the traffic density prediction, the last machine learning model will estimate the entry flight level and exit flight level of the ATC sector studied. This machine learning prediction model will be based on the classification of aircraft in the flight levels of the sector based on general data of the operations presented in the timetable that the ATC service uses. In this case there will be two models, the first one predicting FLmin and the second one predicting FLmax. For consistency, the Random Forest algorithm has been used again, keeping the same model parameters:

- Number of trees: 200
- Function to measure the quality of a split: Gini
- Maximum depth of the trees: 10
- Minimum samples required to split a node: 2
- Minimum samples in a leaf: 1

It is also important to state the input and output variables in these models. Table 6 presents the structure of the first model, and Table 7 presents the structure of the second model.

Before presenting the results of the machine learning models, it is important to understand the operation within the sectors in terms of aircraft flight levels. LECPMUA is an en-route sector. The lower limit through which aircraft can operate is the flight level FL345. The upper limit is FL660. Within these limits, aircraft may operate at any flight level they prefer. As for GCCCRNE, this sector contains en-route and climbing/descending traffic. It will therefore have no restrictions on the lower flight level at which they can fly. The upper limit will be FL660. Aircraft will again have freedom of operation within these limits. In this model, the input information is the flow, but in reality the flow through which the aircraft operates will not be a restriction on the flight levels at which the aircraft can operate. This restriction is given by the sector in which it is located. The flow will only give information on which flight levels are most commonly operated in this air traffic stream, depending on the type of aircraft that tend to make up the flow and the length of the route. It is therefore considered that flow may be a useful variable for predicting flight levels.

In order to have a complete coherent traffic density prediction model, a Random Forest algorithm will again be used in these two machine learning models. The training of both models will again be with 80% of the sample and the testing with 20%. The flight levels are too many, it has been decided to group the flight levels in tens. The flight levels will be rounded to the nearest ten and the aircraft will be reassigned. This is expected to reduce the flight levels in order to facilitate the classification and interpretation of the results.

Once this information on the flow level prediction model has been presented, the results are presented in LECPMUA and GCCCRNE. The indicators to evaluate the performance of the models will be Accuracy, Precision, Recall and F1-Score. The first results of the models are presented in the following Accuracies of the 4 models:
Accuracy\_FL\_GCCRNE = 0.94

Accuracy\_FL\_LECMPAU = 0.99

Accuracy\_FL\_GCCCRNE = 0.98

After having analysed the overall performance of each of the models, it is clear that:

- FL\_{out} prediction models perform better than FL\_{in} prediction models.
- In LECMPAU, a less accurate result is obtained when trying to predict the entry flight level than in GCCCRNE, although the improvement is greater when predicting the exit flight level.

Table 8 provides a summary of the results for the FL\_{in} prediction, and Table 9 for the FL\_{out} prediction. In particular, mean Precision, Recall and F1-Score are displayed, standard deviation of the indicators and the indicators’ values for the best-case and worst-case scenario of each model.

The Entry Flight Level model exceeds the 0.85 limit for their indicators in most cases, obtaining results which are considered correct. But the standard deviation suggests that there is a large variability in the results. Turning to the specific results, it can be seen that from flight level FL360 to FL390 in LECMPAU the indicators are below 0.85. This performance of the model can be due to many reasons. The most likely reason is that the operation at these flight levels is more varied, and aircraft operate them indiscriminately. This is why the algorithm finds it difficult to differentiate in this range. In GCCCRNE the indicators are very high, and the standard deviation is lower than in LECMPAU, so it can be said that the performance of the model in this case is correct.

The exit flight level prediction model has an excellent performance. Its mean values are very high, and its standard deviation is very low. Moreover, the worst cases have indicators above the 0.85 limit.

After knowing the values of the indicators for the entry flight level and exit flight level models, the K-fold cross Validation is presented. This indicator will allow to evaluate the model in different situations, strengthening the evaluation of the model and generalising its results. As in the flow prediction model, in these cases the sample is divided into 5 to be able to compare with the Accuracy obtained. Table 10 and Table 11 show the results for the entry flight level and exit flight level respectively.

For the FL\_{in} model, the results are slightly lower than the Accuracy results obtained. In LECMPAU, despite having an Accuracy of 0.88, there are cases of K-fold cross Validation in which it reaches 0.84, below the marked limit. Although these are specific situations and not general ones, they can lead to a loss of reliability. However, this only happens in one split, so it is not a cause for concern. In GCCCRNE, the results are above 0.9 on all occasions, although there is also considerable variability. In this case, the prediction of FL\_{in} is correct.

The FL\_{out} models have a very high K-fold cross Validation in all cases, with the lowest score being 0.96 in GCCCRNE. The dependence on the FL\_{in} entry makes this model very robust, so the scores will be very high in all situations. This is the most robust model of the three due to the high dependence, so it will be the one with the highest overall indicators.

As with the air traffic flows model, an emphasis will also be placed on the relative importance of the input variables in the prediction. The aim of this discussion is to be able to validate the output, and that the models do not arrive at them in a random way.

In Fig. 13 the relative importance of the FL\_{in} prediction model for the LECMPAU and GCCCRNE sectors will be presented. Fig. 14 will present the relative importance of the FL\_{out} predictions in the same two sectors.

In the FL\_{in} prediction, the origin and destination have the strongest influence in both cases. This is meaningful as the type of flight (cruise, arrival, departure) will largely depend on this. Added to this geographical component is the relative importance of the flows, which will also be significant, especially in the case of GCCCRNE where there is a greater variety of possible flight levels, and therefore the flow may be more decisive. In addition, aircraft information is also important, especially the aircraft type. GCCCRNE is a sector that combines cruise and departure/arrival flights. LECMPAU is a sector that will only have cruise flights, even if they are climbing or descending. Here, the flight level the aircraft fly will strongly influence the fuel consumption of the aircraft, so it is logical that the aircraft type will be more influential. In turn, airlines also influence the flight level of aircraft. The least influential for prediction will be the temporal analysis. This is also logical, as it is the other factors that operationally determine the flight level, not the time of the flight.

In the prediction of FL\_{out}, there are variations in the trend. Firstly, the entry flight level is decisive. In LECMPAU more than half of the prediction is determined by the entry flight level. This is because, being cruise flights, flight level variations will in most cases be null, since aircraft will always try to remain at the optimal flight level, and in case of flight level variation, this variation will be bounded. In these prediction models, the most important geographical component is the flow, ahead of the origin and destination airports. As it is a prediction of an outbound variable of the sector, the specific path it has travelled since it entered is important, so the flow will be more decisive than the origin and destination of the aircraft. The next most important variables are related to the aircraft, and finally the temporal analysis, being this again logical from an operational point of view.

After testing the validity of these models, it is proven that, following the proposed methodology, traffic density can be assessed three-dimensionally in an ATC sector.

3.4. Air traffic density prediction

Following the work done, a priority is to be able to make direct applications of the methodology developed in this paper. To this end, traffic density predictions with the machine learning algorithms will be developed in this section. These predictions will consist of the allocation of flows and flight levels of aircraft operating in the airspace at a given time. To check that the models work correctly in a real scenario, two tests have been performed:

- The first test serves to test the veracity of the models in the general operation. For this purpose, the entire 2019 operation in the LECMPAU sector was selected and the models were run consecutively. As there is a large amount of data, the models are evaluated using the parameters Accuracy, Precision, Recall and F1-score. If the results are correct, it is possible to get an idea of the feasibility of using the model in general.
- The second test consists of testing the models on three different operating days in 2019 at LECMPAU. The aim of this is to be able to compare the results in a more specific way.

3.4.1. Evaluation of traffic density in 2019

First, the overall evaluation of the model that is tested with the entire LECMPAU operation in 2019 is shown (Table 12).

Accuracy\_FL\_in = 0.97

Accuracy\_FL\_out = 0.85

Accuracy\_FL\_tot = 0.99

The results are, in general, above 0.85. This means that the model works satisfactorily. In the worst cases, as in the first model test, the indicators are below the 0.85 limit. The labels that are classified incorrectly belong to classes that are hardly used in the actual operation. For this reason, the performance in these cases is not of concern and does not impede the good performance of the overall model.

With the parameters obtained in the overall evaluation, it can be
concluded that the model can be used in general. Next, the traffic density predicted by the models on specific days will be evaluated to see the feasibility of the application of the models in specific cases of daily operation.

### 3.4.2. Evaluation of traffic density in operation days

In addition, a forecast will be made for all aircraft operating in the LECMPAU sector on three different arbitrary days. These days are 3rd April, 25th July and 15th October 2019. For each of the aircraft operating in the sector on these days, a prediction has been made of their air traffic flow, and subsequently of their entry and exit flight levels. At LECMPAU, 4972 aircraft operated throughout that day, so only overall graphs will be presented for this evaluation of air traffic density.

On operating days, 430 aircraft flew on 3rd April, 506 aircraft on 25th July, and 441 on 15th October. A total of 1,377 aircraft were used for the prediction of the different traffic density prediction models.

Specifically, an individual assessment will be made for each of the days. The analysis of the results will be done quantitatively and qualitatively. First, a quantitative analysis will be performed using the Accuracy parameter. It has been considered to use this parameter as it will give an overall picture of the model. It will be checked whether the operating cases have similar Accuracy tiers to those calculated in section 3. Then, a qualitative study will be carried out, comparing the estimated traffic density with the real one. For this purpose, comparative graphs of the three models are presented in Figs. 15, 16 and 17.

First, a study of the horizontal traffic density will be made in Fig. 15, specifying the aircraft flow allocation in the different scenarios, and comparing the predicted and actual flow density. The Accuracy parameters of the flow prediction models on each of the selected days are also indicated.

\[
\text{Accuracy}_{\text{flow}_{\text{FLin}}} = 0.97
\]

\[
\text{Accuracy}_{\text{flow}_{2\text{FLin}}} = 0.96
\]

\[
\text{Accuracy}_{\text{flow}_{\text{FLout}}} = 0.98
\]

The overall accuracy of the flow prediction model in LECMPAU is 0.97. In these three days of operation, the Accuracy is of the same order of magnitude, with a tolerance of ±0.1. This means that the flow allocation model works correctly. Looking at Fig. 15, it can be seen that in most of the flows there is the same number of aircraft in the real case (blue) and in the predicted case (red). The case with the lowest Accuracy is the 25th July. In this case, the allocation of flows 1, 4, 1EECMAPU_CL, 2, 2_LECMPAU, 2, LECMPAU_CL, and 5, LECMPAU_CL are flows with high aircraft occupancy, while flows 17, LECMPAU_CL, 19, LECMPAU_CL, 27, LECMPAU_CL, and 30, LECMPAU_CL are flows with low or no occupancy. These patterns hold in the actual and predicted cases, albeit with some differences in the amounts of aircraft flows on each of the days analysed. Differences in operating days are also captured by the model. So, this flow prediction model works according to the standards previously defined.

In terms of traffic density, the same patterns can be seen over the three days of operation. Flows 1, LECMPAU_CL, 11, LECMPAU_CL, 2, LECMPAU, 4, LECMPAU_CL, and 5, LECMPAU_CL are flows with high aircraft occupancy, while flows 17, LECMPAU_CL, 19, LECMPAU_CL, 27, LECMPAU_CL, and 30, LECMPAU_CL are flows with very low or no occupancy. These patterns hold in the actual and predicted cases, albeit with some differences in the amounts of aircraft flows on each of the days analysed. Differences in operating days are also captured by the model. So, this flow prediction model works according to the standards previously defined.

Once the horizontal density has been analysed by assigning aircraft to air traffic flows, the vertical density is evaluated. To do so, the entry and exit flight levels of the aircraft are studied. Fig. 16 shows the results of the FLout analysis, comparing the actual and predicted cases for the three days. Accuracy is added for the three days analysed.

\[
\text{Accuracy}_{\text{FLin}_{\text{FLin}}} = 0.85
\]

\[
\text{Accuracy}_{\text{FLin}_{\text{FLout}}} = 0.81
\]

\[
\text{Accuracy}_{\text{FLin}_{\text{FLout}}} = 0.85
\]

The Accuracy obtained in the evaluation of the model in LECMPAU was 0.88. On the three days analysed, the Accuracy is below this 0.88. Specifically, there are two days of 0.85 and one of 0.81. This error is the accumulated error of the flow prediction model, and the error intrinsic to the difficulty of traffic on the day. July 25th is the day that is seemingly the most complex, and the day on which the Accuracy of the flow prediction model is slightly lower. For these reasons, the Accuracy of the FLout model is 0.81, below even the permitted standards.

As for the trends, in this case, Fig. 16 shows the violin plots indicating the vertical density of aircraft per flow at the entrance to the ATC sector. The predicted and actual scenarios are shown for the three days studied. For this model, the trends are broadly unchanged, but there are larger differences than for the previous model. In general, the violins of the actual and predicted scenarios have similar shape distributions, but in which the width of the figures is different. This indicates that, although most flights are correctly predicted, small deviations make the shapes more or less uniform than the actual ones. This tendency occurs on all three days analysed, although the case of 25th July is the most noticeable.

Even so, and despite the accuracy being lower than that of the previous evaluation, the model still captures the sector’s trends well on the different days. So the performance of the model, despite being just at or below the standard, seems to work well in predicting vertical traffic density at the entrance to the sector.

Finally, the vertical traffic density at the exit of the sector is analysed. This is done by using the FLout prediction model. Fig. 17 analyses the densities of the actual and predicted scenarios for the three selected days, and also determines the Accuracy of the three models.

\[
\text{Accuracy}_{\text{FLout}_{\text{FLin}}} = 0.99
\]

\[
\text{Accuracy}_{\text{FLout}_{\text{FLout}}} = 0.98
\]

\[
\text{Accuracy}_{\text{FLout}_{\text{FLout}}} = 0.99
\]

In this case, the Accuracy of the global model is 0.99. In the three days analysed, the Accuracy is 0.98 in the worst case, which is again on July 25th.

In this case, the model again captures the trends correctly. The shape of the actual and predicted scenarios is much more similar in the sector entry. This is because it has a much higher Accuracy than the FLin model. In this case, the model is again very accurate in capturing vertical traffic density at the sector exit.

With the two tests carried out during this section, it has been possible to test the generalisability of the use of the model, as it has been validated with large samples. But we have also evaluated the feasibility of using it on specific operating days. It should be noted that these tests have been carried out with real operating data, so the results obtained are of a real application.

### 4. Discussion

Having presented the final results in the previous chapter, a discussion of these results will be made in the present section.

First, an analysis of the results obtained after the air traffic flows characterisation process is described. The methodology, based on three steps, allows for continuous feedback of the process and for optimal results to be attained in the end for two different factors.

- The number of flows obtained covers practically all operations, except for a few flights which route diverges from the norm.
However, the number of air traffic flows identified is still manageable, so that subsequent machine learning models can be used.
- The identified flows correspond to routes defined by the Spanish service provider ENAIRE. This validates the methodology. The obtained results also ensure that, except in exceptional cases, the data show that only some of these defined routes are actually operated.

The first test in a simple route sector such as LECMPAU allows the methodology to be validated in simpler circumstances, as the operation in this type of sector is quite structured. This greatly facilitates the process of identifying and clustering operational air traffic flows. It is in the GCCCRNE sector where it is possible to see how the algorithm performs under more adverse circumstances. GCCCRNE has three airports within it, with operations in all of them. This means that, in addition to cruise traffic, there is a large percentage of climbing/descending flights, making this ATC sector a much more dynamic one. The algorithm has not only correctly captured the en-route traffic, which it had already done in LECMPAU, but has been able to see through the climbing/descending operation. In addition, these flows still correspond to routes or procedures defined by ENAIRE. With this, the clustering methodology is also validated in this context.

The air traffic flow prediction model, based on Random Forest, has made it possible to classify aircraft according to the flow they are expected to occupy during their operation. The results, except in very minor cases, are well above the established 0.85 limit. This makes the developed model reliable. It has been tested in the LECMPAU and GCCCRNE ATC sectors, obtaining similar and above-standard results. Furthermore, when trying to explain the learning process of the model, it has been found that the algorithm has given much importance to the geographical component, although the time of operation and the type of aircraft or airline are also important in the prediction. Comparing with the real operation scenarios, this learning process has been found to be reasonable.

The entry and exit flight level prediction models have been as correct as the flow prediction model. In most cases, the results were well above the 0.85 limit. Moreover, the learning process has also been logical compared with real operation. Particularly, the exit flight level prediction model has performed very well due to its strong dependence on the entry flight level.

The performance of all the machine learning models developed makes the evaluation of the traffic density quite accurate, and the results obtained are very reliable.

The study of the models in different real operating scenarios has led to the following conclusions:

- The flow prediction model allows accurate estimation of horizontal traffic density. This model can capture common and sector-specific trends, but it is also capable of adapting to the variability of different days.
- The entry and exit flight level prediction models, despite containing the cumulative error of flow prediction, perform acceptably. The results of the FLin prediction model are worse than those of the full model evaluation. This is due to the cumulative error, and the higher variability of the model input parameters on a given day. The FLout prediction model gives very accurate results despite the accumulated error. This is due to the high dependence on the input flight level, which makes it a very accurate model.

Overall, except for the FLin prediction model, which needs a slight improvement, it has become clear that the approach followed and the models studied are correct for studying the traffic density of ATC sectors.

5. Conclusions

Finally, the conclusions obtained during the course of this paper and the proposals for improving and future work related to this research project are presented.

Two fundamental results have been obtained from this research process:

- The algorithm for identifying and clustering air traffic flows based on historical traffic data. This algorithm is capable of identifying both routes and defined procedures, and its implementation brings several advantages:
  o Low computational cost. The algorithm only uses entry and exit points, and aircraft attitude at these locations. This actually makes the algorithm computationally inexpensive, as it requires very little information. The requirement of little information has the associated advantage of facilitating the data extraction process. In order to specify the concrete computational cost of the flow clustering algorithm, the elapsed time of the two simulations is calculated. This computing time is compared with the flow clustering algorithm used in the company “ATM Research and Development Reference Centre (C.R.I.D.A.)”. This comparison is made in Table 13. The simulations are made in all 2019 in both cases in order to have a fair comparison.

Table 13 shows the great reduction in computational cost of the algorithm presented in this paper with respect to the algorithm developed by C.R.I.D.A. Specifically, there is a reduction in computational cost of 78% in LECMPAU and 76% in GCCCRNE. These algorithms have been executed both in high computational capacity computers that C.R.I.D.A. has at its disposal.

- Accurate and logical results. The algorithm has been tested in situations of greater and lesser operational difficulty, giving satisfactory results in both cases. The algorithm is therefore validated.
- The algorithm provides results that can be easily and visually assessed. This facilitates the validation process and the generation of results. In addition, it allows a first estimation of the complexity of the sector to be obtained based on the number of clustered air traffic flows and their interaction.
- Machine learning models are able to predict, firstly, the operational flow that the aircraft will occupy, and then the flight levels it will actually use. These models provide very satisfactory results, and the learning process proves to be correct. The high quality of the models makes the prediction a priori accurate. Thanks to the combination of these models, it is possible to structure the aircraft that will operate in an ATC sector horizontally and vertically.

With the developed methodology, the traffic density, a fundamental area in the determination of the complexity of the sector, can be assessed. Moreover, by maintaining an approach based on the simplicity of the models and algorithms, this complexity assessment will be kept, but without compromising accuracy.

Although the results obtained have been very satisfactory, there is always room for improvement. Moreover, this line of research is very interesting for future work on the topic. For this reason, future work is proposed to improve the methodology proposed in this paper.

- Improvements should be made to the methodology: As for the flow identification process, although it is quite robust, further work can be done to improve it.
  o The first step is the automation of the calculation of $t_{fe}$ and $t_{fw}$. In this first proposal, the formation tolerances of elementary and clustered flows are determined manually according to a series of criteria. It would be interesting to automate this selection process, thus extending its range of application by not depending on the human factor.
  o Future research into the definition of the main parameters on which the model is based could be established. Adding information...
such as aircraft heading flying into or out of the ATC sector can make the flow formation more accurate.

- Consider applying this algorithm to several contiguous sectors. This could extend the range of use of the algorithm from elementary sectors, where it is currently used, to integrated sectors (consisting of two or more elementary sectors). This would allow the algorithm to be used in larger airspaces, but without compromising the level of detail and accuracy already available.

- If the algorithm for identifying and clustering operational flows were to be changed, the machine learning models for predicting both flows and flight levels would have to be re-evaluated, as their performance may vary. At present, the models appear to perform properly and with high accuracy, but changing the flows on which these predictions are based may lead to different results.

- In flight levels prediction, there are certain situations where the model does not perform satisfactorily. Although these cases are very rare, it is worth studying them in order to improve the performance of the model. Future lines of research on this subject will attempt to advance in the explainability techniques of these machine learning models in order to try to explain this behaviour and confront it.

Furthermore, it would be interesting to be able to apply this methodology to more ATC sectors of Spanish airspace, or of any other country. This would have the advantage of allowing the methodology to be tested in sectors of a different nature to those in which it has been validated. It would be interesting to conduct a real demonstration of the methodology developed here. This methodology should consist of applying the machine learning models on a specific day, and comparing the results obtained with the real solution. Although this has been done in a general way to test the models, it would be interesting to specifically see where the differences between the predictions and the real operation occur to try to improve the model.

CRediT authorship contribution statement

Francisco Pérez Moreno: Methodology, Writing – original draft.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Víctor Fernando Gómez Comendador reports financial support was provided by Enaire.

The remaining authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

Acknowledgement to ENAIRE and CRIDA for the collaboration and funding of the project in which this research paper has been developed. I would also like to express my gratitude to CRIDA for providing the data necessary to carry out the work and obtain the results.

References


