Departamento de Ingeniería Topográfica y Cartografía

Escuela Técnica Superior de Ingenieros en
Topografía, Geodesia y Cartografía

UNIVERSIDAD POLITÉCNICA DE
MADRID

**Estimación de exposición y vulnerabilidad sísmica empleando
técnicas de aprendizaje automático con datos tomados por sensores
remotos: aplicaciones en Lorca (España) y Puerto Príncipe (Haití).**

*Exposure and seismic vulnerability assessment using machine learning
techniques with remote sensing data: Applications in Lorca (Spain) and
Port-au-Prince (Haiti)*

Tesis Doctoral

Autora:

Yolanda Torres Fernández

Máster en Ingeniería Geodésica y Cartografía

Directoras:

María Belén Benito Oterino

Dra. Ciencias Físicas

Sandra Martínez Cuevas

Dra. Ing. Geográfica

2019

Tribunal nombrado por el Sr. Rector Magnífico de la Universidad Politécnica de
Madrid, el día de de 2019.

Presidente:

Vocales:

Secretario:

Suplentes:

Realizado el acto de defensa y lectura de la Tesis el día de de 2019
en la E.T.S.I. en Topografía, Geodesia y Cartografía.

Calificación

El presidente

Los vocales

El secretario

*A Javier, Julia y Candela, por crear mi mundo; por alegrarme la vida.
A mis padres y hermanos, por lo que soy.*

AGRADECIMIENTOS

Me gustaría expresar mi agradecimiento a todas las personas que me han acompañado durante el tiempo de realización de esta tesis y han contribuido a que sea posible.

En primer lugar, a mis directoras. A la Dra, Belén Benito, por acercarme al mundo de la investigación y compartir conmigo la gran experiencia de trabajar en Haití. A la Dra. Sandra Martínez, por orientarme y guiarme (no sólo en esta tesis), y proporcionarme el marco de trabajo en Lorca. Gracias a las dos por apoyarme, por confiar en mí y por el esfuerzo intenso de los últimos meses.

Agradezco también su apoyo a mis compañeros del grupo de investigación, con quienes comparto una buena parte de mi día a día y muchas conversaciones interesantes (e importantes, que es más importante) sobre lo profesional y también lo personal. En especial, a Jorge Gaspar le agradezco su ayuda con las traducciones y sus aportaciones en forma de “referencias”, porque algunas de ellas han cambiado el curso de esta tesis. También agradezco a Alejandra Staller su confianza en mí y su gran apoyo, tanto en la investigación como en la docencia. A mi compañera Pouyé Yazdi, con quien comparto esta experiencia del doctorado y a quien auguro un futuro brillante en el campo de la investigación. A Ligia Quirós, que comenzamos juntas nuestra andadura en el camino de la ingeniería sísmica y que aún me resuelve algunas dudas técnicas. Y finalmente, a Pilar G. Crende y Ana Rita Serna, con quienes me une una gran amistad, que colaboraron durante unos años junto con Raúl García en la investigación en Haití.

A mis compañeros de la Escuela les agradezco su apoyo, su colaboración en el desarrollo de esta investigación y su ayuda con las tareas docentes, especialmente en el último año: Alejandro Solano, Miguel Ángel Manso, José Juan Arranz, Carmen Morillo, Luis Sebastián, Andrés Díez y Antonio Vázquez.

A los alumnos Azade Haghi y Álvaro Arredondo, cuyos TFM se han desarrollado en el marco de esta tesis, así como a David Álvarez, les agradezco su colaboración en las tareas de investigación de Lorca y Puerto Príncipe.

Quisiera dar las gracias también a un grupo de colegas que me han dado apoyo a lo largo de estos últimos años: a Josechu Martínez-Díaz, por su colaboración en el comienzo de la tesis. A Juan Carlos Ojeda, por su contribución a esta investigación y el interés que ha mostrado siempre. A Sergio Molina, por todo lo que me ha enseñado desde que empecé a estudiar el riesgo sísmico, por su apoyo y guía en los proyectos de Haití, en los viajes y en las publicaciones relacionadas; y por hacer posible la última estancia en su Universidad, junto con Juan Reyes e Igor Gómez, a quienes también agradezco su ayuda. A Emrah Erduran y a Dominik Lang, por sus clases de ingeniería sísmica, que consiguieron despertar mí la curiosidad para seguir investigando en este campo, y por su hospitalidad durante mi estancia en NORSAR. A Txomin Hermosilla, por sus consejos y por responder siempre y rápido a todas mis preguntas sobre Teledetección. A Massimiliano Pittore, que me dio la oportunidad de empezar a trabajar en esta línea que une la Teledetección y la Ingeniería Sísmica y de realizar la estancia en Potsdam, que

fue el verdadero inicio de esta investigación; y por seguir brindándome su apoyo, siempre con enfoque positivo. A Marc Wieland, con quien empecé a aprender las técnicas de análisis de imagen y *Machine Learning*, y por su contribución al estudio en Puerto Príncipe. A los demás colegas del GFZ y a Natalia Zamora y Tasca Santimano, que aportaron el cariz personal y familiar a la estancia en Potsdam que, en principio, iba a ser solo de investigación.

Querría hacer un agradecimiento especial a mis colegas y amigos de Haití, con quienes he compartido tantas experiencias durante casi cuatro años: Dwinel Belizaire, Haendel Dorfeuille, Gaspard Pierristal, Junie Moise, Bobby Piard. Estas personas me han ayudado con la investigación y me han acompañado y guiado en los viajes a Haití, pero también me han abierto las puertas a su mundo, bastante diferente al mío, en el que he podido ver lo injusto de esa diferencia.

Esta tesis no habría sido posible sin la gran cantidad de datos que han estado disponibles, sin coste alguno, o que han sido cedidos para la investigación. Se han utilizado datos abiertos del Instituto Geográfico Nacional de España (IGN), de la Sede Virtual del Catastro de España, del Servicio Geológico de EE.UU. (USGS) y del Rochester Institute of Technology (RIT). Agradezco especialmente la implicación personal en la cesión de datos de las siguientes personas: Donald Makeown y Nina Requeno, del RIT; Christina Corbane, del Joint Research Center of the European Commission; Charles Damien Baguidy, del Ministère des Travaux Publics, Transports et Communications de Haïti; Juan Camilo, del GFZ; y a los expertos que han colaborado en la evaluación de costes: Jorge Gaspar-Escribano, Dimitra Kalantoni, Dominik Lang, Paco Martínez, Sandra Martínez-Cuevas, Abdelghani Meslem, Sergio Molina, Massimiliano Pittore, Antonios Pomonis, Ligia Quirós, Robin Spence y Marc Wieland.

Agradezco también a mi familia, tanto la de Granada como la de Madrid, su ayuda, su apoyo y su comprensión en los momentos que he tenido que dedicar a esta tesis y que no he podido compartir con ellos.

Y por último y muy especialmente, le doy las gracias a Javier, que tanto me ha ayudado siempre y, sobre todo, en estos últimos meses; por encargarse de Julia y de todo lo demás; por creer en mí y valorar lo que hago más que yo misma. Y con todo mi corazón, a súper-Julia, que me echó tanto de menos y le afectó mucho mi ausencia, pero venció la distancia como lo que es: ¡toda una campeona!

RESUMEN

Según el informe de CRED/UNISDR (2016), más de la mitad de las víctimas mortales causadas por los desastres naturales en los últimos 20 años se debieron a terremotos. A través del Marco de Sendai, la ONU promueve la implantación de medidas orientadas a la reducción del riesgo de desastres (DRR) que está justificada por dos razones, principalmente: (1) las amenazas naturales están siendo más frecuentes y mortales con el paso de los años; y (2) la vulnerabilidad de las áreas urbanas va en aumento. Se insta, por tanto, a los países a trabajar por una mejora de la resiliencia de sus ciudades y comunidades.

En ese sentido, con esta tesis se pretende avanzar en la explotación de los datos disponibles para optimizar los procesos de comprensión y caracterización de la vulnerabilidad sísmica del entorno urbano. El desarrollo que las tecnologías geoespaciales han experimentado en las últimas décadas ha hecho posible incorporarlas a estos estudios. Sin embargo, a pesar de que se obtienen resultados satisfactorios en la caracterización de la edificación, aun no se ha valorado, de una manera robusta, la aplicabilidad de esos resultados en los estudios de riesgo sísmico.

Por tanto, en la presente investigación se ha evaluado si es viable técnica y económicamente estimar la exposición y la vulnerabilidad sísmica de una ciudad por medio de técnicas de observación terrestre e incorporar el resultado a un estudio de riesgo. Se han realizado sendas aplicaciones en dos escenarios muy distintos: Lorca y Puerto Príncipe. En ambas ciudades, se han analizado datos tomados por sensores remotos siguiendo las fases diseñadas en el procedimiento metodológico: (1) se ha estratificado la ciudad en patrones urbanos; (2) se han extraído las huellas de los edificios y los atributos que conforman la base de datos de exposición; y (3) se les ha asignado un modelo de vulnerabilidad empleando técnicas de aprendizaje automático. Se ha prestado especial atención a la cuantificación de las incertidumbres inherentes a todos los procesos de análisis y se ha estudiado su impacto en un cálculo de daños por sismo. Finalmente, se ha comparado el coste de este tipo de estudios con el de los procedimientos tradicionales basados en realizar inspecciones de edificios in situ.

Toda la investigación se ha realizado con datos abiertos que no han supuesto ningún coste. Se han empleado imágenes de satélite Landsat, ortofotos de alta resolución, LiDAR y datos catastrales. Para la verificación de los resultados de todas las fases del análisis se ha contado con bases de datos de referencia generadas en investigaciones anteriores de otros autores y cedidas para este estudio.

En el escenario de Lorca, se ha trabajado en una zona del centro histórico y otra de ensanche. Se han identificado un total de 5 tipologías constructivas: 2 de edificios de mampostería y 3 de hormigón armado. La clasificación de más de 800 edificios en estas tipologías se ha conseguido con precisiones globales del 82% (en las zonas del centro) y del 90% (en el ensanche). Tras asignar a cada tipo de edificio el modelo de vulnerabilidad que mejor reproduce su desempeño sísmico, se ha calculado un escenario de daños correspondiente a un sismo simulado de $M_w = 5.2$ con epicentro a 4 km al norte de la ciudad. La comparación de estos daños con los que se obtienen con la base de datos de vulnerabilidad de referencia arroja un MAPE del 4%.

En Puerto Príncipe, se ha trabajado con más de 6000 edificios en áreas de muestra de toda la ciudad, ubicadas en zonas Residenciales, Urbanas, Rurales y asentamientos Informales. Se han identificado 4 tipologías constructivas: 2 de edificios de mampostería y madera y otras 2 de hormigón armado. Los modelos predictivos de clasificación de tipologías alcanzan precisiones globales entre el 74%-88%, según la zona. A cada tipología se le ha asignado un modelo de vulnerabilidad y se ha calculado un escenario de daños tras simular un sismo de $M_w = 7.0$ a 20 km de la ciudad. La comparación de los daños obtenidos con el mejor modelo de exposición y los que se obtienen al usar los datos de vulnerabilidad de referencia deja un MAPE también del 4%.

Estos valores bajos de error parecen demostrar que las incertidumbres encadenadas en los procesos previos de generación de huellas, cálculo de atributos y clasificación de tipologías tienen un impacto mínimo en la estimación de daños por sismo. Son resultados altamente satisfactorios, que permiten responder positivamente a la hipótesis de investigación sobre la viabilidad técnica del uso de las bases de datos creadas por teledetección en estudios de riesgo.

Finalmente, se ha llevado a cabo un análisis comparativo del coste de este tipo de estudios para evaluación de la exposición y la vulnerabilidad sísmica y el coste que conlleva la realización de una campaña de campo para el mismo fin. La información sobre el coste de las campañas de campo ha sido proporcionada por 12 expertos internacionales con base en su propia experiencia. De la comparación se deduce que el análisis de datos tomados por sensores remotos permite obtener un producto comparable al que se obtendría en campo, pero por un 10%-30% de su precio y empleando, como mucho, un tercio del tiempo, lo que avala la viabilidad económica de estos estudios.

Como aporte final de esta investigación, se propone un procedimiento de 5 fases para generar bases de datos de exposición y vulnerabilidad de alta calidad a nivel de edificio, que combina teledetección y campañas de campo dirigidas optimizando recursos y coste.

ABSTRACT

More than half of the casualties caused by disasters (related to natural hazards) during the last 20 years were due to earthquakes (CRED/UNISDR, 2016). The UN, through the Sendai Framework, fosters the implantation of measurements for Disaster Risk Reduction (DRR). This is based on two factors: (1) the increase in the frequency and mortal impact of natural hazards and (2) the rise of the vulnerability conditions in urban areas. Hence, all countries are urged to work toward the improvement of the resilience of cities and communities.

In this sense, this dissertation aims at expanding knowledge on the exploitation of available data to optimize the processes of understanding the earthquake vulnerability of the built environment. Earth Observation (EO) technologies have been successfully incorporated in these studies during the past decade. However, despite good results are achieved in building characterization, it has not been robustly valued yet the applicability of these results to a seismic risk study.

Thus, in the present research it is examined the technical and economic viability of assessing the urban exposure and seismic vulnerability by means of Earth Observation techniques and incorporating the results in a risk study. Two applications are performed in cities that present significant differences regarding extension and complexity: Lorca (Spain) and Port Prince (Haiti).

In both cities, data acquired by remote sensors are analyzed according to the stages designed in the methodological procedure: (1) stratification of the city into homogeneous urban patterns; (2) building footprint extraction and attribute calculation to implement the exposure database; and (3) vulnerability model allocation using machine learning techniques followed by an estimation of earthquake damage. Special attention is paid to the quantification of the uncertainties inherent in all the steps of the analysis and their impact on the final results of damage. Finally, the cost of this procedure is compared to the cost of traditional approaches based on in-field building inspections.

The entire research is developed with open data that have do not imply any monetary cost. The data includes Landsat satellite images, high-resolution orthophotos, LiDAR and cadastral data. Reference databases generated in previous studies by other authors (and provided to this study) are used for the verification of the results obtained in all the phases of the analyses.

In the scenario of Lorca, the work is carried out in one area of the historic center and a sprawl of the city. A total of five building typologies are identified: two of masonry and three of reinforced concrete. The classification of more than 800 buildings is achieved with global accuracies of 82% and 90% for the historic center and the sprawl, respectively. After assigning to each building the vulnerability model that best reproduces its seismic performance, it is calculated the damage scenario corresponding to a Mw 5.2 earthquake with epicenter located 4 km to the North of the city. The comparison of these result with those obtained with the reference vulnerability database gives a MAPE of 4%.

In Port Prince, the work is done over more than 6000 buildings in sampling areas of the entire city, located in residential, urban, rural and informal settlements. Four building typologies are identified: two of masonry and timber and another two of reinforced concrete. The predictive models for classification of typologies reach global accuracies of 74%-88%, depending on the zone. Each typology is assigned a vulnerability model and it is computed a damage scenario after simulating a $M_w = 7.0$ earthquake located 20 km East of the city. The comparison of the damage obtained with the best exposure model and those obtained using the reference vulnerability data gives also a MAPE of 4%.

These low error values seem to demonstrate that the uncertainties involved in the previous processes of footprint generation, attribute determination and typologies classification have a minimum impact on the estimation of earthquake damage. They are highly satisfactory results, which allow giving an affirmative answer to the research hypothesis on the technical viability of using exposure and vulnerability databases created by means of EO techniques in seismic risk studies.

Finally, the cost of an in-field survey to collect exposure and vulnerability data is compared to the cost of creating the same database using remote sensing techniques. The latter has been derived from this study, while the former has been provided by twelve international experts based on their own experience. From the comparison, it is deduced that the analysis of remotely sensed data provides a product that is comparable to the one obtained in the field, but for a 10%-30% of its monetary cost and consuming about one third of the total time, which avails the economic feasibility of these studies.

As a final contribution of this research, it is proposed a five-stage procedure to generate high-quality, building-scale exposure and vulnerability databases that combines remote sensing and directed in-field surveys that allows for optimization of resources and cost.

ÍNDICE DE CONTENIDOS

Bloque 1	15
1 Introducción	17
1.1 Marco conceptual	21
1.2 El problema de la evaluación de la vulnerabilidad sísmica	23
1.3 Motivación y alcance	27
1.4 Contexto y antecedentes	29
2 Estado del arte	33
2.1 Modelos de exposición	33
2.2 Modelos de vulnerabilidad	38
2.3 Técnicas de teledetección aplicadas a la caracterización de edificios	41
3 Hipótesis de la investigación y objetivos	49
Bloque 2	53
4 Propuesta metodológica	55
4.1 Introducción	55
4.2 Procedimiento metodológico	55
4.2.1 Fase 1: Estratificación de la ciudad	55
4.2.1.1 Segmentación de la imagen	57
4.2.1.2 Clasificación de los segmentos	58
4.2.1.3 Verificación de los resultados de clasificación	63
4.2.2 Fase 2: Creación de la base de datos de Exposición	66
4.2.2.1 Generación de las huellas de los edificios	66
4.2.2.2 Verificación del número de huellas	68
4.2.2.3 Cálculo de los atributos	68
4.2.2.4 Verificación de los atributos	72
4.2.3 Fase 3: Clasificación de tipologías y asignación de la vulnerabilidad sísmica	72
4.2.3.1 Elección del modelo de vulnerabilidad y clasificación de tipologías constructivas	72
4.2.3.2 Verificación de los resultados de clasificación de los edificios en MBT	73
Bloque 3	75
5 Aplicación del procedimiento diseñado en dos escenarios	77
5.1 Aplicación en Lorca, España	77
5.1.1 Área de estudio y datos	78

5.1.2	Proceso de cálculo diseñado para Lorca. Resultados y análisis.....	82
5.1.2.1	Fase 1: estratificación de la ciudad de Lorca.....	82
5.1.2.2	Fase 2: creación de la base de datos de exposición de Lorca.....	86
5.1.2.3	Fase 3: clasificación de tipologías constructivas y asignación de la vulnerabilidad sísmica a los edificios de Lorca.....	98
5.1.3	Estimación de tiempos y costes.....	105
5.1.4	Estimación del daño a partir de los datos obtenidos.....	107
5.1.5	Sumario y conclusiones de la aplicación en Lorca.....	110
5.2	Aplicación en Puerto Príncipe, Haití.....	115
5.2.1	Área de estudio y datos.....	117
5.2.2	Proceso de cálculo diseñado para Puerto Príncipe. Resultados y análisis....	123
5.2.2.1	Fase 1: estratificación del escenario de Puerto Príncipe.....	125
5.2.2.2	Fase 2: creación de la base de datos de exposición de Puerto Príncipe....	137
5.2.2.3	Fase 3: clasificación de tipologías constructivas y asignación de la vulnerabilidad sísmica a los edificios de Puerto Príncipe.....	151
5.2.2.4	Extrapolación al resto del área de estudio.....	160
5.2.3	Estimación de tiempos y costes.....	163
5.2.4	Estimación del daño a partir de los datos obtenidos.....	164
5.2.5	Sumario y conclusiones de la aplicación en Puerto Príncipe.....	173
6	Análisis comparativo de costes: estudio por teledetección Vs. campaña de campo	181
Bloque 4		189
7	Sumario, discusión y conclusiones generales.....	191
8	Summary, discussion and main conclusions.....	203
9	Líneas futuras.....	213
Bloque 5		215
10	Referencias.....	217
11	Glosario de términos y acrónimos.....	231
12	Anexos.....	237

Bloque 1

1 Introducción

Más de la mitad de las víctimas mortales causadas por los desastres naturales en los últimos 20 años se debieron a terremotos (Figura 1-1). El informe de CRED/UNISDR (2016) sobre los desastres naturales acaecidos entre 1996 y 2015 saca a la luz estas reveladoras cifras y otras relacionadas con la pobreza, como que la abrumante mayoría de estas víctimas vivían en países en vías de desarrollo.

De los tres megadesastres que marcaron este período (eventos que dejan más de 100.000 víctimas mortales), dos fueron terremotos. En la última década, el evento que más mortalidad causó fue el terremoto de Haití de 2010, de Mw = 7,0 (Hayes et al., 2017), con una cifra estimada de 223.000 fallecidos (CRED/UNISDR, 2016).

En la década anterior, fue el tsunami de Sumatra (Indonesia) de 2004, causado por un terremoto submarino de gran magnitud (Mw = 9,1, Duputel et al., 2012), que dejó 160.000 muertes (NatCatSERVICE, 2019). El tercer megadesastre fue el ciclón Nargis que azotó Birmania en 2008 (140.000 fallecidos, NatCatSERVICE, 2019).

El terremoto de Haití, que ocurrió el 12 de enero 2010, inspiró la presente investigación. Fue un evento que marcó un antes y un después en la conciencia de la población mundial por la tragedia humana que supuso. La comunidad humanitaria internacional lanzó una respuesta masiva en ayuda del país (Bhattacharjee y Lossio, 2011). Haití es el estado más pobre de América (Banco Mundial - PIB2017) y, hasta ese momento, era un gran desconocido (UN, 2011); este desastre lo colocó, desafortunadamente, a la vista de todos. Las cifras de afectados han sido materia de debate en la sociedad y en la comunidad científica. La estimación del número de fallecidos ha oscilado entre los 70.000 y los 316.000 (DesRoches et al., 2011). La ONU cifraba en más de un millón a las personas que quedaron sin hogar y que tuvieron que instalarse en campamentos (UN, 2011). Tras el terremoto, durante 2010 y 2011, los haitianos sufrieron, además, la época de huracanes y un brote de cólera.

Un mes después, se produjo un fuerte seísmo en Chile que fue 500 veces mayor en términos de energía liberada y, sin embargo, la cifra de mortalidad fue mucho menor (medio millar). En Nueva Zelanda, tuvo lugar otro terremoto en septiembre de ese año, con la misma intensidad

Figura 1-1. Número de terremotos, tormentas, temperaturas extremas e inundaciones ocurridos entre 1996 y 2015. Extracto de CRED/UNISDR (2016), Figura 2. El 3'2% restante corresponde a otros tipos de desastres menos frecuentes.

(IMM) que el de Haití, y no causó muertes. Esta comparación invita a preguntarse cuáles fueron las causas que elevaron el caso de Haití a la mayor tragedia humanitaria de la historia reciente (Bhattacharjee y Lossio, 2011). Parte de la explicación se encuentra en la baja frecuencia con que ocurren grandes seísmos en la región. Ello hace que los gobiernos sean menos proclives a iniciar políticas de prevención de riesgos y, razonablemente, deciden invertir en otras materias sociales de primera necesidad, como educación o sanidad (CRED/ UNISDR, 2016). Por otro lado, la inestabilidad política mantenida durante décadas en un sistema democrático débil ha ido minando la confianza de la población en los gobiernos y ha favorecido la corrupción (UN, 2011). Además de esta vulnerabilidad social, Haití presenta una alta vulnerabilidad en sus construcciones (Bilham, 2010; Mix et al., 2011). Hasta el momento del terremoto, en el país no se seguía un código sísmico para la edificación, a pesar de estar en una zona de alta sismicidad.

Tras sufrir eventos como los tres megadesastres citados, normalmente los países ponen en marcha medidas para mitigar el riesgo en el futuro. En el caso de Indonesia, se incluyeron políticas para la Reducción del Riesgo de Desastres (DRR, *Disaster Risk Reduction*) en la estrategia de desarrollo del país justo después del tsunami de 2004. Birmania, en su caso, se comprometió a implementar el Marco de Acción del Hyogo (CRED/UNISDR, 2016). En Haití, sin embargo, no ha habido una política sólida en este sentido y se han ido dando pequeños pasos, hasta 2017, año en que la ONU y el gobierno haitiano han redactado conjuntamente el Documento Nacional para la Reducción del Riesgo de Desastres 2016-2020 (UNDRR-AM, 2017). Cuando los países no actúan rápido, se ven sumidos en catástrofes posteriores sin haberse recuperado.

La necesidad de implantación de medidas DRR está justificada por dos razones, principalmente (CRED/UNISDR, 2016): (1) la tendencia reciente, que indica que las amenazas naturales están siendo más frecuentes y mortales con el paso de los años; y (2) las previsiones sobre el incremento de la vulnerabilidad de las áreas urbanas. La ONU estima que la mayor parte del crecimiento de las ciudades que se espera para 2030 tendrá lugar en países que no tienen capacidad para garantizar un desarrollo urbano sensible al riesgo. No obstante, hay que destacar que también los países más desarrollados son sorprendidos por eventos naturales que sobrepasan sus predicciones. Un ejemplo claro es el terremoto de Tohoku (Japón) de 2011, de Mw=9,0 (Duputel et al., 2011), que fue seguido de un tsunami, ocasionando un desastre nuclear en Fukushima. Eventos de otra naturaleza también dejan casos destacados, como el huracán Katrina a su paso por Nueva Orleans (EE.UU.) en 2005, o la ola de calor que azotó Europa en 2003. Estos casos evidencian que es necesario seguir actuando en materia de reducción de riesgo de desastres a todos los niveles, local, nacional e internacional, con especial énfasis en el análisis y la comprensión del riesgo en todas sus dimensiones (CRED/UNISDR, 2016).

En la inevitable comparación de las consecuencias de los desastres en países más y menos favorecidos, se da una paradoja sobre la que es conveniente reflexionar. Los mayores números de víctimas ocurren en países en desarrollo, como es el caso del terremoto de Haití de 2010 o el tsunami de Sumatra de 2004. Sin embargo, los mayores costes asociados a reposición y reparación se dan en países desarrollados, como en el tsunami de Tohoku de 2011 (con un daño estimado en 210000 millones de dólares) o el terremoto de Kobe de 1995 (con una estimación

de costes de 100000 millones de dólares), ambos en Japón (NatCatSERVICE, 2019). Esto no quiere decir que los daños en los primeros ejemplos mencionados hayan sido menores, sino que el coste de los bienes dañados varía significativamente de unos países a otros. Este hecho puede desvirtuar la percepción de los ciudadanos sobre el alcance de los desastres.

Con el ánimo de comparar una aplicación de estimación de vulnerabilidad sísmica en países de distinto nivel de desarrollo, además de en Haití, en la presente tesis se ha trabajado en un segundo escenario: Lorca (España). El interés por esta localización surge a raíz del terremoto que ocurrió en esta localidad el 11 de mayo de 2011, y que dejó 9 víctimas mortales y unos 300 heridos (Martínez-Díaz et al., 2012). El sismo alcanzó una magnitud $M_w=5.1$ (Cabañas et al., 2011), que se considera moderada en el contexto internacional. Con el objetivo de proteger a la población de este tipo de eventos, en España se cuenta con la Directriz Básica de Planificación de Protección Civil ante el Riesgo Sísmico (DBPC, 2004), que es la norma que establece cómo debe prepararse el sistema de protección civil para hacer frente a este riesgo. Este documento estipula que todos aquellos municipios que se encuentren en una zona donde son previsibles sismos de intensidad VII o mayor (según el mapa de peligrosidad sísmica de la Norma Sismorresistente Española, NCSE-02¹), deben realizar su propia evaluación de riesgo sísmico a nivel local. En total son más de 700 municipios de 15 provincias los que tienen obligación de desarrollar su plan municipal de emergencias ante el riesgo sísmico y, hasta la fecha, son muy pocos los municipios que los han realizado.

A nivel internacional, existen iniciativas que llevan años funcionando. El Marco de Acción de Hyogo (HFA-2005, *Hyogo Framework for Action*), promovido por la Asamblea General de Naciones Unidas, tenía como objetivo global aumentar la resiliencia de los países y las comunidades ante los desastres en la década 2005-2015. Hasta 168 países respaldaron el HFA. La ONU, en su evaluación, concluyó que, a pesar de haber conseguido importantes avances durante esos años en la preparación y la respuesta, así como en los sistemas de alerta temprana, aún queda mucho por hacer. Especialmente, hay que trabajar en los aspectos subyacentes al riesgo de desastres, como la pobreza, la gobernanza o el rápido crecimiento de la urbanización y la población (CRED/UNISDR, 2016). En la evaluación global del riesgo (GAR, 2009), la ONU pone de manifiesto que, además del propio carácter impredecible de la naturaleza, son las limitaciones de los datos las que dificultan la comprensión del riesgo actualmente. En ese sentido, con esta tesis se pretende avanzar en la explotación de los datos disponibles para optimizar los procesos de comprensión y caracterización de la vulnerabilidad sísmica del entorno urbano.

Como continuación del HFA, se definió el Marco de Sendai para la Reducción del Riesgo de Desastres 2015-2030 (Sendai-Framework-2015, *Sendai Framework for Disaster Risk Reduction*). De nuevo, fue una iniciativa que respaldó el conjunto de la Asamblea General de Naciones Unidas. De los siete objetivos globales del Marco de Sendai, merece la pena destacar tres, que están en línea con esta tesis: (1) incrementar el número de países que implementan estrategias DRR para 2020; (2) mejorar la cooperación internacional hacia países en desarrollo

¹ <https://www.ign.es/web/ign/portal/mapas-sismicidad>.

en esta materia, mediante acciones sostenibles; y (3) incrementar la disponibilidad y el acceso de la población a sistemas multi-amenaza y datos relacionados con el riesgo de desastres.

La reducción del riesgo de desastres es tan relevante para los países, que ha sido un asunto clave en la definición de otras acciones internacionales, como los Objetivos de Desarrollo Sostenible (ODS-2015) de la Agenda 2030 de las Naciones Unidas. Concretamente, el objetivo 11 está enfocado en ciudades sostenibles. Con la previsión de que dos tercios de la población mundial vivirá en ciudades en 2030², la ONU insta a los países a trabajar por un aumento de la resiliencia de sus ciudades y comunidades. Una incitativa reseñable y relacionada con esta temática, que ilustra el papel y el compromiso de las Naciones Unidas con el objetivo 11, es el caso de la reconstrucción de la ciudad de Gorkha, el epicentro del terremoto de Nepal de 2015, en el que casi un millón de casas quedaron destruidas. La propuesta local es reconstruir siguiendo los métodos tradicionales, pero aplicando técnicas más seguras. El PNUD y el gobierno de India, de acuerdo con esta propuesta, iniciaron un proyecto de reconstrucción para ayudar a más de 26000 propietarios de Gorkha³.

Estructura de la tesis:

El presente documento se estructura en cinco grandes bloques, divididos a su vez en capítulos. Dentro del **primer bloque**, el Capítulo 1 presenta un marco conceptual en el que se tratan los fundamentos de esta investigación y se establecen algunos criterios de nomenclatura. A continuación, se describe el problema actual de la asignación de la vulnerabilidad sísmica, que justifica la motivación y el alcance. Cierra el capítulo el contexto de investigación en el que se ha llevado a cabo la tesis. En el Capítulo 2, se incluye la revisión del estado del arte, seguida de la hipótesis de investigación y los objetivos concretos que se pretende alcanzar con este estudio, descritos en el Capítulo 3.

El **segundo bloque** corresponde al planteamiento metodológico, en el que se describen las fases diseñadas para conseguir los objetivos, y los métodos y técnicas de análisis empleados en cada una de ellas (Capítulo 4). Seguidamente, en el **tercer bloque**, se describe la aplicación en los dos escenarios elegidos, Lorca y Puerto Príncipe, junto con los resultados y el análisis de los mismos en el Capítulo 5. Para cada escenario, en este capítulo, también se incluye una estimación de tiempo y costes, un cálculo de daño por sismo y las conclusiones extraídas de la aplicación. El capítulo 6 está dedicado a realizar una comparativa de costes de estos métodos con las técnicas tradicionales de elaboración de bases de datos de exposición y vulnerabilidad.

En el **cuarto bloque**, se presenta un sumario y las conclusiones generales de esta investigación (Capítulos 7 y 8), así como las líneas que quedan abiertas tras este estudio, que podrán ser abordadas en futuras investigaciones (Capítulo 9). Por último, en el **quinto bloque**, se incluyen las referencias bibliográficas (Capítulo 10), un glosario de términos y acrónimos (Capítulo 11) y una serie de Anexos (Capítulo 12).

² <http://www.bancomundial.org/temas/cities/datos.htm>

³ <https://medium.com/@pnud/reconstrucci%C3%B3n-y-resiliencia-en-nepal-22c9ada43bae>

1.1 Marco conceptual

Evaluar el riesgo sísmico consiste, básicamente, en predecir las consecuencias que tendrá un terremoto sobre una región, ciudad o infraestructura (e.g. Lang, 2013). Por consecuencias se entiende tanto el daño físico de las estructuras como las pérdidas humanas y económicas asociadas a ese daño. Intervienen, por tanto, dos factores principales: la **peligrosidad** sísmica y la **vulnerabilidad** de los elementos expuestos a ese peligro (TerminologyDRR-2016). La estimación de la peligrosidad sísmica aporta una medida del movimiento del suelo causado por la sacudida sísmica; mientras que la vulnerabilidad de los bienes sometidos a dicha sacudida informa acerca de la capacidad de estos a sufrir daños. Un bien expuesto es todo aquello que se quiere proteger del peligro, ya sean recursos –como un bosque protegido, por ejemplo-, infraestructuras –tanto viviendas convencionales como elementos de singular importancia, tales como puentes, colegios u hospitales-, líneas vitales –como carreteras o redes de gas-, y personas. Los conceptos de **exposición** (bienes expuestos al riesgo) y vulnerabilidad se encuentran habitualmente separados en la literatura. Así, tradicionalmente se ha considerado que los componentes del riesgo sísmico son tres: peligrosidad, vulnerabilidad y exposición (UNDRRO, 1979). En esta definición de UNDRRO, el concepto de vulnerabilidad estaba asociado a los daños o pérdidas directamente, es decir, a los efectos que una sacudida sísmica puede tener sobre los elementos expuestos. Inicialmente se expresaba en términos de escala, de 0 a 1. Trabajos posteriores proponen que los componentes del riesgo son dos, la peligrosidad y la vulnerabilidad, sin mencionar explícitamente la exposición (e.g. Coburn y Spence, 2002; figura 9.2). Incluso el coste de reconstrucción de los bienes se trata por separado.

Figura 1-2. Definición de riesgo sísmico

Sin embargo, en esta tesis, se considera que la exposición es un componente relevante del riesgo que debe tener entidad por sí solo. La vulnerabilidad es una característica intrínseca a los bienes expuestos, al ser un indicativo de su capacidad para resistir el movimiento sísmico; también lo es el valor económico. La ecuación del riesgo queda, por tanto, expresada en términos del peligro y la exposición (Figura 1-2). La exposición tiene dos componentes principales: las personas y las estructuras. Cuando la exposición se refiere a estructuras (como bienes inmuebles, infraestructuras o instalaciones especiales), estas han de venir caracterizadas por su vulnerabilidad y, si es posible, también por su valor económico, como se indica en la Figura 1-2. Cuanto mejor se caracterice la exposición, más indicadores de riesgo se

podrán calcular. En la Figura 1-3 se presenta un esquema del procedimiento de cálculo de **indicadores del riesgo sísmico**. De la vulnerabilidad de las estructuras, en combinación con los parámetros del movimiento del suelo, se derivan los daños físicos. Estos vienen expresados en términos de número de estructuras dañadas y del grado de los daños. Con esta estimación de daño y el valor económico de los bienes materiales expuestos, se calculan las pérdidas económicas. Análogamente, con el daño y el dato de población, se estiman las pérdidas humanas.

Figura 1-3. Esquema del cálculo de los indicadores de riesgo sísmico

El concepto probabilista del riesgo sísmico va asociado al concepto probabilista de la peligrosidad, que se define como la probabilidad de excedencia de un determinado nivel de movimiento fuerte del suelo como consecuencia de la ocurrencia de terremotos en un emplazamiento o región durante un cierto periodo de tiempo (Coburn y Spence, 2002). Sin embargo, un estudio de riesgo sísmico se aborda generalmente considerando un sismo concreto asociado a una falla (Lang, 2013), limitando de esta manera el espacio y el tiempo. El sismo se caracteriza por su magnitud, localización y profundidad, así como por la geometría y el mecanismo de la falla que lo genera. Es lo que se conoce como escenario sísmico determinista. El nivel de movimiento del suelo causado por un escenario determinista se calcula utilizando modelos de predicción del movimiento fuerte (GMPE, *Ground Motion Prediction Equation*), también llamados leyes de atenuación.

Dado que esta tesis tiene como escenarios de aplicación dos ciudades, en lo sucesivo se va a tratar el riesgo sísmico según este enfoque determinista. Las entidades gubernamentales locales encargadas de la gestión del riesgo trabajan, generalmente, con estudios de riesgo deterministas en lugar de probabilistas, pues aportan una visión concreta de las consecuencias que un sismo potencial puede tener en su ciudad (Lang, 2013). De estos estudios deterministas se obtienen predicciones de daños y estimaciones de las pérdidas humanas y económicas asociadas a los daños. Con base en estos resultados, se elaboran los **planes de emergencias** ante este riesgo, en los que se definen medidas de prevención y acciones de contingencia para mitigar las consecuencias negativas de terremotos futuros. Dado que la ocurrencia de terremotos (peligro) no se puede evitar, las políticas de mitigación actúan sobre la exposición, bien reduciéndola o controlándola, o bien aumentando su resistencia a la sacudida por medio de la construcción sismorresistente (es decir, disminuyendo su vulnerabilidad). De esta manera se consiguen **ciudades y sociedades más resilientes** ante el riesgo sísmico.

La mayoría de los estudios de riesgo sísmico a nivel local centran la exposición en la caracterización de los edificios de viviendas convencionales, dejando otras estructuras, como edificios de especial importancia (colegios, hospitales...), infraestructuras de transporte (puentes, carreteras...) e instalaciones (presas, centrales, suministros...), para estudios adicionales de carácter particular. Es por ello que esta tesis se centra en la caracterización de los edificios residenciales como elemento principal de la exposición.

1.2 El problema de la evaluación de la vulnerabilidad sísmica

Estimar la exposición y determinar su vulnerabilidad para llevar a cabo un estudio de riesgo sísmico en una ciudad es una tarea ardua, compleja y costosa en términos de recursos y tiempo. El proceso consta de tres pasos principales. En primer lugar, se deben identificar las tipologías constructivas más representativas del parque inmobiliario de la ciudad, es decir, los modelos de edificios tipo en los que se podrían clasificar todas las construcciones que hay en el momento de la realización del estudio. El objetivo es simplificar la realidad para facilitar el análisis de riesgo posterior. Para ello, es necesario contar con conocimiento de experto sobre esta materia (ingeniería civil, arquitectura...) y, además, sobre las particularidades relacionadas con los métodos de construcción de la zona (materiales habituales, técnicas constructivas, calidad de la edificación, normativas aplicadas y fechas de entrada en vigor, procesos de inspección, etc.). El segundo paso es recabar toda la información posible sobre todos y cada uno de los edificios de la ciudad de estudio. Esta información debe ser de utilidad para conocer la vulnerabilidad del edificio ante un sismo: localización y aspectos geométricos (en planta y alzado), materiales de construcción y sistema estructural, edad y estado de conservación, etc. Toda esta información se almacena en forma de atributos, generando así la base de datos de exposición. A continuación, y con base en estos atributos, se clasifican los edificios en las tipologías identificadas en el paso anterior. De esa manera, se consigue agrupar los edificios que guarden ciertas similitudes entre sí en un número manejable de clases. El último paso es asignar un modelo de vulnerabilidad a cada una de las clases anteriores que informe acerca de cómo se comportará cada tipología ante un sismo.

Para recopilar toda la información necesaria para acometer un estudio de vulnerabilidad sísmica, generalmente se parte de bases de datos espaciales de catastro, de censo o cartográficas. En estas fuentes se puede encontrar información sobre la localización de los edificios y algunas características relacionadas con su vulnerabilidad, como el año de construcción. Sin embargo, no todas las ciudades tienen catastro oficial; y con frecuencia, en aquellas que lo tienen, se presentan contratiempos como que el catastro está desactualizado o no recoge todos los datos necesarios para el estudio (como ocurre generalmente con los materiales de construcción de los edificios e infraestructuras); o simplemente no es accesible debido a la falta de una política de libre distribución de los datos. En estos casos, para poder llevar a cabo el estudio de riesgo sísmico, se ha de recurrir a campañas de campo. En estas campañas, un grupo de ingenieros y arquitectos expertos en vulnerabilidad sísmica recorren la ciudad para recoger datos sobre los edificios y otras estructuras de interés. Uno de los primeros documentos que prescribe cómo realizar una inspección de edificios fue publicado por la *US Federal Emergency Management Agency* (FEMA-154, 1988). Este documento estableció el

concepto de inspección rápida de edificios o RVS (*Rapid Visual Screening*) que permite a expertos y no expertos en edificación (como, por ejemplo, los propietarios) discernir si un edificio es vulnerable o no a terremotos, para identificar aquellos en los que es necesario una inspección posterior de mayor detalle. Más tarde, la Sociedad Americana de Ingenieros Civiles presentó un manual (ASCE/SEI 31-03, 2003) para una inspección mucho más rigurosa, en tres niveles, que comienza con una RVS. Otros países con elevada actividad sísmica, como Japón (JBDPA, 1990) o Turquía (Ozdemir et al., 2005), han elaborado sus propios métodos de inspección de edificios en sucesivos niveles que permiten estimar un índice de vulnerabilidad global de cada edificio.

En las ciudades de estudio de la presente tesis ya ha habido experiencias previas de recolección de datos de la edificación relacionadas con la gestión del riesgo sísmico. En Lorca, tras el terremoto de 2011 y a raíz de una iniciativa voluntaria local, se confeccionó un formulario de recogida de datos sobre los edificios, en cuya definición participó la doctoranda. El objetivo era que los propietarios pudieran generar una base de datos de exposición pre-evento introduciendo los datos referentes a su vivienda o edificio en una aplicación móvil. La aplicación se llamó SGE 2.0⁴ y acaba de presentarse la última versión desarrollada en el seno del proyecto POCRISC (Martínez y Goula, 2019). En Haití, el Ministerio de Obras Públicas, Transportes y Comunicaciones (MTPTC, *Ministère des Travaux Publics, Transports et Communications*) realizó una evaluación de los edificios tras el terremoto de 2010⁵. Las tareas de inspección, que se alargaron durante un año, involucraron a 554 ingenieros y abarcaron casi 400.000 edificios. El formulario que emplearon estos ingenieros y el implementado en la aplicación SGE 2.0 están incluidos en el Anexo I.

Como se ve, cada institución puede generar sus propios formularios de recogida de datos, ya sean analógicos (fichas en papel) o digitales (para usar en *smartphones* o *tablets*), si bien, en general, son todos muy similares. Entre los datos más relevantes que se suelen recoger se encuentran la localización, altura, área del solar, materiales de construcción, método constructivo, año de construcción, coste de construcción y personas que lo habitan o frecuentan. Todos estos atributos están definidos y justificados en las diferentes taxonomías para caracterización y clasificación de edificios que se han desarrollado a lo largo de las últimas décadas. Como describen Pittore et al. (2016), la toma de datos puede realizarse para todos los elementos expuestos (*full enumeration*) o para una muestra representativa que después se extrapola (*statistical sampling*). En cualquier caso, parece evidente que esta toma de datos expofeso para un estudio de riesgo sísmico es una buena opción para caracterizar la vulnerabilidad del parque inmobiliario de una ciudad; no obstante, hay que valorar la ratio coste/beneficio de llevarla a cabo, puesto que una exhaustiva campaña de campo de vulnerabilidad conlleva un coste económico de personal cualificado muy alto y un tiempo de ejecución que no todos los estudios se pueden permitir (en el capítulo 6 se presentan datos al respecto proporcionados por expertos internacionales en la materia).

⁴ <https://www.sge2.com/>

⁵ https://www.mtptc.gouv.ht/accueil/actualites/article_7.html

Además, la elaboración de estas bases de datos se complica si se considera el carácter dinámico del entorno urbano (Wieland y Pittore, 2017), que cambia y se expande rápidamente, especialmente en países emergentes. Según cifras del Banco Mundial, se estima que en 2030 el 60% de la población mundial vivirá en entornos urbanos. Generalmente, tasas altas de crecimiento urbano dan lugar a asentamientos informales y/o no planificados, como demuestra el hecho de que un tercio de la población de países en desarrollo viva en poblados chabolistas¹. Esta realidad hace a las ciudades menos resilientes (Bilham 2009). Los asentamientos informales no suelen ser incluidos en registros oficiales, como el catastro o el censo, lo que dificulta a las autoridades gubernamentales locales las tareas de mantenimiento de las bases de datos oficiales de población y vivienda. Estas bases de datos han de contener información actualizada para poder ser de utilidad en estudios de riesgo sísmico y otras investigaciones.

El análisis de datos tomados por sensores remotos puede ser una herramienta muy valiosa para complementar las bases de datos oficiales y las campañas de campo anteriormente descritas, que ayuda a solventar estos problemas de actualización, coste y alcance. En los últimos años hemos presenciado un avance importante en las tecnologías de Observación de la Tierra (OT. La referencia en inglés es EO, *Earth Observation*), tanto en los tipos de sensores, como en las resoluciones espacial, espectral y temporal de los datos, y también en las técnicas de análisis. Tan importante como el avance tecnológico es la aparición de programas y normativas de distribución libre de los datos, a través de Infraestructuras de Datos Espaciales (IDES) o de geo-portales de consulta y descarga, ya sean nacionales o internacionales.

En estudios de exposición y vulnerabilidad sísmica suele emplearse una combinación de datos tomados por sensores remotos (SYNER-G D8.9, 2013), lo que obliga a trabajar con diferentes escalas y técnicas de análisis. Es común el empleo de imágenes de satélite multispectrales de media resolución espacial, con tamaños de pixel (GSD, *Ground Sampling Distance*) de varias decenas de metros, para hacer una primera aproximación a la caracterización de la trama urbana de la ciudad (e.g. Wieland et al., 2016). Un segundo nivel de análisis requiere más detalles en los datos para poder caracterizar los edificios de manera individual. Para ello, se emplean imágenes de satélite (e.g. Chaudhuri et al., 2015) o imágenes aéreas de alta resolución (GSD sub-métrico), tomadas desde avión y, más recientemente, incluso desde RPAs (*Remotely Piloted Aircraft*). E.g. Zhuo et al., 2018; He et al., 2019), como ha descrito recientemente Sameh Wahba, director de Gestión de Riesgos, Urbanismo y Desarrollo Territorial en el Banco Mundial⁶. También es común el uso de Radar (*RA*dio *D*etection and *R*anging) y de LiDAR (*LI*ght *D*etection and *R*anging) para incorporar la componente vertical al estudio, aunque éste último es menos frecuente.

La técnica de análisis de imagen más utilizada es la clasificación, que permite identificar los objetos de forma automática. En lugar de la tradicional clasificación por píxel, se tiende a realizar clasificación por objeto (OBIA, *Object-Based Image Analysis*) dado su alto potencial respecto a la anterior (Yan et al., 2006). La clasificación OBIA conlleva una fase previa de

⁶<https://www.observatorio-empresas.vodafone.es/articulos/administraciones-publicas/big-data-gestion-desastres-naturales-sameh-wahba/>

segmentación de imágenes (Blaschke et al., 2010) en un proceso automático que agrupa píxeles de valores similares creando regiones homogéneas. El umbral de similaridad de los píxeles y el tamaño de los segmentos resultantes son parámetros que ha de fijar el usuario, generalmente en un proceso de ensayo y error, hasta conseguir un resultado aceptable y acorde con sus objetivos. La clasificación consiste en identificar qué objetos son esos segmentos, a partir de una muestra recogida por el usuario y empleando técnicas de aprendizaje automático (u otras técnicas más tradicionales, como el aprendizaje basado en reglas). Al utilizar segmentos, se puede aportar más información al clasificador que si trabaja con píxeles únicamente, ya que, además de la información espectral, cada segmento tiene indicadores geométricos (derivados de su forma y tamaño) y de la textura formada por los píxeles internos (Haralick et al., 1973). El resultado es una imagen nueva, formada por polígonos (segmentos) con una etiqueta que indica si son edificios, vegetación, carretera, masa de agua, etc.

El tratamiento de los datos Radar o LiDAR está orientado, principalmente, a obtener información de la altura de los edificios y otros elementos urbanos. Para ello, se generan Modelos Digitales del Terreno (MDT) y se estiman las alturas de los objetos (edificios, árboles...), obteniendo Modelos Digitales de Superficie Normalizados (MDSn, conocidos en inglés como nDSM).

La combinación de todos estos resultados permite identificar y caracterizar los edificios individualmente, con información acerca de su localización, su geometría, algunos materiales y la configuración de su entorno (área ocupada por jardines, densidad de edificación, presencia de aparcamientos, configuración geométrica de los viales, etc.). Hay otros atributos de los edificios que no pueden obtenerse por observación directa, sino que se recurre a técnicas de inferencia (SYNER-G D8.9, 2013). A partir de los atributos observables, se crean modelos matemáticos de predicción en un conjunto de muestras que permiten inferir otros atributos, como el sistema estructural o los materiales constructivos de la estructura y los cerramientos. Todos estos atributos forman la base de datos de exposición, en la que se basa la asignación de vulnerabilidad.

Normalmente se dispone de series temporales de datos, tanto de imágenes ópticas como de Radar y LiDAR. Esto permite realizar análisis de control de cambios por comparación de los resultados obtenidos en distintas fechas, manteniendo así las bases de datos de exposición actualizadas, que es uno de los retos antes mencionados.

Recientemente, otras técnicas como los *Mobile Mapping Systems* (MMS) y los *Building Information Modelling* (BIM) están siendo incorporadas también a los procesos de caracterización de edificios en estudios de riesgo sísmico y en evaluación de daños post-sismo (e.g. Zeibak-Shini et al., 2016; Ma et al., 2015; Brumana et al., 2018). Los MMS son sistemas integrales de toma de datos que registran imagen panorámica (360º) y LiDAR, y permiten crear modelos tridimensionales (Arranz, 2013). En estos modelos, es posible identificar objetos, caracterizarlos y almacenarlos en verdadera posición y magnitud. Cuando estos modelos se realizan de edificios individuales, se generan los BIM. En el futuro, sería deseable que todos los

BIM que se están creando ahora puedan ser incorporados a las bases de datos de exposición de las ciudades.

Estas técnicas no pretenden, por sí solas, dar una solución holística a la tarea de crear y mantener bases de datos de exposición con información que pueda ser útil para caracterizar la vulnerabilidad de un parque inmobiliario. En esta tesis se plantean como una vía más para generar datos útiles, que se debe combinar con otras bases de datos ya existentes y con las campañas de campo. El objetivo final es conseguir generar la información de forma óptima, es decir, minimizando el coste, el tiempo y los recursos. En muchos países existen bases de datos oficiales (catastrales, censales, cartográficas) que pueden y deben incorporarse a estos estudios como información auxiliar o complementaria (*ancillary data*). Incluso existen otras fuentes de datos abiertas, como OpenStreetMap o GoogleEarth, de las que se pueden extraer imágenes para hacer fotointerpretación o capas vectoriales de elementos urbanos ya creadas y a disposición de los usuarios.

Respecto a las campañas de campo, probablemente nunca podrán ser sustituidas completamente por una toma de datos por medio de sensores remotos. Las inspecciones in-situ de los edificios son necesarias, ya sea para adquirir conocimiento de experto sobre los métodos de construcción locales (*expert knowledge*), que es esencial a la hora de tomar decisiones en este tipo de estudios, o para generar las muestras de entrenamiento y validación de los modelos de predicción. En la sección 7 se hace una propuesta de procedimiento combinado de EO y campaña de campo para la generación de bases de datos de exposición y vulnerabilidad optimizando las actividades y los recursos.

Es importante señalar que en esta tesis se propone mantener la base de datos de exposición, con todos los atributos posibles sobre los edificios, independiente de la consiguiente asignación de vulnerabilidad sísmica. Es decir, la base de datos de exposición es un resultado muy valioso, por sí solo, dentro del proceso de estimación del riesgo. De esta manera se facilitan las tareas de mantenimiento y actualización y se optimiza su aplicación, ya que se puede utilizar para asignar la vulnerabilidad de los edificios ante terremotos, o ante cualquier otro peligro natural. Esto le confiere el carácter multi-riesgo o multi-amenaza que reclama el Marco de Sendai.

1.3 Motivación y alcance

Todo lo expuesto anteriormente pone de manifiesto que las ciudades sometidas a un nivel de actividad sísmica tal que pueda causar daños y pérdidas considerables, se encuentran ante el reto de crear y/o mantener sus bases de datos de exposición para ser utilizadas en estudios de riesgo sísmico. Ante este reto se concibió la presente tesis, con la motivación de dotar a las entidades encargadas de la gestión del riesgo sísmico de un procedimiento eficaz y sostenible para la elaboración y actualización de estas bases de datos, de manera que puedan dar cumplimiento a los objetivos establecidos en el Marco de Sendai.

Otros autores están trabajando en esta línea, haciendo sus propuestas particulares. Este tipo de procedimientos debería cumplir una serie de requisitos:

1. *Sobre la vigencia y el acceso a los datos:* que los datos que se empleen se actualicen con una periodicidad acorde al ritmo de crecimiento o desarrollo de las ciudades. Es conveniente que en dichas actualizaciones se registren tanto las ampliaciones oficiales de la estructura urbana como los asentamientos informales. Por supuesto, estos datos deberían ser de libre acceso para estudios de riesgo sísmico (y por extensión, para cualquier otro tipo de riesgo y acción sin ánimo de lucro), cuyo fin último es proteger a la población de las consecuencias adversas de los terremotos, generando así comunidades resilientes.
2. *Sobre las campañas de campo:* que el procedimiento ayude a minimizar los costes de las campañas de campo para toma de muestras. Ello se consigue, bien reduciendo el número de campañas (su frecuencia), o bien optimizando la planificación de las mismas.
3. *Sobre los actores:* que permita incorporar al estudio el conocimiento de los expertos de la zona, dado que se trata de información muy valiosa y, además, permite adaptar el procedimiento a diferentes escenarios. La transferencia de los resultados debería satisfacer las expectativas de los gestores del riesgo. Para ello, las decisiones de nivel de detalle, escala o nivel de agregación de los resultados deberían ser tomadas conjuntamente con los receptores, así como las de incertidumbres asumibles.
4. *Sobre la replicabilidad del procedimiento:* que sea escalable para poder aplicarse en cualquier ciudad. Para ello, el procedimiento debe ser flexible, ya que cada escenario tiene sus particularidades que introducen variaciones, tanto en los datos como en los métodos (Torres et al., 2019b).
5. *Sobre los costes:* que el coste del procedimiento sea asequible para los gobiernos a todos los niveles, central, regional y local, ya que de la coordinación de todos ellos depende la realización de estudios de riesgo sísmico y el consiguiente diseño de medidas DRR.

El desarrollo de este procedimiento integral para estimación de la exposición y asignación de la vulnerabilidad sísmica es de particular interés en países en vías de desarrollo, puesto que se dan tres factores adversos muy relevantes: (1) la autoconstrucción y los asentamientos informales son más frecuentes; (2) la carencia de datos y recursos es más acusada; y (3), precisamente, es en estos países donde se presentan mayores amenazas, dando lugar a los mayores desastres naturales. En el escenario de Puerto Príncipe de esta tesis se dan los tres factores. Esto permitirá evaluar la aplicabilidad del método propuesto en casos de particular dificultad. Como se ha mencionado, la aplicación de estos estudios en países en desarrollo es una de las prioridades establecidas en el Marco de Sendai.

En línea con otro de los objetivos del Marco de Sendai, para el escenario de Lorca, se cuenta con datos oficiales de libre acceso del Instituto Geográfico Nacional y de la Dirección General del Catastro. Su uso permitirá valorar el papel de las bases de datos abiertas en este tipo de estudios, así como las posibilidades de escalado del procedimiento a otras ciudades, de cara a la futura implementación de los planes locales de emergencia ante el riesgo sísmico.

1.4 Contexto y antecedentes

La presente tesis se ha enmarcado en un flujo de trabajo que comprende los siguientes hitos:

- Realización del Trabajo Fin de Máster (TFM) en 2012, titulado: *Estimación del riesgo sísmico en Puerto Príncipe (Haití)*. Este TFM pertenece a la titulación de Máster en Ingeniería Geodésica y Cartografía de la E.T.S.I. Topografía Geodesia y Cartografía de la UPM, y tiene carácter investigador. En esta investigación, la doctoranda calibró un conjunto de espectros de capacidad para tipologías edificatorias típicas de Puerto Príncipe empleando datos de daño del terremoto de 2010. Además, probó la aplicabilidad de este método híbrido realizando un cálculo completo de riesgo sísmico determinista en la ciudad. El TFM obtuvo una mención de honor en la XV edición del Premio Internacional Francisco Coello en 2014. Fruto de este trabajo, se publicó el siguiente artículo en revista científica:
 - *Using the damage from 2010 Haiti earthquake for calibrating vulnerability models of typical structures in Port-au-Prince (Haiti)*. S. Molina, Y. **Torres**, B. Benito, M. Navarro, D. Belizaire. *Bulletin of Earthquake Engineering*. August 2014, Vol 12, Issue 4, pp 1459-1478 (**Q1 en 2014, Factor de Impacto 1.884**).
- Serie de proyectos SISMO-HAITI I-IV. Desde la ocurrencia del terremoto de Haití en 2010, y hasta el año 2015, la Universidad Politécnica de Madrid financió una serie de cuatro proyectos en sucesivas convocatorias de Subvenciones y Ayudas para Acciones de Cooperación Universitaria para el Desarrollo. Los proyectos estaban centrados en estudiar la peligrosidad y el riesgo sísmico de la región haitiana, y en crear capacitación para hacer posible y sostenible una política de reducción de riesgo de desastres. Relacionadas directamente con esta tesis, se llevaron a cabo una serie de acciones concretas en las que la doctoranda participó activamente: medidas del periodo natural de vibración del suelo para una propuesta de zonificación sísmica en Puerto Príncipe; estudio de las fuentes sismogénicas de la región; cálculo de la peligrosidad sísmica de la Española; campañas de campo en colaboración con ingenieros locales para recogida de información sobre edificios en Puerto Príncipe y en Cabo Haitiano; estudio de fuentes catastrales y censales de ambas ciudades; cálculo de escenarios de riesgo sísmico deterministas; cursos de capacitación en materia de ingeniería sísmica (peligrosidad, exposición, vulnerabilidad y riesgo) de una semana de duración a ingenieros locales⁷. Los proyectos fueron realizados por un equipo multidisciplinar de investigadores integrado por miembros de las siguientes instituciones: UPM, Observatorio Nacional de Medioambiente y Vulnerabilidad de Haití (ONEV), Universidad Complutense de Madrid (UCM), Consejo Superior de Investigaciones Científicas (CSIC), Universidad de Alicante (UA), Universidad de Almería (UAL), Red Sísmica de Puerto Rico (SRSPR), Centro Sismológico de Rep. Dominicana (UASD) y la Universidad de Kanagawa (Japón). Fruto de la investigación desarrollada en estos proyectos, se publicaron los siguientes artículos:

⁷ <http://proyectos.topografia.upm.es/cursosismohaiti/>

- *A first approach to earthquake damage estimation in Haiti. Advises to minimize the seismic risk.* Y. **Torres**, S. Molina, S. Martínez-Cuevas, M. Navarro, J.J. Martínez-Díaz, B. Benito, J.J. Galiana-Merino, D. Belizaire. *Bulletin of Earthquake Engineering*. January 2016, Vol 14, Issue 1, pp 39-58 (**Q2 en 2016, Factor de Impacto 1.899**)
- *Aplicación SIG en riesgos naturales: calibración del modelo de cálculo del riesgo sísmico de Puerto Príncipe (Haití).* Y. **Torres**, S. Molina, B. Benito. *Mapping*. June 2013, Vol. 22. Nº159. pp. 46-55.
- *SISMO-HAITÍ: proyecto de cooperación para el cálculo de la peligrosidad y el riesgo sísmico en Haití.* Y. **Torres**, B. Benito, D. Belizaire, Grupo de trabajo Sismo-Haití. *Proceedings del 4CNIS 2011, Granada*. ISBN: 978-84-15261-27-8. Editorial: Copicentro. 111pp. <http://datos.bne.es/edicion/a5091892>.
- Contribución a la base de datos global de acceso libre CEQID (*Cambridge Earthquake Impact Database*)⁸: los resultados de los artículos Molina et al. (2014) y Torres et al. (2016), antes mencionados, han sido incluidos en la base de datos de datos CEQID, implementada y gestionada por el Cambridge Architectural Research Ltd. de la Universidad de Cambridge.
- Beca/Contrato de Personal Investigador en Formación (PIF): durante la duración de los proyectos SISMO-HAITI, la UPM concedió a la doctoranda una Beca/Contrato PIF de cuatro años de duración (2012-2016). En el seno de esta Beca/Contrato, la doctoranda realizó dos estancias de investigación: (1) tres meses (octubre-diciembre de 2013) en el centro GFZ-Potsdam, financiada por la UPM y por el Servicio Alemán de Intercambio Académico (DAAD); (2) un mes (junio de 2014) en NORSAR (*Norwegian Seismic Array*, Noruega), financiada por NERA (*Network of European Research Infrastructure for Earthquake Risk Assessment and Mitigation*). El objetivo de las estancias fue investigar en las técnicas de análisis de imagen aplicadas a la estimación de la exposición y la vulnerabilidad sísmica. Fruto de la estancia en GFZ-Potsdam se firmó un Memorandum of Understanding (MoU) entre la UPM y el GFZ-Potsdam, y se publicó un artículo en revista científica:
 - *Object-based urban structure type pattern recognition from Landsat TM with a Support Vector Machine.* Wieland, M., **Torres**, Y., Pittore, M., & Benito, B. (2016). *International Journal of Remote Sensing*, 37(17), 4059-4083. (**Q2 en 2016, Factor de Impacto 1.724**).
- Proyecto MERISUR: Metodología para la evaluación Efectiva del Riesgo Sísmico Urbano. Durante los años 2014-2017, se llevó a cabo este proyecto, financiado por el Programa Nacional de I+D+I orientado a los Retos de la Sociedad del Ministerio de Economía y Competitividad. Ref.: CGL2013-40492-R. En este proyecto se buscó innovar en las fases de estimación del riesgo sísmico a escala local. Como parte de los objetivos de esta tesis, se desarrollaron tareas de integración de datos LiDAR y ortofotos para evaluar la exposición y la vulnerabilidad de la ciudad de Lorca. Fruto de este trabajo, se publicaron los siguientes artículos:

⁸ <http://ceqid.org/CEQID/Home.aspx>

- *Integration of LiDAR and multispectral images for rapid exposure and earthquake vulnerability estimation. Application in Lorca, Spain. **Torres, Y., Arranz, J.J., Gaspar-Escribano, J.M., Haghi, A., Martinez-Cuevas, S., Benito, B., Ojeda, J.C. (2019) International Journal of Applied Earth Observation and Geoinformation. 81, 161-175. DOI: <https://doi.org/10.1016/j.jag.2019.05.015> (Q1 en 2017, Factor de Impacto 4.433)***
 - *Evaluación de los modificadores por comportamiento que controlan el daño en los terremotos. Aplicación al terremoto de Lorca. Martínez-Cuevas, S.; Morillo, M.C., Cervera, J.; Benito, B.; San Millán Escribano, M.; Gaspar-Escribano, J.; **Torres, Y.** Proceedings en el congreso internacional sobre Intervención en Obras Arquitectónicas tras Sismo: L' Aquila (2009), Lorca (2011) y Emilia Romagna (2012), 13-14 mayo 2015, Murcia. ISBN: 978-84-92986-89-7. pp 237-256. Editorial: Universidad Católica San Antonio de Murcia. 616 páginas.*
- Estancia de investigación de tres meses (junio-agosto 2019) en la Facultad de Ciencias de la Universidad de Alicante. El objetivo de esta estancia fue validar la definición de tipologías constructivas de Puerto Príncipe empleadas en la última fase de la tesis, realizar un último cálculo de riesgo sísmico comparativo y entablar foros de discusión sobre el estado del arte de esta temática y el posicionamiento de la tesis en relación al mismo.
 - Dirección de dos TFM de alumnos de la E.T.S.I. Topografía Geodesia y Cartografía de la UPM relacionados con la tesis. En 2015-2016, la doctoranda co-dirigió el TFM titulado “Elaboración de una metodología de actuación y documentación para el estudio de vulnerabilidad sísmica a partir de imágenes y datos LiDAR”, de la alumna Azadeh Haghi, que estaba enmarcado en el proyecto MERISUR. En 2018-2019, ha dirigido el TFM titulado “Aplicación de técnicas de aprendizaje automático para la estimación de la vulnerabilidad sísmica en Puerto Príncipe (Haití)” del alumno Álvaro Arredondo, que ha desarrollado un software que integra el análisis completo de imágenes y LiDAR del escenario de Puerto Príncipe.
- De la posterior comparación de ambos escenarios, se ha publicado el siguiente artículo de *Proceedings* en congreso internacional:
- *The new methods for rapid exposure and seismic vulnerability assessment. How do they adapt to different scenarios? **Torres, Y.** Arranz, J.J., Martinez-Cuevas, S., Gaspar-Escribano, J.M., Benito, B., Staller, A., Gómez, I. In Multidisciplinary Digital Publishing Institute Proceedings (Vol. 19, No. 1, p. 14).*

2 Estado del arte

A continuación, se presenta una revisión de los trabajos más relevantes de los últimos años sobre taxonomías de clasificación de edificios -como elementos expuestos- y sobre los modelos de vulnerabilidad que pueden asignarse a esos edificios. La revisión de la literatura se ha centrado, principalmente, en Europa y América, que son las regiones donde se encuentran los dos escenarios elegidos para el análisis en esta tesis.

Asimismo, se presenta una revisión de las técnicas de Observación de la Tierra aplicadas a la estimación de la exposición y la vulnerabilidad sísmica, en un enfoque comparativo con la aplicación que se realiza en esta tesis.

2.1 Modelos de exposición

La evaluación de la vulnerabilidad sísmica de los edificios en una zona de estudio se realiza a partir de la información contenida en la base de datos de exposición (*exposure database*), también llamada inventario de la edificación (*building stock inventory*). Es una base de datos espacial, que contiene un registro para cada edificio y un número variable de campos, que son atributos relacionados con la vulnerabilidad. Campos habituales son la localización (absoluta y/o relativa), la geometría, la altura y/o número de plantas, los materiales de construcción o el tipo de tejado.

Existen diferentes taxonomías que presentan una propuesta sobre qué atributos de los edificios afectan a su comportamiento cuando se ven sometidos a una sacudida sísmica y, por tanto, son los que debe recoger una base de datos de exposición. La primera taxonomía moderna orientada a caracterizar y clasificar los edificios según su desempeño sísmico es la propuesta por el *Applied Technology Council* de California en la década de los ochenta (ATC-13, Rojahn y Sharpe, 1985). El sistema de clasificación del ATC-13 describe 78 tipos de estructuras edificatorias (40 edificios y 38 estructuras de otro tipo, como puentes o depósitos de agua) y también considera el método constructivo, aspectos de forma y altura. Rápidamente se extendió y se convirtió en un estándar del país, cuando la FEMA basó en este sistema la definición de una nueva clasificación de los edificios de EE.UU., a los que denominó *Model Building Types* (MBT), para la publicación de una serie de documentos sobre diseño antisísmico en 1988 (FEMA-154, 1988; FEMA-154, 2002). Aunque las MBT se definieron para describir construcciones típicas de esta región norteamericana, lo cierto es que son utilizadas en otros muchos países del mundo para realizar estudios de riesgo sísmico (Brzev et al., 2013). La taxonomía actual de la FEMA-154 contiene un total de 15 MBT (2 de madera, 5 de acero, 3 de hormigón, 2 prefabricados y 3 de mampostería) que consideran los materiales, método constructivo, área, uso y altura. En 2003, la FEMA presentó HAZUS (HAZUS-MH, 2003), una metodología completa de estimación del riesgo sísmico con una nueva clasificación de edificios basada en el sistema de FEMA-178 (1992), que contiene 15 clases: 1 de madera, 5 de acero, 3 de hormigón, 2 de hormigón prefabricado, 2 de mampostería reforzada, 1 de mampostería sin reforzar y 1 de casas móviles. Estas clases se subdividen según el número de plantas.

Recientemente, también en América, Yepes-Estrada et al. (2017) han presentado un modelo de exposición para los edificios típicos de siete países andinos en el marco del proyecto SARA (*South America Risk Assessment*), liderado por la iniciativa internacional *Global Earthquake Model* (GEM). Para caracterizar cada edificio consideran atributos como los materiales de construcción, la altura, el área, el valor económico, o los ocupantes. Las fuentes de datos principales que proponen los autores para crear estas bases de datos de exposición son los censos estadísticos de vivienda y población.

En Europa, encontramos las primeras clasificaciones de edificios en las escalas empíricas de intensidad macrosísmica. La escala vigente más utilizada en Europa (que es también de aplicación internacional) es la *European Macroseismic Scale*, EMS-98 (Grünthal, 1998). Es una evolución de las escalas MSK-64 (Medvedev et al., 1967) y EMS-92 (primera versión de la Escala Macrosísmica Europea, creada por G. Grünthal, R. M. W. Musson, J. Schwarz y M. Stucchi en Potsdam, en junio de 1992). La EMS-98 presenta, además de una clasificación de edificios según su tipología y su vulnerabilidad sísmica, definiciones de grados de intensidad macrosísmica (en 12 grados) y grados de daño (desde el 1 –sin daño– al 5 –destrucción).

La clasificación de vulnerabilidad sísmica de la EMS-98 se propuso en función de la respuesta que los edificios habían tenido ante numerosos terremotos registrados a lo largo de la historia. Para una misma intensidad de la sacudida sísmica en diferentes partes del mundo, se evaluaron los daños que habían sufrido los edificios y se clasificaron como más o menos vulnerables, desde la A a la F. Se consideraron 15 tipos de estructuras, que son las que permite clasificar esta escala: 7 de mampostería, 6 de hormigón armado, 1 de acero y 1 de madera. Por tanto, el primer dato necesario para poder utilizar este método de clasificación de edificios es el material de construcción de su estructura. La escala indica que edificios del mismo tipo, es decir, contruidos con los mismos materiales, pueden tener vulnerabilidades ligeramente distintas. En la Tabla de Vulnerabilidad, la escala presenta la clase más probable a la que pertenece cada tipo de edificio (representada por un círculo) e incluye además un margen de error (representado por una línea) que indica otras clases plausibles en un rango verosímil. Esta variabilidad en la asignación de vulnerabilidad es realista, y se debe a que existen una serie de modificadores que hay que tener en cuenta: calidad de los materiales de construcción y del proceso de fabricación, estado de preservación, regularidad en planta y en alzado, ductilidad (capacidad de sufrir pequeños daños controlados en la fase post-elástica que disipan energía del terremoto y evitan daños mayores), localización relativa dentro de la manzana, elementos de refuerzo y nivel de diseño sismorresistente. La incorporación de estos modificadores al análisis permite, de manera subjetiva, afinar la clase de vulnerabilidad que a priori se asigna a cada edificio.

Entre 1999 y 2004 se desarrolló en Europa el proyecto Risk-UE (Mouroux et al., 2004; Mouroux y Le Brun, 2006) con el objetivo de crear una metodología modular para calcular escenarios de riesgo sísmico, siguiendo la línea de otras iniciativas similares precedentes, como HAZUS (2003) y RADIUS (1999). Los escenarios de riesgo se calcularon a nivel local en 7 ciudades europeas: Barcelona, Bitola, Bucarest, Catania, Niza, Sofia y Tesalónica. Dos de los objetivos más relevantes del proyecto Risk-UE eran: (1) integrar y homogeneizar todos los proyectos de

riesgo sísmico desarrollados en Europa hasta la fecha; y (2) identificar las características típicas de las construcciones europeas (no solo edificios de viviendas, sino también monumentos e infraestructuras vitales) y la filosofía de sus sistemas de organización y respuesta ante terremoto. En este proyecto se diseñaron dos métodos de cálculo del riesgo, LM1 y LM2, que conllevan sendas evaluaciones de vulnerabilidad. Una está basada en la definición de clases de la EMS-98 (LM1) y otra en la definición de una nueva taxonomía de clasificación de edificios propia de Europa (LM2). Esta clasificación de tipologías (Milutinovic y Trendafiloski, 2003) presenta 23 tipos de edificios: 10 de mampostería (8 sin reforzar y 2 reforzada o confinada), 7 de hormigón armado (2 de ellas prefabricado), 5 de acero y 1 de madera. Se distinguen por el tipo de sistema estructural y los materiales constructivos; adicionalmente, se subdividen en rangos de altura, hasta un total de 65. En esta clasificación de edificios de Risk-UE destaca un mayor número de edificios de mampostería que en la clasificación de HAZUS (10 frente a 3), dada la extendida presencia de esta tipología en Europa. Posteriormente, Lagomarsino y Giovinazzi (2006) propusieron un avance sobre los dos métodos de evaluación de vulnerabilidad propuestos por Risk-UE.

El proyecto colaborativo europeo SYNER-G (SYNER-G D1.1, 2009) se desarrolló entre 2009 y 2012, con el objetivo de unificar los métodos y herramientas desarrollados hasta la fecha para la evaluación de la vulnerabilidad sísmica en Europa, considerando no solo los elementos aislados (edificios, líneas vitales e infraestructuras), sino el sistema urbano en que se encuentran y las relaciones con otros sistemas. Propusieron nuevas taxonomías para clasificación de edificios de hormigón armado (SYNER-G D3.1, 2011) y mampostería (SYNER-G D3.2, 2011), que son los más frecuentes en Europa. Los atributos que consideran estas taxonomías son: el sistema estructural y sus materiales, irregularidades en planta y alzado, elevación, revestimiento, solera, tejado, altura y diseño antisísmico.

Aunque está fuera de las regiones de aplicación de esta tesis, merece la pena citar la iniciativa local de Lang et al. (2017) porque, aun teniendo un enfoque similar, incorpora novedosas variables relacionadas con la vulnerabilidad de los edificios. Los autores han trabajado en cinco países de las regiones de Asia central y Asia del sur, tratando de identificar los distintos tipos de construcciones con el objetivo de crear una taxonomía manejable y unificada de clasificación de edificios en 29 tipos. Los datos que han empleado comprenden tanto una base de datos catastral global de la región de estudio como campañas de campo en los distintos países para inspección de edificios. En la definición de su taxonomía proponen considerar, no solo los atributos relacionados con el sistema estructural (combinación de los materiales de construcción y el método constructivo) y la configuración geométrica del edificio (altura, área, etc.), sino también otros relacionados con la disponibilidad de materiales, conocimientos tradicionales de construcción, estado de desarrollo de la región, condiciones socio-económicas, condiciones climáticas e historia de las amenazas acaecidas en la zona. Resulta interesante analizar esta lista de factores que efectivamente influyen en el modo de construir y en la calidad de las construcciones.

Partiendo de los conocimientos adquiridos durante décadas de trabajo en regiones localizadas, y con la intención de aprovecharlos y extenderlos, se están desarrollando iniciativas globales

para crear taxonomías de clasificación de edificios internacionales. La *World Housing Encyclopedia* (WHE, 2000) es un proyecto promovido por el *Earthquake Engineering Research Institute* (EERI) y la *International Association for Earthquake Engineering* (IAEE) que se inició en enero del año 2000. Voluntarios de todo el mundo hacen sus aportaciones para generar informes (con fotografías) sobre los diferentes tipos de construcciones. Actualmente, la WHE cuenta en su base de datos con unos 130 informes y cubre 43 áreas sísmicamente activas.

Más tarde, el *U.S. Geological Survey* (USGS) se propuso crear una base de datos global de tipologías edificatorias dentro de su programa *PAGER* (*Prompt Assessment of Global Earthquakes for Response*) destinado a estimar el impacto, en términos de víctimas, de los terremotos en tiempo real y en cualquier parte del mundo. La base de datos *PAGER* (Jaiswal y Wald, 2008) contiene una distribución de tipos de edificios en cada país, su uso y sus horarios de ocupación, diferenciando entre entornos urbanos y rurales. Ha sido creada a partir de la integración de datos depurados de distintas partes del mundo, aportados por expertos locales y por organizaciones internacionales, como WHE o UN-HABITAT. La última actualización data de abril de 2014 y actualmente cuenta con 105 tipos de edificios (*PAGER inventory database v2.0*). Es de libre distribución y accesible a través de internet⁹.

La base de datos empírica *CEQID* (*Cambridge Earthquake Impact Database*), creada por el *Cambridge Architectural Research Ltd* (CAR), recoge datos de terremotos de ámbito internacional. Incorpora datos de 76 eventos que han dado lugar a la definición de 68 tipos de estructuras edificatorias (que pueden subdividirse según altura y año de construcción). (Spence et al., 2011). Estas tipologías estructurales finalmente pueden relacionarse con las clases de vulnerabilidad de la escala EMS-98.

El grupo internacional GEM introdujo en 2012 una taxonomía (GEM v2.0) que pretende resolver los inconvenientes que presentan las taxonomías desarrolladas hasta la fecha (Brzev et al., 2013). Estudiaron de manera comparativa diez clasificaciones y elaboraron un ranking en función de criterios relacionados con la sencillez, completitud, facilidad de uso y alcance geográfico, entre otros. El ranking de clasificaciones ha sido actualizado y completado en esta tesis, y se muestra en la Figura 2-1 (algunas taxonomías subdividen, posteriormente, el número de tipologías -Nr. MBT- en rangos de altura). Consideran que algunas taxonomías son vagas o demasiado sencillas (FEMA-154, EMS-98), tienen alcances locales (ATC-13, FEMA-154, HAZUS, EMS-98) o presentan una cierta falta de rigor (WHE). La taxonomía GEM v2.0, que ha sido validada por el EERI (*Earthquake Engineering Research Institute*), nació con la intención de cumplir una serie de objetivos: que sea completa a la vez que sencilla, que resulte familiar a todos los usuarios y que sea extensible a otros peligros naturales y estructuras diferentes de edificios (como puentes o presas). Comprende 13 atributos sobre los edificios que afectan a su desempeño sísmico: dirección, material del sistema estructural, altura, fecha de construcción o del refuerzo sísmico (si lo hubiere), uso, posición relativa en la manzana, forma en planta, irregularidad estructural, paredes exteriores, tejado, solera y cimientos. Algunos de estos atributos tienen más de un nivel de valores. Por ejemplo, el tejado admite forma (nivel 1),

⁹ https://pubs.usgs.gov/of/2008/1160/downloads/PAGER_database/

material de la cobertura (nivel 2), material del sistema constructivo (nivel 3), sistema constructivo (nivel 4) y conexión con los paramentos verticales (nivel 5). Otros, como la orientación, solo tienen un nivel. Esta taxonomía viene acompañada de recursos complementarios, como la herramienta TaxT para su uso en una plataforma informática, que permite al usuario crear un código único de cada tipo de edificio después de haber ido seleccionando los valores de los atributos que vienen en listas desplegables. Este código es de tipo *string*, y está compuesto por el valor de todos los atributos separados por “/” (*slash*). El código también puede crearse a mano, siguiendo las instrucciones que proporcionan los autores. Un aspecto importante es que permite definir restricciones a la hora de combinar los atributos, de forma que no puedan darse combinaciones inaceptables como rascacielos contruidos en madera.

	Nr. MBT	1. Differentiates seismic performance	2. Observable	3. Complete	4. Simple and collapsible	5. Nearly exhaustive	6. Familiar	7. Treats non-buildings	8. Extensible to other hazards	9. User-friendly	10. International in scope	Score	Comments
ATC-13	78	1	1	1	1	0	2	2	0	2	0	10	California focused
FEMA-154 (ATC-21)	15	1	1	0	1	0	2	0	0	2	0	7	For US construction
EMS-98	15	1	1	0	1	0	2	0	0	2	2	9	Too broad
WHE	130	1	1	2	1	2	2	0	0	2	2	13	More than structural
Coburn & Spence (2002)	13	1	1	1	2	2	2	0	0	2	2	13	Both engineered and non engineered bdngs.
HAZUS	15	2	1	1	1	0	2	0	2	2	0	11	For US construction
Gunel and Ilgin (2007)	6	1	1	1	1	0	1	0	0	2	0	7	Tall buildings only
CEQID	68	2	1	1	1	1	2	0	0	1	2	11	EQ damage database
PAGER -STR	105	2	1	2	2	2	2	0	0	2	2	15	Most comprehensive to date
SYNER-G	-	2	1	2	2	2	1	2	1	2	2	17	Best potential
LM2 Risk-UE	23	2	1	1	1	0	2	2	0	2	0	11	For European construction
SARA	26	2	2	2	1	0	2	0	0	1	0	10	For seven Andean countries
Lang et al. (2017)	29	2	1	1	1	0	2	0	0	2	0	9	For Central and South Asia

Notes: true = 2 points, somewhat true = 1 point and untrue = 0 points.

Figura 2-1. Comparación de varias taxonomías de clasificación de tipologías edificatorias.
Actualizada a partir de la tabla publicada en GEM v2.0 (Silva et al, 2018)

En línea con el Marco de Sendai, las taxonomías para clasificación de edificios más recientes tienen un enfoque multi-riesgos, dada la clara separación de los conceptos exposición (común a todos los riesgos) y vulnerabilidad (particular para cada riesgo). La evolución de la taxonomía GEM v2.0 es GED4ALL (Silva et al, 2018), fruto de un proyecto para crear una base de datos de exposición multi-riesgo: terremotos, inundaciones, volcanes, vientos, tsunamis y sequías. Este proyecto ha sido desarrollado por el consorcio formado por GEM Foundation, ImageCat Inc. y

el equipo humanitario de OpenStreetMap (HOT), y financiado por el *Department for International Development of the United Kingdom*. La base de datos está pensada para almacenar información, no solo de los edificios, sino socio-económica, de infraestructuras, líneas vitales, cultivos, bosques y ganado. Respecto a los atributos, incluye los mismos que la GEM v2.0 (habiendo fusionado posición relativa en la manzana y forma en planta) y añade dos más, hasta un total de 14: sistema de protección contra incendios y características hidrodinámicas de la planta baja (este último orientado a inundaciones).

Actualmente, en el marco del proyecto RIESGOS¹⁰ (*Multi-Risk Information System for South America's Andes ReGiOnS*) se está trabajando en una taxonomía global y multi-riegos formulada a partir de las taxonomías GEM v2.0 y GED4ALL (comunicación personal de los autores). Introduce otros dos atributos nuevos con respecto a GED4ALL: condición física y dispositivos internos no estructurales. Además, vuelve a separar los atributos posición relativa del edificio en la manzana y forma en planta, haciendo un total de 17 ítems. Se trata de un sistema de clasificación facetado (*faceted taxonomy*), que presenta tablas para todos los atributos, en las que se especifica un código único para cada valor y los distintos niveles que pueden tener en función de la cantidad de información disponible. Esta taxonomía recomienda que se distingan claramente los enfoques de edificio individual y bloque de edificios (lo que se conoce como manzana) para caracterizar mejor cada ente por separado. A posteriori, si se desea, se podrán relacionar entre sí a través de la localización espacial para mejorar la comprensión del entorno urbano.

Como puede verse, existen diferentes taxonomías para clasificar los edificios con el objetivo de realizar una base de datos de exposición. La elección de la más adecuada dependerá de la zona de estudio (extensión y características edificatorias) y del enfoque y los objetivos del análisis (Martínez-Cuevas, 2014). Cada sistema clasificatorio conlleva el uso de una serie de datos concretos que será necesario conseguir, y la disponibilidad de estos o la posibilidad de generarlos también condiciona la elección. Las taxonomías más modernas permiten incorporar más atributos e incrementar el nivel de detalle en la caracterización de los edificios; pero esto también exige un trabajo más exhaustivo a la hora de recopilar toda la información, lo que conlleva una alta inversión en recursos y costes que es necesario valorar. En esta tesis, se propone adoptar la taxonomía más reciente y completa para caracterizar los edificios, que es la del proyecto RIESGOS.

2.2 Modelos de vulnerabilidad

Una vez generada la base de datos de exposición, la evaluación de la vulnerabilidad consiste en asignar a cada edificio un modelo de vulnerabilidad basado en sus atributos. El modelo de vulnerabilidad sísmica describe el comportamiento que el edificio tendrá en caso de terremoto. La elección de este modelo depende, principalmente, del método de cálculo de daño que se vaya a emplear en el análisis del riesgo sísmico. Aunque existen otras clasificaciones (una descripción amplia puede encontrarse en Quirós, 2017), los métodos de cálculo de daño por

¹⁰ <https://www.gfz-potsdam.de/en/section/seismic-hazard-and-risk-dynamics/projects/riesgos-multi-risk-information-system-for-south-americas-andes-regions/>

sismo generalmente se dividen en dos grandes grupos: métodos empíricos y métodos analíticos; ambos se combinan para dar lugar a métodos híbridos (Calvi et al., 2006). De ahí que al modelo de vulnerabilidad elegido para cada caso se suele llamar también modelo empírico o analítico. En esta tesis se considera que la vulnerabilidad es propia del edificio y, por tanto, es independiente del método de cálculo de daño en que se vaya a emplear. En consecuencia, no se harán referencias a vulnerabilidad empírica o analítica, sino a clasificación empírica o analítica de la vulnerabilidad.

Todos los métodos de cálculo de daño por sismo, ya sean empíricos o analíticos, tienen un objetivo común: estimar una distribución de los distintos grados de daño que sufrirán los edificios de un determinado nivel de vulnerabilidad ante un determinado nivel de movimiento del suelo. La principal diferencia entre ellos es la fuente de datos que emplean para estimar y describir el comportamiento sísmico de los edificios. Los métodos empíricos (o estadísticos) utilizan una base de datos de daños causados por terremotos reales ocurridos en el pasado, mientras que los analíticos (o mecánicos) emplean modelos numéricos de cada tipo de edificio para realizar un análisis estático no lineal o *pushover* (Lawson et al., 1994, ATC-40, 1996).

Los métodos empíricos comenzaron a desarrollarse en la década de los setenta, cuando Whitman et al. (1973) presentaron las conocidas matrices de probabilidad de daño (MPD). Para cada clase de vulnerabilidad, estas matrices expresan la relación directa entre el daño y el movimiento del suelo, dado en una escala discreta de intensidad. Para el uso de las MPD, los edificios deben agruparse en clases según una escala de vulnerabilidad, como la EMS-98 (e.g. Dolce et al., 2003).

Otra forma de caracterizar la vulnerabilidad en los métodos empíricos es a través del índice de vulnerabilidad, *I_v*. Este índice permite distinguir edificios de una misma tipología atendiendo a diferentes aspectos relacionados con su vulnerabilidad, como la separación con el edificio adyacente o la presencia de determinados elementos no estructurales. Es una medida cuantitativa que se normaliza entre 0 y 100, y es particular para cada edificio. El primer método de clasificación de edificios basado en el índice de vulnerabilidad fue propuesto por Benedetti y Petrini (1984). Desde entonces, se han realizado numerosas adaptaciones, como Lagomarsino y Giovinazzi (2006), Lantada (2007) o Feriche et al. (2012).

Los métodos empíricos presentan numerosas desventajas (e.g. Calvi et al., 2006; Lang, 2013), entre las que cabe destacar la subjetividad inherente a la toma de datos sobre los daños; la imposibilidad de aislar el efecto del suelo; la incertidumbre que se introduce al usar relaciones entre la intensidad (medida usual del impacto de terremotos ya registrados) y la aceleración del suelo; y la asimetría en la distribución de la intensidad, pues la mayoría de los terremotos que ocurren son de intensidades bajas.

Algunas mejoras se han conseguido con el desarrollo de las curvas de vulnerabilidad continuas (Calvi et al., 2006). Estas emplean como medida del movimiento fuerte la escala continua de intensidad *PSI* (*Parameterless Scale of Intensity*, de Spence et al., 1992), o la aceleración espectral o el desplazamiento espectral (Rossetto y Elnashai, 2003) calculados con modelos de predicción del movimiento fuerte del suelo (GMPEs).

Para contrarrestar las desventajas de los métodos empíricos, en la década de los noventa empezaron a desarrollarse los métodos analíticos. En estos métodos, se modela cada tipología de edificio, identificada por un código o acrónimo de una taxonomía, y se somete a un análisis *pushover*, aplicando distintos niveles de fuerza lateral (que simulan una sacudida sísmica). Calculando el desplazamiento lateral que se registra en la estructura con cada nivel de fuerza aplicado, se crean las llamadas curvas de capacidad, que relacionan la fuerza cortante en la base del edificio con el desplazamiento lateral del tejado (Freeman, 1998). Estas curvas permiten representar el desempeño sísmico de un edificio. Para simplificar el modelo, se relaciona el comportamiento de la estructura con el de un sistema equivalente de un grado de libertad (SDOF, *single degree of freedom*), y se obtiene la curva de capacidad del SDOF, que es el conocido como espectro de capacidad de la estructura. Estos espectros relacionan la aceleración espectral correspondiente a la fuerza cortante generada por el movimiento sísmico en la base del edificio (S_a) con el desplazamiento máximo espectral (S_d). Al cruzar este espectro de capacidad del edificio con el espectro de demanda sísmica, se obtiene el punto de desempeño, que viene dado por el desplazamiento máximo espectral que sufre el edificio a causa del terremoto representado en el espectro de demanda. Para calcular este punto de desempeño, se han desarrollado distintos métodos. Entre los más utilizados, destacan el método capacidad-demanda (CSM, *Capacity Spectrum Method*, ATC-40, 1996) y su evolución MADRS (*Modified Acceleration-Displacement Response Spectrum*; FEMA-440, 2005); los métodos basados en desplazamiento únicamente (DBM, *Displacement-Based Methods*, e.g. Calvi, 1999; Crowley et al., 2004), dado que es el desplazamiento el parámetro que presenta mejor correlación con los daños estructurales, y no la aceleración (Crowley et al., 2004); o el método coeficiente desplazamiento mejorado (I-DCM, *Improved Displacement Coefficient Method*; FEMA-440, 2005; Poland y Mitchell, 2007). Una vez determinado el punto de desempeño, el desplazamiento espectral correspondiente se utiliza para determinar la distribución de los distintos grados de daño a través de la función de fragilidad.

En los métodos analíticos, la vulnerabilidad de los edificios se indica a través del código o acrónimo antes mencionado. Por ejemplo, en el método analítico LM2 de Risk-UE (Milutinovic y Trendafiloski, 2003), el acrónimo M12L representa a todos los edificios de mampostería de piedra simple con 1 o 2 plantas de altura. Además, el acrónimo se complementa con una indicación del nivel de diseño antisísmico del edificio mediante los sufijos -pre, -low, -mod, -high, que indican, respectivamente, si el edificio carece de diseño sismorresistente (previo a cualquier código sísmico) o, si lo tiene, cuál es su nivel (bajo, moderado o alto). Este código, a su vez, lleva asociados un espectro de capacidad y una función de fragilidad. En esta tesis se emplea el término “funciones de daño” para hacer referencia al par de curvas [espectro de capacidad-función de fragilidad] (tomado de Molina et al., 2014).

Los métodos analíticos también presentan desventajas (Lang, 2013). Básicamente todas ellas derivan de la complejidad de realizar el análisis *pushover*, en referencia a la dificultad de modelar los edificios y al coste en términos de tiempo. Por ello, no hay funciones de daño para todas las tipologías de edificios definidas en las diferentes taxonomías. A la hora de realizar un análisis de riesgo sísmico en una determinada localización, es frecuente tener que recurrir a

funciones de daño creadas en otros países para tipologías similares, o a emplear métodos híbridos.

Los métodos híbridos se utilizan cuando no es posible el empleo de un método puramente empírico o analítico, por falta de datos, o para calibrar o comparar ambos métodos. Por ejemplo, Molina et al. (2014), emplearon un método híbrido en el que calibraban sus funciones de daño analíticas con datos empíricos del terremoto de Haití de 2010. Kappos (1998) y Barbat (1996) utilizaron el método analítico para generar funciones de vulnerabilidad asociadas a valores de intensidad para las que no tenían datos empíricos. Cuando se combinan ambos métodos, hay que tener en cuenta que la definición de los grados de daño no es única, sino que cada metodología tiene la suya propia. Hill y Rossetto (2008) hicieron una comparativa de estas definiciones de grados de daño, estableciendo equivalencias entre las propuestas por EMS-98, Risk-UE y HAZUS.

Finalmente, cabe mencionar que a la hora de realizar un análisis de riesgo sísmico empleando cualquiera de los métodos descritos, es necesario incorporar el juicio o la opinión de expertos. Hay decisiones que tomar en las que es necesario contar con la experiencia de ingenieros, arquitectos, geofísicos y analistas de datos. Esta experiencia se refiere, no solo a los conocimientos técnicos, sino a los conocimientos sobre el área de estudio: sismicidad de la zona, sistemas de vigilancia sísmica, prácticas constructivas, organismos locales involucrados en la gestión del riesgo, programas de prevención del riesgo, aspectos sociales y económicos, etc.

La decisión de qué método utilizar para estimar el daño por sismo depende, principalmente, de los datos disponibles para caracterizar los edificios (nivel de detalle y fiabilidad) y de si existen funciones de daño para las tipologías edificatorias de la zona de estudio o datos empíricos de sismos pasados en la región. El caso ideal, de acuerdo con Lang et al. (2012b), sería aplicar un método híbrido en el que se calibrasen las funciones de daños analíticas con datos empíricos, como realizaron Molina et al. (2014) en Haití.

2.3 Técnicas de teledetección aplicadas a la caracterización de edificios

Como se ha expuesto anteriormente, las técnicas de teledetección pueden contribuir a solucionar el problema de generar datos de exposición y vulnerabilidad de manera rápida y económica. Desde hace aproximadamente 15 años, se vienen empleando estas técnicas para extraer las huellas de los edificios y caracterizarlas con atributos relacionados con su vulnerabilidad sísmica (o su vulnerabilidad ante otros tipos de desastres naturales). Los datos tomados por sensores remotos son cada vez más variados, más precisos y accesibles a través de programas nacionales, como el Plan Nacional de Ortofotografía Aérea de España (PNOA); o internacionales, como COPENICUS (de la Unión Europea) o EarthExplorer (del Servicio Geológico de EE.UU.).

Tabla 2-1. Resumen comparativo de estudios destinados al análisis de vulnerabilidad sísmica con técnicas de teledetección en los últimos 10 años.
La tabla completa con más datos se presenta en el Anexo II.

Author	Data	Comments to the main topics addressed	Accuracy or reliability
Optical imagery			
Wieland et al. (2012a)	Time-series of Landsat TM and MSS; QuickBird; manual digitization.	Age of construction from Landsat time series analysis. LULC classification from OBIA of Landsat image. QuickBird image analysis for footprint delineation: segmentation + post-processing for over-segmentation solution + SVM for classification.	Accuracy of OBIA of Landsat is 81%. Footprints from QuickBird (built-up mask): Accuracy = 90%.
Dell'Acqua et al. (2013)	QuickBird image.	Design of BREC-4-GEM software for extracting the number of buildings.	Buildings are detected with accuracy around 90%.
Erlich et al. (2013)	QuickBird image; DB of damage assessment after the Haiti 2010 EQ; Ground truth footprint DB provided by the municipality.	Alpha-tree of multi-level segmentations. Building segments were selected using parameters related to shape, area, width, length, spectral rage and presence of shadow. Height from shadow.	Sana's: 78.86% of accuracy in built-up mask generation. Height not validated. Port Prince: nr. bldgs. estimated with $R^2=0.64$.
Optical imagery and 3D data			
Matsuka et al. (2012)	VHR panchromatic image; DSM; building DB from in-field survey; footprints available.	Roof type (flat/gable) from panchromatic band and DEM with already existing footprints. Building height using DSM and footprints already available.	Correlation between estimated vulnerability and ground truth data is 75%.
Polli & Dell'Acqua (2011)	QuickBird panchromatic band; SAR.	W-filter for footprint extraction from image. Analysis of two SAR images of the same building for height extraction.	Not mentioned
Mück et al. (2013)	QuickBird image; DEM from Radar; building DB from in-field survey; manual digitization.	QuickBird for OBIA and building extraction; DEM for height. Building vulnerability to earthquake and tsunami hazards is considered.	Accuracy of footprint comparison is 84 %. Building classification in testing phase: accuracy = 81%
Wieland et al. (2012b); Pittore & Wieland (2013)	Time-series of Landsat images; ground-based omnidirectional imaging; OSM footprints; building DB from in-field survey; manual digitization.	OBIA of images for extracting building data at both, block and building levels. Height is extracted from 360° images. Nr. buildings estimated through the building density per stratum.	Accuracy of OBIA of built-up mask creation ranges between 80%-90%. Validation of height not provided. Visual validation of MBT classification.
Geiss et al. (2015)	IKONOS VNIR; Time-series of Landsat TM and ETM+; nDSM from Radar; vulnerability/damage DB from in-field survey.	Calculation of features at both, building and block level. Multi-temporal analysis of Landsat images to extract the period of construction. Vulnerability allocation is done following a (1) scoring method (expert knowledge), (2) the EMS-98 scale, and (3) MBT definition. Damage calculation is performed using fragility functions derived with data from the 2009 earthquake.	(1) Vulnerability based on scoring: MAPE = 10.6%. When compared with (2) EMS-98, accuracy = 65.4% and kappa=0.36 For (3) MBT classification, only visual validation is done for the 561 bldgs. The remaining 2221 bldgs. are validated by searching within a radius of 50m, with accuracy =84%. Damage is not validated.

Author	Data	Comments to the main topics addressed	Accuracy or reliability
Optical imagery and 3D data (cont.)			
Riedel et al. (2014; 2015)	Aerial images; DEM from airborne data; footprints from Cadaster; building DB from in-field survey; INSEE database.	Vulnerability proxy developed with the vulnerability DB. Calculation of 15 morphological indicators using cadastral data (footprints) and Remote Sensing. Two predictive models created (1) with all the attr., and (2) with only 2 attr., i.e., nr. stories and date, since these are the attr. present in the INSEE DB. Estimation of vulnerability is extended to the whole country using the INSEE French DB. A damage estimation is done and compared to a reference from the Risk-UE project.	Accuracy in training phase is 62.4% when only date and nr. stories are used. Maximum accuracy in training phase is 94.5% when all attr. are included. and the 6 vulnerability classes are merged into 3. Accuracy is not provided for the MBT classification of the entire French territory. Accuracy of the damage comparison is not quantified.
Costanzo et al. (2016)	LiDAR and IPERGEO 503-band VNIR images; cadastral data for verification.	DTM and DSM, together with image classification to characterize terrain, buildings and road network.	Roof type classification in testing phase: accuracy = 95% and kappa = 0.92. The rest is not validated.
Geiss et al. (2016)	RapidEye RGB-NIR images; DEM from TanDEM-X; Time series of Landsat; Expected damage DB; GPS-photos from Google Panoramio.	Multi-temporal analysis of Landsat images to extract the period of construction. Multi-scale segmentation to delineate the urban structures. Direct estimation of damage grades with predictive modelling.	Accuracy of 3 urban structures classification is 85.6%, and kappa = 0.77. MAPE of damage assessment is 13%.
Geiss et al. (2017)	WorldView 2 RGB+NIR; OSM complete DB.	Multi-level segmentation and classification with 270 features to obtain LULC, with samples from OSM. An empirical relation between the nr. buildings in OSM and the segments classified as buildings is calculated for nr. building estimation.	LULC classification accuracy is 86.7% with kappa=0.81. MAPE of the best Nr. bldg. estimation is 26.21%.
Optical imagery, 3D data and other useful information			
Sarabandi et al. (2008)	QuickBird images; Tax assessor database; building DBs from in-field surveys; manual digitization.	Rational Polynomial Coefficients (RPC) are used to obtain the building height by measuring the image coordinates of the ground and roof points of a building corner. Footprint digitization for extraction of geometric and shape-related attributes.	Accuracy of height extraction by comparison with independent data survey: MAPE = 1.42%. Accuracy in testing phase for 6 MBT classification ranges between 63% - 84%.
Borfecchia et al. (2010)	LiDAR; AISA-Eagle RGB-NIR images Landsat ETM+; QuickBird; building DB from in-field survey; census data.	Footprints from local cartography 1:2000. LiDAR for nr. stories. AISA-Eagle images RGB-NIR for spectral signatures of roof inside the footprints. Landsat ETM+ and QuickBird images to obtain the date of construction (> or < 2000) for RC buildings. High-dimension feature vector for MBT classification with deep-learning.	Accuracy of built-up areas: before 2000 is 84.5%; after 2000 is 70.5%. Accuracy in testing phase for 3 MBT classification is 85%. If RC classes are grouped, then accuracy for 2 MBT is 90%.
Wu et al. (2014)	WorldView-2 images; building DB from in-field survey; vulnerability database.	OBIA for building detection (shape and shadow). Height extracted from shadow for 48 isolated buildings.	Bldg. detection accuracy from OBIA is 81%. Height error < 2 m.

Author	Data	Comments to the main topics addressed	Accuracy or reliability
Optical imagery, 3D data and other useful information (cont.)			
Su et al. (2015)	VHR Optical images; building-related local knowledge (Br-LK); building DB from in-field survey; manual digitization.	Digitization of footprints for area extraction. Algorithm for height extraction from optical images by measuring the building SIPD (shadow length). MBT classification using a simple rule based only on height.	MAPE of height is 3.23%. Estimated area is 2.99% lower than real. Accuracy of 3 MBT classification using rules is 96% and kappa = 0.88
Qi et al. (2017)	Google Earth images; Tencent/Baidu Street View; WeChat; VGI; Br-LK; Real state databases; institutional websites; local informants providing in-field data; manual digitization.	Improvement of previous study by Su et al (2015). For MBT classification, authors incorporate a great amount of information from internet sources and in-field informants in order to improve their previous accuracy.	MAPE of area is 4.64%. Accuracy of nr. stories is 89% and kappa = 0.86 Accuracy of 3 MBT classification using rules is 98% and kappa = 0.94

NOTE: Accuracy refers to global or overall accuracy. MAPE stands for Mean Absolute Percentage Error. MBT stands for Model Building Type, as a synonym of SBSTs (Seismic Building Structural Types). Abbreviations: attr. for attributes; bldgs. for buildings; DB for database.

En el Anexo II, se presenta una tabla que compara las características más relevantes de una veintena de estudios realizados en la última década. Actualiza y completa, añadiendo más trabajos y detalles, las comparaciones hechas por Su et al. (2015) y Qi et al. (2017). La Tabla 2-1 es un resumen, donde se muestran los datos empleados y las precisiones obtenidas (una descripción de estas se encuentra en la sección 4.2.1.3).

Cuando se aplican técnicas de EO para caracterizar la vulnerabilidad del entorno urbano, el primer paso es siempre detectar las huellas de los edificios y generar una capa vectorial con sus contornos. Lo deseable es que cada edificio sea representado por un (y solo un) polígono. Para ello, se recurre a una de estas cuatro fuentes:

- Bases de datos oficiales: cartografías existentes o registros de tipo catastral o censal. En Torres et al. (2019a) se tomaron las huellas de los edificios de la Sede Virtual del Catastro de España para el estudio de vulnerabilidad de Lorca; Borfecchia et al. (2010) usaron cartografía local oficial a escala 1/2000.
- Otras bases de datos de cartografía existente, como OpenStreetMap empleada por Geiss et al. (2017); o portales inmobiliarios (Goolzoom, Idealista, Fang Holding Limited, etc.) o GoogleEarth, como combinaron Qi et al., 2017.
- Segmentación de imágenes para detectar objetos, seguida o no de clasificación en un proceso de OBIA. Este método fue el elegido por Wieland et al. (2012a).
- Digitalización manual, como proponen Su et al. (2015).
- En algunos casos se emplea una combinación de varias fuentes (e.g. Wieland et al., 2012b).

La fuente más fiable e inmediata es la base de datos oficial; no obstante, no siempre existe o no es accesible. En ese caso, es recomendable recurrir a otras fuentes abiertas ya elaboradas en proyectos locales previos o globales de tipo colaborativo. Si existiesen estos datos (algo que no está garantizado para todas las regiones del mundo), habría que realizar un proceso de validación para comprobar su exactitud, precisión y completitud. Esto podría realizarse, por ejemplo, empleando técnicas de fotointerpretación de imágenes aéreas o satelitales.

En caso de que estas fuentes no existan en el área de estudio de interés, o existan, pero no sean válidas, habría que recurrir a generar exprofeso las huellas de los edificios empleando datos tomados por sensores remotos, como imágenes multiespectrales, RADAR y LiDAR. Para ello, se emplean técnicas de segmentación de imágenes o digitalización manual. La segmentación es la técnica más adecuada para cubrir amplias zonas de estudio, como grandes ciudades o regiones enteras de un país, ya que es un proceso automático que permite analizar mosaicos de imágenes de gran extensión. No obstante, la segmentación rara vez aporta un resultado vectorial satisfactorio en el que un polígono representa de manera exhaustiva y excluyente a un edificio, algo que sí se consigue con las demás técnicas y que es necesario para obtener una correcta estimación del número de edificios expuestos. Incluso realizando complejos procesos de segmentación anidada o jerárquica (e.g. Soille, 2007; 2008; 2009; Bruzzone y Carlin, 2006; Esch et al., 2008; Geiss et al., 2017), localizando sombras adyacentes a los segmentos que representan edificios (Ehrlich et al., 2013; Hermosilla, 2011) o completando la segmentación

con una clasificación de la imagen basada en objeto (e.g. Wieland et al., 2012a), que aumenta considerablemente los tiempos de procesamiento, los resultados finales no son del todo satisfactorios (Su et al., 2015). Por ese motivo, algunos autores como Qi et al. (2017) defienden la digitalización manual, que, a pesar de ser el método más rudimentario para obtener las huellas de los edificios, es muy preciso y consume menos tiempo que las complejas segmentaciones. Los autores reportaron un tiempo de 4-5 horas para la digitalización de unos 1200 edificios por una sola persona (de 4 a 5 edificios por minuto). Su principal desventaja es la limitada capacidad de escalado, ya que el tiempo crece de manera lineal con el número de huellas a digitalizar, por lo que puede resultar inviable analizar grandes extensiones.

Una vez que se tienen las huellas de los edificios en un archivo vectorial, se procede a dotar a esas huellas de atributos que estén relacionados con la vulnerabilidad sísmica del edificio al que representan. Las huellas y los atributos constituyen la base de datos de exposición, en la que se basará la posterior clasificación de los edificios según su tipología constructiva y, finalmente, se les asociará un modelo de vulnerabilidad. Se recomienda que se siga alguna taxonomía para la selección de atributos y para su definición en términos de rango de valores, intervalos, nomenclatura, etc. Existen taxonomías mundialmente aceptadas y utilizadas, como las propuestas por GEM (Brzev et al., 2013) y la que se está desarrollando actualmente en el Proyecto RIESGOS. Cuantos más atributos se incluyan en la base de datos de exposición, mejor caracterizados quedarán los edificios y mejor definida estará su vulnerabilidad. Una interesante revisión de todos los atributos relacionados con la vulnerabilidad de los edificios es la aportada por Harb et al. (2015). No obstante, no siempre será posible ni necesario calcularlos todos. Para cada estudio se elegirá un conjunto de ellos en función de los datos disponibles, de las características de la zona de estudio y de la metodología de estimación de la vulnerabilidad que se emplee. Suelen ser comunes a la mayoría de los estudios (e.g. Costanzo et al., 2016; Pittore and Wieland, 2013; Riedel et al. 2015) los atributos siguientes: área del edificio, coordenadas planimétricas de su centroide, año de construcción, altura y/o número de plantas, tipo de tejado (plano o inclinado) y materiales de este. Otros autores consideran también parámetros relativos a la forma, como índices de forma o compacidad, o indicadores de regularidad en planta y/o alzado (Sarabandi et al., 2008; Ehrlich et al., 2013). Finalmente, en otros estudios se ha incluido información relativa a los materiales de construcción, el uso y la ocupación (Qi et al., 2017; Sarabandi et al., 2008).

El área, la posición del centroide y los indicadores de forma y regularidad son parámetros geométricos que se derivan de la huella del edificio y se calculan en el SIG de manera más o menos directa. El resto de los atributos, si no se encuentran en fuentes de datos ya existentes, pueden obtenerse por teledetección. El año de construcción es relevante en ciudades que se han desarrollado de forma ordenada (es decir, cuyo crecimiento se ha apoyado en planes de desarrollo urbanístico) y/o que han aplicado códigos de construcción sismorresistente a lo largo de su historia. En estos casos, la fecha de construcción de los edificios aporta información acerca de los materiales, métodos constructivos y nivel de diseño antisísmico, aspectos relevantes ambos para la asignación de la vulnerabilidad sísmica. Para conseguir el año de construcción, se llevan a cabo análisis de control de cambios empleando series temporales de imágenes (Wieland et al., 2012a). Esta técnica da buenos resultados con imágenes de media

resolución (GSD de 20 m - 30 m) cuando la ciudad de estudio ha crecido por expansión, pues se pueden identificar claramente las áreas de nuevos desarrollos en las sucesivas imágenes de la serie. De hecho, este análisis se probó en el escenario de Puerto Príncipe y se pudieron datar los edificios, pero finalmente este dato no se ha utilizado porque en este escenario el año de construcción no ha resultado ser un atributo relevante para la clasificación de tipologías. En el caso de ciudades que han experimentado un crecimiento lento y, por tanto, su extensión ha permanecido casi inalterable con el paso del tiempo, se debe recurrir a imágenes de mayor resolución que permitan identificar los edificios y las calles individualmente. Este fue el caso de Lorca, donde también se intentó datar los edificios. Las imágenes de media resolución no dieron un buen resultado, por lo que se utilizaron ortofotos del PNOA (GSD = 15 cm). Sin embargo, al no haber datos disponibles previos al 2003, no se pudo asignar el periodo de construcción a todos los edificios, sino solo a los más modernos. Por tanto, tampoco se ha utilizado el dato de año de construcción en el escenario de Lorca. Hay que tener en cuenta que con esta técnica basada en análisis de imagen solo se pueden datar edificios construidos a partir de la década de los sesenta, aproximadamente, que es cuando empieza a haber imágenes de satélite disponibles de forma global y abierta. Todos los edificios que aparezcan en la primera imagen de la serie temporal serán datados en un único intervalo previo a la fecha de la imagen.

La altura del edificio es uno de los atributos más complicados de conseguir por las dificultades de acceso a datos 3D que hay en determinadas regiones del mundo; si bien estas dificultades tienden a desaparecer con el paso del tiempo. El uso de sombras en imágenes, tanto aéreas como satelitales, ha sido uno de los métodos más empleados para calcular la altura de los edificios; y lo sigue siendo actualmente, a pesar de las limitaciones que presenta. Los algoritmos más tradicionales requieren que la sombra esté completamente proyectada en el suelo y se pueda medir (Ehrlich et al., 2013), y otros más actuales proponen medir solo un segmento de la sombra del edificio (Su et al., 2015). Ambos algoritmos presentan una doble limitación: no pueden emplearse en zonas con alta densidad de edificación (porque las sombras de los edificios quedan proyectadas sobre otras fachadas, deformándose) ni en terrenos con pendiente. Sarabandi et al. (2008) desarrollaron un nuevo algoritmo que no emplea sombras, sino un eje vertical de la fachada de los edificios para calcular su altura. Pero de nuevo, no puede aplicarse a zonas densamente edificadas porque los edificios quedan ocluidos por sus vecinos. Pittore y Wieland (2013) realizaron una toma de datos que hasta el momento había sido poco habitual, basada en un Mobile Mapping System (MMS) montado en un vehículo propio. Con las imágenes panorámicas (360º) recogidas por el sistema calcularon la altura de los edificios de una zona usada como muestra. Esta técnica, aunque precisa, requiere de un diseño de toma de datos exprofeso con un hardware específico. Otra alternativa es el uso de pares estereoscópicos para tomar un punto con coordenada altimétrica en cada tejado y otro en el suelo cercano, derivando a partir de ellos la altura del edificio. Esta técnica no suele emplearse en este campo debido a que se necesitan conocimientos muy específicos de fotogrametría y un equipo de restitución que no está al alcance de la mayoría.

Como alternativa al análisis de imágenes provenientes de sensores pasivos se están utilizando datos tridimensionales tomados por sensores activos, como SAR o LiDAR (e.g. Sohn y Dowman, 2007; Geiss et al., 2015; 2016). Los datos SAR son cada vez más accesibles y presentan

resoluciones altas, con GSD de hasta 1 m, adecuadas para este tipo de estudios (Polli y Dell'Acqua, 2011). Las nubes de puntos LiDAR superan estas precisiones proporcionando densidades de varios puntos por m². La disponibilidad de datos LiDAR también va aumentando con el tiempo gracias a programas nacionales como en España (PNOA) o en Italia (Borfecchia et al., 2010), iniciativas globales como OpenTopography, o a través de campañas de campo específicas para toma de datos con RPAS.

Mención especial merece el proyecto PNOA, que es uno de los tres componentes de la iniciativa PNOT del Instituto Geográfico Nacional (IGN¹¹). Es un proyecto cooperativo y descentralizado, en el que participan todas las comunidades autónomas. En el marco de este proyecto, se registran imágenes aéreas y LiDAR de todo el territorio español con actualización de 2-3 años. Estos datos se distribuyen libremente a través del portal web del Centro Nacional de Información Geográfica (CNIG) del IGN. Lleva casi dos décadas funcionando y fue galardonado con el *Public Service Award of the United Nations* en 2013, reconocimiento de prestigio internacional. Se trata, por tanto, de un proyecto ejemplar, no solo para España, sino para otros países. Los datos del PNOA se utilizan en esta tesis, lo que supone un reto, porque no fueron concebidos originariamente con este propósito, sino para proporcionar un MDT continuo que diera apoyo a estudios ambientales, como la hidrología, usos del suelo, control de cambios, etc.

¹¹ <https://pnoa.ign.es/>

3 Hipótesis de la investigación y objetivos

De la revisión del estado del arte se desprende que ya hay técnicas de análisis de datos tomados por sensores remotos lo suficientemente probadas como para garantizar que algunos atributos relacionados con la vulnerabilidad sísmica de los edificios se puedan obtener con buenas precisiones. Los numerosos estudios comparados en el Anexo II y en la Tabla 2-1 acreditan que el área construida (*built-up mask*), la altura de los edificios, el material del tejado y algunos indicadores geométricos se pueden conseguir con precisiones globales de alrededor del 75% y mayores y con errores porcentuales (MAPE) menores al 5%. El desafío para conseguir estas precisiones está asociado a la disponibilidad de datos y a la cantidad de trabajo que los presupuestos y cronogramas de los estudios puedan permitirse. Se han detectado algunas cuestiones que aún están abiertas, relacionadas con aspectos técnicos, como la estrategia de integración de los datos tomados por sensores remotos, el proceso de muestreo, la flexibilidad del procedimiento metodológico y la verificación de los resultados.

Lo que tampoco está resuelto aún de manera fiable es la aplicabilidad de estos resultados a un análisis de riesgo sísmico que sea la base para el diseño de políticas de prevención y mitigación, que es el fin último de estos estudios. Hasta ahora, los estudios que han llegado a estimar un escenario de daños no han podido verificar sus resultados de manera robusta ni han llegado a dar información sobre los costes asociados.

Con base en todo lo anterior, en esta tesis se plantean las siguientes hipótesis de investigación:

¿Es viable, desde el punto de vista técnico y económico, el uso de una base de datos de exposición y vulnerabilidad creada a partir de datos tomados por sensores remotos en un estudio de riesgo sísmico?

Esta pregunta conlleva otras dos asociadas a la viabilidad técnica y económica, respectivamente:

- i) ¿Cómo influyen las incertidumbres acumuladas en el proceso de cálculo de los atributos de la base de datos de exposición en los resultados de daño por sismo?***
- ii) ¿Cuánto permite ahorrar, en términos de tiempo y presupuesto, la elaboración de una base de datos de exposición por técnicas de teledetección con respecto a los métodos tradicionales de inspección in situ?***

Para responder a estas cuestiones, se plantean los siguientes objetivos, que se desarrollarán en las sucesivas fases de análisis de cada uno de los escenarios elegidos en el presente estudio.

- 1. Acceso a fuentes de datos disponibles.** El primer paso para abordar cualquier estudio territorial es la descarga y preparación de los datos espacio-temporales que se van a necesitar y configurar el espacio de trabajo en un SIG. Este paso, común a estudios de muchos otros campos, es de especial importancia en este caso, ya que las estrategias de análisis, el software, los tiempos, el coste y las precisiones de los resultados dependen de la

disponibilidad de los datos. Concretamente, se pretende evaluar el coste de los datos empleados, su accesibilidad, su idoneidad para este estudio y sus limitaciones, si las hay.

2. **Extracción de atributos.** Se trata de conseguir todos los atributos relacionados con la vulnerabilidad de las construcciones que resulten de interés, a escala de edificio individual, con el fin de generar la base de datos de exposición. La caracterización del entorno de los edificios en este estudio se realiza a través de la identificación de estratos urbanos, que son regiones de la ciudad en las que la edificación, la vegetación y la red viaria tienen configuraciones homogéneas. La extracción de los atributos conlleva la integración del mayor volumen de datos del estudio, tanto de imágenes ópticas, como de datos tridimensionales y otras fuentes auxiliares. Para realizar esta integración, se propone emplear técnicas que estén en línea con el estado del arte, e incluso, plantear algunas innovaciones.

Es importante que el formato de los atributos siga las recomendaciones de alguna de las taxonomías más actuales, para garantizar su interoperabilidad y que pueda ser compartida.

3. **Clasificación de tipologías constructivas.** Es necesario identificar qué tipologías constructivas existen en la ciudad y es posible reconocer con los atributos obtenidos. La clasificación de tipologías es un paso muy importante en el proceso de evaluación del riesgo sísmico y se realiza por medio de modelos predictivos (*predictive modeling*) que se crean con algoritmos de aprendizaje automático. Lo deseable es llegar a detallar al máximo la composición del parque inmobiliario de una ciudad, pero los atributos y las técnicas de análisis empleadas no siempre lo permiten. Por tanto, hay que llegar a un acuerdo entre el número de tipologías a identificar y el coste del estudio. El estado del arte de estas técnicas de clasificación deja ver que se suele recurrir a los algoritmos más sofisticados, como los SVM, que son también los más complicados. Rara vez se emplean técnicas sencillas, como los clasificadores Bayesianos (Pittore y Wieland, 2013) o la regresión logística (Sarabandi et al., 2008), que son basados en probabilidad. Ni siquiera se prueban. Esto ha inspirado algunos experimentos de la tesis para comparar varios clasificadores y tratar de demostrar que no siempre hay que recurrir a técnicas complejas.

4. **Asignación de vulnerabilidad y cálculo de daños por sismo.** Los objetivos anteriores llevan a este, en el que se evalúa la capacidad de las técnicas de análisis empleadas para generar datos fiables que puedan utilizarse en cálculos de riesgo. Se trata de asignar a cada una de las tipologías anteriores un modelo de vulnerabilidad sísmica que describa cómo se comportaría un determinado tipo de edificio en caso de sismo. Esto permite evaluar el posible daño que sufrirían las estructuras, que es el indicador de riesgo sísmico más importante y el primero que se calcula, ya que de él dependen los demás indicadores (víctimas, coste económico, escombros, etc. Ver Figura 1-3). En este punto, es necesario realizar una verificación de resultados robusta e independiente, para garantizar la viabilidad técnica de estos métodos de generación de bases de datos de exposición con aplicación al riesgo. Es importante poder cuantificar las incertidumbres de la cadena de cálculo de atributos a través de verificaciones exhaustivas y evaluar su impacto en los resultados finales de daño.

5. **Evaluación de costes.** El objetivo final del desarrollo de procedimientos para la caracterización del entorno urbano con técnicas de teledetección en este campo es optimizar la toma de datos tradicional, que consiste en visitar los edificios para recoger la información relativa a sus atributos de manera visual. Si el coste de estas técnicas no es menor que el coste de la visita a campo, entonces no merece la pena incorporarlas a los procesos de cálculo de riesgo sísmico. Además, es necesario identificar e indicar de manera concreta cómo pueden ayudar estas técnicas modernas a las campañas de campo para reducir su coste, su complejidad o su duración.

Para la consecución de estos objetivos, se plantea realizar la aplicación en los dos escenarios elegidos: Lorca y Puerto Príncipe. Ambas ciudades son lo suficientemente distintas como para tener que realizar análisis complementarios que muestren la variabilidad de datos, métodos y técnicas de que se puede hacer uso hoy día. Cada escenario tiene su enfoque particular, que deriva de la disponibilidad de fuentes de datos, de su extensión y del tipo de edificación, principalmente (Torres et al., 2009b). Con estas aplicaciones tan dispares se pretende, además, justificar la necesidad de flexibilizar los flujos de análisis, pues una técnica que ha funcionado bien en un escenario puede no funcionar, o simplemente puede no ser aplicable, en otro. Es importante mencionar que el objetivo no debería ser obtener siempre la mejor precisión final, sino obtener una precisión final aceptable considerando el coste del análisis y la relevancia del resultado que este aporta, de modo que la relación coste-beneficio sea positiva.

Bloque 2

4 Propuesta metodológica

4.1 Introducción

A continuación, se presenta el procedimiento metodológico propuesto en esta tesis para crear una base de datos de exposición y realizar la consiguiente asignación de vulnerabilidad a los edificios a escala municipal. A la hora de definir las fases del método y las técnicas de análisis propuestas, se han tenido en consideración los cinco requisitos planteados en el alcance (sección 1.3): que los datos empleados puedan actualizarse con frecuencia y queden publicados en abierto; que se minimice el trabajo de campo para inspección de edificios; que incorpore conocimiento de experto sobre las construcciones locales; que sea lo suficientemente flexible como para poder aplicarse en distintos escenarios; y que minimice los costes.

4.2 Procedimiento metodológico

El procedimiento metodológico diseñado en esta tesis consta de tres fases principales: (1) estratificación de la ciudad en patrones urbanos; (2) creación de la base de datos de exposición; y (3) clasificación de tipologías constructivas y asignación de vulnerabilidad sísmica. En la Figura 4-1 se muestra un esquema del mismo, en el que se diferencian estas fases y se detalla qué técnicas se emplean en cada una de ellas. En algunas fases se proponen varias técnicas alternativas, de forma que se podrá elegir cuál aplicar en cada escenario concreto en función de los objetivos y los datos disponibles, principalmente. Esto confiere flexibilidad al procedimiento, en línea con el cuarto requisito. La integración de todos los datos necesarios y de los resultados parciales que se obtienen, así como la realización de algunos geo-procesos, se hace posible gracias al uso de un Sistema de Información Geográfica (SIG).

En los siguientes apartados, se describen las tres fases de manera detallada.

4.2.1 Fase 1: Estratificación de la ciudad

La primera fase está destinada a la estratificación de la ciudad. Esta consiste en dividir la ciudad en zonas con patrón urbano homogéneo, es decir, en zonas donde los edificios sean similares y exista homogeneidad en la configuración de la trama urbana, zonas ajardinadas y espacios abiertos. Tener la ciudad dividida en estratos permite realizar las subsiguientes operaciones de análisis estratificado. Esto es, la estratificación hace posible establecer los parámetros de configuración que han de definirse en las fases dos y tres, como área mínima de los edificios o altura máxima de la vegetación, ya que son valores característicos de cada estrato. Son pocos los autores que comienzan el estudio con la estratificación de la ciudad (e.g. Wieland et al., 2012a) por no considerarlo necesario. Pero la experiencia propia ha demostrado que tiene sentido trabajar por estratos (Wieland et al., 2016; Torres et al., 2019a) y, de hecho, se obtienen mejores resultados. Además, se pueden identificar las áreas donde se localizan los errores para hacer un mejor diagnóstico y acotar la zona de acción en la que implementar las mejoras. Ejemplos de esto se dan en cada uno de los escenarios de aplicación de esta tesis.

Figura 4-1. Esquema metodológico

4.2.1.1 Segmentación de la imagen

Si no se encuentra una base de datos en la que basar la definición de los estratos, como límites administrativos infra-municipales (distritos, barrios, etc.), su delimitación puede hacerse manualmente, en un proceso de fotointerpretación empleando imágenes aéreas o satelitales. Esto puede ser útil cuando la ciudad de estudio no es muy extensa y el trabajo resulta abordable. En caso contrario, habría que recurrir a clasificar la imagen. En alguna ocasión se ha preferido la clasificación por píxel (Costanzo et al., 2016), pero generalmente se recurre al análisis OBIA, ya que los contornos de los objetos obtenidos en el proceso de segmentación se aprovechan para generar las huellas de los edificios. Todas las clasificaciones de imágenes realizadas en esta tesis se han hecho empleando esta técnica.

El uso de la técnica OBIA implica segmentar y clasificar las imágenes de manera semi-automática. Para la segmentación, se pueden emplear algoritmos de crecimiento de regiones, si se van agrupando píxeles de similar nivel de gris; o de búsqueda de bordes, si se buscan discontinuidades, es decir, cambios bruscos de nivel de gris en píxeles vecinos (Hermosilla, 2011). En este trabajo se emplean ambos tipos de algoritmos. Uno es el algoritmo *watershed* de regiones crecientes de Jin (2012), que agrupa píxeles vecinos cuyos valores de gris se parecen (se basa en la definición algorítmica de Vincent y Soille, 1991). Este algoritmo comienza por los valores más bajos y va fusionando píxeles de similar valor hasta que encuentra diferencias significativas. Ese umbral de diferenciación lo establece el usuario, de manera que los segmentos resultantes sean polígonos que representen objetos reales de su área de estudio (edificios, árboles, carreteras, etc.). Es el parámetro de escala. Cuanto mayor sea este parámetro, menor detalle tiene la segmentación, y mayores resultan los segmentos. Otro parámetro que debe definir el usuario es el de unión, que establece el tamaño mínimo de los objetos, de manera que, a mayor valor del parámetro, mayores serán los segmentos. Los algoritmos de regiones crecientes trabajan con un solo valor por píxel. Cuando se trabaja con imágenes multispectrales, como ocurre en la mayoría de los casos en este campo de aplicación, es necesario convertir una imagen multibanda en una sola banda. Para ello, se puede aplicar el método de bordes o el método de intensidad. El primero aplica un filtro Sobel a la imagen original para calcular la imagen gradiente, en la que quedan resaltados los bordes de los objetos. El algoritmo Watershed se aplica entonces a la imagen gradiente y va agrupando píxeles desde los menores valores hasta los más altos, que son los bordes de los objetos. Este es el que se propone en esta tesis, ya que los objetos que se quiere identificar en la imagen tienen bordes bien definidos (manzanas urbanas, edificios, viales, jardines, etc.). El método de intensidad promedia los valores de gris de cada banda para crear una única banda; el algoritmo Watershed trabaja en este caso con valores promediados de intensidad. Un uso muy extendido de este método es el cálculo de cuencas hidrográficas a partir de un MDT.

El segundo tipo de algoritmos de segmentación de imágenes, que se basa en detección de bordes, también se emplea en esta tesis. Concretamente, se utiliza el algoritmo de Felzenszwalb y Huttenlocher (2004). Es un potente algoritmo basado en grafos, que compara el vector de características de los píxeles en distintas direcciones hasta que localiza un borde por similitud entre los píxeles que lo componen, y va comprobando que se mantiene esa similitud y, además, existe una diferencia (disimilitud) entre píxeles de cada uno de los segmentos que separa dicho borde. El vector de características lo crea usando los valores de los píxeles de cada banda de la imagen, y emplea la distancia euclídea como medida de

disimilitud. Este algoritmo de segmentación basado en detección de bordes también requiere que el usuario establezca valores para los parámetros de escala y de unión.

Estos parámetros se establecen, generalmente, mediante un proceso de ensayo y error, hasta que el usuario comprueba que los segmentos obtenidos representan el contorno de los objetos de su área de estudio lo más fielmente posible. En otros casos, cuando no se conoce bien el área de estudio o cuando se prefiere llevar a cabo un procedimiento menos dependiente del criterio del usuario, se realizan análisis comparativos de homogeneidad intra-grupo (dentro de cada segmento) y heterogeneidad inter-grupos (entre segmentos) (Dey et al., 2010). Para la primera opción, en esta tesis se propone utilizar la desviación estándar de los valores de los píxeles de cada segmento ponderada por el tamaño de este (Johnson y Xie, 2011); y el índice I de Moran para la segunda. El par de parámetros de escala y unión que maximice ambas, es decir, la homogeneidad intra-grupo y la heterogeneidad inter-grupos, es considerado el óptimo, ya que proporciona la mejor división de la imagen en segmentos homogéneos (Wieland et al., 2016).

El resultado de la segmentación es una capa vectorial poligonal, en la que cada segmento o polígono es un objeto de interés. En este caso, cada polígono es el contorno de un estrato. Esta capa se incorpora al SIG. Antes de pasar a la clasificación de los segmentos en patrones urbanos, es común realizar procesos de depuración geométrica de los mismos, como suavizado de las formas o eliminación de polígonos de área menor que un umbral (establecido por el usuario) por fusión con vecinos mayores.

4.2.1.2 Clasificación de los segmentos

Los segmentos depurados son ahora clasificados con el fin de etiquetar cada uno con la clase de objeto a la que pertenecen. En este caso, las clases son los patrones urbanos típicos de la ciudad de estudio, como casco histórico, zona industrial, ensanche, etc. En el proceso de clasificación de imagen se utilizan técnicas de aprendizaje automático para generar modelos predictivos. Las técnicas de aprendizaje automático se emplean cuando no se tiene un algoritmo para conseguir un resultado a partir de unos datos de entrada, pero se tienen datos de los que se puede extraer información (Alpaydin, 2009). Estos datos iniciales son denominados datos o muestras de entrenamiento (*training dataset*). Estas técnicas estudian los datos y buscan patrones espaciales, temporales o comportamentales (en forma de similitudes y diferencias entre valores, rangos que se repiten, tendencias claras, etc.) y son capaces de predecir un valor o comportamiento para datos nuevos. Estas predicciones se obtienen por medio de modelos predictivos (Kuhn y Johnson, 2013).

Los modelos predictivos creados deben ser validados antes de probarse en datos nuevos. Esto suele hacerse por medio de una validación cruzada de k iteraciones (*k-fold cross validation*), que divide la base de datos de entrenamiento en k sub-grupos de igual número de muestras. Dejando un sub-grupo para verificación, se crea un modelo con los $k-1$ restantes y se repite el proceso dejando para verificación un sub-grupo distinto cada vez. El promedio de las precisiones obtenidas con los k modelos creados en esta validación cruzada es un indicador, a priori, del poder predictivo del modelo. Una vez que el usuario obtiene una precisión aceptable, crea el modelo final con todas las muestras de entrenamiento.

Tras comprobar que el modelo es capaz de predecir con la precisión requerida, se procede a evaluar su desempeño con datos nuevos, llamados datos de verificación (*testing dataset, ground truth dataset, yet-to-be-seen data* o *unseen data*). De esta verificación se obtiene la precisión a posteriori o final del modelo. Si es aceptable, se asume que el modelo puede realizar predicciones en cualquier otro conjunto de datos con precisión comparable. La calidad de los modelos predictivos depende fuertemente de los datos de entrenamiento, es decir, de la técnica de muestreo, del número de muestras, del tipo de muestra, del usuario que creó el conjunto de datos y de las características del escenario de aplicación (configuración de la ciudad, tipo y calidad de la imagen, etc.). Es de esperar que los resultados de la predicción sean comparables si se mantienen estas condiciones en el nuevo conjunto de datos, y viceversa. Por tanto, modelos de clasificación de imágenes creados para una ciudad podrían ser aplicados en otras ciudades similares (ciudades cercanas, dentro del mismo país); pero difícilmente se podrán aplicar en ciudades o países diferentes de manera satisfactoria.

Las técnicas de aprendizaje automático tienen una particularidad, y es que cuando se añaden más datos de entrenamiento, los modelos se actualizan de manera automática y aprenden nuevas reglas de los nuevos datos. Los modelos predictivos son divididos, habitualmente, en dos grandes grupos según el problema que resuelven: regresión o clasificación. Los primeros se usan cuando se quiere predecir un valor numérico de una variable cuantitativa continua y, los segundos, cuando la variable a predecir es cualitativa o discreta, es decir, es una clase o categoría. La identificación de los objetos de una imagen o la predicción de la vulnerabilidad sísmica de un edificio son problemas de clasificación que se presentan en esta tesis.

Elección del modelo de clasificación

Para la clasificación de los segmentos en patrones urbanos se utiliza Support Vector Machines (SVM) como clasificador. La elección de esta técnica de aprendizaje se basa en los resultados de un amplio estudio realizado por Wieland y Pittore (2014) con objetivo similar a este. Los autores compararon el desempeño de varios clasificadores (*Naïve Bayes, SVM, Random Trees* y *K-Nearest Neighbors*) para estratificar tres ciudades diferentes, usando hasta seis imágenes de satélite distintas. Llegaron a la conclusión de que el clasificador más estable era SVM, es decir, que el modelo calculado con SVM permitía obtener las mismas precisiones de clasificación cuando se probaba en datos nuevos. Esto le confiere robustez y mayor capacidad de escalado y transferencia.

SVM es un modelo estadístico no paramétrico desarrollado por Vladimir Vapnik en la década de los sesenta. Una revisión completa puede encontrarse en Vapnik (2000). Es un algoritmo ampliamente usado y probado en problemas de clasificación, por ser una de las técnicas de aprendizaje automático más flexibles y eficaces (Kuhn y Johnson, 2013). Un SVM calcula la superficie matemática (hiperplano) que mejor separa dos clases (*decision boundary*) empleando los datos de un sub-grupo de muestras, denominadas vectores soporte. Por tanto, no necesita todo el conjunto de muestras de entrenamiento (Figura 4-2 (A)). Los vectores soporte elegidos son las muestras que quedan más cerca del hiperplano, y la distancia entre este y las muestras se llama margen. El objetivo del SVM es maximizar el margen para

garantizar menos errores de clasificación cuando entren datos nuevos, es decir, para garantizar que el modelo no queda sobreajustado. En estos casos, se dice que el margen es suave (*soft margin*). Sin embargo, este margen grande (suave) puede implicar errores en la clasificación de las muestras de entrenamiento. Puede verse en la Figura 4-2 (B), donde la representación de la derecha muestra un margen grande, pero este da lugar a un error de clasificación de la clase azul, indicado por la distancia ξ . Por el contrario, cuando se crea un hiperplano muy sinuoso o ajustado al espacio que queda entre clases, el modelo resulta sobreajustado, y el margen se dice que es duro (*hard margin*. Figura 4-2 B, izquierda). El equilibrio entre los errores en la clasificación de las muestras y el tamaño del margen lo rige el parámetro de coste, C . Valores bajos de C dan lugar a márgenes suaves, y viceversa. El valor óptimo de C se establece, generalmente, por medio de una validación cruzada.

Figura 4-2. Clasificador SVM (Support Vector Machine).

(A) Vectores soporte y margen de un hiperplano creado por SVM para separar dos clases de muestras, representadas por los puntos verdes y rojos. (B) Dos hiperplanos creados por SVM para diferentes valores de C .

Fuente: modificada de <https://mubaris.com/posts/svm/>

En los problemas reales, lo normal es no tener clases que puedan separarse con funciones lineales. Para superar esa limitación, SVM utiliza funciones kernel (Boser et al., 1992) para representar los datos de entrada en otro espacio de dimensión superior, el espacio de características, donde las clases sí son linealmente separables. Invirtiendo el camino, la frontera lineal creada en el espacio de características se transforma en una frontera no lineal en el espacio de datos de entrada original. En este trabajo se prueban diferentes kernels para crear los modelos con SVM: kernel polinómico, *Radial Basis Function* (RBF) y *Pearson VII Universal Kernel* (PUK, Üstün et al., 2006). El kernel polinómico requiere establecer la potencia del mismo, a través de parámetro d . El parámetro clave del RBF es gamma, la anchura del kernel (γ). Para el PUK, hay que ajustar sigma (σ), para la mitad del ancho de la función de Pearson VII (*half-width*), y omega (ω), para el factor de cola (*tailing factor*). Este último kernel tiene dos parámetros que establecer, lo que puede significar más tiempo de procesamiento. Sin embargo, con el tamaño de las bases de datos que se manejan en este estudio, este posible incremento de tiempo no supone ningún riesgo. Al contrario, construir un SVM con un PUK puede suponer un

beneficio, ya que permite crear modelos más flexibles y potentes (Üstün et al., 2006). El uso del kernel Universal de Pearson VII no es habitual en el campo de la ingeniería sísmica, como sí lo es en otros relacionados con las ciencias químicas o la informática (e.g. Raccuglia et al., 2016; Stanimirova et al., 2010; Santos et al., 2013). La innovación que supone su uso en este campo aumenta el interés por incorporarlo al estudio.

Toma de muestras

Como se ha mencionado antes, el desempeño de un clasificador se ve claramente influenciado por las muestras de entrenamiento. Para eliminar cualquier sesgo introducido por la configuración de la base de datos de entrenamiento, se llevan a cabo varios análisis en paralelo para comparar la influencia de los siguientes factores en la precisión de la clasificación:

1. La técnica de muestreo, realizando muestreo aleatorio y muestreo manual, dirigido por el usuario (ambos estratificados).
2. El número de muestras de cada clase (también abordado por otros autores, como Matsuka et al., 2012; Riedel et al. 2015).
3. La distribución de las muestras de cada clase, utilizando datos de entrenamiento con el mismo número de muestras por clase (*balanced*) y con un número distinto (*unbalanced*).
4. La forma de las muestras, empleando los segmentos directamente y otro juego de muestras cuadradas (tomadas dentro de cada segmento).

Además de para eliminar un posible sesgo introducido por la configuración de la base de datos de entrenamiento, el objetivo de trabajar con diferentes tipos de muestras es tratar de optimizar el proceso de muestreo. La toma de muestras es un proceso que requiere tiempo y esfuerzo, pues hay que crear un gran número de ellas (dependiendo del número de clases y del tamaño de la población, se pueden necesitar cientos de muestras) y hay que etiquetarlas manualmente, empleando imágenes de alta resolución que permitan la fotointerpretación. Por tanto, con estas pruebas se pretende estimar cuál sería el mínimo número de muestras que se necesitan para conseguir precisiones aceptables de los modelos predictivos, y si es necesario hacer un muestreo aleatorio o se puede hacer manual, siempre que el usuario sea capaz de tomar una muestra que recoja toda la variabilidad presente en el área de estudio.

La propuesta de utilizar muestras cuadradas tiene como objetivo independizar el proceso de muestreo de la segmentación, ya que es frecuente cambiar la segmentación a lo largo del proceso, para mejorarla. Si se demuestra que es posible trabajar con muestras cuadradas que son independientes de los segmentos, y se obtienen precisiones aceptables, se podría generar un único dataset de entrenamiento que se podría reutilizar con cualquier segmentación. De esta forma, se agilizaría mucho el proceso completo de clasificación de imágenes. Incluso, se podrían reutilizar los vectores de características de estas muestras para entrenar otros modelos de áreas de estudio similares.

Cálculo del vector de características y reducción de su dimensión

El vector de características (*feature vector*) es un conjunto de atributos calculados para cada muestra (segmento o cuadrado) a partir de su geometría y de los valores de los píxeles que contienen. Se dividen en tres categorías:

- a. Espectrales: son estadísticas calculadas con los valores de los píxeles de cada muestra, para cada banda de la imagen o combinación de ellas (como el NDVI). Se calcula la media, desviación típica, máximo, mínimo, rango, etc.
- b. De textura: basadas en la matriz de co-ocurrencia de niveles de gris, GLCM (Haralick et al. 1973). Se calcula contraste, ordenación, homogeneidad, entropía, etc.
- c. Geométricas: derivadas de su forma, como área, perímetro, compacidad, etc.

Un listado completo y la definición de todas las características pueden encontrarse en Wieland y Pittore (2014).

El número de características aumenta cuando la imagen que se trabaja tiene muchas bandas, que es lo habitual, llegando a sobrepasar el centenar. Es necesario, por tanto, reducir la dimensión del vector de características para eliminar del análisis las que tienen menor poder discriminante (menor capacidad de predecir la variable dependiente) y las que son combinación lineal de otras. Para ello, se emplean tres métodos con el fin de comparar los resultados:

- a. RFE (*Recursive Feature Elimination*, Guyon et al., 2002), que ejecuta un clasificador (SVM en este caso) de manera iterativa, y va ponderando los atributos en cada iteración según su poder discriminante. Cada vez utiliza solo los atributos que mayor peso tienen, y elimina los de menor peso, hasta estabilizar el proceso. Para evitar sesgos, se ha aplicado RFE mediante validación cruzada.
- b. ReliefF (Kononenko, 1994), que establece un ranking con los atributos en función de su capacidad para separar muestras vecinas que pertenecen a diferentes clases. Para ello, selecciona de manera recursiva una muestra de una clase, y aleatoriamente, varias muestras vecinas de esa misma clase y de otra, y compara los valores de cada atributo para todas las muestras seleccionadas.
- c. Análisis discriminante lineal (Huberty, 1994), que es una técnica estadística multivariante que crea funciones de clasificación de la variable dependiente por medio de una combinación lineal de otras variables predictoras (las características o atributos). Calcula el poder discriminante de estas por comparación de las varianzas intra-grupos e inter-grupos usando los índices F de Snedecor y lambda de Wilks. Los atributos con menor varianza intra-grupos y mayor varianza inter-grupos son los de mayor poder discriminante. Mediante un procedimiento *stepwise*, los atributos son introducidos uno a uno para crear las funciones de clasificación de la variable dependiente, en orden, de mayor a menor poder discriminante. El proceso termina cuando la introducción de un nuevo atributo no mejora significativamente la calidad del modelo.

4.2.1.3 Verificación de los resultados de clasificación

Se emplean diferentes métricas para evaluar la precisión del SVM a la hora de clasificar los segmentos de la imagen en estratos. Por un lado, se calculan las métricas habituales derivadas de la matriz de confusión. La matriz se crea colocando por columnas (o filas) el número de instancias asignadas por el clasificador a cada clase, y por filas (o columnas), la instancias que realmente pertenecen a cada clase, según el dataset de verificación. De la comparación de filas y columnas se puede saber cuántas instancias han sido correctamente clasificadas (son los valores de la diagonal principal de la matriz de confusión), y cuántas se han asignado erróneamente a cada una de las demás clases.

Estas métricas (Fawcett, 2006) se expresan en porcentaje o en una escala equivalente de 0 a 1, siendo siempre mejor un modelo que presente métricas cercanas al 100% o a 1.

- Precisión (*precision*), también conocida como precisión del usuario, es una medida de los errores por comisión. Se calcula como el cociente entre el número de muestras de una clase clasificadas correctamente y el número total de muestras de otras clases que han sido etiquetadas incorrectamente como pertenecientes a ella. Informa acerca de cuántas de las muestras asignadas a una clase pertenecen verdaderamente a ella. Esta métrica se utiliza para elegir modelos que minimicen los errores por falsos positivos.
- TP Rate: índice de verdadero positivo. También se conoce como sensibilidad, *recall* o precisión del productor. Indica los errores por omisión y se calcula como el cociente entre el número de muestras de una clase clasificadas correctamente y el número total de muestras de esa clase. Informa acerca de cuántas muestra se han asignado a una clase respecto a todas las que pertenecen a ella, es decir, cuántas muestras de una clase se han identificado correctamente. Esta métrica se utiliza para elegir modelos que minimicen los errores por falsos negativos.
- FP Rate: índice de falso positivo. Se calcula como el cociente entre el número de muestras asignadas erróneamente a una clase y el número total de muestras de las demás clases. También se identifica como 1-especificidad y, junto con el TP-rate, se emplea para crear la curva ROC (*Receiver Operating Characteristics*), que se explica más adelante.
- Precisión global (*overall* o *global accuracy*), se calcula como el cociente entre el número total de muestras correctamente clasificadas y el número total de muestras. Esta es una de las métricas más utilizadas para comparar la calidad de varios modelos y elegir el mejor. Pero es una medida adecuada cuando se trabaja con bases de datos equilibradas (*balanced datasets*), es decir, con números similares de muestras por clase. En caso contrario, puede dar lugar a decisiones poco acertadas. Para salvar esta limitación, se emplea el F1-Score.
- F1-Score: es la media armónica de la precisión y el recall. Es una medida de precisión muy completa y encuentra un balance entre el coste de los errores por falso positivo y falso negativo. Es la métrica más adecuada para comparar modelos cuando las muestras están desequilibradas. Se calcula con la Ecuación 1:

$$F1\text{-Score} = (2 * \text{precisión} * \text{recall}) / (\text{precisión} + \text{recall})$$

Ecuación 1

Cuando se realiza una validación cruzada, se calcula el promedio de las métricas de precisión de todas las clases. Este promedio será ponderado cuando se lleva a cabo una clasificación multi-clase, utilizando como peso el número de muestras por clase.

Otro indicador que se calcula también a partir de la matriz de confusión es el estadístico kappa de Cohen, κ (*coefficient of agreement*). Su uso está muy extendido a la hora de comparar el desempeño de diferentes modelos de clasificación. Es un indicador que compara la precisión observada con la precisión esperada, siendo esta última una medida de lo diferentes que son los resultados obtenidos por un clasificador y los que se hubieran conseguido clasificando la muestra al azar. Sin embargo, el rigor con el que se utiliza kappa ha sido fruto de discusiones en la literatura en los últimos años. De hecho, no hay una escala estándar que permita interpretar los valores de kappa en este campo de aplicación. En la literatura se encuentran dos referencias: la escala propuesta por Landis y Koch (1977) y la de Fleiss et al. (1981), colega de Cohen y co-autor de varios de sus artículos. Los primeros establecen seis intervalos: $\kappa < 0.00$, acuerdo de mala calidad; $0.00 \leq \kappa \leq 0.20$, acuerdo leve; $0.21 \leq \kappa \leq 0.40$, acuerdo aceptable; $0.41 \leq \kappa \leq 0.60$, acuerdo moderado; $0.61 \leq \kappa \leq 0.80$, acuerdo alto; $0.81 \leq \kappa \leq 1.00$, acuerdo casi perfecto. La escala de Fleiss, más simplificada, propone sólo tres intervalos: $\kappa < 0.4$ indica una clasificación de mala calidad; $0.4 \leq \kappa \leq 0.75$ indica una calidad entre aceptable y buena; $0.75 < \kappa \leq 1$ indica que hay un acuerdo excelente entre la clasificación y la realidad. En esta tesis se utilizará como referencia la escala de Landis y Koch (1977) por ser más detallada.

Foody (2004) ofrece una reflexión muy completa sobre este tema y alerta de que hay que tener en cuenta que la comparación directa del coeficiente kappa de Cohen debe hacerse únicamente cuando los modelos se han obtenido con muestras independientes. Esto no suele ocurrir en la mayoría de los estudios de clasificación en el campo de la Teledetección y tampoco ocurre en muchas de las comparaciones de esta tesis, pues los diferentes modelos se entrenan y se verifican usando las mismas bases de datos. En estos casos, debería hacerse una comparación de los coeficientes kappa de cada modelo a través de un contraste de hipótesis que permita analizar el nivel de significación de las diferencias de kappa. Como este proceso es costoso y probablemente no arroja un veredicto diferente a la hora de elegir el mejor modelo, los autores suelen hacer la comparación directa, a pesar de que sus modelos no cumplen la condición de independencia. En este trabajo se sigue la recomendación de Foody (2004) de utilizar otras métricas de verificación para comparar modelos y elegir el mejor (como la precisión global o el F1-Score). En cualquier caso, kappa es un buen indicador de la fiabilidad de un modelo y es correcto emplearlo de forma directa si no se compara con otros.

También se calculan las curvas ROC, que representan el TP Rate frente al FP rate para diferentes umbrales de discriminación. Aunque solo considera dos métricas sencillas, es un indicador interesante porque la curva ROC es insensible a distribuciones desequilibradas (*unbalanced datasets*) de muestras por clase (Fawcett, 2006). Derivada de la curva ROC, se calcula el área bajo la curva (*Area Under the Curve*, AUC) para tener un escalar que facilite la interpretación.

Igual que las otras métricas, cuanto más próxima a 1 sea el AUC, mejor calidad tiene el modelo de clasificación.

Finalmente, se calculan métricas de paisaje para cuantificar las características espaciales de los segmentos de cada clase, lo que proporciona una potente herramienta de comparación (Herold et al., 2003). Las métricas se calculan empleando datos LiDAR de alta resolución previamente clasificados. Para clasificar el LiDAR se utiliza el algoritmo TOD (*TIN for Object Detection*, Torres et al., 2019a), que es la evolución de Superficie Mágica (Arranz, 2013). TOD, además de filtrar los puntos del suelo, es capaz de detectar y clasificar puntos que pertenecen a objetos, como edificios, árboles, vegetación (en distintas alturas), puentes, cuerpos de agua, etc. El algoritmo aplica técnicas de densificación de TIN para calcular y refinar de manera iterativa el MDT con los puntos del suelo; y técnicas de comparación de TIN para clasificar los diferentes objetos. Los parámetros más importantes que se deben establecer son la altura mínima de vegetación para cada tipo (baja, media y alta), la altura máxima de vegetación y el área mínima de los edificios. La calidad de la clasificación de puntos LiDAR que proporciona TOD ha sido evaluada a través de una plataforma creada a tal efecto por la *International Society of Photogrammetry and Remote Sensing* (ISPRS). El algoritmo queda segundo en el ranking, en el que participan los mejores algoritmos de clasificación LiDAR a nivel internacional, con un indicador de error porcentual del 5.4% (de cada 1000 puntos, clasifica erróneamente 54).

Con los puntos clasificados, se crean imágenes ráster de los segmentos en las que se pueden distinguir edificios, vegetación y suelo (terrenos, caminos, calzadas, jardines, etc.). Sobre estas imágenes se calculan tres métricas de paisaje (McGarigal et al., 2012):

- Densidad de clases (parches), que es el número de objetos de una clase dividido entre el área de estudio total.
- Distancia euclídea a los vecinos más cercanos de la misma clase, que informa acerca de la conectividad de los objetos.
- Índice de forma (Ecuación 2), que relaciona el perímetro de cada objeto (p) con el perímetro que tendría un cuadrado (forma regular) con su misma área (A):

$$I_f = p / 4 \cdot \sqrt{A} \quad \text{Ecuación 2}$$

La evaluación de la clasificación de los segmentos en patrones urbanos se realiza a través de estas métricas de paisaje para eliminar cualquier sesgo que haya podido introducir al etiquetar manualmente los segmentos de la base de datos de verificación. Las métricas proporcionan una descripción independiente, cuantitativa y fiable de cada segmento, que sirve para comparar entre ellos y dilucidar si dos segmentos clasificados como pertenecientes al mismo patrón urbano (por ejemplo, Residencial), verdaderamente lo son al tener métricas similares. La calidad de la clasificación se evalúa por medio del coeficiente de silueta (*silhouette coefficient*) que mide cómo han quedado separados los segmentos de una clase respecto a los de otra, comparando la distancia entre agrupaciones o clases (*clusters*). El coeficiente de silueta compara la distancia media entre una muestra y las demás muestras de la misma clase, con la

distancia media entre esa muestra y las demás del *cluster* más cercano. El coeficiente varía entre -1 y 1; los valores altos indican separaciones más correctas, y viceversa.

4.2.2 Fase 2: Creación de la base de datos de Exposición

La segunda fase consiste en crear la base de datos de exposición. Se trata de una base de datos espacial, donde cada registro almacena información sobre un edificio, tanto su geometría como otros atributos relevantes en los que basar la posterior asignación de vulnerabilidad. Como ya se ha comentado, esta base de datos es un resultado valioso por sí mismo, ya que proporciona una caracterización de los edificios que, si es lo suficientemente amplia, puede usarse para asignar la vulnerabilidad ante otros tipos de peligros naturales (inundaciones, incendios, tsunamis, etc.). Se crean, de esa manera, bases de datos de exposición multi-riesgo o multi-amenaza. En esta tesis, los atributos que se calculan están relacionados directamente con la vulnerabilidad sísmica, ya que es el objetivo del estudio.

Esta fase se divide, a su vez, en dos: generación de las huellas de los edificios y cálculo de los atributos. Ambos resultados han de ser validados. A continuación, se explica con detalle cada una de ellas.

4.2.2.1 Generación de las huellas de los edificios

La huella de un edificio (*building footprint*) es el contorno de su vista en planta. Por tanto, la generación de las huellas de los edificios de una ciudad consiste en generar un polígono (y solo uno) que reproduzca la geometría de cada edificio lo más fielmente posible. Como se expuso en la sección 2.3, en algunos escenarios es posible encontrar bases de datos oficiales, generalmente de catastro o cartográficas, que ya contengan esta información (*ancillary data*). Si no es así, se pueden consultar otras bases de datos cartográficas abiertas, como OpenStreetMap.

Si no es posible conseguir las huellas de otras fuentes de datos, es necesario generarlas. Para ello, se cuenta con dos alternativas, que se analizan comparativamente en esta tesis: digitalización manual y segmentación de imagen. La digitalización manual se realiza teniendo como referencia ortofotos o imágenes satelitales de alta resolución, que permitan distinguir con claridad los objetos que configuran el entorno urbano. Estas deben corregirse previamente de posibles distorsiones y de falta de alineamiento con otros datos, como LiDAR.

En cuanto a la segmentación de imagen, se utilizan los dos algoritmos descritos en la sección 4.2.1.1. aplicados a imágenes de alta resolución. El proceso de segmentación, que se realiza en cada estrato por separado, es muy delicado cuando se trata de delinear objetos tan definidos como los tejados. En pocas ocasiones, el resultado es un polígono o segmento por objeto, que sería lo deseable. Incluso habiendo aplicado un meticuloso proceso de ajuste de parámetros de escala y unión, es difícil evitar problemas de sobre-segmentación (más de un polígono por edificio) o infra-segmentación (un polígono engloba dos o más edificios). Ante este hecho, es inevitable el proceso de depuración de segmentos en el SIG para suavizar los contornos y eliminar polígonos de áreas más pequeñas que el área mínima de los edificios del estrato correspondiente.

Una vez que las imágenes están segmentadas, es necesario seleccionar los segmentos que se corresponden con tejados y desechar aquellos que representan cualquier otro objeto (árboles, viales, jardines, etc.). La forma tradicional de acometer esta selección es clasificar la imagen, lo que da lugar a un proceso completo de OBIA, como el descrito en la sección 4.2.1.2. Clases habituales son tejado (distinguiendo entre los materiales de la cubierta, como tejado de hormigón, tejado de teja, tejado de metal, etc.), vegetación, agua, sombra, asfalto y suelo desnudo. Este proceso de clasificación de imágenes de gran detalle puede resultar costoso, tanto por la necesidad de crear bases de datos de entrenamiento y verificación para cada clase y en cada estrato, como por el propio análisis predictivo, que requiere ajustar todos los parámetros de cada modelo que se desea calcular por medio de validaciones cruzadas.

En este estudio se realiza una fusión del LiDAR clasificado y la imagen para conseguir mejorar los resultados de la segmentación y la posterior clasificación. Para ello, se añaden dos bandas más a la imagen con información continua, una con las alturas de los objetos sobre el suelo y otra con la intensidad. Además, con el objetivo de aprovechar la información de las bandas del espectro visible (RGB, *red, green, blue*), se incluyen otras 3 bandas con las diferencias entre ellas, tomadas de dos en dos.

Para clasificar las imágenes, en esta fase se opta por crear el modelo predictivo con bosques aleatorios (*Random Forests*. Breiman, 2001). Este método de ensamblado consiste en crear un gran número de sub-grupos de atributos a partir del vector de características. Con cada sub-grupo se crea un árbol de decisión para clasificar los segmentos de la imagen en las distintas clases (explicado con más detalle en la sección 4.2.2.3). El número máximo de atributos que se seleccionan es el parámetro k , que establece el usuario. La selección se hace aleatoriamente para que los árboles no estén correlacionados. El resultado de cada árbol, que es la probabilidad de pertenencia a las distintas clases, se promedia para dar un resultado de clasificación final. Visto desde otra perspectiva, se dice que cada árbol emite un voto para la clase a la que cree que es más probable que pertenezca una muestra. La clase final será aquella que haya conseguido más votos entre todos los árboles del bosque. Breiman (2001) recomienda establecer este parámetro en la raíz cuadrada de la dimensión del vector de características, D . Otros autores recomiendan probar valores equiespaciados desde 2 hasta D , o crear un número inicial de 1000 árboles e ir aumentando este número hasta estabilizar las precisiones de la clasificación (Kuhn y Johnson, 2013). Otros dos atributos clave que es conveniente ajustar son la profundidad de los árboles (f) y el mínimo número de muestras (mM) que quedan separadas en un nodo al cumplirse la condición que este establece.

Una de las ventajas del bosque aleatorio es que es sencillo de implementar; pero sin duda, como demostró Breiman (2001), su potencial radica en que es inmune al sobreajuste y reduce la varianza del modelo ensamblado si se compara con la de los modelos creados individualmente por cada árbol. Tiene sentido aplicar este tipo de modelos cuando el número de atributos del vector de características es muy grande, ya que el proceso de computación es más rápido que usando otros algoritmos, como SVM.

Se propone también una alternativa a la clasificación de la imagen, que consiste en utilizar el LiDAR ya clasificado para la selección de los segmentos que corresponden a edificios. La superposición del LiDAR sobre la imagen segmentada permite esta selección y, además, la identificación o clasificación del resto de objetos. De esta manera, es posible evitar el proceso de clasificación de la imagen, lo que conlleva un importante ahorro de tiempo. Además de esta ventaja, usar el LiDAR presenta otras mejoras con respecto a la clasificación por OBIA que merece la pena destacar. Por un lado, los segmentos correspondientes a vegetación quedan subclasificados por alturas, en vegetación baja, media y alta. Por otro lado, no se pierden segmentos por ser clasificados como sombra, ya que esta clase no existe cuando se trabaja con LiDAR. Y, por último, se identifican segmentos correspondientes a objetos singulares, como puentes o cables.

La superposición del LiDAR sobre los segmentos puede dar lugar a imprecisiones si no se ha tenido la precaución de alinearlos previamente. Un error común es que haya puntos de edificio que caigan sobre segmentos adyacentes a los mismos, como aceras o sombras (Figura 5-8). Si esto ocurre, habrá segmentos seleccionados como edificio que, en realidad, no lo son. Se trata de falsos positivos. En estos casos es necesario hacer cumplir a los segmentos seleccionados una serie de requisitos de forma y altura. Para ello, se calcula el índice de forma con la Ecuación 2. En el caso de edificios cuyo perímetro presente numerosos entrantes y salientes, emplear este cociente puede dar lugar a eliminaciones indeseadas de segmentos que verdaderamente son edificios. En ese caso, se plantea sustituir el perímetro real del edificio por el de su *convex hull*. Con este dato, se desechan aquellos segmentos cuyo coeficiente de forma pase un umbral, alejándose del 1, que representa formas compactas. Este parámetro se combina también con un área mínima que deben tener los segmentos para ser considerados edificios. Por último, se calcula la densidad de puntos que caen en cada segmento y se excluyen los que tienen un valor por debajo de un umbral. Ambos umbrales se estiman por ensayo y error en el SIG.

4.2.2.2 Verificación del número de huellas

La verificación del número de huellas se realiza por objeto, comparando el número de segmentos clasificados como edificios con una base de datos de referencia que contenga las huellas reales de los mismos. Se evalúan los resultados obtenidos tanto por OBIA como con el LiDAR clasificado. La comparación consiste en calcular el número de polígonos segmentados que se superponen espacialmente con cada huella real. Una huella bien determinada arrojaría un resultado de 1. Resultados mayores indicarían sobre-segmentación, y menores, infra-segmentación.

4.2.2.3 Cálculo de los atributos

Una vez que se tienen las huellas de los edificios, ya sea directamente tomadas de una base de datos, o digitalizadas, u obtenidas por segmentación seguida de clasificación o selección, se pasa a dotarlas de los atributos de interés. En esta tesis, como ya se ha mencionado, se propone adoptar la taxonomía que se está definiendo en el marco de proyecto RIESGOS, que es la más

reciente y completa. Esta taxonomía considera 17 atributos relacionados con la vulnerabilidad de un edificio a diferentes riesgos: terremoto, inundación, incendio... (enfoque multi-riesgo).

A partir de observaciones aéreas no es posible conocer algunos de los atributos recogidos en la taxonomía, como, por ejemplo, los materiales de la estructura o los cimientos, ni el sistema estructural de un edificio (*lateral load-resisting system, LLRS*). Sin embargo, sí es posible obtener un gran número de ellos. Los que pueden obtenerse cuando se aplican técnicas de EO y que, de hecho, se han obtenido en esta tesis, son el área, la ubicación, su posición relativa en la manzana, la altura (o número de plantas), los materiales del tejado y la inclinación del mismo, así como otros atributos geométricos derivados de su forma en planta.

Todos los atributos que se calculan forman la base de datos de exposición y se utilizan para asignar la vulnerabilidad sísmica de cada edificio en la fase 3. Además, la incorporación de los atributos a la base de datos permite: (1) cumplir con las indicaciones de las taxonomías de clasificación de edificios, y (2) estimar modificadores de la vulnerabilidad por comportamiento (Martínez-Cuevas, 2014).

A continuación, se exponen todos los atributos calculados junto con su definición, el modo de calcularlos en cada huella y el código de la taxonomía de RIESGOS (en su caso):

1. **Área:** el área se calcula directamente en el SIG.
2. **Localización:** son dos atributos en realidad, ya que son las coordenadas planimétricas (X,Y) del centroide.
3. **Orientación:** en el proyecto RIESGOS, la orientación de un edificio viene definida para cada uno de sus ejes principales con relación a la calle en la que se encuentra (paralelo o perpendicular). En su lugar, en esta tesis se propone utilizar una medida cuantitativa: el acimut del lado más largo del rectángulo que envuelve a cada huella. Varía entre 0° y 180°. Esta es una información sencilla y rápida de interpretar que independiza al edificio del dato de la calle (que puede tenerse o no), aunque podría fácilmente relacionarse con él para adaptarse a la nomenclatura de la taxonomía de RIESGOS.
4. **Forma en planta:** RIESGOS propone varios indicadores cualitativos y cuantitativos. Aquí se han considerado dos de este último grupo:
 - 4.1. **Elongación:** se obtiene como la relación entre el ancho y el largo de la huella. Si esta tiene una forma irregular, que es lo común, se utiliza el rectángulo envolvente orientado. Varía entre 0 y 1: las huellas muy elongadas tienen valores próximos a 0, y las formas más parecidas a un cuadrado, tienen valores próximos a 1.
 - 4.2. **Grado de compacidad** de la forma en planta: el grado de compacidad se mide comparando el área del edificio con el área de un círculo que tenga el mismo perímetro (Cerri, 2018; Schröter et al., 2018) lo denomina *elongation index*). Varía entre 0 y 1, y se calcula con la Ecuación 3:

$$G.C. = \text{Área} \cdot 4 \cdot \pi / \text{perimetro}^2$$

Ecuación 3

5. **Posición relativa en la manzana:** considera si un edificio es exento o está adosado por un lateral (pareado), por dos o por tres. Los códigos correspondientes de la taxonomía de RIESGOS son, respectivamente: BPD, BP1, BP2 y BP3. Para la búsqueda de edificios adyacentes, se realiza una consulta espacial empleando como distancia de búsqueda el 4% de la altura del edificio (según recomiendan los autores).
6. **Altura (y número de plantas):** se obtiene como la mediana de todos los puntos del MDSn que caen dentro de la huella. A partir de la altura (HHT, *total height from ground*), se deriva el número de plantas (HAPP, *approximate nr. stories*) dividiendo aquella entre la altura típica de una planta en cada estrato. El dato de número de plantas, además de la altura, es importante y así lo reflejan las taxonomías de clasificación de edificios, ya que la mayoría de los modelos de vulnerabilidad dividen las construcciones en rangos de altura en *low-*, *mid-*, y *high-rise*. Dependiendo del modelo de vulnerabilidad a adoptar, el número de plantas que corresponde a cada rango varía. En el escenario de Lorca, este dato también se podría haber tomado del Catastro, pero se ha preferido probar este método por si fuera necesario realizar actualizaciones.
7. **Tejado:** RIESGOS propone varios niveles de información para caracterizar el tejado, ya que es un elemento crucial a la hora de evaluar el comportamiento de un edificio ante la mayoría de los riesgos naturales. Aquí se han considerado dos:
 - 7.1. **Inclinación del tejado:** es la media de las pendientes de todos los triángulos de un TIN que se forma con los puntos LiDAR. Las pendientes se ponderan por el área de los triángulos para minimizar la influencia de triángulos pequeños (Ecuación 4).

$$Pte_t = \frac{\sum_{i=1}^n Pte_i \cdot A_i}{\sum_{i=1}^n A_i}$$

Ecuación 4

Donde Pte_t es la pendiente del tejado calculada con los n triángulos de su TIN. Pte_i y A_i son la pendiente y el área de cada triángulo, i .

La pendiente del tejado permite derivar el atributo **forma del tejado (roof shape)**, que en la taxonomía de RIESGOS puede tomar hasta 10 valores diferentes. En este estudio se han considerado dos: tejado plano (RSH1) y tejado inclinado sin definir (RSH0, que corresponde a “otros”).

- 7.2. **Material del tejado (roof covering):** de los doce valores propuestos en RIESGOS, en los dos escenarios de aplicación de esta tesis se dan tres: losa de hormigón (RMN), teja de arcilla (RMT1) y lámina metálica (RMT6). Para obtener este atributo hay dos alternativas, dependiendo del proceso seguido en el paso anterior. Si se ha clasificado

la imagen para extraer los segmentos que representan edificios, ya se tiene el material de la cubierta del tejado. Si se han seleccionado los segmentos usando el LiDAR, este atributo aún no está en la base de datos. En este caso hay que derivarlo por correlación con los otros atributos, empleando alguna técnica de análisis predictivo, de manera que la variable dependiente es el material del tejado y las variables predictivas son la inclinación del mismo, la altura del edificio, la localización, etc. Las técnicas predictivas que se emplean en esta tesis son SVM, un árbol de decisión y una red Bayesiana, que representan tres conceptos diferentes en el campo del aprendizaje automático: basado en funciones, basado en árboles y probabilístico, respectivamente. SVM se describe en la sección 4.2.1.2.

El árbol de decisión utilizado es el creado con el algoritmo C4.5 de Quinlan (1993). Un árbol de decisión es una representación jerárquica de condiciones definidas a partir de los valores de los atributos de las muestras de entrenamiento (Figura 4-3 (A)). Esta jerarquización subdivide la base de datos de entrenamiento en grupos más pequeños y puros. Los nodos que se forman tras una subdivisión contienen una proporción mayor de instancias de una misma clase (Kuhn y Johnson, 2013). La pureza de un nodo se puede medir con distintas métricas; en esta tesis se emplea el índice de impureza de Gini, que varía de 0 a 1, siendo menor cuando mejor separa las muestras. Cada una de las clases cumple las condiciones establecidas en el nodo, que están organizadas desde la “raíz” del árbol, pasando por todos los nodos “hijos”, hasta las “hojas”, que son las clases. En los nodos raíz e hijos se evalúan los valores de los atributos y en los nodos hoja se expresan las probabilidades finales de pertenecer a una determinada clase. El parámetro clave para configurar un árbol de decisión es el factor de confianza, f , que regula el tamaño (la profundidad o frondosidad) del árbol. Cuanto mayor es f , más profundo es el árbol.

Figura 4-3. Esquema de un árbol de decisión (A) y de una red Bayesiana (B).
En (A), el nodo azul es la raíz, los verdes son los nodos hijos y, los granates, son las hojas del árbol.
En (B), el nodo granate es el nodo raíz y los verdes son nodos hijos y padres.

Una red Bayesiana es un modelo estadístico multivariante, que se representa por medio de un conjunto de nodos y arcos que forman un grafo acíclico dirigido, DAG (*Directed Acyclic Graph*. Pearl, J. 1985; Ben-Gal, 2007. Figura 4-3 (B)). El nodo raíz es la MBT y los demás nodos son los atributos de los edificios. Los arcos son dirigidos, es decir, van uniando los nodos dos a dos siempre y cuando exista una relación de influencia entre ellos, siendo esta influencia la que marca el sentido del arco. El nodo del que parte el arco es padre del nodo al que llega, que es su hijo. En la Figura 4-3 (B), el nodo del atributo 2 es padre del nodo del atributo 3. Que el grafo sea acíclico significa que un camino formado por arcos nunca parte y llega al mismo nodo. Las relaciones entre los atributos que forman la red se expresan en términos de probabilidad condicionada de los nodos, dados sus padres. Esta probabilidad se puede establecer a partir de los valores de los atributos de las muestras de entrenamiento (previamente discretizados) y se representa por medio de tablas de probabilidad condicionada. La probabilidad de que el nodo raíz tome un valor se calcula por el Teorema de Bayes a través de las probabilidades de los demás nodos que están conectados hasta llegar a él. Para crear la red, en esta tesis se prueban dos algoritmos: K2 y TAN. K2 es un algoritmo de tipo *hill climbing*, pero que añade arcos con un orden de atributos fijado (Cooper y Herskovits, 1992). El parámetro clave para crear la red con K2 es el máximo número de padres que puede tener un nodo, p . Cuando $p=1$, se está creando un clasificador Bayesiano ingenuo (también llamado *naïve* o simple). TAN (*Tree Augmented Naïve Bayes*) es un clasificador *naïve* aumentado a árbol. El árbol se crea calculando el máximo árbol de expansión con los arcos ponderados (Friedman et al., 1997). Este algoritmo no requiere parametrización. Las redes Bayesianas han demostrado ser un clasificador potente en este campo de aplicación, a la vez que flexibles y sencillas (e.g. Li et al., 2010; Pittore y Wieland, 2013).

4.2.2.4 Verificación de los atributos

Una vez que se han calculado los atributos, pueden verificarse aquellos que están contenidos en la base de datos de referencia: área, número de plantas y material del tejado. El área se verifica comparando los valores de área total construida estimada y observada (la observada la proporciona la base de datos de referencia). Se da el error porcentual que supone esta diferencia. El número de plantas se verifica a través del histograma de los residuos, de nuevo, comparando el número de plantas de cada edificio estimado y observado. También se proporciona la precisión global. Y, finalmente, el material del tejado se valida mediante las métricas de calidad derivadas de la matriz de confusión, tras aplicar los modelos predictivos.

4.2.3 Fase 3: Clasificación de tipologías y asignación de la vulnerabilidad sísmica

4.2.3.1 Elección del modelo de vulnerabilidad y clasificación de tipologías constructivas.

La asignación de vulnerabilidad sísmica consiste en asociar cada tipología edificatoria (MBT) con el modelo de vulnerabilidad que mejor represente su desempeño sísmico. Para ello, los

edificios de la base de datos de exposición creada en el paso anterior son clasificados en MBT según los valores de sus atributos. La elección del modelo de vulnerabilidad que se va asignar a los edificios (empírico, analítico o híbrido, sección 2.1) depende, principalmente, de los objetivos finales del estudio y de las características de los edificios presentes en el escenario de aplicación.

En esta tesis, el problema de clasificación de los edificios en MBT se resuelve con los siguientes modelos predictivos, aplicados por estrato: regresión logística, redes Bayesianas, SVM, y árboles de decisión. Excepto la regresión logística, los demás modelos han sido explicados en apartados anteriores. La regresión logística predice la presencia o ausencia de un valor de una variable (dependiente, que es la MBT) a partir de los valores de otras variables (independientes o predictores, que en este caso son los atributos de los edificios). El término de la izquierda del modelo de regresión logística (Ecuación 5) es el *logit* y su argumento es la ODDS ratio, que es la probabilidad de que la variable MBT tome un valor frente a que tome el otro (Kuhn y Johnson, 2013; Field, 2013). El término de la derecha es similar al del modelo de regresión lineal y es sencillo de interpretar.

$$\ln \left[\frac{P[Y = 1 | x_1, x_2, \dots, x_n]}{1 - P[Y = 1 | x_1, x_2, \dots, x_n]} \right] = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n \quad \text{Ecuación 5}$$

El cálculo de los coeficientes de la regresión logística ($\beta_0, \beta_1, \dots, \beta_n$) se hace por el método de máxima verosimilitud, que requiere de un proceso iterativo. Este proceso permite obtener, además de los coeficientes, sus errores estándar y las covarianzas entre las variables independientes del modelo (atributos de los edificios, X_i). La decisión de cuántos atributos introducir en el modelo y en qué orden, se hace, en este estudio, a través del método por pasos hacia delante (Wald), que contrasta la entrada de un atributo basándose en la significación del estadístico de puntuación y decide si lo elimina o lo mantiene basándose en la probabilidad del estadístico de Wald (IBM SPSS Regression 22, 2014). Puede ocurrir que el mejor modelo (el que ofrece más precisión a la hora de clasificar los edificios en MBT) no sea aquel en el que intervienen todos los atributos, sino el que se forma con los atributos que mayor influencia tienen en la variable dependiente (MBT).

Este clasificador no tiene parámetros que ajustar, lo que facilita su implementación. El uso de un modelo de regresión logística no es muy común en este campo, a pesar de su sencillez, robustez y facilidad de interpretación (Kuhn y Johnson, 2013). Un trabajo de referencia es el realizado por Sarabandi et al. (2008) para clasificar edificios según su tipología estructural en California, empleando regresión logística.

4.2.3.2 Verificación de los resultados de clasificación de los edificios en MBT

Los resultados de clasificación son verificados a través de la matriz de confusión cuando se aplica un modelo predictivo. Esto es posible cuando la base de datos de verificación tiene edificios etiquetados con su MBT en su verdadera posición. De esta forma, puede establecerse una correspondencia unívoca entre el edificio de esa base de datos y el edificio de la base de datos generada en el estudio, para comparar la MBT real u observada con la estimada. Sin

embargo, no es extraño encontrar bases de datos de verificación donde la posición de los edificios es aproximada debido a, por ejemplo, haberse creado a partir de registros tomados con dispositivos GPS de mano (e.g. Geiss et al., 2015).

Bloque 3

5 Aplicación del procedimiento diseñado en dos escenarios

5.1 Aplicación en Lorca, España

Como se ha mencionado en la introducción (sección 1), el interés por este escenario surgió a raíz del terremoto de 2011, de Mw 5.1, que ha sido el más dañino del último medio siglo en España (Cabañas et al., 2011). Fue precedido por otro de Mw 4.5 que debilitó las estructuras y alertó a la población. La mayoría de los edificios de la ciudad sufrieron daños ligeros o moderados (Figura 5-1), y se registró el colapso de un edificio moderno (Feriche et al., 2012). El hipocentro, muy superficial (5 km), se localizó en la falla de Alhama de Murcia (Martínez-Díaz y Álvarez-Gómez, 2011), a unos 5 km al noreste de Lorca. Según Álvarez-Cabal et al., 2013, el terremoto de Lorca produjo pérdidas económicas de más de 460 M €.

El terremoto causó daños graves en un 4% de las edificaciones de la ciudad, y daños moderados en el 6% (Quirós, 2017); muchas de esas edificaciones estaban construidas al amparo de normativas sismorresistentes. España publicó el primer código sísmico en el año 1962, y desde entonces ha habido actualizaciones en 1968, 1974, 1994 y 2002 (Cabañas et al., 2011). A pesar de que esto ha permitido dotar a las construcciones de la calidad y los sistemas modernos necesarios para resistir los terremotos a los que se iban a ver sometidas a lo largo de su vida útil, el terremoto de Lorca, con su patrón de daños y las víctimas, cuestiona si el nivel de diseño sismorresistente de las edificaciones expuestas resulta (o no) adecuado, máxime teniendo en cuenta que se trató de un sismo de magnitud moderada. De alguna forma, el sismo nos enseñó que aún queda mucho por hacer.

Figura 5-1. (A) Derrumbamiento de la nave de crucero de la iglesia de Santiago. (B) Edificio colapsado en la calle Infante Juan Manuel.

(A) Fotografía tomada de Cabañas et al. (2011). (B) Fotografía propia.

La aplicación en Lorca está enfocada a diseñar un procedimiento que sea rápido de implementar, replicable y escalable, de cara a la realización de los más de 700 planes de riesgo sísmico municipales que se deben completar en los próximos meses-años para cumplir con el requerimiento de la Directriz Básica de Planificación de Protección Civil (DBPC, 2004). Por tanto, se propone incorporar datos existentes que sean de utilidad y que estén disponibles para todo el territorio español. Asimismo, se pretende crear flujos de trabajo eficaces y rápidos, minimizando el refinamiento manual en post-proceso de los resultados intermedios.

5.1.1 Área de estudio y datos

El municipio de Lorca tiene una población de casi 93.100 habitantes¹². Está localizado en una zona con actividad sísmica moderada, que ha sufrido terremotos de intensidad VIII-IX EMS-98 en el periodo histórico, y de magnitud máxima registrada de 5.0 en el periodo instrumental (Cabañas et al., 2011). Un análisis de todos los terremotos registrados en los alrededores de Lorca muestra que se han producido 53 sismos de intensidad EMS-98 mayor de V, es decir, que han causado algún grado de daño. Otros terremotos ocurridos en la provincia de Murcia que han producido daños en las últimas décadas son Mula (1999), Bullas (2002) y La Peca (2005). Una posible explicación para esta alta frecuencia de ocurrencia de sismos próximos es un mecanismo de disparo entre eventos (Benito et al, 2007). Este hecho hace aumentar la probabilidad de ocurrencia de terremotos futuros, lo que a su vez aumenta la peligrosidad sísmica y, por tanto, el riesgo.

Tras el terremoto de 2011, el área de Lorca ha sido, y sigue siendo, foco de investigación. Los trabajos de Martínez-Cuevas y Gaspar-Escribano (2016) y de Martínez-Cuevas et al. (2017) son una referencia para esta tesis, ya que de ellos se ha tomado la base de datos de vulnerabilidad de 826 edificios que se ha empleado como dataset de verificación o *ground truth*. Los edificios están localizados en tres áreas distintas de la ciudad (Figura 5-2, zonas Norte, Centro y Sur), y están caracterizados por su vulnerabilidad siguiendo el modelo LM2 de Risk-UE (Milutinovic and Trendafiloski, 2003). De acuerdo con los autores, los edificios de Lorca pueden clasificarse en seis tipologías constructivas (MBT, *Model Building Types*): una con estructura de hormigón armado (*reinforced concrete*, RC) y cinco de mampostería (*masonry*, M). De estas seis, cuatro se encuentran en las zonas de estudio (Figura 5-3). Los edificios de RC predominan en el barrio de San Diego (zona Norte) y en el barrio de La Viña (zona Sur), que son de construcción más reciente (Figura 5-2). Estos edificios tienen un sistema estructural de pórtico de hormigón armado con muros de mampostería sin reforzar y, por tanto, se asocian con la tipología RC1 de Lagomarsino y Giovinazzi (2006). Generalmente, son edificios multifamiliares con más de 3 plantas.

Los edificios de mampostería, que son escasos en dichos barrios, predominan, sin embargo, en la zona del centro histórico (Figura 5-2, zona Centro). Aquí se encuentran tres de las cinco tipologías de mampostería que hay en la ciudad. En las zonas del centro hay edificios antiguos con muros portantes de mampostería de piedra irregular (tipología M1.1), y otros de mampostería sin reforzar con forjado de madera o de hormigón (tipologías M3.1 y M3.4, respectivamente). Una característica de este tipo de edificios es que no se han hecho intentos para mejorar los elementos horizontales de su estructura, siendo los pisos generalmente de madera y, por lo tanto, flexibles. Son edificaciones unifamiliares, de baja altura (2-3 plantas), regulares en planta y en alzado, que se asociarían al modelo de vulnerabilidad M3 de Lagomarsino y Giovinazzi (2006).

¹² Censo INE 2008: <http://www.ine.es/jaxiT3/Datos.htm?t=2883>.

Figura 5-2. Zonas de estudio en la ciudad de Lorca.

Martínez-Cuevas et al. (2017) agruparon estas tres MBT de mampostería en un único grupo, motivados por dos factores: por un lado, las tres tipologías corresponden a edificios similares de mampostería sin reforzar con geometría regular; y por otro, el número de edificios de cada tipología no era suficiente para crear tres muestras separadas en cada área de estudio, como puede comprobarse en la distribución de tipologías de la Tabla 5-1. Finalmente, para este estudio quedan dos tipologías, RC3.1 (que se abrevia a RC) y M. Esta se puede considerar una primera clasificación de edificios que está en concordancia con las propuestas de SYNER-G (SYNER-G D3.1, 2011; SYNER-G D3.2, 2011) y Risk-UE (Milutinovic y Trendafiloski, 2003). En estas metodologías, la clasificación de edificios empieza por diferenciar de manera genérica ambos tipos de estructuras. A partir de ahí, se puede ir afinando la clasificación de cada edificio a medida que se van teniendo otros atributos, como la altura, los materiales de los muros, el sistema estructural, etc.

Tabla 5-1. Distribución de MBT en las áreas de estudio. Datos tomados de la base de datos de referencia.

Nº Edificios	Norte	Sur	Centro	Total
M1.1	3	3	37	43
M3.1	4	0	92	96
M3.4	2	1	36	39
RC1	79	360	209	648
Total	88	364	374	826

A. Edificio de muros portantes de mampostería de piedra irregular (M1.1)

B. Edificio de mampostería sin reforzar con forjado de madera (M3.1)

C. Edificio de mampostería sin reforzar con forjado de hormigón (M3.4)

D. Edificio de pórtico de hormigón armado con muros de mampostería sin reforzar (RC1)

Figura 5-3. Construcciones típicas en Lorca.
Fotografías cedidas por Sandra M. Cuevas.

Otro trabajo relevante realizado en estas mismas zonas de estudio es el de Quirós (2017), que planteó un procedimiento metodológico para estimar el riesgo sísmico a nivel municipal, realizando diversas combinaciones de escenarios sísmicos, modelos de atenuación y métodos de cálculo de daños. Simuló dos sismos: uno cercano de Mw 5.2, para un periodo de retorno PR=475 años, muy similar al ocurrido en 2011; y otro a 15 km, de Mw 6.3, para un PR = 975 años. En la estimación de daños del primero, obtuvo que podrían colapsar unos 14 edificios en las tres zonas de estudio (Norte, Centro y Sur) y 75 experimentarían daño extenso. Estos resultados concuerdan con los daños reales registrados por el Ayuntamiento de Murcia y por la Delegación de Gobierno de la Región de Murcia tras el sismo.

En esta aplicación, se trabaja en áreas de estudio con diferente extensión (Figura 5-2). La fase 1 de estratificación se realiza en toda la ciudad (1100 Ha), mientras que el resto del análisis (fases 2 y 3) se ciñe a las tres zonas que cubre la base de datos de verificación, con el objetivo de poder comprobar los resultados: zonas Norte (9 Ha), Centro (15 Ha) y Sur (18 Ha). El resto de la ciudad queda estratificado para continuar con la clasificación de edificios en el futuro.

Datos

Además de la base de datos de verificación, se emplean en esta aplicación datos incorporados de fuentes oficiales y otros tomados por sensores remotos. A continuación, se describen todos ellos.

1. Una imagen Landsat8 de 11 bandas, tomada el 3 de marzo de 2015 por los sensores OLI y TIRS, que registran las bandas 1-9 y 10-11, respectivamente. Esta imagen, que se utiliza en la Fase 1 de estratificación de la ciudad, se ha descargado del repositorio del USGS (*EarthExplorer*). Está corregida radiométrica y geométricamente, además de por las distorsiones del relieve. Todas las bandas están en formato GeoTiff de 16 bits y tienen una resolución geométrica de 30 m, excepto la banda pancromática (banda 8), cuya resolución es de 15 m (Data-Landsat). El RMSE en X es 4.621 m, y en Y es 4.501 m. El sistema de referencia original de esta imagen es WGS-84.
2. En las fases 1 y 2 también se utilizan ortofotos capturadas en 2013 por el PNOA, con una cámara digital de alta resolución, equipada con un sensor pancromático y cuatro sensores multiespectrales. Estas imágenes tienen 4 bandas, 3 en el rango del visible y 1 en el infrarrojo cercano (RGB-NIR). La resolución radiométrica es de 16 bits, la geométrica es de 50 cm y el RMSE planimétrico es menor de 1 m. Estas imágenes están en el sistema de referencia ETRS89-UTM30N.
3. En la fase 2, se incluye también una nube de puntos LiDAR registrada en el mismo vuelo fotogramétrico. La densidad media es de 0.5 - 1 pto/m². Los RMSE planimétrico y altimétrico de estos puntos son 30 cm y 20 cm, respectivamente. Los puntos están en el sistema de referencia ETRS89-UTM30N, y las altitudes son ortométricas (transformadas desde el elipsoide GRS80 con el modelo EGM2008-REDNAP) (Data-PNOA).
4. Finalmente, en la fase 2, se han incorporado las huellas de los edificios tomadas de la Sede Virtual del Catastro.

Todos estos datos han sido integrados al SIG diseñado para la aplicación y transformados al sistema de referencia oficial de España: ETRS98-UTM30N. El software de SIG utilizado en esta tesis es ArcGIS de ESRI.

5.1.2 Proceso de cálculo diseñado para Lorca. Resultados y análisis.

La Figura 5-4 muestra en rojo las técnicas seleccionadas para el escenario particular de Lorca. Esta selección, además de otras decisiones tomadas en referencia a los procesos de cálculo, se ha hecho teniendo en cuenta el objetivo del estudio en este escenario, los datos disponibles, las tipologías de edificios existentes y la configuración de la ciudad. A continuación, se presentan los procesos de cálculo, así como los resultados y su análisis, de cada fase del estudio.

5.1.2.1 Fase 1: estratificación de la ciudad de Lorca

Se han identificado seis patrones urbanos en la ciudad de Lorca, que se muestran en la Figura 5-5:

1. Residencial: en este patrón se encuentran, por lo general, viviendas unifamiliares adosadas, pareadas o independientes, de no más de 2 o 3 alturas. Su huella tiene un área grande, de más de 100 m², y están rodeadas de zonas verdes con o sin piscina. Sus tejados, de teja en su mayoría, son complejos, formados por numerosas limas, lo que da lugar a combinaciones de faldones iluminados y en sombra. Su trama viaria es regular, con calles anchas y rectas.
2. Centro A: es un patrón de casco histórico, típico de las ciudades españolas del sur. Está formado por edificios multifamiliares adosados de más de 3 alturas, formando manzanas urbanas. Los tejados suelen ser azoteas planas con elementos de tipo cobertizo, aunque también se encuentran tejados de teja. El patrón Centro A se diferencia del Centro B, principalmente, en que el área de sus construcciones es mayor. La red viaria en este patrón sigue siendo regular, aunque sus calles empiezan a hacerse más sinuosas hasta fusionarse con el Centro B.
3. Centro B: es también un patrón de casco histórico, pero con prevalencia de viviendas adosadas unifamiliares antiguas y pequeñas, que forman manzanas irregulares. La mayoría de los tejados son azoteas planas de geometría irregular, con áreas que apenas llegan a los 100 m². Las calles son sinuosas y estrechas, y suelen estar en pendiente.
4. Ensanche: son las áreas de ampliación de la ciudad. Representa un patrón urbano más moderno, con edificios grandes y regulares, que forman manzanas rectangulares. En estas áreas se ubican los edificios más altos de Lorca y, por tanto, es común encontrar sombras proyectadas de unos edificios sobre los tejados de otros. Se pueden ver tejados inclinados de teja, pero los de construcción más reciente son planos. Los viales son anchos y rectos. Las áreas de ensanche se disponen alrededor del casco histórico, como pasa en otras muchas ciudades.

Figura 5-4. Proceso metodológico aplicado en el escenario de Lorca.

5. Rural: está formado por construcciones aisladas rodeadas de campos de labor o terreno desnudo. Algunas de estas construcciones son viviendas ocupadas ocasionalmente, de 1 o 2 plantas, y con tejados inclinados de teja. Suelen disponerse dentro de parcelas valladas de formas geométricas regulares. Aunque algunas tienen áreas pequeñas, la mayoría supera los 100 m². Las diferencias con el patrón Residencial son que las construcciones están más distantes entre sí y que el terreno que las rodea no está verde en todas las épocas del año debido a la estacionalidad de los cultivos. La red viaria es escasa y está formada principalmente por caminos sin asfaltar.
6. Industrial: es un patrón de construcciones rectangulares que se caracterizan por tener áreas grandes (más de 500m²) y cubiertas de materiales metálicos altamente reflectantes. Se disponen en las afueras de la ciudad, formando manzanas grandes y rectangulares separadas por calles muy anchas y rectas. A pesar de ser fácilmente identificables, las instalaciones industriales no suelen incluirse en estos estudios de riesgo sísmico a nivel municipal, como ya se adelantó en el Marco Conceptual (sección 1.1); el análisis de vulnerabilidad de estas construcciones se aborda por separado y exprofeso.

1 - Residencial

2 - Centro A

3 - Centro B

4 - Ensanche

5 - Rural

6 - Industrial

Figura 5-5. Patrones urbanos identificados en la ciudad de Lorca.

Segmentación y clasificación del área de estudio

El proceso de estratificación comienza con la segmentación de la imagen OLI8 de Landsat. Se ha empleado el software ENVI con el algoritmo *Watershed*. Los parámetros elegidos, tras un

proceso de ensayo y error, son escala = 30 y unión = 80, por generar los segmentos que mejor dividen la ciudad en los patrones urbanos identificados. El siguiente paso de clasificación de los segmentos se ha realizado manualmente, dado que la extensión de la ciudad resulta manejable. Realizar una clasificación automática llevaría más tiempo y es innecesaria en este caso.

El resultado se muestra en la Figura 5-6, donde se puede ver una clara diferenciación de zonas por patrones. Las zonas más antiguas, Centro A y Centro B, se localizan en el centro y en el norte de la ciudad. Están rodeadas por los patrones de Ensanche, al noreste y al sur, y por áreas Residenciales, al este. Las zonas Rurales se entremezclan con Residenciales, aunque se identifica un gran *cluster* en el noreste. El tejido industrial está muy localizado, al suroeste.

En la Figura 5-6 se han representado en amarillo las áreas de estudio para las siguientes fases, que se corresponden con la extensión de los datos de referencia. Se aprecia que están distribuidas sobre tres patrones urbanos diferentes: Ensanche (las zonas Norte y Sur), Centro A y Centro B. Por tanto, estos son los patrones con los que se trabaja en esta tesis.

Estratificación de Lorca en Patrones Urbanos

Figura 5-6. Resultado de la estratificación de Lorca en patrones urbanos.

5.1.2.2 Fase 2: creación de la base de datos de exposición de Lorca

Como se ha descrito en la sección 4.2.2, la creación de la base de datos de exposición es un proceso en dos pasos. Primero se generan las huellas de los edificios y, a continuación, se calculan los atributos para cada una de ellas. Esto se realiza por separado en cada estrato: Ensanche (Norte y Sur), Centro A y Centro B.

Generación de las huellas de los edificios

Las huellas de los edificios en este escenario tienen dos orígenes: incorporación de la geometría de las parcelas catastrales y generación de las mismas siguiendo un proceso de segmentación de imagen y clasificación del LiDAR. Para diferenciarlas, los polígonos resultantes de la primera opción se han llamado regulares, y los generados con datos tomados por sensores remotos, irregulares. El cálculo de atributos y la subsiguiente asignación de vulnerabilidad se ha hecho en paralelo para cada conjunto de huellas, con el fin de comprobar hasta qué punto es viable trabajar con huellas generadas en procesos semi-automáticos.

La incorporación de las huellas de catastro ha sido posible en este escenario dado que en España esta información está disponible y es de libre acceso. Además, cubre todo el territorio nacional, lo que permitiría el escalado del flujo de trabajo diseñado en esta tesis a otras ciudades del país. Incorporar una base de datos oficial ya generada es la opción más rápida y precisa. Se ha planteado la opción de incorporar la altura de los edificios del catastro, pero el LiDAR del PNOA se actualiza con más frecuencia que este, y el dato de altura de los objetos se puede derivar con LiDAR con fiabilidad, como se verá posteriormente.

Para la generación de las huellas irregulares, se ha segmentado la imagen del PNOA del 2013 en el software ENVI, con el algoritmo *Watershed*. Para las zonas de Ensanche, una inspección visual permite identificar las mejores segmentaciones, que son las obtenidas con parámetros de escala = 60 y de unión = 90. En las zonas del Centro, la combinación óptima ha sido: escala = 70 y unión = 75. Un rápido post-proceso en el SIG ha permitido suavizar los contornos de los polígonos y eliminar aquellos cuya área no superaba los 5 m², por fusión con sus vecinos mayores.

La selección de los segmentos que representan tejados se ha realizado por superposición del LiDAR, en lugar de clasificando los segmentos mediante la técnica OBIA. Esta decisión se fundamenta en que para Lorca se pretende diseñar un proceso rápido, y como el LiDAR debe ser clasificado para calcular las alturas y las pendientes de los tejados, se aprovecha ese trabajo sin tener que invertir más tiempo en la clasificación de la imagen. Por otro lado, el CNIG ha publicado recientemente un nuevo producto derivado del LiDAR del PNOA con una clasificación por defecto de alta calidad, llamado Mapa LIDAR¹³, lo que permitiría ahorrar el paso de clasificación del LiDAR en sucesivos estudios.

¹³ http://www.ign.es/web/resources/acercaDe/Boletines/Boletin_IGN_mayo_2019.pdf

La clasificación del LiDAR se ha hecho con el software MDTopX que incorpora el algoritmo TOD. Las clases de interés son suelo, edificio y vegetación en tres alturas (baja, media y alta). Las combinaciones de parámetros clave para conseguir una clasificación óptima en cada área de estudio se muestran en la Tabla 5-2. La máxima distancia entre puntos se ha fijado en 3 m.

Tabla 5-2. Combinación de parámetros del algoritmo TOD para clasificar el LiDAR en las distintas áreas de estudio.

Zona	mAVA	mAVM	mAVB	MAV	mArE
Norte	5	1.5	0.5	15	70
Centro A	3	1.5	0.5	10	50
Centro B	3	1.5	0.5	10	30
Sur	5	1.5	0.5	15	70

mAVA, mAVM, mAVB son los parámetros de mínima altura de vegetación alta, media y baja, respectivamente; MAV es la altura máxima de vegetación; y mArE es la mínima área de los edificios, en metros cuadrados. Las alturas están expresadas en metros.

Una vez clasificados los puntos LiDAR, se superpone la nube a la segmentación y se seleccionan aquellos segmentos en los que caen puntos clasificados como edificio (Figura 5-7). Para eliminar los segmentos seleccionados erróneamente (Figura 5-8), se calcula el índice de forma, el área y la densidad de puntos. En la zona Sur ha sido necesario calcular el índice de forma con el perímetro de la *convex hull* en lugar de con el del segmento real. En la Tabla 5-3 se muestran los umbrales de estos indicadores de geometría que han motivado la exclusión de los falsos positivos.

Figura 5-7. Proceso para selección de segmentos correspondientes a edificios. (A) segmentación de la imagen completa; (B) superposición con los puntos LiDAR clasificados como edificios; (C) segmentos que pertenecen a tejados, que han sido seleccionados por superposición con el LiDAR de los edificios.

Tabla 5-3. Valores umbrales de los indicadores de geometría utilizados para refinar la selección de huellas de edificios.

Zona	Índice Forma	Área (m ²)	Densidad Ptos.
Norte	$\geq 1,7$	≤ 70	$\leq 0,2$
Centro A	$\geq 1,5$	≤ 60	$\leq 0,7$
Centro B	$\geq 1,5$	≤ 30	$\leq 0,6$
Sur	$\geq 1,6$	≤ 70	$\leq 0,2$

Figura 5-8. Ejemplos de segmentos seleccionados erróneamente como edificios. (A) segmentos que se desechan utilizando el índice de forma; (B) segmentos que se desechan empleando la densidad de puntos.

Verificación del número de huellas

Una vez generadas las huellas, se han comparado en número con las huellas de los edificios de la base de datos de verificación. Solo se han verificado las huellas obtenidas por segmentación e integración con el LiDAR. La verificación de las huellas tomadas de Catastro carece de interés, al ser oficiales y coincidir con las de referencia. La Tabla 5-4 muestra estas cifras. Como se ve, en todas las zonas de estudio, el número estimado de edificios supera al real. La zona Sur presenta la mejor aproximación, con 72 edificios sobreestimados respecto de un total de 364. En el resto de las zonas, el número estimado de edificios es 3 (Norte), 5 (Centro B) y hasta 9 (Centro A) veces mayor que el número real. Esto se debe a casos de sobre-segmentación, como se muestra en los ejemplos de la Figura 5-9. En el ejemplo E, hay 20 polígonos en el tejado que ocupa la parte central de la imagen; en el F, hay 5 en un tejado y 6 en el otro.

Tabla 5-4. Comparación del número de huellas obtenidas por procedimientos semiautomáticos usando datos tomados por sensores remotos, con el número de huellas reales contenidas en la base de datos de referencia.

	Ensanche		Centro	
	Norte	Sur	Centro A	Centro B
Nº huellas real	88	364	179	195
Nº huellas estimado	237	436	1661	1037
Residuo	149	72	1482	842
Coef. Comparac.	2.7	1.2	9.3	5.3

Figura 5-9. Ejemplos de huellas de edificios obtenidas por segmentación.

A, B, C, D son ejemplos aceptables de segmentaciones en tejados de las zonas Norte, Centro A, Centro B y Sur, respectivamente. E, F son ejemplos de sobre-segmentación en las zonas Centro A y Centro B.

Los resultados de la segmentación se podrían mejorar, pero eso requiere procesos de depuración que pueden resultar largos y delicados (e.g. Geiss et al., 2017). La experiencia demuestra que, incluso después de dedicar tiempo y recursos a refinar la segmentación, o llevando a cabo complejos procedimientos de segmentación anidada, (e.g. Bruzzone y Carlin, 2006; Geiss et al., 2016), nunca se consigue tener un único polígono por edificio, que es lo deseable (Qi et al., 2017). Por otro lado, en este escenario ya se cuenta con huellas oficiales para los edificios, que están disponibles para todo el territorio español en caso de tener que aplicar este método a otras ciudades. Por tanto, para poder terminar la aplicación, no es necesario dedicar más tiempo a la generación de contornos de tejados.

Ante estos resultados, se ha decidido no continuar trabajando con los segmentos irregulares de las zonas Centro A y Centro B, pero sí con los del patrón Ensanche (zonas Norte y Sur), que resultan menos distorsionados. De esta forma se pueden comparar en estas zonas los resultados finales de asignación de vulnerabilidad con los que se obtengan con las huellas oficiales del Catastro. Hay que tener en cuenta que no se podrá utilizar el número de segmentos directamente como el número de edificios, algo que sí puede hacerse con las huellas catastrales. En cambio, sí sería posible trabajar con el área construida (Lang, 2013; Ehrlich et al., 2013).

Cálculo de los atributos

Para cada una de las huellas, se calculan todos los atributos planteados en la sección 4.2.2.3. Algunos atributos geométricos, como el área de cada huella y las coordenadas planimétricas de su centroide (X, Y), son de cálculo directo. La orientación y la elongación se han derivado a partir del ancho y el largo del rectángulo envolvente de cada huella. Para el grado de compacidad, se ha utilizado la expresión de Cerri, 2018 (Ecuación 3). La posición en la manzana se ha calculado mediante una consulta espacial para buscar el número de edificios adyacentes.

La altura se ha calculado en MDTopX. Para ello, se han proyectado los puntos LiDAR clasificados como edificio sobre el MDT, de manera que se restan las altitudes. La mediana de las alturas de los puntos que caen dentro de cada huella es el estadístico más robusto para estimar la altura del edificio, ya que, en estos casos, suelen darse distribuciones asimétricas. El número de plantas se ha derivado de la altura del edificio, dividiéndolo entre la altura típica de una planta en los edificios de Lorca. Este valor se ha fijado en 3 m, según el conocimiento que se tiene de la zona de estudio. Estos valores se han verificado, como se explica en el apartado siguiente, por comparación con la base de datos de referencia. De los segmentos irregulares de Ensanche, se han validado los que tienen un área mayor de 50 m²; además, en este patrón urbano, no se han podido validar 14 instancias por tener valores perdidos.

La inclinación del tejado se ha calculado según la Ecuación 4. Los tejados con inclinaciones menores de 10° se han considerado planos, y el resto, inclinados. El material de la cubierta se ha obtenido por medio de un modelo predictivo que proporciona las dos categorías de tejados que se pueden encontrar en Lorca: teja y losa de hormigón armado. Se ha creado una base de datos para entrenamiento de cada modelo, seleccionando muestras de los datos de referencia de manera aleatoria; el resto se han dejado para verificación de su desempeño. Las muestras de entrenamiento seleccionadas se unen en el SIG con la base de datos de exposición de este estudio, por medio de una unión espacial. La base de datos de exposición tiene las huellas y el resto de los atributos ya calculados. De esta forma, es posible añadir a las huellas la etiqueta de tipo de tejado tomada de las muestras de entrenamiento.

En el estrato de Ensanche, se ha calculado un modelo para las huellas regulares y otro para las irregulares. En el caso de los estratos del Centro, se han fusionado en uno solo, ya que por separado no se reunían suficientes muestras de tejado de hormigón. En este estrato solo se trabaja con las huellas regulares. La Tabla 5-5 presenta el número de muestras seleccionadas para entrenamiento y para verificación. Como puede verse, el número de instancias de entrenamiento está equilibrado por clases. El resto de las muestras de la base de datos de referencia se han dejado para verificación, de la cuales se ha excluido una instancia más de Ensanche y 12 del Centro, por tener valores perdidos.

Tabla 5-5. Número de muestras en los datasets de entrenamiento y verificación para la clasificación del tipo de tejado.

	Nº muestras				
	Entrenamiento		Verificación		Total
	Losa H	Teja	Losa H	Teja	
ENSANCHE – H.R.	50	50	147	190	437
ENSANCHE – H.I.	30	30	213	35	308
CENTRO – H.R.	50	50	48	214	362

Nota: H.R. y H.I. indican huellas regulares e irregulares, respectivamente.

Con las instancias de entrenamiento, se crean los modelos predictivos utilizando las técnicas explicadas en la sección 4.2.2.3. Previamente se realiza una selección de atributos con el

algoritmo ReliefF implementado en el software WEKA para minería de datos. Los que presentan mayor poder discriminante son la altura del edificio, las coordenadas (X,Y) del centroide, el área y la pendiente del tejado, quedando los atributos relacionados con la forma geométrica al final de la tabla de clasificación. Se calculan los modelos en WEKA incorporando los atributos en dos fases: primero se calculan con los cinco atributos mencionados, y después se incorporan los restantes. Mediante una validación cruzada de 10 iteraciones, se obtienen los parámetros óptimos para los algoritmos de aprendizaje automático empleados, variando sus valores en los siguientes rangos: $f = \{0.05...1\}$, $p = \{1,2\}$, $C = \{1...20\}$, $d = \{1,2,3\}$, $\gamma = \{0.01...50\}$, $\sigma = \{0.01...10\}$, $\omega = \{0.01...10\}$. Se comprueba que añadir el segundo grupo de atributos no mejora sustancialmente la precisión de la clasificación y, por tanto, se decide incorporar sólo los primeros atributos del ranking. Esta decisión se apoya en uno de los objetivos planteados para este escenario: que el proceso sea ágil y rápido, y que los modelos creados sean sencillos y fáciles de entender para poder conocer cuáles son los atributos más importantes a la hora de clasificar los tejados. Los mejores modelos obtenidos se muestran en la Tabla 5-6, ordenados por su precisión global y F1-score, y los parámetros finales.

Tabla 5-6. Ranking de los modelos predictivos para clasificar tejados que mayor precisión han alcanzado en la fase de entrenamiento.

ENSANCHE – huellas regulares					
Algoritmo Aprendiz. Autom.	F1-Score	Precis. Glob.	AUC	kappa	Parámetros
SVM (NormPolyK)	85%	85%	0.70	0.70	$C=10, d=2$
Red Bayesiana K2	81%	81%	0.69	0.62	$p=2$
J48 Árbol Decisión	79%	79%	0.67	0.58	$f=0.25$
ENSANCHE – huellas irregulares					
Algoritmo Aprendiz. Autom.	F1-Score	Precis. Glob.	AUC	kappa	Parámetros
SVM (PUK)	95%	95%	0.75	0.90	$C=10, \omega=5, \sigma=1$
J48 Árbol Decisión	90%	90%	0.73	0.80	$f=0.50$
Red Bayesiana K2	88%	88%	0.72	0.77	$p=1$
CENTRO - huellas regulares					
Algoritmo Aprendiz. Autom.	F1-Score	Precis. Glob.	AUC	kappa	Parámetros
SVM (PUK)	84%	84%	0.70	0.70	$C=10, \omega=1, \sigma=0.5$
J48 Árbol Decisión	80%	81%	0.68	0.60	$f=0.50$
Red Bayesiana K2	72%	72%	0.65	0.44	$p=2$

Nota: NormPolyK indica Kernel Polinómico Normalizado.

En todos los estratos hay al menos un modelo que proporciona resultados de clasificación muy satisfactorios, con precisiones globales y F1-Scores que superan el 84% en la fase de entrenamiento. En los tres casos, SVM supera a los demás algoritmos, que son más sencillos, pero con diferencias poco significativas (4% - 5%). Los valores de kappa de los mejores modelos en cada zona indican acuerdos entre moderados y casi perfectos ($0.70 \leq \kappa \leq 0.90$).

A. Huellas regulares de estrato Ensanche

B. Huellas irregulares de estrato Ensanche

Figura 5-10. Ejemplo de edificio con los atributos calculados en el SIG.
(A) Huellas regulares. (B) Huellas irregulares.

La Figura 5-10 muestra una representación de las huellas, regulares e irregulares, obtenidas para una de las zonas de estudio (Norte), a modo de ejemplo. Cada huella almacena los atributos calculados, que pueden verse en la tabla informativa, asociada al edificio de color granate. Se observa que hay dos edificios en la esquina inferior que aparecen en las huellas regulares, pero no en las irregulares, seleccionadas con el LiDAR. Se trata de dos edificios que fueron demolidos tras el terremoto de 2011. Estas variaciones son detectadas con precisión por el LiDAR.

Verificación de los atributos

La verificación del atributo área se ha realizado en las huellas irregulares, puesto que las regulares tienen el área correcta, al provenir de una fuente de datos oficial. Se ha comparado el

área total de todos los edificios de la base de datos de referencia (área observada) con el área total de todas las huellas obtenidas en este estudio (área estimada), en cada zona de estudio y por estrato. Con esto, se ha calculado el error porcentual (E.P.) que supone el desacuerdo entre ambos valores (por defecto o por exceso). Adicionalmente, se ha calculado la intersección espacial entre ambos conjuntos de huellas y el porcentaje de detección (P.D., relación entre el área de intersección y la observada). En la Tabla 5-7 se presentan todos estos resultados.

Tabla 5-7. Resultado de la verificación del atributo área.

Zona de estudio	ENSANCHE		CENTRO	
	Norte	Sur	Centro A	Centro B
Área observada (m ²)	39458.55	65483.50	51912.75	19844.49
Área estimada (m ²)	39737.59	65959.16	48470.19	19830.49
Error Porcentual (E.P.)	0.70%	0.70%	7.10%	0.07%
E.P. agregado	0.70%		3.59%	
Área de intersección (m ²)	35822.56	50780.35	38822.16	16910.11
Porcentaje de Detección (P.D.)	91%	78%	75%	85%
P.D. agregado	85%		80%	

En las zonas de Ensanche, Norte y Sur, el error porcentual es menor de 1 (0.70%) y los porcentajes de detección son altos (91% y 78%, respectivamente). En la zona Centro B se obtiene resultados igualmente satisfactorios, con un E.P. de casi cero y un P.D. del 85%. Estos resultados parecen indicar que el procedimiento diseñado en esta tesis para conseguir el área construida es fiable en estas zonas.

En cambio, la zona Centro A presenta el mayor E.P., que sube al 7%, y el P.D. es del 75%. Esto indica que la segmentación en esta zona no es buena, lo que avala la decisión anterior de no continuar con estas huellas en las siguientes fases del estudio. La zona Centro B fue descartada también por presentar un número de segmentos 5 veces mayor que el número de edificios, sin embargo, la estimación del área construida es bastante precisa.

El error porcentual agregado en el estrato de Ensanche es del 0.70% y en las zonas del Centro es del 3.6%. Esto indica que el procedimiento propuesto estimaría el área de un edificio de Ensanche de 100 m² en 101 m²; y la de un edificio del Centro de 100 m² en 104 m². Estas diferencias no son significativas cuando se trata de analizar la vulnerabilidad sísmica. En general, se puede decir que se ha conseguido crear una máscara de área construida muy precisa. En la Figura 5-11 se presentan las huellas irregulares obtenidas. Al representarse sobre la ortofoto del PNOA, puede verse cómo cubren los edificios con precisión. El uso del LiDAR para seleccionar los edificios ha permitido detectar aquellos que han sido demolidos después del terremoto y que, en 2013 (fecha de registro de la imagen), aún no habían sido reconstruidos. En la zona Norte y en la Sur pueden encontrarse sendos ejemplos.

Figura 5-11. Huellas irregulares obtenidas para cada zona de estudio.

La validación del número de plantas se realiza también comparando el valor obtenido en este estudio (estimado) y el valor contenido en la base de datos de referencia (observado). Se validan las alturas de todos los edificios en todas las zonas. En Ensanche, se han calculado para las huellas regulares y las irregulares. Los residuos se muestran en la Figura 5-12.

Figura 5-12. Histograma de los residuos del número de plantas obtenidos para cada zona de estudio.

En todos los casos se obtiene una mayoría de residuos cero. En Ensanche, se hace una predicción correcta del número de plantas para 378 edificios de huella regular, que son el 86% del total (precisión global, OA). Los restantes 60 edificios tienen un error de solo 1 planta. Cuando se analizan las huellas irregulares, se observa que la precisión disminuye ligeramente. El número de plantas se estima correctamente para 215 edificios (OA del 69%), mientras que 65 (el 21%) tienen un error de 1 planta. En Centro A, 134 estimaciones son correctas (OA del 74%) y 43 (el 24%) presentan errores de 1 planta. En Centro B, se estima correctamente la altura del 72% de los edificios (140), y en el 11% los errores son de 1 planta. El error máximo en la estimación de este atributo es de 4 plantas en las huellas irregulares de Ensanche, y de 3 plantas en Centro B. Sin embargo, en ambos casos, los errores mayores de 2 plantas no superan el 5% de las huellas. Estos satisfactorios resultados evidencian la efectividad del algoritmo TOD para la clasificación del LiDAR.

Hay que considerar que los modelos de vulnerabilidad hacen divisiones de la altura de los edificios en rangos de 2 o 3 plantas. Es decir, se considera que dos edificios de la misma tipología que tengan alturas entre 1 y 2 plantas, por ejemplo, o entre 3 y 5, se comportan igual ante un sismo. Por lo que un error de 1 o 2 plantas puede no resultar significativo a la hora de asignar la vulnerabilidad sísmica a un edificio.

Finalmente, se verifica el material del tejado utilizando el resto de las huellas que quedaban en la base de datos de referencia tras crear el dataset de entrenamiento (Tabla 5-5). La Tabla 5-8 presenta las matrices de confusión obtenidas en la fase de verificación de la clasificación del tipo de tejado, para el modelo de mayor precisión, y la Tabla 5-9 muestra las precisiones

obtenidas para cada modelo predictivo, en términos del F1-Score promedio, la precisión global y kappa. Se observa que en las zonas de Ensanche (H.I) y Centro, donde los datasets de verificación estaban desequilibrados, la precisión global y el F1-Score difieren. Por tanto, la indicación sobre la calidad de los modelos queda más completa si se dan ambas métricas que si se da solo la precisión.

Tabla 5-8. Matrices de confusión obtenidas en la verificación de la clasificación del tipo de tejado

	ENSANCHE – H.R. Predicción con BAYES K2 2P			ENSANCHE – H.I. Predicción con SVM (PUK)		CENTRO – H.R. Predicción con SVM (PUK)	
	Losa H Teja			Losa H Teja		Losa H Teja	
	Losa H	Teja		Losa H	Teja	Losa H	Teja
Real	Losa H	121	26	171	42	44	4
	Teja	15	175	0	35	32	182

Tabla 5-9. Ranking de los modelos predictivos creados para estimar el tipo de tejado en cada área de estudio.

		F1-Score	Precis. Global	Kappa
ENSANCHE (H.R.)	Red Bayes. K2 2P	88%	88%	0.75
	J48 Árbol Decis.	86%	86%	0.72
	SVM (NormPolyK)	83%	83%	0.65
ENSANCHE (H.I.)	SVM (PUK)	76%	83%	0.53
	Red Bayes. K2 1P	72%	81%	0.47
	J48 Árbol Decis.	69%	77%	0.42
CENTRO (H.R.)	SVM (PUK)	81%	86%	0.63
	J48 Árbol Decis.	77%	83%	0.55
	Red Bayes. K2 2P	74%	80%	0.50

En comparación con las precisiones obtenidas a priori (Tabla 5-6), se confirma el potencial del SVM creado con el PUK kernel, pues ofrece las mejores precisiones también a posteriori en las zonas de Ensanche, con huellas irregulares, y de Centro. En las huellas regulares de Ensanche, en cambio, el SVM creado con un kernel polinómico es superado por la red Bayesiana K2, que ofrece la precisión más alta (F1-Score de 88%), e incluso, por el árbol de decisión. Las precisiones finales de ambos clasificadores aumentan un 7% con respecto a las obtenidas en la fase de entrenamiento. En cambio, en las huellas irregulares de Ensanche, todos los modelos parecían estar sobreajustados, pues pierden alrededor de un 10% de precisión global en la fase de verificación.

En general, la precisión de la clasificación del tipo de tejado es alta, especialmente para las huellas regulares. En cualquier caso, todos los coeficientes kappa obtenidos indican una buena calidad de los modelos.

Figura 5-13. Resultados de la obtención del atributo “material del tejado” en comparación con la base de datos de referencia.

Todas las figuras están orientadas al norte.

(Continuación) Figura 5-13. Resultados de la obtención del atributo “material del tejado” en comparación con la base de datos de referencia.

Para complementar la información de las matrices de confusión y las métricas de precisión, en la Figura 5-13 se presenta una comparación del material del tejado obtenido con los modelos de predicción y la información referente al tipo de tejado contenida en la base de datos de referencia. El atributo estimado se representa por medio de un símbolo puntual en el centroide de cada huella. En las zonas de Ensanche (Norte y Sur), se presentan los resultados para huellas regulares e irregulares. Como puede verse, a excepción de algunos errores, la coincidencia es bastante alta, en línea con las altas precisiones que indica la Tabla 5-9. Los errores parecen concentrarse, razonablemente, en zonas heterogéneas; es decir, en zonas donde hay manzanas con construcciones similares que, sin embargo, presentan tejados diferentes. Ejemplos de esto pueden verse en las cuatro zonas, pero especialmente se ven en la zona Sur y en el Centro.

5.1.2.3 Fase 3: clasificación de tipologías constructivas y asignación de la vulnerabilidad sísmica a los edificios de Lorca.

Creación de los modelos predictivos para la clasificación de tipologías constructivas.

Para la clasificación de los edificios en tipologías constructivas, se crean modelos predictivos empleando las técnicas descritas en la sección 4.2.3. Con estos modelos, a cada instancia de la base de datos de exposición se le asignará una MBT a través de la etiqueta RC, para edificios con estructura de hormigón armado, y M, para mampostería, utilizando los atributos calculados en la fase anterior. La Tabla 5-10 muestra las bases de datos de entrenamiento y verificación

generadas para esta fase, a partir de los datos de referencia. Las instancias de entrenamiento se han seleccionado por muestreo aleatorio estratificado, creando muestras equilibradas. El resto de las instancias se dejan para verificación. De las 12 instancias excluidas al clasificar los tejados de las zonas del Centro en la fase anterior por tener valores perdidos, se han recuperado 8, que sí contienen el dato de vulnerabilidad.

Como se mencionó en la sección 5.1, en las zonas de Ensanche predominan los edificios de hormigón armado. Esto tiene repercusión a la hora de generar las bases de datos de entrenamiento, pues el número de muestras de esta clase es muy limitado, como se puede ver en la Tabla 5-10. Entre las huellas regulares, solo hay 14 edificios de RC; y 6 en las irregulares. Todas deben tomarse para entrenamiento, porque un número menor no permitiría calcular los modelos. De hecho, calcular un modelo con 14 muestras es ya muy arriesgado, pero se ha hecho para comprobar la capacidad de los algoritmos de manejar un número tan pequeño de muestras. Para las huellas irregulares, que solo tienen 6 instancias, no se han podido calcular los modelos. En cuanto a la verificación, en Ensanche no quedan muestras de mampostería, por lo que no se ha podido evaluar la capacidad de generalización (escalado) de los modelos creados en este estudio. Se plantea esta comprobación para futuras actividades de investigación. En la zona Centro se han calculado los modelos predictivos y se han verificado.

Tabla 5-10. Número de muestras en los datasets de entrenamiento y verificación para la clasificación de la vulnerabilidad.

	Nº Muestras				
	Entrenamiento		Verificación		Total
	RC	M	RC	M	
ENSANCHE – H.R.	14	14	409	0	437
ENSANCHE – H.I.	6	6	296	0	308
CENTRO – H.R.	60	60	93	157	370

Nota: H.R. y H.I. indican huellas regulares e irregulares, respectivamente.

Como paso previo a la creación de los modelos, se seleccionan los atributos con ReliefF en WEKA, obteniéndose que los de mayor poder discriminante son la altura de los edificios, las coordenadas (X,Y) del centroide, el área, el tipo de tejado y la pendiente de este. Estos atributos son los que se utilizan para predecir la tipología.

Para el entrenamiento de los modelos predictivos, se lleva a cabo en WEKA una validación cruzada de k iteraciones, con k=10 para la zona Centro y k=4 para Ensanche. Esta validación cruzada permite establecer los valores de todos los parámetros del mismo modo que para la clasificación de los tejados. La regresión logística se ha calculado en SPSS. La Tabla 5-11 presenta un ranking de los cuatro mejores modelos ordenados por su F1-Score, precisión global, AUC (área bajo la curva ROC) y kappa, así como los parámetros que han dado lugar a estos resultados.

Tabla 5-11. Ranking de los modelos predictivos que mayor precisión han alcanzado en la fase de entrenamiento.

ENSANCHE – huellas regulares					
Algoritmo Aprendiz. Autom.	F1-Score	Precis. Glob.	AUC	kappa	Parámetros
J48 Árbol Decis.	89%	90%	0.82	0.79	$f=0.25$
Reg. Logística	89%	89%	0.71	0.79	
SVM (PUK)	86%	86%	0.86	0.71	$C=5, \omega=2, \sigma=0.2$
Red Bayes. (TAN)	79%	79%	0.72	0.57	
CENTRO – huellas regulares					
Algoritmo Aprendiz. Autom.	F1-Score	Precis. Glob.	AUC	kappa	Parámetros
J48 Árbol Decis.	87%	87%	0.81	0.73	$f=0.05$
Reg. Logística	84%	84%	0.84	0.68	
Red Bayes. (TAN)	80%	80%	0.84	0.60	
SVM (PUK)	79%	81%	0.79	0.58	$C=1, \omega=1, \sigma=1$

De todos los modelos, el árbol de decisión es el que mejores métricas de precisión alcanza. Es de destacar el desempeño de la regresión logística que, a pesar de ser un clasificador sencillo, obtiene precisiones comparables al árbol, superando incluso al SVM con el kernel PUK. En cualquier caso, las precisiones son todas altas, por encima del 80% (en su mayoría) y rozando el 90%. Estas precisiones a priori han de ser contrastadas con los datos nuevos de verificación, para comprobar la estabilidad de los modelos a la hora de generalizar.

En el Anexo III se presentan los gráficos de los árboles de decisión y las redes Bayesianas, así como la tabla de los coeficientes de la regresión logística. De los seis atributos utilizados para crear los modelos, se observa que los árboles sólo utilizan la altura de los edificios y la coordenada X del centroide en ambas zonas (Ensanche y Centro). Es decir, que con sólo dos atributos se pueden conseguir muy buenas precisiones en la predicción de la tipología constructiva. Por su parte, la regresión logística, en el Centro incluye también el material del tejado, y en el estrato de Ensanche, utiliza la Y del centroide en lugar de la X. De nuevo se ve que con pocos atributos se pueden obtener resultados precisos de clasificación. Por el contrario, el SVM y la red Bayesiana utilizan los seis atributos y arrojan precisiones algo más bajas.

Verificación de los resultados de clasificación de MBT

La verificación de estos resultados de clasificación de tipologías constructivas en la zona Centro se hace a través de las métricas derivadas de la matriz de confusión, que se presentan en la Tabla 5-12. Con respecto a las métricas de precisión en la fase de entrenamiento, se observa que, excepto el árbol de decisión, los modelos experimentan pérdidas de F1-Score y precisión global de sólo un 2%, aproximadamente. Esta es una indicación de la robustez de los modelos, que son capaces de generalizar sin una pérdida significativa de precisión. El árbol, en cambio, parece presentar sobreajuste, ya que ha perdido el 8% de F1-Score y el 7% de precisión global al aplicarse a los datos nuevos. Aun así, queda segundo en el ranking. El modelo de regresión logística es el que alcanza mayores métricas de calidad en esta fase de verificación, siendo capaz

de clasificar las tipologías de la zona Centro con una precisión del 82% y un F1-Score del 80%. Kappa indica un acuerdo alto entre los resultados de este modelo y la realidad.

Tabla 5-12. Matrices de confusión obtenidas en la verificación de la clasificación de vulnerabilidad en la zona Centro, y métricas de precisión.

Matrices de confusión					FP rate	F1-Score	Precisión global	Kappa
Predicción: Reg. Logística						80%	82%	0.60
Real		RC	M					
	RC	62	31	9%				
	M	14	143	33%				
Predicción: J48 Árbol Decis.						79%	80%	0.57
Real		RC	M					
	RC	67	26	42%				
	M	24	133	28%				
Predicción: SVM (PUK)						77%	80%	0.54
Real		RC	M					
	RC	57	36	10%				
	M	15	142	39%				
Predicción: Red Bayes. (TAN)						77%	78%	0.53
Real		RC	M					
	RC	67	26	18%				
	M	29	128	28%

Figura 5-14. Resultados de la obtención de la tipología (MBT) en comparación con la base de datos de referencia. Todas la figuras están orientas al norte.

(Continuación) Figura 5-14. Resultados de la obtención de la tipología (MBT) en comparación con la base de datos de referencia.

Predicción de MBT en la zona Centro (Red Bayesiana)

(Continuación) Figura 5 14. Resultados de la obtención de la tipología (MBT) en comparación con la base de datos de referencia.

Para complementar la información detallada por las matrices de confusión y las métricas de calidad, se presenta la distribución espacial de los resultados de clasificación de MBT en la Figura 5-14. La MBT estimada se ha representado con un punto en el centroide de cada huella para poder superponerse a la geometría de las huellas de verificación. Con el fin de analizar con más detalle los resultados obtenidos con los cuatro modelos, se presenta a continuación una comparativa en tres manzanas de especial dificultad, dada la variabilidad contenida dentro de las mismas (en columnas). Por filas, se presentan los resultados de cada modelo.

Los resultados obtenidos con la regresión logística muestran que es posible captar la variabilidad dentro de la manzana. Esto mismo ocurre en muchos otros casos de la zona de estudio.

Con el árbol de decisión, los resultados son muy similares, aunque en algunos casos se ve superado por la regresión logística (caso C).

SVM también muestra dificultades para diferenciar tipologías dentro de una misma manzana, cuando las instancias guardan grandes parecidos.

La red Bayesiana es capaz de resolver correctamente algunos casos complicados, como el detalle C, pero incurrir en errores que otros modelos superan, como en las muestras A y B.

En general, en las manzanas donde la regresión logística no obtiene un 100% de acierto en la clasificación, los demás modelos tampoco la alcanzan. Una excepción es la manzana abierta localizada al este de la zona de estudio. En ella, el árbol de decisión y SVM clasifican los 23 edificios correctamente como hormigón armado, mientras que la red Bayesiana y la regresión logística comenten 2 fallos, lo que supone un acierto de más del 90%.

En vista de estos resultados, se puede concluir que los cuatro modelos podrían predecir la tipología de los edificios de Lorca con un acierto global del orden de 80%. Que todos los modelos arrojen precisiones similares es un indicativo de la robustez de la predicción en este escenario, ya que parece que los datos siguen unas reglas que todas las técnicas empleadas han sido capaces de descubrir. Además del dato global de precisión, es posible distinguir en qué zonas las tasas de acierto son mayores, como ocurre en la manzana abierta antes citada, o menores, como en la parte oeste de la zona de estudio, donde los edificios son más pequeños y a todos los modelos les cuesta predecir correctamente los casos aislados dentro de las manzanas. Esta distinción es importante y se puede utilizar para orientar las campañas de campo de recogida de datos de los edificios hacia estas áreas de la ciudad más complicadas de estudiar mediante técnicas de teledetección. Esta reducción del área a visitar permitiría optimizar el trabajo, bien reduciendo el tiempo en campo, o bien reduciendo los recursos.

Asignación de un modelo de vulnerabilidad

El resultado de esta fase es una clasificación del parque inmobiliario de las zonas Centro A y Centro B de Lorca en las tipologías constructivas (MBT) mampostería (M) y hormigón armado (RC). A cada MBT se le ha asignado un modelo de vulnerabilidad de Lagomarsino y Giovinazzi (2006), que es M3 para la tipología M y RC1 para la tipología RC. El detalle de estos resultados de clasificación de la vulnerabilidad se ha aumentado al subdividir las categorías según el número de plantas, siguiendo los criterios de los autores. Las clases son low-rise (L-rise), para

edificios con 1-2 plantas de M3 y 1-3 plantas de RC1; mid-rise (M-rise), para edificios con 3-5 plantas de M3 y 4-7 plantas de RC1; y high-rise (H-rise) para alturas mayores.

La Figura 5-15 muestra la distribución de vulnerabilidad en Lorca. Los datos del patrón urbano Centro son los resultados de esta fase. Se han incluido, a efectos de comparación, los datos de referencia para las zonas de Ensanche. Se aprecia una clara diferencia en la distribución de tipologías constructivas entre ambos patrones. En la zona del Centro hay presencia significativa de todas las tipologías, si bien es cierto que predominan los edificios de mampostería. Lo contrario ocurre en el Ensanche, donde la gran mayoría de los edificios son de hormigón armado de baja y media altura. Los edificios más altos de este tipo están en las zonas del Centro debido a que hay una ordenanza distinta que afecta a unas determinadas calles en las que se permite construir edificios más altos que en el resto del municipio (de hasta 10 alturas).

Figura 5-15. Distribución de vulnerabilidad en Lorca

5.1.3 Estimación de tiempos y costes

A continuación, se presenta una estimación del tiempo que se emplearía en repetir el proceso diseñado para el escenario de Lorca, en una ciudad similar, al margen de todos los experimentos realizados. Este sería un planteamiento muy realista, en el que se aplicaría el procedimiento en una ciudad de la misma región, donde las tipologías constructivas y la configuración urbana son parecidas a las analizadas en este estudio. En la estimación de tiempos y costes, se considera que trabaja un equipo de dos personas con experiencia y que se cuenta con las huellas catastrales, que son las que han permitido realizar el proceso completo. El tiempo se mide en jornadas laborales estándar, de 8 horas (Tabla 5-13).

Se estima que dos personas con experiencia en esta materia y con conocimiento de la zona de estudio podrían obtener la vulnerabilidad de unos 1000 edificios en 5 jornadas laborales, siguiendo el proceso diseñado para este escenario. En la Tabla 5-14 se presenta la estimación de los costes asociados. Se ha considerado que el trabajo lo realizaría un especialista en datos (*data scientist or engineer*) y un colaborador con experiencia (*assistant*), que trabajarían durante los 5 cinco días. Además, se propone contar con un experto en la rama de ingeniería

civil o arquitectura, que tenga conocimiento de las tipologías constructivas de la ciudad de estudio y que pueda hacer sus aportaciones en reuniones de trabajo durante 2 días. Como recursos materiales se han considerado 2 ordenadores y material de oficina. Los datos no suponen un coste adicional, porque serían de libre acceso (como en esta tesis). Si fuera necesario adquirir datos de la misma resolución que los utilizados en este estudio, se podrían necesitar unos 2 km² de ortofoto y LiDAR, cuyo precio se estima en 1000 € (Dr. Juan Carlos Ojeda, Jefe del Servicio de Fotogrametría del IGN entre los años 2013-2017 y actual Jefe de Área de Programa del S.G Procedimientos y Atención al Ciudadano de la Dirección General del Catastro. Comunicación personal).

En total, se obtiene que el trabajo tendría un coste total de casi 4500 €. Ese valor se ha calculado añadiendo un 30% al coste de los recursos, como margen para contingencias. En los recursos materiales se ha imputado el coste completo, no un porcentaje de amortización.

Tabla 5-13. Estimación de tiempos para la realización de un análisis de vulnerabilidad por teledetección (a partir del estudio en Lorca)

Proceso	Tiempo (jornadas laborales)
1. Estratificación de la ciudad	½
2. Generación de la base de datos de exposición	
2.1. Huellas catastrales	1
2.2. Clasificación del LiDAR	1
2.3. Obtención de los atributos relacionados con el LiDAR: altura, número de plantas, inclinación del tejado.	½
Obtención de los atributos relacionados con la geometría de las huellas.	
2.4. Obtención del atributo “material del tejado”	1
3. Obtención de la tipología	1

Tabla 5-14. Estimación de costes para la realización de un análisis de vulnerabilidad por teledetección (a partir del estudio en Lorca)

Human Res.	Number	Time (days)	Unit cost (€)	Total cost
Data Sci./Eng.	1	5	150	750
Assistant	1	5	100	500
Civil Eng./Arch.	1	2	150	300
Material Res.	Number	-	Unit cost (€)	Total cost
PC/laptop	2		600	1200
Printer	1		300	300
Office supplies	1		400	400

Cost 3450 €

Total cost 4485 €

5.1.4 Estimación del daño a partir de los datos obtenidos

En este apartado se ha tratado de evaluar hasta qué punto los errores cometidos en la clasificación de tipologías constructivas pueden afectar a los resultados de daño que, en definitiva, es el objetivo último de los estudios de riesgo sísmico. En la zona Centro de Lorca, con la clasificación obtenida con regresión logística (que proporciona las mejores métricas de precisión, con una precisión global del 82%), se tienen 31 edificios clasificados erróneamente como tipología de mampostería y 14 como hormigón. Estos errores suponen un índice de falso positivo del 33% y del 9%, respectivamente (Tabla 5-12).

Para saber qué influencia tienen esos porcentajes de error en la estimación del riesgo sísmico, se ha calculado un escenario de daños utilizando las dos bases de datos de vulnerabilidad: la de referencia (datos reales tomados en campo) y la estimada en este estudio con Regresión Logística. Los resultados se han comparado para analizar la sensibilidad del cálculo ante estos dos inputs, manteniendo todos los demás constantes. El cálculo se ha realizado con los datos de la zona Centro, empleando el software especializado SELENA¹⁴, que ha sido desarrollado conjuntamente por la Universidad de Alicante y NORSAR, y es de libre distribución.

Se ha simulado un terremoto de Mw 5.2 en la zona donde ocurrió el de mayo de 2011 (Martínez-Díaz et al. (2012)). Los parámetros del escenario sísmico se presentan en la Tabla 5-15. La Figura 5-16 muestra el epicentro del sismo y la localización de la ciudad de Lorca, sobre la que se marca el área de estudio, la zona Centro. Se ha utilizado la ley de atenuación de Campbell y Bozorgnia (2014), uno de los modelos de la nueva generación (NGA) que en el estudio de Quirós (2017) fue el que mejor reprodujo los daños causados por el sismo de Lorca de 2011. De Navarro et. al (2014) se ha tomado el valor de la V_s ³⁰ que caracteriza el tipo de suelo. Con estos datos, se ha obtenido el espectro de demanda que representa el movimiento fuerte del suelo en la zona Centro. Las funciones de daño correspondientes a las MBT son M3 y RC1 de Lagomarsino y Giovinazzi (2006), con sus respectivas curvas para diferentes rangos de altura (sección 5.1.2.3).

Mediante el método I-DCM (FEMA-440, 2005) implementado en SELENA se ha calculado el daño que se espera que el terremoto simulado cause en los edificios del centro de la ciudad. La Tabla 5-16 y la Figura 5-17 muestran los resultados obtenidos con las bases de datos de vulnerabilidad de referencia y estimada en este estudio. Un MAPE de solo el 4% es un indicador de la alta calidad del resultado. Las diferencias entre las dos estimaciones no son significativas y, por tanto, los resultados de daño obtenidos con la base de datos elaborada en este estudio se podrían utilizar en la definición de un plan de políticas ante el riesgo sísmico. El patrón de daños que se obtiene indica que el 49% de los edificios experimentaría daño ligero, y el 29% permanecería sin daño. Del resto, el 16% alcanzaría daño moderado y una mínima parte de los edificios (6%) sufriría daños severos o colapso.

¹⁴ <https://www.norsar.no/r-d/safe-society/earthquake-hazard-risk/the-selena-open-risk-software/>

Tabla 5-15. Parámetros del escenario sísmico simulado en Lorca

Latitud	Longitud	Profundidad	Mw	Orientación	Buzamiento	Mecanismo
37.7175º	-1.7114º	4 km	5.2	240º	54º	Normal

Epicentro del escenario sísmico de Lorca

Figura 5-16. Escenario sísmico simulado en Lorca

Tabla 5-16. Resultados del cálculo de daños en el escenario de Lorca

N. Edificios dañados						
	Nulo	Ligero	Moderado	Extensivo	Completo	Total
Referencia	106	182	61	19	3	371
Estimación	101	181	65	19	4	370

NOTA: Se han comparado 370 edificios. El efecto del redondeo en las probabilidades de daño suma un edificio más para la base de datos de referencia

Figura 5-17. Gráfica comparativa de los daños totales

Con el interés de evaluar con más detalle el impacto de la incertidumbre en la asignación de vulnerabilidad en los resultados de riesgo, se presentan los daños desagregados por tipología constructiva en la Figura 5-18, y por altura, en la Figura 5-19. Con la base de datos de vulnerabilidad generada en este estudio, se obtienen más edificios, tanto dañados como no dañados, en la tipología de mampostería y menos en la de hormigón. Estas diferencias, que son ligeras (MAPE menores del 10%), están en consonancia con el error de falso positivo de mampostería en que todos los modelos predictivos incurrierán, y que aportaba a la base de datos más edificios de este tipo de los que realmente hay. Esto hay que tenerlo en cuenta a la hora de derivar otros indicadores de riesgo sísmico, como el coste económico. Al haber más edificios dañados de mampostería que de hormigón, es probable que el coste de reposición se vea ligeramente infraestimado.

Figura 5-18. Gráfica comparativa de los daños obtenidos desagregados por tipología

Si se hace la comparación por edificios de diferente altura (Figura 5-19), se observan de nuevo ligeras diferencias entre los edificios de altura baja y media, entre los cuales se encuentran las construcciones de mampostería. Los edificios, de más de 8 plantas, son de hormigón armado y las predicciones de daños son iguales para ambos datasets de vulnerabilidad.

Figura 5-19. Gráfica comparativa de los daños obtenidos desagregados por número de plantas.

En cualquier caso, es importante destacar que las mayores diferencias se producen en los grados de daño nulo y ligero, que son los que menos influyen a la hora de estimar el número de víctimas y el grado de habitabilidad de los edificios. Estos son los indicadores más relevantes de los estudios de riesgo sísmico y dependen, directa y casi exclusivamente, de los grados de daño más severos, que son los que mejores estimaciones presentan.

5.1.5 Sumario y conclusiones de la aplicación en Lorca

La aplicación en Lorca está enfocada a diseñar un proceso rápido de implementar para poder atender a la demanda de la Directriz Básica de Planificación de Protección Civil de analizar el riesgo sísmico en más de 700 municipios españoles en los próximos años. Para ello, se cuenta con imágenes de satélite Landsat, con ortofotos aéreas y LiDAR del programa PNOA del IGN y con datos de la Sede Virtual del Catastro. El objetivo es identificar los edificios de tres zonas de estudio de la ciudad y clasificarlos según su tipología edificatoria (MBT) en dos clases: hormigón armado (RC) y mampostería (M). A estas clases se les asigna un modelo de vulnerabilidad, que es RC1 y M3, respectivamente, definidas en Lagomarsino y Giovinazzi (2006). Estas tipologías son subdivididas según su altura en bajas (low-rise), medias (mid-rise) y altas (high-rise), siguiendo las indicaciones de los autores. Y, finalmente, se realiza un cálculo de daños para ver hasta qué punto las incertidumbres acumuladas a lo largo del proceso afectan a la estimación del riesgo. La Figura 5-20 presenta un resumen de todas las métricas de precisión.

NOTE: All the metrics correspond to the testing phase, except for the MBT verification of Sprawl, that could not be tested in unseen data. P.E. is percentage error; P.D. is percentage of detection; OA is overall accuracy; F1 is F1-Score; k is kappa coefficient; FP(M)-FP(RC) are False Positive rates of M and RC typologies; MAPE is Mean Absolute Percentage Error; Reg.F/Ir.F. are regular/irregular footprints.

Figura 5-20. Resumen de las métricas de calidad del procedimiento aplicado en Lorca

- El primer paso es estratificar la ciudad en patrones urbanos Residencial, Centro A (casco histórico con edificios de tamaño moderado a grande), Centro B (casco histórico con edificios más pequeños), Ensanche, Rural e Industrial. Esta estratificación se ha realizado a través de un proceso de segmentación automático seguido de una clasificación manual, dado el tamaño reducido del área de estudio (Figura 5-6). Las fases posteriores se han llevado a cabo en las tres áreas de estudio en las que Martínez-Cuevas et al. (2017) generaron una base de datos de vulnerabilidad de 826 edificios, por observación in situ, que se ha tomado como referencia (*ground truth*). Las tres áreas de estudio se encuentran ubicadas en estratos de Ensanche (zona Norte y Sur), Centro A y Centro B (Figura 5-6). Esta distinción en patrones urbanos permite continuar el proceso de forma estratificada, lo que facilita la toma de decisiones en relación con las técnicas y parámetros que se utilizan en cada fase.
- El siguiente paso es implementar la base de datos de exposición y vulnerabilidad.
 - Se comienza generando las huellas de los edificios. Estas provienen de dos orígenes: parcelas de Catastro (huellas regulares) y objetos obtenidos con un proceso semiautomático que combina segmentación de imagen y clasificación de LiDAR

(huellas irregulares). Los puntos LiDAR clasificados como edificio se utilizan para seleccionar los segmentos que pertenecen a tejados. El resto del estudio se ha realizado en ambos conjuntos de huellas, para comparar los resultados. A diferencia del escenario de Haití (como se verá en los próximos capítulos), el proceso de generación de las huellas irregulares no ha sido muy exhaustivo, porque en España se tienen parcelas de catastro para todo el territorio, que serían la fuente principal de aporte de este dato. De ahí que la obtención de las huellas irregulares presente un error significativo, pues superan en número a las regulares en tres zonas de estudio (Tabla 5-4): Norte (coef. comp. = 2.7), Centro B (coef. comp. = 5.3) y Centro A (coef. comp. = 9.3). Estas grandes diferencias, especialmente en las zonas del Centro, ponen de manifiesto que los tejados han resultado sobre-segmentados, y que el número de huellas irregulares no representa al número de edificios reales. Por ello, se ha decidido no continuar trabajando con las huellas irregulares de las dos zonas del Centro.

- 2.2. Para completar la base de datos de exposición, las huellas son dotadas de atributos relacionados con su vulnerabilidad sísmica: material y pendiente del tejado, número de plantas, área y otros parámetros derivados de su geometría (dirección, ancho, largo, elongación, compacidad, posición relativa en la manzana y coordenadas planimétricas del centroide). El cálculo de algunos de estos atributos conlleva incertidumbres, que se han cuantificado por comparación con los datos de referencia. El material del tejado se ha calculado mediante un modelo predictivo que utiliza los atributos de mayor poder discriminante: número de plantas, coordenadas del centroide, área y pendiente del tejado. Los mejores modelos que predicen el tipo de tejado son (Tabla 5-9): Red bayesiana, para las zonas de Ensanche con huellas regulares (F1-Score y precisión global de 88%); SVM (PUK) para las huellas irregulares de Ensanche, donde las métricas de calidad decrecen ligeramente (F1-Score de 76% y precisión global de 83%); también SVM (PUK) para las zonas del Centro (F1-Score de 81% y precisión global de 86%). En todos los casos, el coeficiente kappa indica buena calidad de los resultados ($0.53 \leq \kappa \leq 0.75$).

El número de plantas de los edificios se ha obtenido con un MDSn derivado del LiDAR, dividiendo la altura sobre el suelo de los tejados entre 3 m, que es la altura media de los pisos en el área de estudio. La verificación del número de plantas por comparación con la base de datos de referencia (Figura 5-12) ha permitido cuantificar que el 86% de las huellas regulares de Ensanche tienen una altura correcta, porcentaje que baja al 69% en las irregulares. En el Centro A, el 74% de los edificios tienen un número de plantas estimado correctamente, y en el Centro B, el 72%. En cuanto a los errores, la mayoría son de 1 planta, por lo que no tendrán mucha incidencia en la clasificación de la tipología constructiva de la siguiente fase.

Por último, se ha verificado el área de las huellas irregulares para comprobar si el proceso semiautomático de generación de huellas de edificios es capaz de identificar con precisión la zona construida de una ciudad. La verificación se ha realizado por comparación con los datos de referencia (Tabla 5-7). Las huellas regulares no se han

verificado porque provienen de la misma fuente que estos: el Catastro. Se ha obtenido que, en las dos zonas de Ensanche (Norte y Sur), el error porcentual en la estimación del área es solo del 0.70%, con porcentaje de detección del 91% y 78%, respectivamente. En la zona Centro A, el área estimada presenta un error porcentual del 7.10%; mientras que en el Centro B, es prácticamente igual a la observada, con un error del 0.07%. Los porcentajes de acierto son del 75% (Centro A) y 85% (Centro B). Se ve que, aunque el número de huellas irregulares no puede utilizarse como estimador del número de edificios debido a la sobre-segmentación, sí es posible conseguir una máscara de área construida de alta precisión.

3. Con los atributos verificados, se procede a clasificar los edificios en las dos tipologías identificadas en la ciudad: hormigón armado (RC) y mampostería (M). Se calculan cuatro modelos predictivos en los estratos Ensanche y Centro (uniendo Centro A y B para conseguir un mayor número de muestras). Los algoritmos utilizados son regresión logística, redes Bayesianas, árboles de decisión y SVM. Los atributos incluidos son los de mayor poder discriminante en este caso: número de plantas, coordenadas del centroide, área y tipo de tejado. Para ambos patrones urbanos se consiguen precisiones en la fase de entrenamiento cercanas al 90% (Tabla 5-11). Se han verificado los de la zona Centro (en la de Ensanche no han quedado muestras para verificación), obteniéndose F1-Score de 80%, precisión global de 82% y $\kappa = 0.6$ para el mejor modelo, el creado con Regresión Logística (Tabla 5-12). El error en la clasificación es principalmente por falso positivo de mampostería (33%), es decir, edificios RC clasificados erróneamente como M, lo que aumenta el número de edificios de este tipo en la base de datos generada. El falso positivo de RC es de 9%. Los demás modelos obtienen métricas de precisión algo más bajas, pero siempre por encima del 77%. Este equilibrio de los modelos es un indicador de la robustez de la predicción de la tipología edificatoria en la zona centro.

El tiempo estimado en realizar esta clasificación de tipologías constructivas para asignación de la vulnerabilidad sísmica a unos 1000 edificios es de 5 días para dos personas expertas en la materia y en la zona (Tabla 5-13). El coste asociado sería de 4500 € (Tabla 5-14).

Finalmente, se ha realizado un cálculo de daños por sismo en la zona Centro, para la base de datos de vulnerabilidad de referencia y para la obtenida en este estudio con las huellas regulares. Se ha simulado un sismo de Mw 5.2 en la región donde se produjo el terremoto de 2011, a unos 4 km al norte de la zona de aplicación, y se han comparado los resultados de daño obtenidos para ambas bases de datos. En la comparación por tipología (Figura 5-18) se observa que, como se esperaba, se sobrestima el número de edificios dañados de mampostería, pero también el de no dañados. Lo mismo se aprecia cuando se comparan los daños desagregados por altura de las construcciones (Figura 5-19). Esto se justifica por el índice de falso positivo de mampostería de la clasificación de tipologías, pero es importante señalar que los patrones de daños obtenidos son iguales. Las discrepancias más notables se obtienen en los grados de daño nulo y ligero, que son los que menos influencia tienen a la hora de calcular los indicadores de riesgo más relevantes. Esto permite concluir que ese error por falso positivo de la clase M no afecta significativamente al cálculo del daño. Es decir, el hecho de clasificar el 33% de los

edificios de hormigón como mampostería, que es más vulnerable, no aumenta los daños graves, lo que habría sesgado el patrón de daños. De hecho, en el cómputo de daños totales (registrados en edificios de los dos tipos), los daños calculados para las dos bases de datos de vulnerabilidad son prácticamente iguales (Figura 5-17), con un MAPE del 4%.

Estos resultados parecen indicar que las incertidumbres inherentes a las distintas fases del proceso de cálculo en la zona Centro (clasificación del tipo de tejado y cálculo el número de plantas, principalmente), no tienen una influencia significativa en la determinación de la tipología constructiva de las huellas regulares y permiten obtener la vulnerabilidad con una precisión global del 82%. Estos datos de vulnerabilidad dan lugar a un cómputo de daños que apenas difiere del obtenido con una base de datos de exposición y vulnerabilidad generada en campo, y podrían emplearse en la definición de políticas de prevención y mitigación ante el riesgo sísmico, que es el fin último de estos estudios.

5.2 Aplicación en Puerto Príncipe, Haití

Puerto Príncipe es la ciudad capital de Haití, el país que forma la Isla de la Española junto con República Dominicana. Haití arrastra una política inestable, tanto interna como externa, desde la caída de la dictadura de Duvalier en 1986 (Taft-Morales, 2017). Desde entonces, se han sucedido periodos de luces y sombras en forma de procesos electorales y golpes de estado, que han dado lugar a gobiernos militares y civiles. Estos gobiernos han creado instituciones democráticas débiles, que, junto con la corrupción asociada, (también) tanto interna como externa, han impedido que el país se desarrolle adecuadamente.

Según el Banco Mundial¹⁵, Haití es el país más pobre del hemisferio oeste, con un PIB per cápita de 870 \$ en 2018. Esto implica que 6 millones de haitianos subsisten con 2,5 \$ al día, y 2,5 millones quedan por debajo de la línea de extrema pobreza (con 1,2 \$ al día). El PIB ha subido unas décimas en el último año, pero a costa de aumentar su deuda con el Banco Central, lo que deprecia su moneda local, la *gourde*. Esta desesperante situación, unida al desamparo político, ha hecho que la población se revele a través de manifestaciones y huelgas, con más frecuencia en los últimos años.

El contexto político, económico y social de Haití ayuda a comprender la lentitud de los avances en materia de prevención de desastres. Por su situación geográfica en el límite norte de la Placa Caribe, el país se ve expuesto a desastres naturales de tipo geofísico, climático y meteorológico, como terremotos, huracanes e inundaciones, que dan lugar a pérdidas recurrentes de las que es difícil recuperarse. Desde 2017, Haití está comprometido con la ONU a trabajar para reducir estos riesgos a través del Documento Nacional para la Reducción del Riesgo de Desastres 2016-2020 (UNDRR-AM, 2017).

El interés por trabajar en este escenario surgió tras el terremoto de 2010. Este se produjo a unos 25 km al suroeste de Puerto Príncipe (Calais et al., 2010), y afectó a numerosas ciudades de la región, como Carrefour, Delmas o Lèogâne. La extrema gravedad de los daños se explica, principalmente, por la alta vulnerabilidad de las construcciones. Así lo justifican los expertos que visitaron la zona afectada en sus informes, como Bilham (2010), Mix et al. (2011), Fierro y Perry (2010) o Eberhard et al. (2010). En estos informes, se describen malas prácticas constructivas, como que las construcciones tenían configuraciones constructivas incorrectas, con pilares esbeltos soportando grandes y pesadas losas; muchas de las varillas de armado del hormigón eran lisas en lugar de corrugadas; había evidencias de materiales reutilizados y, por tanto, fatigados; el hormigón, en muchos casos, era de mala calidad, con impurezas y restos de arena de playa y salitre; se comprobó que no se cumplían unos requisitos mínimos de calidad de la construcción, por ejemplo, en fallos de la rótula plástica de la unión viga-pilar, que se hacía en la cara interna de los pilares en lugar de en la externa. La Figura 5-21 muestra algunos ejemplos.

¹⁵ <https://www.worldbank.org/en/country/haiti/overview>

A. Pilares esbeltos que no soportaron el peso de las grandes losas (Fierro y Perry, 2010)

B. Fallo de la rótula plástica en la unión viga-pilar (Eberhard et al., 2010)

C. Varillas de armado lisas (Eberhard et al., 2010)

D. El hormigón de mala calidad se desmorona (Eberhard et al., 2010)

Figura 5-21. Ejemplos de prácticas constructivas típicas de Haití.

No hay garantías de que en la reconstrucción de las zonas afectadas se estén erradicando estas malas prácticas, pues se sigue dando la autoconstrucción o construcción espontánea (Lallemant, 2014), que aplica las mismas normas constructivas transmitidas de padres a hijos durante generaciones. Esto hace que la vulnerabilidad se mantenga en niveles similares a los de hace 10 años. Por otro lado, la peligrosidad sísmica es muy elevada en la región, según las últimas investigaciones de Symithe y Calais, (2016). Los autores afirman que el movimiento del suelo esperado en Puerto Príncipe y sus alrededores es el doble que el que aparece en el actual mapa de la Norma Sismorresistente. Esta afirmación se basa en nuevas observaciones geodésicas que parecen indicar que el estrés acumulado en la región presenta una componente de acortamiento en una falla inversa que había pasado desapercibida hasta el momento, además del componente de desgarre que ya se conocía. La combinación de valores altos de vulnerabilidad y peligrosidad da lugar a un elevado riesgo sísmico, que como se ve, permanece en la región.

El estudio de Torres et al. (2016) tiene como objetivo, precisamente, calcular este riesgo ante las pruebas de que hay fallas aún cargadas en los alrededores de Puerto Príncipe. Y es que, en un principio, se pensó que la falla responsable del evento de enero de 2010 debía ser la falla de Enriquillo. Sin embargo, tanto las observaciones de campo (Prentice et al., 2010), como los

modelos de deslizamientos obtenidos a partir de datos telesísmicos, GPS e InSAR (Calais et al., 2010, Hayes et al., 2010) muestran un tipo de rotura diferente al esperado si el sismo hubiera ocurrido en la falla de Enriquillo. Por tanto, ésta sigue cargada, como ya mostraron Manaker et al. (2008). Torres et al. (2016) simularon un escenario de Mw 7.0 en el segmento de Dumay, en la falla de Enriquillo, y calcularon el riesgo de la ciudad de Puerto Príncipe. Emplearon unas funciones de daño (espectro de capacidad-función de fragilidad) calibradas con datos del terremoto de 2010 por los mismos autores en un trabajo previo (Molina et al., 2014). En su estudio concluyen que casi la mitad de los edificios de la ciudad quedarían inhabitables (unos 30.000). Los datos sobre el terremoto provienen de la intensiva campaña de campo llevada a cabo por el Ministerio de Obras Públicas, Transportes y Comunicaciones de Haití (MTPTC), que fue proporcionada para este estudio por el equipo de la Oficina Técnica de Evaluación de los Edificios (BTEB, *Bureau Technique d'Évaluation de Bâtiments*) del propio ministerio.

Por su parte, Lallemand (2014) calculó un escenario de daños comparable a este, pero extendiendo el estudio a las ciudades de alrededor de puerto Príncipe, las mismas que fueron dañadas en 2010. Simuló un terremoto de M 7.0 en el mismo segmento de la falla de Enriquillo. En su escenario, Lallemand predice que colapsarían al menos 60.000 edificios en toda la región, el doble que en el terremoto de 2010. Para ello, empleó curvas de fragilidad desarrolladas exprofeso para su estudio. David Lallemand fue el diseñador y coordinador de la campaña de campo para la recogida de información sobre edificios con la que se elaboró la base de datos del BTEB, que financió el Banco Mundial (Banco Mundial – TF99644, 2017).

El enfoque de la aplicación en Puerto Príncipe se plantea con base en todos estos antecedentes. El objetivo es diseñar un procedimiento de captura de información que ayude a actualizar su base de datos de la edificación y la de otras ciudades del país, en vista de que el riesgo sísmico se mantiene alto y de que la construcción sigue aumentando en número, por la migración de la población desde las zonas rurales¹⁶, pero no en calidad. El procedimiento debe contemplar la posibilidad de que no se tenga acceso a bases de datos oficiales y de que haya que recoger información de áreas urbanas extensas.

5.2.1 Área de estudio y datos

Torres et al. (2016) clasificaron las construcciones de Puerto Príncipe en seis categorías principales (Tabla 5-17): 4 de estructura reforzada, ya sea de hormigón o de mampostería; 1 de mampostería sin reforzar; y 1 de madera. Cada una de ellas ya representa un modelo de vulnerabilidad, puesto que tienen asociado el par de curvas [espectro de capacidad-función de fragilidad] que representan el desempeño de cada tipo de edificio ante un sismo y los daños que experimentaría (Molina et al., 2014). Las cifras correspondientes al número de edificios que aparecen en la tabla provienen de la base de datos del MTPTC de Haití.

¹⁶ <http://www.worldbank.org/en/news/infographic/2015/09/29/towards-a-new-road-map-for-inclusive-and-sustainable-growth-haiti>

Tabla 5-17. Tipologías constructivas de Puerto Príncipe. Tomada de Torres et al. (2016).

Materiales constructivos		Tipología edificatoria		Num. Edif.	Num. Edif. Totales
Estructura	Cerramientos	MBT desagregada	MBT		
Hormigón Armado	Hormigón armado	RC-SW		966	
Hormigón Armado	Mampostería de bloques	RC-CB	RC	47301	49667
Hormigón Armado	Mampostería	RC-UM		1207	
Mampostería reforzada	Mampostería de bloques	RL-BM		193	
Mampostería confinada	Mampostería	CM-UM	MAndW	10507	13401
Madera	Metal/Mampostería	W-UM		2894	

En la Figura 5-22 se muestran fotografías de construcciones típicas de Puerto Príncipe. La predominante es la (A), que se caracteriza por tener pórticos estructurales (vigas y pilares) de hormigón armado con paredes de mampostería sin reforzar. Se correspondería con la MBT RC-CB de la Tabla 5-17. Este tipo de construcción presenta una gran variación en calidad, como puede verse por comparación con la fotografía (B), lo que afecta a su vulnerabilidad. Es la tipología predominante en la ciudad y en el país, y se puede encontrar prácticamente en cualquier zona del centro o de las afueras. En la tabla puede verse que, del total de 63068 edificios, más de 47000 son de este tipo. En general, las edificaciones más grandes y altas de esta tipología suelen ser más resistentes; y, por el contrario, las más pequeñas suelen ser más vulnerables (Dr. D. Lallemand, comunicación personal). Por tanto, una subdivisión de esta categoría para diferenciar distintos grados de vulnerabilidad podría basarse en una combinación de área y altura. Hay que tener en cuenta que existen otros factores urbanísticos que modifican la vulnerabilidad, y que tienen que ver con la relación de la posición de un edificio con respecto a su vecino. Esto puede apreciarse en la fotografía (C), donde se muestra un claro ejemplo de configuración urbana sin planificar. Se ve cómo la separación entre edificios no es adecuada, ni los forjados de los edificios vecinos están a la misma altura, lo que puede dar lugar a daños por efecto golpeteo. También se ven irregularidades en planta y en alzado creadas por los grandes balcones o terrazas contruidos, que además están soportados por pilares demasiado esbeltos.

Otra categoría predominante en la ciudad es de mampostería sin reforzar, CM-UM. La fotografía (D) es un ejemplo de viviendas de este tipo, donde se ve que los ligeros tejados son de madera y láminas metálicas, y la mampostería puede ser de roca y de bloque. Finalmente, se encuentran edificaciones con estructura de madera, que presentan cerramientos de láminas metálicas o de mampostería de ladrillo. Las primeras (fotografía E) son las más vulnerables y suelen encontrarse en áreas ocupadas por asentamientos informales. Las segundas (fotografía F) representan un tipo de construcción de herencia colonial francesa. Según la información contenida en la base de datos del MPTTC, fue una de las tipologías que peor se comportó en el sismo de enero de 2010, pues, a pesar de que la madera dota a la estructura de ductilidad, la gran masa que aporta la mampostería contrarresta el buen desempeño que suele caracterizar a las ligeras viviendas de madera.

A. Edificio de estructura de hormigón armado con muros de mampostería. Construcción de buena calidad que resistió al terremoto. Fotografía propia.

B. Edificio de estructura de hormigón armado con muros de mampostería. Construcción de mala calidad. Fotografía propia.

C. Disposición de los edificios sin planificar. Fotografía de Haendel Dorfeuille.

D. Vivienda de mampostería. Fotografía cedida por Christina Corbane.

E. Viviendas de madera y láminas metálicas. Fotografía propia.

F. Vivienda colonial con estructura de madera y muros de mampostería de ladrillo. Fotografía propia.

Figura 5-22. Construcciones típicas en Puerto Príncipe.

Para el presente estudio, se han considerado dos grandes grupos de tipologías edificatorias, de acuerdo con las recomendaciones del Dr. D. Lallemand (comunicación personal). Por un lado, las 4 categorías con estructura armada (RC-SW, RC-CB, RC-UM y RL-BM) se han unido en 1 sola,

denominada RC, ya que, en la práctica, son muy parecidas entre sí, especialmente las dos de hormigón armado y muros de mampostería sin reforzar (sean bloques, RC-CB, o rocas, RC-UM). Por otro lado, se ha considerado otra tipología en la que se engloban las construcciones de mampostería y las de madera, que son de peor calidad que las de RC. Su vulnerabilidad podrá variar en función de su área y altura, y de su localización (siendo más vulnerables si se encuentran en áreas deprimidas de la ciudad). A esta categoría se le denomina MAndW. Esta agrupación de categorías está en acuerdo con el trabajo de Hancilar et al. (2013), que desarrollaron curvas de fragilidad para dos tipologías, hormigón armado y madera, empleando datos tomados en campo de más de 6.400 edificios tras el terremoto de 2010 (Corbane et al., 2011). El modelo de vulnerabilidad se asigna a estas clases mediante las funciones de daño desarrolladas y calibradas en Molina et al. (2014) con la siguiente correspondencia: las funciones asociadas a CM-UM para la tipología MAndW, y las funciones de RC-CB para la tipología RC.

En esta aplicación se trabaja en dos zonas de estudio. La más extensa cubre Puerto Príncipe y toda el área poblada de sus alrededores (Figura 5-23), y alcanza 15800 Ha. Las zonas del centro y el suroeste están ocupadas, principalmente, por las áreas urbanas de Puerto Príncipe y Carrefour. Hacia el norte y al este, se encuentran extensas áreas rurales (Marin y Croix-des-Bouquets, respectivamente). Pétionville es una zona residencial ubicada al este de la capital. Esta extensa zona es la que se estratifica en la primera fase. La geometría de esta zona se ha conseguido por medio de una clasificación binaria de la imagen en área construida y no construida. El proceso se describe en la sección 5.2.2.1. El resto del estudio se centra en el área urbana de Puerto Príncipe (Figura 5-23), que tiene una extensión de 2500 Ha y está dividida en 36 distritos.

Áreas de estudio de Puerto Príncipe

Figura 5-23. Áreas de estudio para la aplicación de Puerto Príncipe.

En el momento de la ocurrencia del terremoto, era difícil encontrar información espacial pública de Haití. Todas las personas que quisieron prestar ayuda de una u otra manera (cooperantes, voluntarios, ONGs, investigadores...), no encontraban la manera de obtener la información clave para diseñar su estrategia. Se necesitaba conocer el estado de las carreteras, la ubicación y el tamaño de las áreas urbanas y las rurales, las edificaciones, cómo localizar instalaciones e infraestructuras de interés para la gestión de la emergencia y el aporte de las primeras unidades de ayuda... Ante esta necesidad, las instituciones comenzaron a publicar información geográfica en internet. Movidos por el Banco Mundial, grandes compañías proveedoras de imágenes de satélite, como Digital Globe¹⁷, Geo Eye y SPOT, liberaron imágenes de alta resolución de toda la zona afectada, en las horas posteriores al terremoto¹⁸. Estas imágenes sirvieron a muchos investigadores para llevar a cabo su plan de ayuda. Pero, sin duda, un hecho relevante que marcó un antes y un después en el campo de la Información Geográfica Voluntaria (IGV), fue la creación del equipo HOT (*Humanitarian OpenStreetMap Team*). Justo después del sismo, HOT inició una misión global para digitalizar todos los elementos posibles de la zona afectada en la plataforma OSM (Figura 5-24). En cuestión de días, unos 600 voluntarios de más de 30 países crearon el mapa más detallado de Haití que jamás había existido¹⁹. Hoy en día, HOT sigue abriendo misiones cada vez que ocurre una catástrofe de alcance internacional en zonas del mundo donde no hay información espacial disponible o accesible.

Figura 5-24. Evolución de la información geográfica digitalizada por HOT en Haití tras el terremoto (Waters, 2010)

El Banco Mundial lanzó otras iniciativas, como la recolección de datos LiDAR para tener información tridimensional de primer nivel. Los datos fueron registrados por el *Rochester Institute of Technology*, RIT, y *Kucera International*, con subcontrato a *ImageCat Inc.* (Banco Mundial – TF99644, 2017). El vuelo se realizó entre los días 21 y 27 de enero de 2010, justo después del terremoto, y se cubrió toda el área afectada. El LiDAR y otros productos derivados, como un MDT, están accesibles en el sitio web de OpenTopography. Además de LiDAR, se registraron imágenes con las que se confeccionaron ortofotos de alta resolución, que también son públicas y pueden solicitarse al RIT.

¹⁷ <https://www.digitalglobe.com/ecosystem/open-data/haiti>

¹⁸ <https://blogs.worldbank.org/latinamerica/4-years-looking-back-openstreetmap-response-haiti-earthquake>

¹⁹ https://www.hotosm.org/projects/haiti_

Gracias a este fenómeno social, se han podido recopilar datos para realizar la aplicación en este escenario.

- Imagen Landsat TM (Data-Landsat) para la fase de estratificación, registrada el 26 de noviembre de 2009. De acuerdo con los metadatos, esta imagen fue puesta a disposición del público en el repositorio del USGS con el propósito especial de ayudar en la respuesta al terremoto de Haití (<<Purpose: Emergency Response for event "201001_Earthquake_Haiti">>). Está en formato GeoTiff de 8 bits. Tiene un GSD de 30 m para las seis bandas espectrales ente 0.45 – 12.5 μm , y un GSD de 120 m para la banda del infrarrojo térmico. Esta ha sido remuestreada a 30 m para poder utilizarla junto con las demás bandas en el análisis. La imagen no ha necesitado ningún tipo de corrección. Su sistema de referencia es WGS84-UTM18N. El RMSE en X es 3.458 m; y en Y, 3.022 m.
- Ortofotos de la ciudad de Puerto Príncipe para todas las fases del estudio, descargadas a través de un FTP proporcionado por el Laboratorio de Imagen Digital y Sensores Remotos del RIT. Hay dos conjuntos de imágenes: uno con imágenes de 3 bandas en el rango visible, RGB; y otro con registro de infrarrojo de onda corta en 1 banda, SWIR. Las imágenes RGB están en formato GeoTiff de 16 bits y tienen un GSD de 15 cm; las SWIR están en el mismo formato, pero tienen un GSD de 83 cm. Su sistema de referencia es WGS84.
- Nube de puntos LiDAR del RIT, descargado de OpenTopography, para todas las fases del estudio. La nube se registró con un sensor Leica ALS60 sn/6133 y tiene una densidad de 3.4 pto/m². Estos datos están en el sistema de referencia WGS84-UTM18N y las altitudes ortométricas se han obtenido usando el modelo EGM96.
- Una base de datos que se ha utilizado como referencia es la elaborada por el BTEB del Ministerio de Obras Públicas, Transportes y Comunicaciones de Haití. Esta base de datos contiene 88205 registros en el área de estudio de este escenario. Tras una depuración para excluir registros incongruentes (como, por ejemplo, edificios de madera con más de 5 plantas), el número total de edificios con el que se ha trabajado es 63068. Cada registro contiene las coordenadas de la localización aproximada de un edificio, y una serie de atributos relacionados con su tipología constructiva, su uso, su vulnerabilidad y el daño que sufrió en el terremoto. En total, hay 28 atributos que pueden consultarse en el formulario en papel de recogida de datos del Anexo I. Hay que tener en cuenta que en esta base de datos no están recogidos todos los edificios del área de estudio y que las localizaciones no son exactas (Figura 5-25 (A)), pues fueron tomadas por posicionamiento absoluto por medida de código registrada con un GPS autónomo de mano (habitual en este tipo de campañas de campo). Esto hace que no puedan utilizarse para verificación, como se verá en las secciones 5.2.2.1 y 5.2.2.3). El sistema de referencia de estos datos es WGS84.
- Se ha utilizado otra base de datos de referencia de edificios, sólo para verificación. Es la compilada en el marco del proyecto realizado conjuntamente por el Centro Nacional de Información Geoespacial de Haití (CNIGS, *Centre National de l'Information Géo-Spatiale*) y tres entidades internacionales referentes en esta materia: el Banco Mundial, el

Programa de Aplicaciones Satelitales del Instituto de Formación e Investigación de la ONU (UNOSAT-UNITAR, *Operational Satellite Applications Programme of the United Nations Institute for Training and Research*) y el Centro de Investigación de la Comisión Europea (JRC, *European Commission's Joint Research Centre*). Corbane et al. (2011) explican el proceso de elaboración de esta base de datos y otros detalles de este ejemplar proyecto colaborativo. La base de datos contiene 6492 edificios con información sobre su tipología constructiva, el número de plantas, el uso y el daño sufrido por el terremoto. En el área de estudio de este escenario quedan contenidos 2728 edificios. A diferencia de la base de datos anterior, la localización de los edificios sí es exacta, de manera que se puede relacionar inequívocamente cada edificio con su información (Figura 5-25 (B)). Esta base de datos es la utilizada también por Hancilar et al. (2013) en su estudio de las tipologías constructivas de Haití. El sistema de referencia de estos datos es NAD1983-UTM18N.

Todos los datos se han integrado en el SIG, y se ha homogeneizado su sistema de referencia a WGS84-UTM18N, siguiendo las indicaciones del director del CNIGS, Bobby Emmanuel Piard (comunicación personal).

Figura 5-25. Ubicación de los edificios (A) de la base de datos del MTPTC de Haití y (B) de Corbane et al. (2011).

5.2.2 Proceso de cálculo diseñado para Puerto Príncipe. Resultados y análisis.

La Figura 5-26 muestra en rojo las técnicas seleccionadas para el escenario particular de Puerto Príncipe. Esta selección, además de otras decisiones tomadas en referencia a los procesos de cálculo, se ha hecho teniendo en cuenta el objetivo del estudio en este escenario, los datos disponibles, las tipologías de edificios existentes y la configuración de la ciudad. A continuación, se presentan los procesos de cálculo, así como los resultados y su análisis, de cada fase del estudio.

Figura 5-26 Proceso metodológico aplicado en el escenario de Puerto Príncipe.

5.2.2.1 Fase 1: estratificación del escenario de Puerto Príncipe

El proceso de estratificación de Puerto Príncipe se describe en Wieland et al. (2016). Se han identificado cinco patrones urbanos, que se muestran en la Figura 5-27 y se detallan a continuación.

1. Residencial: se caracteriza por tener viviendas unifamiliares aisladas de áreas superiores a 100 m². Las viviendas son de buena calidad y están rodeadas de jardines con mucha vegetación y espacios abiertos. Las calles son anchas y forman tramas regulares. Este tipo de edificación se localiza, generalmente, en los alrededores del área central de las principales ciudades, donde parece que se ha respetado el planeamiento urbanístico.
2. Urbano Regular: es un patrón propio del centro de las ciudades, donde la densidad de construcción es más alta y los edificios son más pequeños que en las zonas residenciales. Entre los edificios se puede encontrar algo de vegetación aislada, y las calles forman una trama regular.
3. Urbano Irregular: es también un patrón del centro de las ciudades, pero con edificios algo más pequeños que los del patrón Urbano Regular (con áreas de alrededor de 50 m²) y de peor calidad. Lo que mejor caracteriza a este patrón es la alta densidad de construcción y la irregularidad de su red viaria, formada en su mayoría por calles sin asfaltar.
4. Rural: se corresponde con pequeñas poblaciones o zonas de campo a las afueras de las grandes ciudades. La densidad de vegetación es más alta que en cualquiera de los otros patrones, y las viviendas están más esparcidas. Se identifican algunas estructuras regulares, como vallas o muros, que separan las propiedades familiares y los campos de cultivo. Se pueden encontrar edificaciones pequeñas (menores de 50 m²) y otras más grandes (entre 50 m² y 100 m²), estando el tamaño directamente relacionado con su calidad constructiva.
5. Industrial: se identifica claramente por el gran tamaño y la forma regular de sus construcciones. No suele haber vegetación en estas zonas y las calles son anchas y regulares.
6. Informal: este patrón se caracteriza por tener viviendas de muy reducido tamaño (desde 10 m²), formas regulares, generalmente cuadradas, y de muy mala calidad. Algunas de estas zonas se encuentran en regiones de topografía accidentada, como laderas de montañas u orillas de ríos, y siempre en las afueras de las ciudades, lindando con las áreas rurales. Una clave para su identificación es la caótica red viaria, compuesta por callejones estrechos que apenas se ven en las ortofotos ni se distinguen en las nubes de LiDAR.

Tras varios intentos de clasificación de los patrones urbanos incluyendo estas seis clases, finalmente se decidió excluir los asentamientos Informales porque los modelos no eran capaces de distinguirlos del patrón Urbano Irregular con una precisión aceptable. Por tanto, esta primera fase de estratificación por procesos semiautomáticos se ha realizado para los cinco primeros patrones. Sobre el resultado final obtenido, se identificaron manualmente

los segmentos de asentamientos Informales que entraban en el área de estudio de las siguientes fases: el área urbana de Puerto Príncipe.

Figura 5-27. Patrones urbanos identificados en la ciudad de Puerto Príncipe

Segmentación del área de estudio

La segmentación de la imagen Landsat TM se ha realizado empleando el algoritmo de Felzenszwalb y Huttenlocher (2004) para buscar los bordes que separan agrupaciones de píxeles homogéneas, como se describe en la sección 4.2.1.1. Los segmentos creados se han clasificado, en primer lugar, en área construida y no construida. Para ello, se ha aplicado el modelo predictivo previamente creado con SVM por Wieland y Pittore (2014) con otras imágenes de Landsat TM, de forma que este primer paso de detección del área construida ha sido totalmente automático. El resultado se ha validado con una muestra de 100 instancias tomadas aleatoriamente y se han obtenido resultados muy precisos: F1-Score y precisión global del 92%. El área construida resultante es la que se estratifica en esta fase, y se muestra en la Figura 5-23 (en blanco).

Esta área ha vuelto a segmentarse con el mismo algoritmo, obteniéndose un total de 1736 segmentos con una combinación de parámetros de escala = 20 y unión = 20. Esta ha resultado la combinación óptima que maximiza la heterogeneidad inter-grupos y la homogeneidad intra-grupos, tras realizar 384 segmentaciones con distintas combinaciones de parámetros. Un detalle del resultado se muestra en la Figura 5-28, donde puede verse cómo los bordes de los segmentos marcan cambios en la configuración de la ciudad.

Figura 5-28. Detalle de la segmentación en P. Príncipe

Clasificación de los segmentos en patrones urbanos

Para la clasificación de esos segmentos en los cinco patrones urbanos antes descritos, se han creado distintos datasets de entrenamiento, con el fin de evaluar la influencia de estos en los resultados de la clasificación. En la sección 4.2.1.2, en el apartado de “Toma de muestras”, se exponen las características de las bases de datos de entrenamiento que se quieren comparar. De acuerdo con ellas, se han creado los siguientes datasets (Tabla 5-18):

Tabla 5-18. Bases de datos de entrenamiento creadas para clasificar los patrones urbanos de Haití.

Dataset	Distribución de las muestras	Número total de instancias	Método de muestreo	Tipo de muestra
1. RUB-SQ-RD	Desequilibrada	75	Aleatorio	Cuadrado
2. RUB-SE-RD	Desequilibrada	75	Aleatorio	Segmento
3. R50-SQ-RD	Equilibrada	250 (50 por clase)	Aleatorio	Cuadrado
4. R50-SE-RD	Equilibrada	250 (50 por clase)	Aleatorio	Segmento
5. R15-SQ-RD	Equilibrada	75 (15 por clase)	Aleatorio	Cuadrado
6. R15-SE-RD	Equilibrada	75 (15 por clase)	Aleatorio	Segmento
7. R15-SQ-GD	Equilibrada	75 (15 por clase)	Guiado	Cuadrado
8. R15-SE-GD	Equilibrada	75 (15 por clase)	Guiado	Segmento

Los dos primeros datasets de referencia tienen una distribución de muestras desequilibrada (RUB, *unbalanced reference*), es decir, contienen 75 instancias en total que se distribuyen en número proporcional a la presencia de cada clase en el área de estudio. Por el contrario, los datasets 3 y 4 tienen el mismo número de instancias por clase, 50, hasta un total de 250 muestras. Para crear las bases de datos 5 y 6, se han seleccionado sólo 15 muestras por clase. La selección de muestras de todos estos datasets se ha realizado de forma aleatoria (RD, *random*). En cambio, las bases de datos 7 y 8 se han creado seleccionando 15 instancias por clase de manera manual, en un proceso guiado por el usuario (GD, *guided*).

En todas las parejas de datasets, hay uno que contiene muestras de segmentos y otro que contiene muestras con geometría regular (SE, *segment* vs. SQ, *square*). Los segmentos completos se han tomado directamente de la segmentación. Las muestras regulares son cuadrados de 200 m de lado, que se han creado dentro de cada segmento muestra (Figura 5-29).

A. Muestras cuadradas

B. Muestras de segmentos

Figura 5-29. Muestras cuadradas y segmentos en los datasets de entrenamiento.

NOTA: imagen de referencia de alta resolución.

A parte, se ha creado también una base de datos para verificación que contiene 20 muestras por clase, lo que hace un total de 100 instancias.

Selección de atributos del vector de características

El primer paso ha sido calcular el vector de características y elegir qué atributos se utilizarán para crear el modelo de predicción con SVM, que serán los de mayor poder discriminante. Para ello se han utilizado los datasets aleatorios con mayor número de muestras, R50-SQ-RD y R50-SE-RD. Se han calculado, para ambos datasets, 29 características espectrales, más 2 del NDVI, otras 56 de textura (derivadas de la matriz de co-ocurrencia de niveles de gris, GLCM) y, además, 10 características geométricas. Estas últimas no se han calculado para los cuadrados, ya que son todos iguales y los atributos geométricos no aportan un valor diferenciador. Hay que tener en cuenta que las características se han calculado sobre la imagen Landsat TM, que no es la que se muestra como referencia en la Figura 5-29.

Para llevar a cabo todos estos procesos de segmentación de imagen y cálculo de características se ha implementado un código en C++ utilizando las librerías GDAL y OpenCV. Los modelos de clasificación de segmentos con SVM se han calculado con la librería scikit-learn en Python. Se han aplicado los tres métodos de selección de atributos detallados en la sección 4.2.1.2: RFE, ReliefF (estos dos incluidos en el código) y análisis discriminante (en SPSS). Mediante una validación cruzada de 5 iteraciones, se han ido obteniendo las curvas ROC y las AUC, como métricas de precisión, para comparar los resultados. Se muestran en la Figura 5-30 (A) para las muestras cuadradas y (B) para los segmentos. Además, bajo las gráficas, se presenta una tabla con el número y el tipo de atributos seleccionados con cada método. La curva azul se ha calculado con todos los atributos (Todos Atr.) para tener la referencia de comparación con los sub-grupos. En todos los casos se obtienen altas precisiones, con AUC cercanas o superando 0.90; si bien es cierto que la precisión mejora ligeramente cuando se usa un sub-grupo de atributos en lugar de todos ellos.

Figura 5-30. Curvas ROC obtenidas para diferentes conjuntos de atributos, con muestras cuadradas (A) y con segmentos (B).

Con RFE, el sub-grupo que se crea está dominado por características espectrales en ambos datasets (cuadrados y segmentos), seguidas de la textura, y las dos derivadas del NDVI. Este método de selección de atributos considera que las características geométricas apenas tienen poder discriminante para estas clases. ReliefF también selecciona un sub-grupo en el que predomina la información espectral, pero es el método que más indicadores de textura y geometría escoge. Análisis discriminante crea un sub-grupo compuesto también por características espectrales, en su mayoría, seguidas de indicadores de textura y geométricos. Las características relacionadas con el NDVI siempre se seleccionan, dado su elevado poder discriminante para separar las clases que tienen vegetación y las que no.

A juzgar por las AUC, el método de selección de atributos no parece que tenga una influencia significativa en el resultado final de la clasificación de patrones urbanos, pues sus valores son muy similares dentro de cada dataset. Aunque sí es significativo que tanto ReliefF como análisis discriminante identifican un grupo de atributos más pequeño que RFE en el dataset de segmentos y, sin embargo, permiten alcanzar las mismas precisiones.

Como conclusión de este resultado, se ha decidido utilizar ReliefF para los siguientes experimentos, ya que ha funcionado mejor que RFE y es más sencillo de implementar en el código que el análisis discriminante.

Tamaño de la muestra

Para evaluar cuál sería el mínimo tamaño muestral que proporciona una precisión aceptable, se han comparado cinco clasificaciones realizadas con sub-grupos de muestras tomados de los datasets completos (R50-SQ-RD y R50-SE-RD). Cada subgrupo tiene un tamaño diferente, con un número total de muestras que va aumentando progresivamente. Para poder comparar las clasificaciones, se han realizado con validación cruzada de 5 iteraciones, de manera que para cada entrenamiento y validación se tiene la media de la precisión global y su desviación estándar. Con estos datos se han elaborado curvas de aprendizaje que se muestran en la Figura 5-31.

Figura 5-31. Curvas de aprendizaje del clasificador para diferentes tamaños muestrales.
(A) muestras cuadradas; (B) segmentos.

Según puede verse en las gráficas (Figura 5-31), cuando se trabaja con un número demasiado pequeño de muestras se obtienen grandes diferencias en la precisión de entrenamiento y de validación. Esto indica que el clasificador tiende a crear modelos sobre-ajustados que no funcionan bien con datos nuevos. Según se va aumentando el número de muestras, se consigue aumentar la precisión de validación y disminuir la varianza, lo que aumenta la capacidad de generalización del clasificador. Para alcanzar una precisión global del 60%, haría falta un mínimo de 100 muestras (20 por clase), en el caso de los cuadrados, y de unas 140 (más de 25 muestras por clase), para los segmentos. La tendencia de las gráficas parece indicar que si se sigue aumentando el número de muestras, la precisión aumentaría. No obstante, en este estudio se está considerando prioritario mantener el tamaño muestral lo más reducido posible.

En lugar de aumentar la muestra, para optimizar el desempeño del clasificador y obtener precisiones por encima del 60%, se ha planteado hacer la selección de atributos y parametrizar el SVM y el kernel que se utilice.

Proceso de muestreo (aleatorio vs. guiado y equilibrado vs. desequilibrado)

Para evaluar la influencia del proceso de muestreo en la calidad de la clasificación de los patrones urbanos, se ha comparado el desempeño que se obtiene empleando los datasets más completos (R50-SQ-RD, R50-SE-RD) con los tres pares de datasets siguientes:

- Los datasets aleatorios que tienen un número de muestras por clase desequilibrado (RUB-SE-RD, RUB-SQ-RD).
- Los otros datasets que contienen un número equilibrado de muestras tomadas de forma aleatoria (R15-SE-RD, R15-SQ-RD).
- Y los que contienen un número equilibrado de muestras, pero tomadas manualmente (R15-SE-GD, R15-SQ-GD).

Utilizar sólo 75 muestras (15 instancias de cada clase) es, sin duda, una limitación importante para el experimento, dado que las curvas de aprendizaje anteriores mostraban que era necesario llegar al menos a 100. Sin embargo, se considera de interés trabajar con un número tan reducido de muestras, ya que es algo habitual en este campo de clasificación de imágenes, especialmente cuando se trabaja con multi-clases y con objetos (en lugar de con píxeles). En estos trabajos, los procesos de muestreo pueden resultar muy tediosos y siempre es interesante reducirlos todo lo posible. Otro objetivo de este experimento es probar el potencial de SVM para trabajar con pocas muestras y con ruido, presente especialmente en los segmentos.

En los cuatro pares de datasets, se ha calculado el vector de características y se han seleccionado las de mayor poder discriminante con ReliefF. Tras realizar la clasificación con SVM en una validación cruzada de 5 iteraciones, se han calculado las curvas ROC, las AUC y otras métricas de precisión, para evaluar la calidad de los resultados. En la Figura 5-32 se muestran en dos columnas: (A) a la izquierda, para las muestras cuadradas; (B) a la derecha, para los segmentos. En la mayoría de los casos se consiguen precisiones altas, con AUC por encima de 0.80, incluso rozando el 1. Las clases 1 (Residencial) y 2 (Urbano Regular) arrojan las AUC más bajas; de hecho, los modelos predictivos suelen confundirlas entre sí. Las menores precisiones se obtienen con los datasets desequilibrados, con F1-Score y precisión global apenas llegando al 70%. Por el contrario, los datasets equilibrados proporcionan mejores resultados, a pesar de tener sólo 15 muestras por clase, igualando los valores de F1-Score y precisión global (ambos ~75%) de los datasets completos que tienen 50 instancias por clase. La diferencia notable entre ambos pares de datasets es que, cuando se trabaja con más muestras, la varianza se reduce significativamente ($0.08^2 - 0.09^2$) con respecto a la que se obtiene cuando se trabaja con 15 muestras por clase ($0.23^2 - 0.28^2$). Esto habrá que tenerlo en cuenta a la hora de aplicar este modelo predictivo a un dataset nuevo, ya que las mayores varianzas indican una menor capacidad de generalización del clasificador.

A. Curvas ROC para muestras cuadradas

B. Curvas ROC para segmentos

Figura 5-32. Curvas ROC y métricas de precisión para 8 datasets distintos. Tomado de Wieland et al. (2016)
Nota: Class 1 a 5 son los patrones urbanos, por este orden: 1-Residencial; 2-Urbano Regular; 3-Urbano Irregular; 4-Rural; 5-Industrial.

De todos los resultados obtenidos en este experimento, las mejores precisiones corresponden al dataset equilibrado de segmentos que tiene 15 muestras por clase seleccionadas de manera manual. Con esta base de datos de entrenamiento, el F1-Score llega al 82%, la precisión global al 83% y las AUC de todos los patrones van desde 0.87 a 0.98. Si se compara con los resultados obtenidos con el dataset aleatorio (R15-SE-RD), las AUC son muy parecidas para todos los patrones, excepto para el Residencial (Clase 1), que baja a 0.60. Esto podría explicarse como algo circunstancial debido a las muestras seleccionadas aleatoriamente de esa clase.

Tipo de muestra

Cuando se evalúa la influencia de la forma de las muestras, se puede ver que esta no influye cuando se trabaja con el dataset completo de 50 instancias por clase. Con los datasets menores, los segmentos funcionan mejor si se utilizan los dataset equilibrados, tanto para selección aleatoria como manual. Por el contrario, los cuadrados arrojan mejores resultados cuando se trabaja con los datasets desequilibrados. Ante esta variabilidad, no se puede concluir que trabajar con cuadrados permite igualar las precisiones que se consiguen con los segmentos, por lo que, en principio, no se podrán crear bases de datos de entrenamiento reutilizables para distintas segmentaciones.

Para superar esto, se plantea crear bases de datos de muestras puntuales etiquetadas con el nombre de cada clase, que puedan superponerse sobre cualquier segmentación para trasladar a los segmentos las etiquetas de los puntos. De esta manera, se puede explotar al máximo el proceso de muestreo.

Verificación de los resultados

Hasta ahora, cada clasificador ha sido entrenado y validado con muestras de la misma base de datos en un proceso de validación cruzada, pero no se ha verificado con datos nuevos. Para ello, se ha utilizado el dataset creado con 100 muestras nuevas (20 por clase). Estas muestras son segmentos, lo que supone un reto más cuando se trabaja con muestras cuadradas, ya que los clasificadores se entrenan con muestras cuadradas que son homogéneas, pero se verifican en segmentos, que tienen más ruido. La Tabla 5-19, contiene la precisión global y el F1-Score de la clasificación de patrones urbanos conseguida con cada dataset. Los mejores resultados se obtienen para el dataset de segmentos equilibrados con 15 muestras por clase, que han sido seleccionadas de forma guiada (precisión global y F1-Score del 68%). La Figura 5-33 muestra el resultado de la clasificación con este dataset. Esto está en línea con los resultados obtenidos en la fase de entrenamiento.

Tabla 5-19. Métricas de calidad de la estratificación del escenario de Haití para cada dataset.

	R50-SQ-RD	R50-SE-RD	R15-SQ-RD	R15-SE-RD	R15-SQ-GD	R15-SE-GD
Precis. global	60%	72%	48%	54%	53%	68%
F1-Score	58%	70%	45%	53%	52%	68%

Figura 5-33. Resultado de la estratificación del área de estudio en patrones urbanos

El dataset de verificación ha sido creado etiquetando los segmentos mediante interpretación visual de una imagen, por tanto, se trata de una evaluación que tiene una base subjetiva. La siguiente comprobación trata de evaluar la calidad de la clasificación de los patrones urbanos mediante el coeficiente de silueta, que es una métrica objetiva de homogeneidad entre segmentos del mismo patrón (similitud) y heterogeneidad entre segmentos de patrones diferentes (disimilitud).

Para obtener el coeficiente de silueta se ha necesitado el LiDAR del RIT como fuente de datos de alta resolución. De la nube LiDAR, compuesta por 975 millones de puntos que cubren toda el área de estudio, se han seleccionado los que se superponían con los 100 segmentos de la base de datos de verificación. El LiDAR se ha clasificado en MDTOPX con el algoritmo TOD, distinguiendo las clases suelo, edificio y vegetación en tres alturas (baja, media y alta) (Figura 5-34). Los parámetros han sido establecidos para cada patrón urbano y se muestran en la Tabla 5-20. La máxima distancia entre puntos se ha fijado en 2 m.

Tabla 5-20. Combinación de parámetros del algoritmo TOD para clasificar el LiDAR en los distintos estratos

Zona	mAVA	mAVM	mAVB	MAV	mArE
Residencial	3	2	1	15	60
Urbano regular	3	2	1	10	30
Urbano irregular	3	2	1	10	20
Rural	3	2	1	15	20
Informal	3	2	1	15	10

Nota: mAVA, mAVM, mAVB son los parámetros de mínima altura de vegetación alta, media y baja, respectivamente; MAV es la altura máxima de vegetación; y mArE es la mínima área de los edificios, en metros cuadrados. Los valores de altura están en metros.

Los resultados se han convertido a ráster de 0.5 m de resolución, y con ellos se han calculado las métricas de paisaje descritas en la sección 4.2.1.3: densidad de parches (*PD*, que son las clases suelo, edificación y vegetación), distancia a los vecinos más cercanos de la misma clase (*ENN*) e índice de forma (*Shape*). Estas métricas se han calculado para cada patrón urbano y para todos los resultados de clasificación obtenidos, además de para el dataset de verificación, que se toma como referencia. Para ello, se ha empleado el software FRAGSTATS v4.2. Asumiendo que estas métricas describen de una forma objetiva los segmentos, se ha calculado con ellas el coeficiente de silueta para examinar si las clases han sido bien separadas. Una buena clasificación debería mostrar un coeficiente de silueta alto, que es un indicativo de altas homogeneidad intra-grupos y heterogeneidad inter-grupos.

Figura 5-34. Ejemplos de LiDAR clasificado en segmentos de los diferentes patrones urbanos.

Figura 5-35. Métricas de paisaje obtenidas para los cinco patrones urbanos (1 a 5) con las seis clasificaciones calculadas y la base de datos de referencia (Ref label).
Tomada de Wieland et al. (2016). PD: densidad de parches/clases; ENN: distancia euclídea entre vecinos de la misma clase; Shape: índice de forma; Comp: composición de área ocupada por cada clase de LiDAR.

Tabla 5-21. Media del coeficiente de silueta calculado para las distintas clasificaciones

Clasificación	Referencia	R15-SQ-RD	R15-SE-RD	RUB-SQ-RD	RUB-SE-RD	R15-SQ-GD	R15-SE-GD
Coef. Silueta	-0.04	-0.07	-0.04	-0.06	-0.06	-0.12	-0.01

La Figura 5-35 muestra las distribuciones de valores de las métricas de paisaje calculadas para los cinco patrones urbanos en cada clasificación. La primera columna corresponde a los resultados obtenidos con la base de datos de referencia para verificación. Así se puede ver que el patrón Urbano Regular (clase 2) muestra la mayor densidad de parches (PD) con distancias (ENN) pequeñas entre ellos. Los patrones Urbano Irregular (clase 3) y Residencial (clase 1) muestran también combinaciones de altas densidades con bajas distancias. Combinaciones opuestas, de bajas densidades y altas distancias se encuentran para los patrones Rural (clase 4) e Industrial (clase 5). Las formas más complejas, con valores altos del índice de forma (*shape*), se encuentran en segmentos Urbanos Irregulares, mientras que las más simples se

corresponden con segmentos de patrón Rural. De la composición de área ocupada por edificación, suelo y vegetación, se observan diferencias entre los cinco patrones urbanos. En zonas Rurales es donde se encuentran las mayores extensiones de suelo desnudo, y en las zonas Urbanas Irregulares predomina la edificación y la falta de vegetación. En las demás clasificaciones (columnas 2 a 6) se observan comportamientos similares.

En la Tabla 5-21 se presenta el coeficiente de silueta calculado con las métricas de paisaje para cada clasificación, y además para la clasificación de referencia. Los valores son todos próximos a cero, lo que indica que las clases (patrones urbanos) presentan algo de solape en el espacio de características. Que el coeficiente de la base de datos de referencia tenga valores similares indica que no se trata de malos resultados en las clasificaciones, sino que la realidad es compleja y las clases no quedan separadas completamente. El signo negativo informa de que hay segmentos que podrían estar mal clasificados. De nuevo, esto ocurre también con la base de datos de referencia, lo que confirma que es complicado, incluso para un operario experimentado y con conocimiento de la zona de estudio, identificar a qué patrón urbano corresponden algunas muestras. Esto se debe a que algunos segmentos caen entre dos patrones urbanos y es difícil asignarlos a una única clase. Es importante recordar que la segmentación y la clasificación se han hecho con una imagen de satélite que tiene una resolución espacial de 30 m. La clasificación que muestra el mayor coeficiente y, por tanto, la mejor separación entre patrones es la obtenida con el dataset de 15 segmentos por clase seleccionados de manera guiada por el usuario (R15-SE-GD). Esta clasificación muestra resultados mejores incluso que la clasificación de referencia. También cuando se han seleccionado aleatoriamente (R15-SE-RD) se consiguen resultados comparables. El resto de las clasificaciones muestra peores resultados, especialmente las conseguidas con las muestras cuadradas.

De estos experimentos se podría concluir que una buena forma de crear un dataset de entrenamiento para clasificar segmentos en patrones urbanos es hacerlo con muestras que sean los propios segmentos, ya que a pesar de que puedan introducir ruido, parece que este es necesario para entrenar los modelos predictivos. En cuanto al número de instancias por clase, es mejor que sea equilibrado (mismo número de muestras de cada patrón urbano) y puede ser un número bajo, aunque debería ser mayor de 15 para poder obtener precisiones aceptables del 70% - 80%, o mayores. Y finalmente, sobre el método de muestreo, un dataset creado de forma guiada por el usuario puede igualar (o incluso superar) la precisión de un dataset aleatorio, siempre que, a la hora de seleccionar las muestras, el usuario recoja toda la variabilidad presente en la población.

5.2.2.2 Fase 2: creación de la base de datos de exposición de Puerto Príncipe

Esta fase se realiza en dos pasos: (1) generación de las huellas de los edificios y (2) cálculo de los atributos en cada una de ellas. Se lleva a cabo en el área de estudio de la ciudad de Puerto Príncipe (Figura 5-23) de manera estratificada, es decir, por separado en cada patrón urbano. Para ello, se parte de la estratificación obtenida en el resultado anterior, de la que se seleccionan los segmentos que caen dentro del área urbana de la ciudad. Con el fin de poder estudiar las zonas de asentamientos Informales, estas se han localizado con ayuda de las ortofotos y se han

reclasificado manualmente los segmentos correspondientes para incluirlos en el estudio. Los segmentos de zonas Industriales no se han considerado en las siguientes fases, dado que las instalaciones industriales suelen quedar fuera del alcance de los estudios de riesgo sísmico a nivel municipal. La Figura 5-36 muestra los patrones urbanos de la ciudad de estudio, habiendo incluido los asentamientos Informales.

Estratificación de la ciudad de Puerto Príncipe

Figura 5-36. Patrones urbanos de Puerto Príncipe y áreas de muestra.

Se ha realizado un proceso de muestreo aleatorio para seleccionar áreas de muestra dentro de cada patrón. Se han creado un total de 24 muestras cuadradas, de 200 m de lado, que se reparten de la siguiente manera: 5 de patrón Residencial; 7 de Urbano Regular; 5 de Urbano Irregular; 2 de Rural; y 5 de Informal. En la Figura 5-36 puede verse la distribución espacial de estas áreas de muestra, dentro de las cuales se ha colocado una etiqueta con su identificador (FID) para poder hacerles referencia. En ellas, se ha llevado a cabo la generación de huellas, el cálculo de los atributos y la asignación de vulnerabilidad, que se explican a continuación. Los resultados se han extrapolado al resto del área de estudio.

Generación de las huellas de los edificios

Las huellas de los edificios en este escenario tienen dos orígenes: digitalización manual y análisis OBIA. En ambos casos se han empleado las ortofotos del RIT y, además, el LiDAR se ha incorporado para el análisis OBIA. Para diferenciarlas, análogamente al caso de Lorca, los polígonos resultantes de la primera opción se han llamado regulares, y a los de la segunda,

irregulares. El cálculo de atributos y la subsiguiente asignación de vulnerabilidad se ha hecho en paralelo para cada conjunto de huellas con el fin de comprobar hasta qué punto es viable trabajar con huellas generadas en procesos semi-automáticos.

Antes de comenzar cualquiera de los dos procesos, se han generado mosaicos de ortofotos para cada una de las 24 áreas de muestra. En primer lugar, las imágenes que tienen solape con cada área de muestra han sido depuradas de píxeles con valores *outliers*, considerándose como tales los que están por encima de $\mu \pm 3.5 \sigma$. También se han recortado los bordes de las imágenes, en una banda de 2 m de ancho, por presentar artefactos que desvirtuaban los mosaicos. Finalmente, con las ortofotos depuradas, se han creado los mosaicos por interpolación cúbica, que están compuestos por las 4 bandas que proporciona el RIT (RGB+SWIR).

La digitalización se ha llevado a cabo en ArcGIS, utilizando como referencia estos mosaicos. La Tabla 5-22 presenta el número de edificios digitalizados y el tiempo empleado en la digitalización de algunas de las muestras, así como los valores promedio por patrón urbano.

Tabla 5-22. Informe de digitalización de huellas de edificios

FID	Patrón urbano	Num. Edificios digitalizados	Tiempo (min)	Num. Edif. en 1 min.	Tiempo en 1 edif. (seg.)	Promedio Num. Edif. en 1 min.	Promedio Tiempo en 1 edif. (seg.)
13	Residencial	45	8	6	11		
23	Residencial	52	10	5	12	5	11
19	Rural	95	12	8	8	8	8
9	Informal	354	35	10	6		
10	Informal	104	10	10	6		
22	Informal	345	25	14	4	11	5
15	Urbano Irreg.	739	64	12	5		
16	Urbano Irreg.	416	35	12	5		
17	Urbano Irreg.	368	21	18	3		
18	Urbano Irreg.	326	28	12	5		
21	Urbano Irreg.	382	34	11	5	13	5
11	Urbano Reg.	332	28	12	5		
12	Urbano Reg.	385	38	10	6		
14	Urbano Reg.	161	14	12	5		
20	Urbano Reg.	434	35	12	5	11	5

En estas 15 muestras, se ha digitalizado un total de 4538 edificios (de los 6414 edificios digitalizados en las 24 muestras). El tiempo empleado ha sido de 6 horas y 37 minutos. En los dos patrones Urbanos (Regular e Irregular) y en los asentamientos Informales, se generan algo más de 10 huellas en 1 minuto, lo que indica que se tarda unos 5 segundos en digitalizar cada tejado. En estos patrones, la tarea de digitalizar es relativamente rápida porque los tejados son muy sencillos (la mayoría son rectángulos) y parecidos, lo que facilita la labor de fotointerpretación. A esto también contribuye la ausencia o baja densidad de vegetación. El hecho de que los tejados sean rectángulos ayuda a mantener muy bajo el tiempo de

digitalización de cada huella, ya que sólo es necesario dibujar tres vértices para obtener su geometría.

En el patrón Rural se tarda algo más (8 segundos), porque el tiempo de interpretación es mayor, al haber otro tipo de estructuras que hay que diferenciar y vegetación que oculta los edificios y produce grandes sombras. En las zonas Rurales se podrían digitalizar unos 8 edificios por minuto (hay que tener en cuenta que estos valores provienen de una sola área de trabajo y podrían variar en otras). En cuanto al patrón Residencial, parece que es el que más tiempo requiere. Esto puede deberse, por un lado, a que la fotointerpretación tampoco es inmediata debido a la presencia de vegetación y a la complejidad y variabilidad de los tejados y, por otro, porque los polígonos resultantes tienen un mayor número de vértices.

Para la generación de las huellas irregulares, se ha llevado a cabo un análisis OBIA utilizando los mosaicos de ortofotos y también el LiDAR. En esta fase, se han clasificado nuevas nubes de puntos para cubrir todas las áreas de muestra. La clasificación se ha realizado igual que en la fase anterior, con el algoritmo TOD de MDTopX y con los parámetros por estrado de la Tabla 5-20. En algunas áreas (FID 0, 5, 8, 18 y 19) ha sido necesario alinear los mosaicos con el LiDAR, lo que se ha hecho mediante un proceso de rectificación de la georreferenciación. El LiDAR de las áreas de muestra se ha rasterizado y se ha añadido como dos bandas más a la composición del mosaico, una con la intensidad de cada punto y otra con sus alturas sobre el suelo. Las alturas han sido obtenidas del MDSn, proyectando cada punto sobre el MDT. Tanto la intensidad como la altura han sido incorporadas en bandas continuas, a partir de los valores de cada punto. Además de estas bandas, se han creado otras 3 con las diferencias entre las coordenadas del color: R-G, R-B, G-B. De esta forma, se tiene una composición de 9 bandas en total: 4 espectrales, 3 combinaciones espectrales y 2 del LiDAR. Para realizar todo el proceso OBIA, se ha creado un código con Python que utiliza las siguientes librerías: Pandas, para manipulación de datos tabulares; Rasterio y Scikit-image, para gestión de imágenes; Scikit-learn, para aprendizaje automático; Matplotlib y Seaborn, para visualización. El código está publicado en GitHub (Arredondo, 2019).

En la fase de segmentación del análisis OBIA se ha utilizado el algoritmo de Felzenszwalb y Huttenlocher (2004) de detección de bordes, con las 3 bandas del visible (RGB) y las alturas del LiDAR. Los parámetros de escala y unión (en píxeles) por estrato se muestran en la Tabla 5-23.

Tabla 5-23. Parámetros de segmentación para P. Príncipe

Patrón urbano	Escala	Unión
Residencial	100	1000
Urbano regular	70	800
Urbano irregular	60	600
Rural	50	600
Informal	50	600

Para la clasificación de los polígonos resultantes, se han incorporado el resto de las bandas: SWIR, intensidad del LiDAR y diferencias de las componentes del color. Se han creado aleatoriamente dos datasets de muestras, uno para entrenamiento y otro para verificación (Tabla 5-24), y se han considerado las siguientes clases: tejado de hormigón, tejado metálico,

pavimento, sombra y vegetación. Para cada instancia, se ha calculado su vector de características compuesto por los siguientes atributos (102):

- Espectrales (7 por cada una de las 9 bandas): mínimo, máximo, media, mediana, moda, varianza y coeficiente de Curtosis.
- De textura (4 por cada una de las 9 bandas): contraste, disimilaridad, homogeneidad y ASM.
- Geométricos (3): área, perímetro e índice de compacidad, calculado con la Ecuación 3.

Tabla 5-24. Datasets de entrenamiento y verificación para la detección y clasificación de huellas de edificios

Clase	Número de muestras		
	Entrenamiento	Verificación	Total
T. Hormigón	221	69	290
T. Metálico	290	56	346
Pavimento	153	45	198
Sombra	116	26	142
Vegetación	116	29	145

Se ha empleado el algoritmo de ensamblado de aprendizaje automático *Random Forest* (RF) para crear bosques de 100 árboles, en los que se han variado los siguientes parámetros: $k = \{\text{raíz}(n), \log_2(n)\}$, siendo n la dimensión del vector de características; $f = \{\text{none}, 5, 3, 1\}$ (“none” indica que cada árbol se construye con todos los nodos que sean necesarios hasta conseguir que todas las hojas estén lo más puras posible); $mM = \{1, 5, 20, 100\}$. Todos estos parámetros se han probado en una validación cruzada de 10 iteraciones hasta conseguir la combinación de parámetros que arrojaba las mejores precisiones. Se han ido probando también varias combinaciones de bandas, desde las 4 utilizadas para la segmentación (RGB y alturas) hasta el total de 9. La precisión global de la clasificación iba mejorando a medida que se introducían las bandas, desde precisiones por debajo del 70% hasta valores por encima del 90%. La banda de alturas derivadas del LiDAR resultó ser determinante en la segmentación, pues ha permitido obtener una huella correcta en la mayoría de los tejados a dos aguas, que suelen resultar sobre-segmentados debido a la diferente respuesta espectral de los faldones (Figura 5-37 (A)). En cuanto a la clasificación, las alturas también han desempeñado un papel importante a la hora de diferenciar el pavimento de los tejados de hormigón, que suelen confundirse si sólo se emplean bandas espectrales (Figura 5-37 (B)). Y para diferenciar los tejados de hormigón de los metálicos, la banda de intensidades del LiDAR ha resultado más útil (Figura 5-37 (C)). Por el contrario, el SWIR no parece haber introducido ninguna mejora significativa, principalmente por tener menor resolución ($GSD = 83$ cm) que el resto de las bandas RGB ($GSD = 15$ cm) y no aportar el detalle necesario. Las métricas de precisión finales se presentan más adelante, en la verificación de los atributos (Tabla 5-27). Los parámetros finales han sido: $k = \text{raíz}(n)$; $f = \text{none}$; $mM = 1$.

De este análisis OBIA se obtienen finalmente los segmentos pertenecientes a tejados, que son un total de 8112 polígonos. Además, este proceso ha permitido obtener el material de su cubierta, que es uno de los atributos relevantes que se incluyen en la base de datos de exposición.

Figura 5-37. Ejemplos de resultados de clasificación con las 9 bandas

Verificación del número de huellas

Los segmentos pertenecientes a tejados han sido validados usando los tejados digitalizados como referencia. Como se mencionó en la sección 5.2.1, la base de datos del MTPTC no ha podido utilizarse en este caso, ya que la localización de cada edificio es aproximada y no es posible establecer una correspondencia unívoca entre los edificios de ambos orígenes de datos. La Tabla 5-25 muestra una comparación del número de edificios contenidos en la base de datos digitalizada (considerados en esta fase de verificación como la realidad) y los segmentos obtenidos por OBIA (considerados aquí como estimación).

En este punto del estudio, se ha decidido excluir del resto del análisis las áreas de muestra de $FID = \{10, 17, 22\}$. Las dos primeras, se ubican en zonas donde el terremoto de 2010 causó graves daños y dejó multitud de edificios colapsados; la última, es una zona de asentamiento informal en la que se encuentran numerosas ruinas (Figura 5-38). Ambas situaciones dan lugar a errores de clasificación en el LiDAR, que considera los escombros como suelo en algunos casos y como edificio en otros, dependiendo de la altura de los escombros y de las ruinas. Los errores en la clasificación del LiDAR se transmiten a la segmentación y clasificación de tejados, no permitiendo extraer las huellas de los edificios de manera fiable.

Figura 5-38. Áreas de muestra excluidas del estudio

Para el resto de las áreas de muestra, se ha realizado la validación de las huellas de los edificios. Como puede verse en la Tabla 5-25, excepto en cuatro áreas de muestra (FID = {3, 15, 11, 20}), el número de edificios obtenido por OBIA (estimado) es mayor que el digitalizado (real), lo que arroja coeficientes de comparación mayores de 1 (Nº tejados estimado/ Nº tejados real). En las áreas de muestra Residenciales, el número de huellas estimadas duplica (y casi triplica en algún caso) el real. Esto era esperable, dada la complejidad de los tejados de edificios residenciales, que suelen resultar sobre-segmentados (Figura 5-40 (A)). Por el contrario, en el estrato de patrón Informal, el coeficiente de comparación está generalmente alrededor de 1, indicando una segmentación de tejados aceptable (Figura 5-40 (B)). En este estrato, los edificios son pequeños y sencillos, compuestos por un único elemento constructivo rectangular, lo que facilita la identificación de los tejados. En los estratos urbano Regular, Irregular y Rural (Figura 5-40 (C)) no predomina un único tipo de cubierta, y esta variabilidad se refleja en los coeficientes de comparación, que oscilan entre 0.94 y 2.52. En la Figura 5-39 se representa gráficamente el coeficiente de comparación de esta verificación.

El número total de edificios estimado y real en las 21 áreas de muestra que han quedado es 6985 y 5597, respectivamente. El coeficiente de comparación entre ellos es 1.25, que indica que hay un 25% más de segmentos que tejados de edificios.

Tabla 5-25. Comparación del número de huellas obtenidas por OBIA con el número de huellas digitalizadas

Área de muestra	Patrón Urbano	Nº Tejados real	Nº Tejados estimado	Residuo	Coef. Comparación
2	Informal	508	706	-198	1.39
3	Informal	668	579	89	0.87
9	Informal	354	388	-34	1.10
0	Residencial	38	95	-57	2.50
6	Residencial	45	97	-52	2.16
8	Residencial	97	179	-82	1.85
13	Residencial	45	126	-81	2.80
23	Residencial	52	111	-59	2.13
5	Rural	103	162	-59	1.57
19	Rural	95	113	-18	1.19
15	Urbano Irregular	739	702	37	0.95
16	Urbano Irregular	416	534	-118	1.28
18	Urbano Irregular	326	557	-231	1.71
21	Urbano Irregular	382	476	-94	1.25
1	Urbano Regular	137	304	-167	2.22
4	Urbano Regular	211	257	-46	1.22
7	Urbano Regular	69	174	-105	2.52
11	Urbano Regular	332	331	1	1.00
12	Urbano Regular	385	417	-32	1.08
14	Urbano Regular	161	269	-108	1.67
20	Urbano Regular	434	408	26	0.94
Total		5597	6985		1.25

Figura 5-39. Coeficiente de comparación del número de huellas regulares e irregulares

Figura 5-40. Ejemplos de huellas obtenidas en diferentes patrones urbanos.

Cálculo de los atributos

Para cada conjunto de huellas, regulares (digitalizadas) e irregulares (segmentadas), se han calculado todos los atributos descritos en la sección 4.2.2.3. Los atributos geométricos área, coordenadas del centroide, orientación, elongación, grado de compacidad y posición en la manzana, se han obtenido de la misma forma que para el escenario de Lorca. La altura se ha obtenido con una consulta zonal al ráster de alturas derivado del LIDAR, empleando la geometría de las huellas como zonas. El estadístico elegido es la mediana. Para convertir el valor de la altura en número de plantas, se ha considerado una altura de piso de 3m. El material de la cubierta del tejado se ha tomado del paso anterior, y la forma del tejado se ha inferido a partir de este dato, sabiendo que los tejados metálicos son inclinados y, los de hormigón, son planos.

Para ilustrar el resultado, la Figura 5-41 presenta un área de muestra de los patrones urbanos Residencial e Informal, los más distintos, tanto para huellas regulares como irregulares. En ella se representan los edificios con su altura y una ventana informativa con todos los atributos de un edificio en cada zona, en color granate. Se observa que, excepto en algunas excepciones, los modelos 3D creados con ambos conjuntos de huellas guardan bastante similitud.

A. Huellas regulares de estrato Residencial

Identify Results

Layers: BD_Expo_Residential_reg

Location: (781568.191026 2051161.174794)

Field	Value
fid	178
geom	Polygon
Area	147.017123
Xcentr	781565.485322
Ycentr	2051162.713423
MBG_Width	7.364959
MBG_Length	19.961736
Orientation	93.83096
HAPP	2
Elongation	0.368954
DegreeCompactness	0.618196
RoofShape	RSH1
RoofCovering	RMN
PositionWithinBlock	BP1
Perimeter	54.65323

B. Huellas irregulares de estrato Residencial

Identify Results

Layers: BD_Expo_Residential_irreg

Location: (781573.776956 2051164.237340)

Field	Value
fid	433
geom	Polygon
Area	159.388515
Xcentr	781566.105388
Ycentr	2051163.077748
MBG_Width	8.19441
MBG_Length	19.450848
Orientation	93.34754
HAPP	2
Elongation	0.421288
DegreeCompactness	0.692633
RoofShape	RSH1
RoofCovering	RMN
PositionWithinBlock	BP1
Perimeter	53.76154

C. Huellas regulares de estrato Informal

Identify Results

Layers: BD_Expo_Informal_reg

Location: (777902.168532 2050044.362683)

Field	Value
fid	839
geom	Polygon
Area	82.945319
Xcentr	777897.768216
Ycentr	2050045.099562
MBG_Width	8.310064
MBG_Length	12.021102
Orientation	99.81345
HAPP	1
Elongation	0.69129
DegreeCompactness	0.685975
RoofShape	RSH1
RoofCovering	RMN
PositionWithinBlock	BP1
Perimeter	38.97056

Figura 5-41. Atributos de la base de datos de exposición. Ejemplos de estrato Residencial e Informal.

D. Huellas irregulares de estrato Informal

(Continuación) Figura 5-41. Atributos de la base de datos de exposición. Ejemplos de estrato Residencial e Informal.

Verificación de los atributos

Los atributos que se han verificado son el área y el material del tejado de las huellas irregulares, así como el número de plantas de ambos conjuntos de huellas. Los dos primeros se han validado por comparación con las huellas regulares. La validación del número de plantas no puede hacerse de esta misma forma, ya que el origen del dato para las huellas regulares e irregulares es el mismo, el LiDAR. Para verificar la altura, se ha utilizado la base de datos de referencia creada en Corbane et al. (2011).

En la Tabla 5-26 se presenta la comparación del área construida obtenida con las huellas segmentadas (estimada) y la obtenida con la digitalización (real), y en la Figura 5-42, el coeficiente de comparación. En el único estrato en que se subestima el área es el Residencial, pero los coeficientes de comparación (área estimada / área real) están muy cerca de 1. Este era el patrón que presentaba mayor número de segmentos respecto a huellas (Tabla 5-25), lo que es razonable, ya que cuanto más se sobre-segmenta un tejado, más se afina en la definición de su forma y su tamaño. En los patrones Informal y Urbano Irregular, parece que se ha podido llegar a un equilibrio entre el número de segmentos y el área construida, pues el coeficiente de comparación de la mayoría de estas áreas está en torno a 1. En el patrón Urbano Regular, se aprecia la misma variabilidad que en el caso del número de huellas, con coeficientes de comparación de áreas que van desde 1.03 a 1.23. En las zonas Rurales, se está sobreestimando el área de las construcciones en más de un 20%.

Se ha calculado, además, la intersección espacial de las huellas segmentadas con las digitalizadas. El área total de la intersección es 237403.27 m², lo que arroja un porcentaje de detección total del 80% (área de la intersección respecto al área real).

Tabla 5-26. Comparación del área construida (m^2) obtenida por OBIA con el número de huellas digitalizadas

Área de muestra	Patrón Urbano	Área real	Área estimada	Residuo	Coef. Comparación
2	Informal	24282.19	28681.71	-4399.52	1.18
3	Informal	24329.29	30090.35	-5761.06	1.24
9	Informal	15494.60	16539.90	-1045.30	1.07
0	Residencial	5457.74	4747.12	710.61	0.87
6	Residencial	7040.41	6150.01	890.41	0.87
8	Residencial	13445.46	11516.59	1928.87	0.86
13	Residencial	5638.84	5460.74	178.10	0.97
23	Residencial	8697.33	7339.79	1357.54	0.84
5	Rural	3791.48	5014.27	-1222.78	1.32
19	Rural	2601.58	3167.11	-565.53	1.22
15	Urbano Irregular	21585.95	27701.17	-6115.22	1.28
16	Urbano Irregular	17182.07	19919.91	-2737.84	1.16
18	Urbano Irregular	14541.83	15916.96	-1375.14	1.09
21	Urbano Irregular	21505.27	20899.58	605.70	0.97
1	Urbano Regular	15569.33	19014.87	-3445.55	1.22
4	Urbano Regular	19815.37	20453.38	-638.01	1.03
7	Urbano Regular	9253.80	9892.70	-638.90	1.07
11	Urbano Regular	16343.29	20072.54	-3729.25	1.23
12	Urbano Regular	18843.02	21511.40	-2668.38	1.14
14	Urbano Regular	13660.75	16483.85	-2823.10	1.21
20	Urbano Regular	17138.82	20285.78	-3146.97	1.18
Total		296218.41	330859.73		1.12

Figura 5-42. Coeficiente de comparación del área construida de las huellas regulares e irregulares

La principal fuente de discrepancia entre el área estimada en el análisis OBIA y la calculada a partir de la digitalización es que la primera no crea espacios entre edificios y la segunda sí. El parámetro de unión de la segmentación se ha establecido para detectar los tejados, por tanto, no permite generar objetos tan pequeños como la separación que queda entre dos edificios. Un ejemplo representativo es el área 3 (coef. Comparación = 1.24), que se muestra en la Figura 5-43 (A). Las áreas de muestra 15 y 5, con coeficientes de comparación de 1.28 y 1.32, respectivamente, son casos particulares también. El área 15 tiene una zona de graves daños, con numerosos edificios colapsados cuyos escombros pueden verse en las fotografías. El algoritmo de segmentación crea polígonos que sobrepasan los límites de los tejados y llegan hasta los escombros de los edificios colindantes, dando lugar a segmentos mucho mayores que los edificios que siguen en pie (Figura 5-43 (B)). El área 5 es una zona Rural en la que algunos árboles han sido clasificados como edificio (Figura 5-43 (C)). El NDVI ayuda a resolver este error fácilmente, pero en este estudio no se ha podido contar con la banda del infrarrojo para calcularlo.

El área total estimada es 33.1 Ha y la real es 29.6 Ha. El coeficiente de comparación entre ellas es de 1.12, es decir, el error porcentual en la estimación del área construida es del 12%. Esto implica que el área de un edificio de 100 m² se estaría estimando en 112 m², lo que no tiene apenas incidencia a la hora de clasificar un edificio en una tipología constructiva.

Figura 5-43. Principales fuentes de discrepancia en la estimación del área construida. Los edificios segmentados, al tener formas más irregulares, se han representado con relleno rojo para diferenciarlos del fondo de la imagen.

A continuación, se ha verificado el material del tejado. Para ello, se han empleado las instancias de verificación de la Tabla 5-24 y se ha calculado una matriz de confusión, de la que se han derivado las métricas precisión, recall y F1-Score. Todos estos resultados se presentan en la Tabla 5-27. Como puede verse, los índices de calidad son altos, la mayoría por encima del 90%. La identificación de tejados, que es el objetivo de esta parte del estudio, tiene un F1-Score del 99% para hormigón y del 95% para metal. La banda de las alturas derivada del LiDAR ha sido clave para distinguir los tejados de hormigón del pavimento, ambos con similar respuesta espectral en las imágenes, hasta el punto de no confundir ninguna de las instancias de estas clases. Por otra parte, la información de intensidad del LiDAR ha permitido distinguir los dos tipos de tejados, que se confundían en numerosos casos. El resultado es que el tejado metálico sólo es confundido en un caso con tejado de hormigón y, en otro, con pavimento. Las sombras y la vegetación son confundidas entre sí, principalmente, pero estas clases no son de interés para este estudio. En general, los resultados de esta clasificación son altamente satisfactorios, con medias ponderadas de las tres métricas que se sitúan en el 95% y un coeficiente kappa de Cohen casi perfecto, de 0.94.

Tabla 5-27. Matriz de confusión y métricas de calidad de la clasificación del tipo de tejado

	Predicciones					Métricas de calidad		
	T. Horm.	Pavimento	Sombra	T. Metál.	Vegetación	Precisión	Recall	F1-Score
T. Hormigón	68	0	0	1	0	99%	99%	99%
Pavimento	0	45	0	0	0	98%	100%	99%
Sombra	0	0	24	1	1	86%	92%	89%
T. Metálico	1	1	0	54	0	93%	96%	95%
Vegetación	0	0	4	2	23	96%	79%	87%
Medias ponderadas:						95%	95%	95%

Las precisiones indicadas en la tabla anterior son el resultado de comparar los segmentos clasificados con muestras de su misma naturaleza. Esta es la forma habitual de hacer la verificación tras una clasificación. Sin embargo, como ya se ha venido haciendo en apartados anteriores, en esta tesis se da un paso más y se validan estos resultados con datos de distinto origen para valorar la verdadera calidad de estos y la capacidad de generalización de los procesos. Con este fin, se ha comparado el tipo de tejado obtenido mediante OBIA con la información recogida durante el proceso de digitalización. En el SIG, se ha hecho una unión espacial para asignar a cada segmento la etiqueta del tipo de tejado que tiene la huella digitalizada dentro de la cual se encuentra su centroide. Comparando la etiqueta de tejado propia del segmento con la etiqueta asignada en la unión espacial, se puede saber cuántos segmentos de tejados han sido bien clasificados. En el Anexo IV se incluye la comparación de los tejados para todas las áreas de muestra. En las 21 figuras se presentan los tejados generados por digitalización, con su verdadera forma, posición y magnitud, sobre los que se superponen los centroides de los tejados obtenidos en el proceso OBIA. Se compara tanto la posición como el número de tejados y el tipo de material de las cubiertas.

De los 6985 segmentos clasificados como edificios en el proceso OBIA, se ha visto que 393 (un 5.6%) no encuentran intersección con las huellas digitalizadas, bien por estar en los bordes de las áreas de estudio, bien por ser otro tipo de construcción (patios, muros, ruinas, etc.). De los 6592 restantes, 5479 segmentos tienen una clasificación correcta en hormigón y metálico, lo que supone un 83% de acierto (precisión global). De los segmentos de tejado metálico, 875 (el 13%) son en realidad tejados de hormigón; y 238 segmentos (el 4%) clasificados como tejado de hormigón, son en realidad tejados de cobertura metálica. Como se verá en la sección 5.2.2.3, el material del tejado tiene una gran influencia en la clasificación de los edificios en tipologías constructivas, por tanto, será necesario evaluar hasta qué punto estos errores influyen en la clasificación de estos segmentos.

En la Figura 5-44 se muestra una comparación de huellas digitalizadas (A) con huellas segmentadas (B) en una manzana complicada de la ciudad, en la que hay tejados de distinto tipo y tamaño. La clasificación OBIA (B) ha conseguido un resultado satisfactorio en la mayoría de los tejados, si bien se aprecian pequeños errores, como los marcados con un círculo amarillo. Se ha calculado el error en términos de área también, y se ha comprobado que los segmentos mal clasificados suman 4 Ha, lo que supone un 12% de las 33 Ha del área total.

Figura 5-44. Comparación de tejados digitalizados (A) y segmentados (B).

Nota: En azul se representan los tejados de hormigón y, en rojo, los tejados metálicos. Los círculos amarillos señalan errores de clasificación.

Por último, se ha verificado el número de alturas empleando la base de datos de Corbane et al. (2011). Esta verificación se ha hecho tanto para las huellas regulares (digitalizadas) como irregulares (segmentadas). De los 6492 edificios que tiene la base de datos de Corbane, solo 356 quedan dentro de las áreas de muestra de este estudio; de esos, 295 encuentran intersección con las huellas regulares y 289 con las irregulares. El resto son edificios colapsados que Corbane et al. (2011) registraron para calcular sus curvas de fragilidad, pero que en este estudio no se han considerado.

Figura 5-45. Histograma de los residuos del número de plantas.

La Figura 5-45 muestra los histogramas de los residuos del número de plantas, para cada dataset de huellas, que se han obtenido como la diferencia entre el número de plantas estimado y el observado por Corbane et al (2011). Como puede verse, la mayoría de los residuos son cero. En las huellas regulares, son 233, lo que supone el 79% de las 295 (precisión global, OA); en las irregulares, son 207 de las 289 (OA de 72%). Del resto de residuos, la mayoría son de 1 planta (56 en las huellas regulares, que son el 18%, y 75 en las irregulares, que suponen el 26%). Los mayores errores que se cometen en la estimación del número de alturas es 2 por exceso y 3 por defecto, pero estos errores suponen menos del 3% del total de los edificios verificados.

Estos resultados están en consonancia con los obtenidos en el escenario de Lorca, lo que parece confirmar la fiabilidad de este procedimiento de estimación de alturas de edificios basado en LiDAR. Como ya se mencionó para Lorca, los modelos de vulnerabilidad agrupan los edificios en rangos de 2 o más pisos, por lo que un error de 1 o 2 plantas puede no resultar muy significativo para el objetivo de este estudio.

5.2.2.3 Fase 3: clasificación de tipologías constructivas y asignación de la vulnerabilidad sísmica a los edificios de Puerto Príncipe

Creación de los modelos predictivos para la clasificación de tipologías constructivas.

Para la asignación de la vulnerabilidad, se han creado modelos predictivos con las técnicas expuestas en la sección 4.2.3. Se ha trabajado con las huellas regulares y las irregulares de las 21 áreas de muestra. Los modelos creados dotan a cada huella de una etiqueta que indica si representa a un edificio con estructura de hormigón armado, RC (*reinforced concrete*), o de mampostería y madera, MAndW (*Masonry and Wood*). Se han creado datasets de entrenamiento y verificación equilibrados y desequilibrados (Tabla 5-28) para comparar el rendimiento de los modelos en cada caso, ya que la proporción de edificios de cada clase en la realidad es desigual (49667 edificios de RC y 13401 de MAndW). Los datasets se han generado por muestreo aleatorio estratificado, seleccionando instancias de la base de datos del MPTTC de Haití que se

encuentran fuera de las áreas de muestra y dentro de cada estrato. Para crear los modelos de predicción se utilizan los atributos área, número de plantas, material del tejado y coordenadas del centroide, que son los comunes entre estos datos de referencia y los obtenidos en este estudio; de esa manera, se podrán aplicar los modelos a los datos generados aquí. En las zonas Residenciales y Rurales, que presentan una menor densidad de edificación, el número de instancias es significativamente menor que en el resto de los estratos.

Previamente al entrenamiento de los modelos predictivos, se ha creado un ranking de los atributos de mayor a menor poder discriminante con ReliefF en WEKA. El algoritmo indica que los atributos, para la mayoría de los estratos, quedan ordenados de la siguiente manera: tipo de tejado, número de plantas, área, coordenada X del centroide, coordenada Y del centroide. En las áreas Rurales, la coordenada Y queda en segundo lugar, y el área en último, indicando que en estas zonas los edificios tienen áreas de muy diversos tamaños y, por tanto, este no es un atributo determinante para clasificar las tipologías edificatorias. Lo mismo ocurre en el patrón Urbano Regular, quedando el área en el último lugar del ranking de atributos.

Tabla 5-28. Número de muestras en los datasets de entrenamiento y verificación para la clasificación de tipologías constructivas.

	Nº Muestras en el dataset equilibrado					Nº Muestras en el dataset desequilibrado				
	Entrenamiento		Verificación		Total	Entrenamiento		Verificación		Total
	RC	M&W	RC	M&W		RC	M&W	RC	M&W	
RESIDENCIAL	240	240	60	60	600	1233	265	305	69	1872
URBANO REG.	1600	1600	400	400	4000	6927	1815	1738	447	10927
URBANO IRREG.	3200	3200	800	800	8000	14029	3496	3482	899	21906
RURAL	120	120	30	30	300	650	124	163	30	967
INFORMAL	3200	3200	800	800	8000	15492	4372	3891	1075	24830

Para decidir si utilizar el dataset equilibrado, el desequilibrado o ambos, se han creado y verificado todos los modelos con regresión logística en SPSS. Se ha podido comprobar que, con las proporciones de muestras desequilibradas, las precisiones se desequilibran también, obteniéndose más porcentaje de acierto en la clase más numerosa, RC (Tabla 5-29). Este resultado está en consonancia con los experimentos anteriores y, por tanto, se ha decidido trabajar solo con los dataset equilibrados.

Tabla 5-29. Precisiones para RC y MAndW obtenidas con los datasets desequilibrados

Estrato	Precisión RC	Precisión MAndW
Residencial	91%	31%
Urbano Regular	91%	28%
Urbano Irregular	96%	20%
Rural	94%	35%
Informal	98%	11%

Para el entrenamiento de los modelos se ha realizado una validación cruzada de 10 iteraciones. Los algoritmos utilizados han sido árboles de decisión, SVM (con kernels RBF y PUK), clasificador Bayesiano naïve y regresión logística. Los parámetros se han variado entre los siguientes valores: profundidad de los árboles = {2, 3, 5}; $C = \gamma = \sigma = \omega = \{0.01, 0.1, 1, 10, 100\}$. En WEKA se han calculado todos los modelos en la fase de diseño de este proceso. Los parámetros finales se han ajustado mediante un código en Python que utiliza la librería Scikit-Learn (Arredondo, 2019).

La Tabla 5-30 presenta un resumen del desempeño de todos los modelos a través de la precisión global obtenida en la fase de entrenamiento, así como los mejores parámetros. Se observa que los mejores modelos son los creados con SVM y, en concreto, con el kernel PUK, que alcanza precisiones entre 88% y 100%. En general, todos los modelos consiguen mejores precisiones en los dos patrones Urbanos y en el Residencial, con valores por encima del 79%. Los edificios en estas zonas parece que siguen unos métodos de construcción más normalizados que los de las zonas Rurales e Informales, cuyas precisiones bajan hasta el 70% (aunque SVM con PUK arroja precisiones muy altas en estos estratos). Es de destacar el desempeño del clasificador Naïve Bayes, que siendo el más sencillo de todos los modelos implementados, obtiene precisiones comparables a los demás.

Tabla 5-30. Resumen de los modelos predictivos creados para clasificar los edificios en MBT.

	Árbol Decisión		SVM (PUK)		SVM (RBF)		Regres. Logística	Naïve Bayes
Estrato	Precis. Global	Parám.	Precis. Global	Parám.	Precis. Global	Parám.	Precis. Global	Prec. Global
Residencial	82%	3	88%	$C = 1; \sigma = 1; \omega = 1$	86%	$C = 1; \gamma = 1$	83%	80%
Urbano Reg.	82%	3	91%	$C = 1; \sigma = 0.01; \omega = 0.1$	81%	$C = 0.01; \gamma = 0.1$	81%	81%
Urbano Irreg.	80%	3	100%	$C = 10; \sigma = 10; \omega = 0.01$	84%	$C = 1; \gamma = 10$	79%	79%
Rural	78%	3	98%	$C = 10; \sigma = 1; \omega = 0.1$	83%	$C = 1; \gamma = 1$	73%	73%
Informal	70%	3	100%	$C = 10; \sigma = 0.01; \omega = 0.1$	79%	$C = 1; \gamma = 10$	70%	70%

En el Anexo V, se incluyen todos los esquemas de los árboles y el listado de las variables seleccionadas para la regresión logística, con sus coeficientes. Los valores de corte de los árboles son los valores tipificados de los atributos (para que tengan media aritmética igual a cero y desviación estándar igual a uno). En todos los árboles, puede verse que el tejado es el primer atributo que se utiliza y, a continuación, se trabaja con las coordenadas; excepto en el estrato Residencial, que se continúa con el área. El número de plantas es utilizado únicamente en los estratos Informal y Rural. Además de en el estrato Residencial, solo aparece el área en el Urbano Regular, lo que resulta una ventaja para las huellas irregulares, que tienen áreas menos realistas. El índice de impureza de Gini está incluido en cada nodo, y también el número de muestras que separa el nodo en cada clase.

En cuanto a la regresión logística, en todos los estratos entran las variables tipo de tejado, número de plantas y área, por ese orden, para generar la ecuación. Esto está de acuerdo con el

ranking de atributos proporcionado por ReliefF. El área tiene un coeficiente próximo a cero, pero se ha decidido mantener en la ecuación porque se considera que sí es influyente en la tipología de un edificio, aunque quizá, con estos datos de entrenamiento, no se pone de manifiesto.

Las precisiones obtenidas en la fase de entrenamiento son validadas a continuación, en la fase de verificación con datos nuevos, para comprobar la robustez de los modelos y su capacidad de generalización.

Verificación de los resultados de clasificación de MBT.

Se han validado todos los modelos creados en el paso anterior empleando las muestras de verificación de la Tabla 5-28 (dataset equilibrado). El ranking de los modelos para cada estrato se presenta en la Tabla 5-31. El modelo calculado con SVM y kernel PUK es el mejor en todos los estratos, seguido del SVM con kernel RBF en la mayoría de los casos. El árbol de decisión y la regresión logística tienen desempeños similares, y también el clasificador naïve Bayes. Como se mencionó en la fase de entrenamiento, merece la pena destacar la capacidad de este último clasificador también en la fase de verificación, que, a pesar de su sencillez, consigue precisiones comparables a otras técnicas más sofisticadas, como los SVM. Donde estos demuestran su potencial es a la hora de trabajar con bases de datos que contienen ruido, como en el estrato Informal, en el que los SVM mantienen precisiones y F1-Scores por encima del 70% y los demás no.

Se aprecia una reducción general de la calidad de las clasificaciones en los patrones Rural e Informal, igual que ocurría en la fase de entrenamiento. En esas zonas, las construcciones no suelen seguir normas, lo que aumenta la componente aleatoria en los datos y dificulta la tarea de los clasificadores. Las mejores precisiones alcanzadas en estos patrones llegan al 78% (iguales al F1-Score) en las áreas Rurales y al 74% en los asentamientos Informales. En las áreas Urbanas, los valores suben al 81% en las tramas irregulares y al 82% en las regulares. Finalmente, los edificios residenciales son los mejor clasificados, con precisiones del 88% y F1-Score del 87%. Los valores de kappa de los estratos Informal y Rural indican un acuerdo moderado entre la predicción del modelo y la realidad, y los de las áreas Residenciales y Urbanas, un acuerdo alto.

Con respecto a la fase de entrenamiento, los modelos calculados con SVM mostraban mejor desempeño que los otros tres modelos en los estratos más complicados (Rural e Informal). Sin embargo, en la verificación, han reducido más sus precisiones (hasta un 26%), lo que indica que podrían haber estado sobre-ajustados. En las zonas Residenciales y Urbanas, los modelos demuestran robustez al presentar pérdidas de precisión mínimas (entre 2% - 4%; excepto el SVM con PUK en el estrato Urbano Irregular, que pierde el 19%) o incluso mejorando su desempeño.

Tabla 5-31. Ranking de los modelos predictivos creados para cada estrato ordenados según las métricas obtenidas en fase de verificación.

RESIDENCIAL			
Algoritmo Aprendiz. Autom.	Precis. Glob.	F1-Score	kappa
SVM (PUK)	88%	87%	0.75
SVM (RBF)	88%	87%	0.75
Reg. Logística	87%	87%	0.73
Árbol Decisión	87%	86%	0.73
Naïve Bayes	87%	86%	0.73
URBANO REGULAR			
Algoritmo Aprendiz. Autom.	Precis. Glob.	F1-Score	kappa
SVM (PUK)	82%	81%	0.64
Árbol Decisión	82%	81%	0.64
SVM (RBF)	81%	81%	0.63
Reg. Logística	82%	81%	0.63
Naïve Bayes	81%	80%	0.62
URBANO IRREGULAR			
Algoritmo Aprendiz. Autom.	Precis. Glob.	F1-Score	kappa
SVM (PUK)	81%	81%	0.62
SVM (RBF)	80%	80%	0.60
Reg. Logística	80%	79%	0.60
Naïve Bayes	80%	79%	0.60
Árbol Decisión	80%	79%	0.59
RURAL			
Algoritmo Aprendiz. Autom.	Precis. Glob.	F1-Score	kappa
SVM (PUK)	78%	78%	0.57
Árbol Decisión	75%	73%	0.50
SVM (RBF)	73%	73%	0.47
Reg. Logística	70%	68%	0.40
Naïve Bayes	70%	68%	0.40
INFORMAL			
Algoritmo Aprendiz. Autom.	Precis. Glob.	F1-Score	kappa
SVM (PUK)	74%	74%	0.49
SVM (RBF)	73%	73%	0.47
Árbol Decisión	70%	68%	0.40
Reg. Logística	70%	67%	0.39
Naïve Bayes	70%	67%	0.39

Las matrices de confusión que dan lugar a estas métricas de calidad se incluyen en la Tabla 5-32. En estas matrices se puede ver que el falso positivo de la clase MAndW (celdas [1,2] de cada matriz) es bastante mayor que el del hormigón (celdas [2,1]), en la mayoría de los casos. Este error indica que un edificio de hormigón se está clasificando como mampostería y/o madera.

Es necesario analizar estas diferencias antes de elegir cuál es el mejor modelo, ya que la elección puede perjudicar el equilibrio de los resultados. Se ha visto que el atributo tipo de tejado es el que más influencia tiene a la hora de clasificar un edificio en hormigón o mampostería. Los tejados de hormigón armado se encuentran en edificios con estructuras también de hormigón armado; sin embargo, un tejado con cobertura metálica puede verse tanto en pequeñas viviendas de madera y/o mampostería (Figura 5-46 (A)), como en edificios de estructura armada (Figura 5-46 (B)). Esto dificulta la distinción de las dos tipologías cuando el tejado es de metal. Una explicación al porcentaje de error que supone el falso positivo de MAndW en todos los modelos puede encontrarse en el hecho de que la mayoría de los edificios de esta ciudad tienen tejados metálicos, y los modelos tienden a clasificarlos como edificios de mampostería y madera.

Tabla 5-32. Matrices de confusión de la fase de verificación de la clasificación de MBT en Puerto Príncipe

		Predicción Residencial		Predicción Urbano Reg.		Predicción Urbano Irreg.		Predicción Rural		Predicción Informal	
		RC	WAndW	RC	WAndW	RC	WAndW	RC	WAndW	RC	WAndW
Árbol Decisión	RC	44	16	266	134	493	307	15	15	364	436
	WAndW	0	60	12	388	19	781	0	30	44	756
SVM (PUK)	RC	47	13	259	141	535	265	24	6	519	281
	WAndW	2	58	4	396	40	760	7	23	131	669
SVM (RBF)	RC	47	13	255	145	550	250	20	10	499	301
	WAndW	2	58	2	398	67	733	6	24	126	674
Regresión Logística	RC	45	15	252	148	487	313	13	17	331	469
	WAndW	1	59	0	400	7	793	1	29	19	781
Naïve Bayes	RC	44	16	251	149	479	321	13	17	331	469
	WAndW	0	60	2	398	0	800	1	29	17	783

Figura 5-46. Edificios de tipología MAndW (A) y RC (B) con tejado metálico

Con todos los modelos generados, se han clasificado las huellas regulares e irregulares en tipologías constructivas. De todos los atributos que tienen las bases de datos de exposición para

ambos conjuntos de huellas, se han utilizado aquellos con los que se han generado los modelos: tipo de tejado, área, número de plantas, y coordenadas del centroide.

Esta clasificación de tipologías constructivas también ha sido verificada por comparación con la base de datos de referencia externa de Corbane et al. (2011), empleando los mismos edificios que se utilizaron para la validación del atributo número de plantas (295 para huellas regulares y 289 para irregulares). Se han calculado las matrices de confusión y las métricas de calidad para los cuatro primeros modelos, que se presentan en la Tabla 5-33.

Se observa que tanto la precisión global (OA) como el F1-Score (F1) están, en la mayoría de los casos, entre el 70% - 80%. Esto indica que, a pesar de la propagación de incertidumbres a lo largo de todo el estudio para generar las huellas, calcular los atributos y entrenar los modelos de clasificación de tipologías constructivas, se pueden obtener precisiones altas en un escenario complicado, como es la ciudad Puerto Príncipe. Esto es así incluso para las huellas irregulares que, a pesar de presentar algunos errores en la identificación de los tejados (un 12% del área total construida), consiguen precisiones, si bien algo más bajas, comparables con las de las huellas regulares.

Tabla 5-33. Matrices de confusión y métricas de calidad de la verificación de las MBT

SVM (PUK)								
Predicción H. Regulares					Predicción H. Irregulares			
	RC	W&W	FP rate	F1-Sc.	RC	W&W	FP rate	F1-Sc.
RC	174	52	14%	85%	175	55	29%	83%
W&W	10	59	23%	66%	17	42	24%	54%
OA=79% F1=75%					OA=75% F1=68%			
SVM (RBF)								
Predicción H. Regulares					Predicción H. Irregulares			
	RC	W&W	FP rate	F1-Sc.	RC	W&W	FP rate	F1-Sc.
RC	183	43	41%	84%	185	45	46%	84%
W&W	28	41	19%	54%	27	32	20%	47%
OA=76% F1=69%					OA=75% F1=65%			
Árbol de decisión								
Predicción H. Regulares					Predicción H. Irregulares			
	RC	W&W	FP rate	F1-Sc.	RC	W&W	FP rate	F1-Sc.
RC	165	61	3%	84%	158	72	7%	81%
W&W	2	67	27%	68%	4	55	31%	59%
OA=79% F1=76%					OA=74% F1=70%			
Regresión Logística								
Predicción H. Regulares					Predicción H. Irregulares			
	RC	W&W	FP rate	F1-Sc.	RC	W&W	FP rate	F1-Sc.
RC	167	59	3%	85%	157	73	7%	80%
W&W	2	67	26%	69%	4	55	32%	59%
OA=79% F1=77%					OA=73% F1=70%			

Mediante la Figura 5-47 se pretende mostrar la dificultad que entraña la clasificación de los edificios para los modelos predictivos en este complicado escenario. Se presentan dos fotografías de edificios cuya clasificación de MBT no coincide con la asignada por el equipo de

trabajo de Corbane et al. (2011). En la fotografía (A) se tiene un edificio de pequeñas dimensiones, una sola planta y tejado de lámina metálica que el modelo ha clasificado como de mampostería y madera. Esta clasificación resulta razonable, sin embargo, la realidad indica que, sorprendentemente, tiene una mínima estructura de pilares de hormigón. Un caso opuesto es el de la fotografía (B), que por sus dimensiones y regularidad podría ser de RC, pero parece que sólo está hecho de mampostería con forjado de madera.

A. Edificio clasificado como MAndW en este estudio, pero etiquetado como RC en Corbane et al. (2011. Foto n° 3023)

B. Edificio clasificado como RC en este estudio, pero etiquetado como MAndW en Corbane et al. (2011. Foto n° 3036)

Figura 5-47. Fotografías de edificios con errores de clasificación.

Respecto a la elección del mejor modelo, como ya se adelantó en el apartado anterior, parece más acertado continuar trabajando con el modelo creado con SVM (PUK). En la Tabla 5-33 se puede ver que hay otros algoritmos que permiten obtener métricas de calidad ligeramente mayores en alguna de las tipologías, especialmente la regresión logística y el árbol de decisión. Sin embargo, SVM con kernel PUK tiene la capacidad de acercar los porcentajes de ambas clases y mejorar las predicciones de RC, que es lo que interesa particularmente en este estudio. Por otro lado, incluso eligiendo el modelo de SVM (PUK), el error por falso positivo de MAndW permanece más alto que el de RC. Aunque lo deseable sería que los errores fueran mínimos, los cierto es que esta distribución de errores es favorable en este caso. El error por falso positivo de MAndW indica que un edificio de hormigón se está clasificando como mampostería y/o madera, es decir, el clasificador dice que es más vulnerable a sismos. Esto probablemente dé lugar a una sobrestimación del daño, que sería un escenario más conservador y justificaría políticas de vigilancia y un estudio más profundo del riesgo sísmico de la ciudad.

La Figura 5-48 presenta un área de muestra de cada estrato con la clasificación hecha con SVM (PUK) para los edificios digitalizados (regulares) en la columna de la izquierda, y los segmentados (irregulares) a la derecha. Como puede verse, las discrepancias para cada grupo de huellas aumentan en los estratos Rural e Informal, que como ya se ha visto anteriormente, son los más complicados de clasificar debido a que no siguen apenas normas de construcción.

Figura 5-48. Resultados de la clasificación de huellas regulares e irregulares en tipologías constructivas

En el Anexo VI se presentan los resultados completos para huellas regulares e irregulares en las 21 áreas de muestra. Además, se incluyen figuras de la comparación de estas clasificaciones con los datos de referencia de Corbane et al. (2011), para las seis áreas de muestra en las que se encuentran los edificios comunes. Hay, al menos, un área de muestra en cada estrato, excepto en Informal.

Asignación de un modelo de vulnerabilidad.

Una vez que se tienen todos los edificios clasificados en alguna de las dos tipologías constructivas, se les asigna su modelo de vulnerabilidad de Molina et al. (2014): RC-CB para RC y CM-UM para MAndW. Cada tipología se ha subdividido en rangos de altura, siguiendo la descripción que proporcionan los autores, en low-rise (1-3 plantas) y mid-rise (4-6) plantas. En la Figura 5-49 se muestra la distribución de vulnerabilidad en los patrones urbanos de Puerto Príncipe. El número de muestras de cada uno es el siguiente: 5 de Residencial; 7 de Urbano Regular; 4 de Urbano Irregular; 2 de Rural; y 3 de Informal. Como se aprecia, la gran mayoría de los edificios son de baja altura (menos de 3 plantas). El estrato Residencial es el que presenta menos densidad de construcción, que con 5 áreas de muestra, apenas suma 270 edificios. Por el contrario, el Informal es el más densamente poblado, con casi 1400 edificios en 3 muestras. En todos los estratos hay más edificios de mampostería y madera que de hormigón, excepto en el Residencial, en el que las construcciones son, en general, de mayor calidad, dado el mayor poder adquisitivos de sus propietarios.

Figura 5-49. Distribución de vulnerabilidad en Puerto Príncipe

5.2.2.4 Extrapolación al resto del área de estudio

Hasta este punto del estudio, se han analizado los edificios de las 21 áreas de muestra, que han quedado identificados y clasificados en dos tipologías constructivas, RC y MAndW. Las áreas están repartidas en 5 patrones urbanos diferentes, o estratos, que cubren la ciudad de Puerto

Príncipe. En este apartado se han extrapolado los resultados obtenidos en las áreas de muestra al resto de la superficie de los estratos, considerada la población. La extrapolación se ha realizado por medio de la densidad de edificios (Pittore y Wieland, 2013) de cada clase calculada para cada estrato utilizando los datos de todas sus áreas de muestra.

Para poder tener una distribución de valores de densidad de edificios en lugar de un único valor por área, se ha realizado un remuestreo empleando la técnica Bootstrap. En cada área de muestra se han creado 200 círculos de 40 m de radio, y se han seleccionado los edificios que quedan dentro de cada uno de ellos. La Figura 5-50 ilustra este proceso, donde se ven las dos clases de edificios y los círculos en el interior de la muestra.

Figura 5-50. Círculos creados para el Bootstrap

Este proceso se ha realizado tanto para las huellas regulares como para las irregulares, y también para los edificios contenidos en la base de datos del MTPTC de Haití. La primera extrapolación se ha hecho para los datos de referencia del Ministerio, para comprobar si este método de extrapolación resulta correcto. La base de datos contiene edificios de toda la ciudad, pero se sabe, por el análisis de las muestras, que está incompleta. Asumiendo que la falta de completitud es homogénea en todo el área de estudio, se ha comprobado si el número total de edificios de cada clase estimado con la extrapolación es comparable al observado (contenido en la base de datos). El estadístico elegido para extrapolar la densidad de edificios es la media. Se ha calculado un intervalo de confianza al 95% para este estimador con la distribución T-Student, ya que la varianza de la población es desconocida (dada la falta de completitud de la base de datos). También se ha calculado la mediana, pero en algunas áreas de muestra se ha obtenido un valor de cero para este estadístico por estar poco pobladas. La Tabla 5-34 contiene todos estos datos por cada tipología constructiva y también los totales. Para obtener la estimación del número de edificios en el resto del área de estudio, se ha multiplicado la densidad de cada estrato por su área, con los siguientes valores en km²: Residencial = 2.524081; Urbano Regular = 5.019572; Urbano Irregular = 5.443670; Rural = 1.044982; Informal = 4.856459.

Los resultados parecen indicar que este método de extrapolación es válido, ya que en la mayoría de los casos se obtiene una estimación del número de edificios comparable al observado, es decir, al contenido en la base de datos del Ministerio de Obras Públicas haitiano. En los estratos Residencial y Urbano Regular, el dato observado está dentro de los intervalos de confianza (I.C.) calculados para la media en la tipología RC. En el Rural las observaciones quedan muy cerca del límite superior del I.C. para RC y del inferior para MAndW. El número de edificios en el estrato Informal queda sobreestimado en un 60%, lo que indica que, o bien las áreas de muestra

analizadas están más pobladas que el resto del estrato, o bien la falta de completitud de la base de datos del ministerio es más acusada en otras zonas Informales de la ciudad. Probablemente el motivo sea una combinación de ambos factores. Esto habrá que tenerlo en cuenta a la hora de analizar los resultados de extrapolación de los edificios obtenidos en este estudio por digitalización y por segmentación.

Tabla 5-34. Resumen de la extrapolación con Bootstrap para la B.D. del MTPTC

Estrato	MBT	Densidad (Nº Edif. / km ²)		Estimación Nº Edif.		I.C. (95%) de la estim.		Nº Edificios Obs.		
		Mediana	Media	Por MBT	Total	Lím. Inf.	Lím. Sup.	Por MBT	Total	Residuo
Residen.	M&W	0.000	111.670	282	1906	264	300	349	1988	-82
	RC	596.831	643.360	1624		1522	1726	1639		(4%)
Urb. Reg.	M&W	397.887	733.960	3538	12692	3351	3724	2446	11606	1086
	RC	1392.606	1899.202	9154		8672	9636	9160		(9%)
Urb. Irreg.	M&W	397.887	1145.170	6234	22124	5799	6669	4629	22658	-534
	RC	3183.099	2919.001	15890		14782	16998	18029		(2%)
Rural	M&W	0.000	140.336	147	1176	132	162	164	1034	142
	RC	795.775	985.040	1029		924	1135	870		(14%)
Inform.	M&W	2785.212	3358.169	16309	42176	14996	17622	5813	25782	16394
	RC	5172.536	5326.385	25867		23785	27950	19969		(64%)

En los demás estratos, el número total de edificios estimados con esta extrapolación está muy cerca del observado. En el estrato Rural, el residuo (valor estimado – valor observado) es de sólo 142 edificios, que suponen un 14% para el total de 1034. En este estrato sólo se tienen dos áreas de muestra, y además es el estrato que menor superficie ocupa en el área de estudio total (solo 1 km²). En el resto de los estratos se han obtenido buenas aproximaciones, con residuos que dan lugar a errores porcentuales entre el 2% y el 10% (indicados entre paréntesis).

Con estas premisas, se ha realizado la extrapolación del número de edificios de cada tipología obtenido en este estudio, para las huellas regulares e irregulares. La Tabla 5-35 muestra los resultados para la estimación con la media y el intervalo de confianza al 95%.

Tabla 5-35. Número de edificios de cada tipología estimado para el área total de los estratos

		Huellas regulares				Huellas irregulares			
		Por MBT	Total	Lím. Inf.	Lím. Sup.	Por MBT	Total	Lím. Inf.	Lím. Sup.
Residen.	M&W	1750	4797	1641	1860	2022	7034	1896	2149
	RC	3046		2856	3237	5012		4698	5326
Urb. Reg.	M&W	16394	28704	15530	17258	16669	36485	15790	17547
	RC	12310		11661	12959	19816		18772	20861
Urb. Irreg.	M&W	34948	60847	32511	37384	48853	74121	45447	52259
	RC	25900		24094	27705	25267		23506	27029
Rural	M&W	1084	1750	973	1196	1172	2121	1051	1292
	RC	665		597	733	949		852	1046
Inform.	M&W	33948	54912	31215	36681	32595	60032	29971	35219
	RC	20964		19276	22652	27437		25229	29646

Si se comparan los resultados de un conjunto de huellas con el otro, se observa que las huellas irregulares proporcionan valores superiores para el número de edificios, debido, principalmente, a la sobre-segmentación. En los estratos Urbanos y en el Rural, la diferencia es de poco más del 20%; pero en el estrato Residencial supera el 40%. Ya se vio que en este estrato los tejados son mayores y más complejos, y eso hace que la segmentación de la imagen genere varios polígonos en cada tejado. En cambio, en las zonas Informales, donde los edificios son más pequeños y con tejados homogéneos, la diferencia entre el número de edificios estimado con las huellas regulares e irregulares apenas supone un 9%. Estos resultados están en consonancia con el coeficiente de comparación calculado en la Tabla 5-25.

5.2.3 Estimación de tiempos y costes

A continuación, se presenta una estimación del tiempo que se emplearía en repetir el proceso diseñado para el escenario de Puerto Príncipe en una ciudad similar, al margen de todas las pruebas realizadas. En la Tabla 5-36 se han incluido dos tareas que son mutuamente excluyentes, la digitalización de huellas y el proceso OBIA; por tanto, en la suma de tiempos habrá que elegir una u otra. Para ambas se ha estimado una duración de 2 días. El tiempo se mide en jornadas laborales estándar, de 8 horas.

Tabla 5-36. Estimación de tiempos para la realización de un análisis de vulnerabilidad por teledetección (a partir del estudio en Puerto Príncipe)

Proceso	Tiempo (jornadas laborales)
1. Estratificación de la ciudad	2
2. Generación de la base de datos de exposición	
2.1. Generación de los mosaicos	1
2.2. Digitalización (huellas y tejado) [excluyente con 2.4]	2
2.3. Clasificación del LiDAR	3
2.4. OBIA en los mosaicos (huellas y tejado)	2
2.5. Obtención de los atributos relacionados con el LIDAR: altura, número de plantas, inclinación del tejado. Obtención de los atributos relacionados con la geometría de las huellas.	½
3. Obtención de la tipología y extrapolación del nº edificios	2

Se estima que dos personas con experiencia podrían repetir este proceso de evaluación de la vulnerabilidad de 6000 edificios (aproximadamente, 20 km²) en unos 11 días. Por comparación con el escenario de Lorca, plantear que la evaluación de 1000 edificios conllevaría un sexto del tiempo no es realista; pero sí lo es la mitad, que es lo que se estimó para ese escenario (5 días).

La estimación de costes no supone ninguna diferencia significativa respecto a la que se hizo para el estudio en Lorca (Tabla 5-14), que ascendía a 4500 € para analizar los 1000 edificios en 5 días.

5.2.4 Estimación del daño a partir de los datos obtenidos

Análogamente al escenario de Lorca, en este apartado se ha tratado de evaluar hasta qué punto los errores cometidos en la clasificación de tipologías constructivas pueden afectar a los resultados de daño. En Puerto Príncipe, dado que se han podido conseguir dos clasificaciones de tipologías edificatorias, una con las huellas digitalizadas (regulares) y otra con las huellas segmentadas (irregulares), se han realizado los cálculos de daño con ambas. La clasificación de tipologías con las huellas regulares realizada con el SVM (PUK) obtuvo un 79% de precisión global, y un 75% la de huellas irregulares, cuando se compararon con la base de datos de referencia de Corbane et al. (2011).

Se han preparado dos juegos de bases de datos de vulnerabilidad para realizar sendas comparaciones. El primero está compuesto por 3 datasets: (1) los datos de referencia de Corbane et al. (2011) y los edificios que estos tienen en común con (2) las huellas digitalizadas y (3) segmentadas, que son 295 y 289, respectivamente. Estos edificios se reparten en 6 áreas de muestra, con FID = {12, 16, 19, 20, 21, 23} que han sido las geounidades de esta parte del estudio. El segundo juego está compuesto por 2 datasets generados tras extrapolar la densidad de edificios de cada tipología edificatoria a toda la ciudad de Puerto Príncipe: (1) con las huellas digitalizadas y (2) con las segmentadas. Para ello, se han creado regiones por intersección de los estratos (patrones urbanos), en los que se tiene el dato de densidad de edificios de cada tipología, con los distritos, que son las unidades administrativas en las que se ha dado el resultado de daño (Figura 5-51). De la intersección, han resultado 147 regiones, que han sido las geounidades del estudio de daños a nivel de ciudad. El número de edificios obtenido tras extrapolar las huellas digitalizadas es 84915 de MAndW y 60614 de RC; con las huellas segmentadas, resultan 98316 edificios de MAndW y 77878 de RC. Las funciones de daño CM-UM y RC-CB, asignadas a estas dos tipologías, se han tomado de Molina et al. (2014), con sus respectivas subdivisiones por altura (sección 5.2.2.3).

Para el cálculo de daños, se ha simulado en SELENA un sismo en el segmento de la falla de Enriquillo analizado por Smithe y Calais (2016), que es el estudio más reciente. La Figura 5-52 muestra la localización del epicentro, a unos 20 km de la ciudad, al que se le ha asignado una Mw 7.0, de acuerdo con los autores. Los parámetros de la fuente se presentan en la Tabla 5-37. Se han utilizado los modelos de atenuación de Boore et al. (2013) y Chiou y Youngs (2014), y se han combinado en un árbol lógico con un reparto del peso de 60%-40%, respectivamente, por mantener los criterios de los estudios previos (Molina et al., 2014; Torres et al., 2016). El valor de V_s^{30} que caracteriza el tipo de suelo se ha tomado de Cox et al. (2011). Con todos estos datos, se ha calculado el espectro de demanda en cada geounidad del estudio.

Estratificación de la ciudad de Puerto Príncipe

Figura 5-51. Superposición de los estratos y los distritos de Puerto Príncipe

Tabla 5-37. Parámetros del escenario sísmico simulado en Puerto Príncipe

Latitud	Longitud	Profundidad	Mw	Orientación	Buzamiento	Mecanismo
18.5042º	-72.1184º	5 km	7.0	270º	90º	Desgarre

Epicentro del escenario sísmico de Puerto Príncipe

Figura 5-52. Escenario sísmico simulado en Puerto Príncipe

Utilizando el método I-DCM (FEMA-440, 2005), se ha calculado en SELINA el daño que el terremoto simulado causaría en los edificios contenidos en las bases de datos de vulnerabilidad. Los resultados de la comparación con la referencia de Corbane se presentan en la Tabla 5-38 y en la Figura 5-53. Para los tres datasets, los daños estimados prácticamente coinciden, con un MAPE del 4% para las huellas regulares y del 8% para las irregulares. A la luz de estos bajos errores se podría concluir que cualquiera de las estimaciones hechas en este estudio es válida como base para definir políticas de gestión del riesgo sísmico. Las distribuciones de daños muestran que la mayoría de los edificios (66%) no experimentarían daños por el terremoto simulado en este estudio. De los edificios dañados, aproximadamente la mitad tendrían daños ligeros (14% del total) y la otra mitad, sufriría daños severos o colapso (16% del total), quedando una minoría con daño moderado (4%).

Tabla 5-38. Resultados de daño en las áreas de muestra comunes con Corbane et al (2011)

	N. Edificios dañados					Total
	Nulo	Ligero	Moderado	Extensivo	Completo	
Referencia	196	40	13	21	25	295
H.Digitalizadas	194	41	15	21	25	296
H. Segmentadas	187	38	15	20	27	287

NOTA: Se han comparado 295 edificios de H. digitalizadas y 289 de H. segmentadas con 295 de la referencia. El efecto del redondeo en las probabilidades de daño suma un edificio más para la base de datos de H. digitalizadas y resta dos a las H. segmentadas.

Figura 5-53. Gráfica comparativa de los daños totales

Si se analizan los daños con más detalle por tipologías (Figura 5-54) se aprecian diferencias, como era de esperar. Las estimaciones de daño con las bases de datos generadas en este estudio son más altas para la tipología de mampostería, pero tanto en los edificios con daños como en casos de daño nulo. En la tipología de hormigón armado ocurre lo contrario. Esto se debe a que

se han clasificado más edificios de mampostería de los que hay en realidad (el índice de falso positivo es de 23% para las huellas digitalizadas y de 24% para las segmentadas). Las diferencias son más notables en el número de edificios sin daño (daño nulo), porque el número de edificios con daños se reparte entre los distintos grados y la incidencia no es apenas considerable.

Figura 5-54. Gráfica comparativa de los daños obtenidos por tipología

Cuando se realiza la comparación por alturas, no se observan diferencias significativas (Figura 5-55). Los edificios altos, con más de 4 plantas, son escasos en esta muestra y las diferencias son de 1 edificio menos (en las huellas regulares) o uno más (en las irregulares) que la referencia.

Figura 5-55. Gráfica comparativa de los daños obtenidos desagregados por altura

Por último, se ha realizado un cálculo de los daños empleando el área construida de cada tipología obtenida con las huellas irregulares. El objetivo de este cálculo es comprobar que se puede realizar un cálculo de daños cuando el número de edificios estimado no es exacto, pero sí lo es la máscara de área construida. El caso que se ha tomado aquí es el de las huellas irregulares, que presentan un error porcentual del 25% en el número de edificios, pero uno menor en el área construida (del 12%, que se corresponde con un 32% por exceso en los edificios de M&W y un 9% por defecto en los de RC). La Figura 5-56 muestra los resultados de este cálculo, donde se aprecia el mismo patrón de daños, con mayoría de daño nulo (22500 m²). Los daños se reparten en, aproximadamente, mitad de ligero (4800 m²) y mitad de extenso y completo (5400 m²); el daño moderado es mínimo (menos de 2000 m²).

Figura 5-56. Estimación de daños empleando el área construida

En la segunda parte del estudio, se han calculado daños para toda la ciudad, con las bases de datos de huellas regulares e irregulares. A continuación, se comparan estos resultados para analizar el impacto de la sobre-segmentación en una evaluación de riesgo. Los resultados agregados de daño (Figura 5-57) permiten ver que existe una diferencia homogénea. Las estimaciones de daño con huellas irregulares son, en todos los grados, aproximadamente un 21% mayores que las calculadas para huellas regulares (MAPE del 21%). Este porcentaje está en consonancia con el coeficiente de comparación del número de edificios global (1.25) obtenido en la Tabla 5-25. Lo interesante de este resultado es que muestra que ese porcentaje del 21% se mantiene aproximadamente constante para todos los grados de daño (variando entre 19%-22%), no hay sesgo hacia un grado concreto. Esto parece indicar que la diferencia en la estimación de los daños se debe únicamente a la sobreestimación del número de edificios. La sobre-segmentación no parece afectar a la asignación de tipologías ni a la evaluación del daño de manera significativa.

Figura 5-57. Gráficas comparativas de daños totales para toda la ciudad de Puerto Príncipe

Figura 5-58. Gráficas comparativas de daños por tipología para toda la ciudad de Puerto Príncipe

Para analizar estos resultados con más detalle, se ha hecho la comparación por tipologías (Figura 5-58), donde puede observarse que, en todos los casos, hay una sobreestimación de los daños con la huellas irregulares, como era de esperar. Sin embargo, esta resulta homogénea, obteniéndose un patrón de daños comparable al calculado para las huellas regulares.

Cuando se separan los resultados por alturas (Figura 5-59), se aprecian evidentes diferencias en la construcciones de más de 3 plantas, con daños en edificios de huellas irregulares que duplican a los de huellas regulares. Las construcciones altas son de hormigón armado, principalmente. Esta diferencia en el número de edificios altos puede deberse a los elementos que hay sobre los tejados, como casetas, cobertizos o techados de terrazas, que son detectados por la segmentación y tienen alturas mayores que el tejado principal, recogido en la digitalización. Este resultado es muy interesante, ya que permite localizar dónde la sobre-segmentación de la imagen tiene un impacto negativo, y se pueden tomar medidas para acotarlo.

Figura 5-59. Gráficas comparativas de daños desagregados por altura para toda la ciudad de Puerto Príncipe

En la Tabla 5-25, atendiendo a la diferenciación por estratos, se veía que el número de huellas segmentadas en las áreas Residenciales duplicaba (o casi triplicaba) el número de tejados digitalizados. En el caso del estrato Informal, ambos números eran similares, y en los demás estratos había, al menos, un 20% más de huellas irregulares que regulares. Para poder apreciar el impacto de estas diferencias, se presentan los resultados desagregados por estrato y por distrito (Figura 5-60). Se ha calculado un coeficiente de comparación para el daño, dividiendo el número de edificios con cada grado de daño estimado para las huellas irregulares por el de

las huellas regulares. En la Figura 5-60(A) se puede ver que el coeficiente de comparación para el estrato Residencial está entre 2.1 y 2.2; para los asentamientos Informales, está cerca del 1, y en los demás, cerca de 1.2. Esto concuerda con el coeficiente de comparación del número de edificios obtenido en la Tabla 5-25.

Figura 5-60. Coeficiente de comparación del daño obtenido para huellas regulares e irregulares por estrato (A) y por distrito (B).

Cuando se analizan los distritos, que es una unidad administrativa en la que usualmente se presentan los estudios de riesgo sísmico a escala local, se ve que se consigue una media del coeficiente de 1.2. Los distritos en los que se alcanzan mínimos cercanos al 1 (distritos 3, 14, 32, 35), tienen la mayor parte de sus edificios en estratos de patrón Informal (Figura 5-51), que es el que presenta un número de huellas segmentadas más cercano a las digitalizadas. Por el contrario, hay distritos con coeficientes de comparación máximos, que son mayoritariamente residenciales (23 y 29), pero en ningún caso llegan al 2. En cada distrito hay una composición distinta de edificios de varios estratos, y esto atenúa las diferencias entre ellos y baja los coeficientes de comparación.

El Residencial es el estrato donde la segmentación duplica el número de edificios, pero también es el que menos edificios aporta a la suma total, porque las áreas residenciales tienen edificios

dispersos, ubicados en parcelas con grandes extensiones de jardín. Por tanto, sería interesante poder evaluar qué porcentaje del número total de edificios supone el error por sobreestimación de edificios en este estrato. En la Tabla 5-39 se presentan los datos correspondientes a la zona Residencial que hay en los distritos 23 y 29, que son los que presentan coeficientes de comparación cerca de 2.

Tabla 5-39. Número de edificios por tipología en los distritos 23 y 29

Distrito	MandW low rise		RC low rise		RC mid rise		Error Total
	H. Digit.	H. Segment.	H. Digit.	H. Segment.	H. Digit.	H. Segment.	
23	183	309	345	785	-	-	0.9%
29	329	555	622	1412	52	202	0.5%

En el distrito 23, la diferencia entre el número de edificios segmentados y digitalizados es de 1348 edificios, que hacen sólo 0.9% del total del área de estudio (más de 145.000). En el distrito 29 es de 715 edificios, que son el 0.5%.

Finalmente, se ha hecho una representación de la distribución espacial del daño por distritos, en términos de edificios inhabitables. Como tales, se han considerado todos los que han alcanzado daño completo y el 90% de los que tienen daño extensivo, siguiendo las expresiones empíricas desarrolladas para Risk-UE (Vacareanu et al., 2004). Los resultados por distrito se presentan en la Tabla 5-40. Como se puede ver, en algunos distritos, las cifras son muy similares (D3, D14, D32, D34, D35); aunque, en general, el número de edificios inhabitables calculado para las huellas segmentadas (irregulares) supera en una media del 20% al de las huellas digitalizadas (regulares). Este resultado era esperable y concuerda con el coeficiente de comparación del número de huellas (Tabla 5-25), que era del 25%.

Tabla 5-40. Número de edificios inhabitables calculado en el escenario de Puerto Príncipe con las huellas regulares (H. Digit.) e irregulares (H. Segm.)

Distrito	Edificios inhabitables		Distrito	Edificios inhabitables	
	H. Digit.	H. Segm.		H. Digit.	H. Segm.
D1	3226.4	3701.4	D19	477.2	611
D2	1687.7	2002.3	D20	338.7	419.9
D3	1432.8	1564.7	D21	590.5	736.6
D4	21.2	27	D22	910.5	1128.1
D5	20.2	27.9	D23	128.7	240.2
D6	161.3	203.7	D24	191.5	248.8
D7	162.5	206.6	D25	181.6	228.3
D8	414.5	524.8	D26	402.2	512.3
D9	809.6	1024.1	D27	306.3	405.9
D10	799.3	1017.4	D28	365.5	532
D11	1128.1	1380.8	D29	331.1	578.5
D12	827.9	1009	D30	3324.2	3998.1
D13	465.6	580	D31	542.7	693.5
D14	861.6	941.1	D32	2054.9	2263.7
D15	298.4	371.7	D33	286.1	367.6
D16	152.4	189.5	D34	1398.4	1545.2
D17	698.5	874.4	D35	672.2	734.5
D18	1611.7	1865.6	D36	96	121.5

ESTIMACIÓN DEL NÚMERO DE EDIFICIOS INHABITABLES

(H.Digitalizadas)

(H.Segmentadas)

NOTA: en coropletas se está representando el porcentaje de edificios inhabitables respecto del total (cifra relativa); y mediante símbolos proporcionales (barras), la cifra absoluta.

Figura 5-61. Mapas de edificios inhabitables.

En la Figura 5-61 se muestran estos datos en mapas, con valores absolutos y relativos al total de edificios. El patrón espacial de los daños es prácticamente igual para ambos conjuntos de huellas, con tres agrupaciones claras. Al norte de la ciudad, donde predominan los patrones Urbanos (Regular e Irregular), se registran los menores porcentajes y también los menores valores absolutos de edificios inhabitables. Por el contrario, en la zona costera (al oeste), se han obtenido los mayores porcentajes. El número absoluto en estos distritos varía, desde valores altos en el distrito D3 (unos 1500 edificios inhabitables), que presenta patrón Informal (Figura 5-51), hasta valores bajos en los distritos D4, D5, D6 y D7 (entre 20 y 206), que son zonas industriales con menos edificios convencionales. Finalmente, al sur de la ciudad, se obtienen porcentajes que van del 16% al 25%, de nuevo con valores absolutos muy variables. Los distritos D23 y D29 corresponden a Pétionville, la zona más adinerada de Puerto Príncipe, donde se encuentran los edificios menos vulnerables; de ahí que estos distritos presenten un número bajo de edificios inhabitables. Por el contrario, los distritos de asentamientos Informales D1, D32, D34, D18 y D30 registran las mayores cifras (entre 1500 y 3300).

En conjunto, el cálculo indica que algo más de 27000 edificios podrían quedar inhabitables en la ciudad del Puerto Príncipe si ocurriera el sismo simulado, lo que supone el 19% del total.

5.2.5 Sumario y conclusiones de la aplicación en Puerto Príncipe

La aplicación en Puerto Príncipe está enfocada a diseñar un proceso que permita la generación de bases de datos de exposición y vulnerabilidad de áreas urbanas de grandes extensiones, y que pueda ser actualizado con regularidad, dado el acelerado ritmo de urbanización de la ciudad, que no siempre queda registrado oficialmente. Para ello, se cuenta con imágenes de satélite Landsat y con ortofotos aéreas y LiDAR proporcionadas por el Rochester Institute of Technology (RIT).

El objetivo es identificar los edificios en 21 áreas de muestra de la ciudad y clasificarlos según su tipología edificatoria (MBT) en dos clases: hormigón armado (RC) y mampostería y madera (MAndW). A estas clases se les asigna un modelo de vulnerabilidad, que es RC-CB y CM-UM, respectivamente, definidas en Molina et al (2014). Estas tipologías son subdivididas según su altura en bajas (*low-rise*), medias (*mid-rise*) y altas (*high-rise*), siguiendo las indicaciones de los autores. Con la clasificación de los edificios de las 21 áreas de muestra, se calcula un índice de densidad para cada tipología constructiva y se extrapola a toda la ciudad de Puerto Príncipe. Finalmente, se realiza un cálculo de daños para ver hasta qué punto las incertidumbres generadas a lo largo del proceso afectan a la estimación del riesgo. La Figura 5-62 presenta un resumen de todas las métricas de precisión.

SUMMARY OF ACCURACY METRICS OF PORT PRINCE	
Stratification	
BUILT-UP MASK	OA-F1.: 92%-92%
STRATA	OA-F1.: 68%-68%
Building footprints	
Nr. FOOTPRINTS	Comparison Coef.: 1.25
Attributes	
AREA	P.E.-P.D.: 12%-80%
Nr. STORIES	OA.: Reg.F.: 79%; Ir.F.: 72%
ROOF TYPE	OBIA: OA=95%; Prec-F1: RC roof: 99%-99%; Tin roof: 93%-95%; 2nd verificat.: OA=83%
MBT	
CONF. MATRIX	OA-F1-k: Resid.: 88%-87%-0.75; Urban Reg.:82%-81%-0.64; Urban Irreg.:81%-81%-0.62; Rural: 78%-78%-0.57; Informal: 74%-74%-0.49
2 nd VERIFICAT.	OA-F1-FP(M&W)-FP(RC): Reg.F.:79%-75%-23%-14%; Ir.F.:75%-68%-24%-29%
Damage estimation	
MAPE: Reg.F.: 4%; Ir.F.:8%	

Figura 5-62. Resumen de las métricas de calidad del procedimiento aplicado en Puerto Príncipe

1. El primer paso es estratificar la ciudad en patrones Residencial, Urbano Regular (alta densidad de población con edificios de tamaño más bien moderado, aunque variable), Urbano Irregular (con edificios más pequeños que en el regular y red viaria sinuosa), Rural, Industrial e Informal (densidad de edificación muy alta, con edificios pequeños y de muy mala calidad). Esta estratificación se ha realizado a través de un procedimiento semiautomático OBIA, que incluye una segmentación de imagen seguida de una clasificación con un modelo predictivo calculado con SVM. Se han realizado varias pruebas para evaluar el impacto del proceso de muestreo en el resultado final de la estratificación (Figura 5-30 a Figura 5-33). Se ha llegado a la conclusión de que se puede realizar un proceso OBIA para clasificar más de 1700 objetos (segmentos) con datasets balanceados de 15 muestras por clase seleccionadas de manera guiada por el usuario, obteniendo una precisión global aceptable del 68% (F1-Score también del 68%). Un dataset aleatorio proporcionaría precisiones parecidas. Para mejorarlas habría que subir el número de muestras al menos a 20 por clase. Las 21 áreas de muestra, que han sido seleccionadas aleatoriamente cubriendo toda la extensión de la ciudad de Puerto Príncipe, han quedado

repartidas en los estratos de la siguiente manera (Figura 5-36): 5 en patrón Residencial; 7 en Urbano Regular; 4 en Urbano Irregular; 2 en Rural; y 3 en Informal (el Industrial no se ha incluido en el resto del estudio, porque para ese tipo de construcciones se suelen realizar estudios exprofeso).

2. El siguiente paso es implementar la base de datos de exposición y vulnerabilidad.
 - 2.1. Se comienza generando las huellas de los edificios. Estas provienen de dos orígenes: digitalización manual sobre mosaicos generados con las ortofotos (huellas regulares) y objetos obtenidos con un proceso semiautomático de OBIA (huellas irregulares), que utiliza los mismos mosaicos, pero añade 2 bandas creadas con el LiDAR (intensidad y altura sobre el suelo) y 3 bandas de diferencias de las componentes del color. Se han digitalizado 6414 edificios, con unos tiempos de entre 5-11 segundos por edificio (Tabla 5-22). Además del contorno, se ha ido registrando el material del tejado (de hormigón y metálico). En cuanto a la clasificación OBIA, se ha utilizado el ensamblado *Random Forest* para la creación del modelo predictivo de clasificación para seis clases: tejado de hormigón, tejado metálico, pavimento, sombra y vegetación. Por tanto, el resultado también aporta el tipo de tejado, además de la geometría. En el modelo se han incluido un total de 102 atributos espectrales, de textura y geométricos.

El número de huellas obtenidas por OBIA se ha verificado por comparación con las huellas regulares (Tabla 5-25), obteniéndose coeficientes de comparación alrededor de 2 para el estrato Residencial, sobre 1.2 para los estratos Urbanos y Rural, y cerca de 1 en el Informal. A diferencia del escenario de Lorca, la mejora de la segmentación que se ha aplicado aquí con la inclusión de más bandas ha permitido obtener coeficientes de comparación menores, cuyo promedio es 1.25. Es decir, que el número de huellas irregulares supera en un 25% al de regulares.

- 2.2. Para completar la base de datos de exposición, las huellas son dotadas de atributos relacionados con su vulnerabilidad sísmica: material del tejado, número de plantas, área y otros parámetros derivados de su geometría (dirección, ancho, largo, elongación, compacidad, posición relativa en la manzana y coordenadas planimétricas del centroide). El cálculo de algunos de estos atributos conlleva incertidumbres, que se han cuantificado por medio de un proceso de verificación.

El material del tejado obtenido por OBIA se ha validado, en primer lugar, con una base de datos de verificación compuesta por segmentos etiquetados por fotointerpretación. Se ha obtenido un F1-Score del 99% para tejados de hormigón y del 95% para los metálicos (Tabla 5-27), con una precisión global del 95%. Se ha visto que la banda de la intensidad del LiDAR ha permitido distinguir entre sí los tejados de cada tipo; y que la banda de alturas es la que ha hecho posible la distinción de los tejados de hormigón y el pavimento. En segundo lugar, la clasificación de tejados por OBIA se ha comparado con los datos registrados en la digitalización, considerados como la referencia en este caso. En un 83% de los tejados, existe coincidencia en el material (precisión global).

Hay un 13% de tejados de hormigón que están erróneamente clasificados como metálicos, y el error contrario ocurre sólo en el 4% de los segmentos.

El área también se ha verificado por comparación de las huellas irregulares con las regulares (Tabla 5-26), obteniéndose un coeficiente de comparación medio de 1.12, o sea, que se está sobrestimando el área en un 12% con las huellas irregulares, lo que probablemente no tendrá un gran impacto en la asignación de tipologías. El porcentaje de detección del área construida es del 80%.

Por último, se ha verificado el número de plantas. Este se ha obtenido con un MDSn derivado del LiDAR, dividiendo la altura sobre el suelo de los tejados entre 3 m. Para la verificación se ha utilizado una base de datos tomada en campo por Corbane et al. (2011), de la que se han podido utilizar 295 edificios que coinciden con las huellas regulares, y 289 con las irregulares (Figura 5-45). Se ha visto que el 79% de los edificios digitalizados tiene una altura estimada correcta (precisión global), y el 72% de los segmentados también. El resto de los errores son, mayoritariamente, de 1 planta, por lo que no se espera que tengan una repercusión notable en la asignación de tipologías.

3. Con los atributos verificados, se procede a clasificar los edificios en hormigón armado (RC) y mampostería y madera (MAndW). Se calculan cinco modelos predictivos en cada uno de los cinco estratos con muestras de la base de datos de la edificación del Ministerio de Obras Públicas, Transportes y Comunicaciones de Haití. Los atributos incluidos son los comunes entre los de esta base de datos y la generada en este estudio: tipo de tejado, número de plantas, área y coordenadas del centroide. En todos los patrones urbanos se consiguen precisiones globales en la fase de entrenamiento que superan el 88% con, al menos, uno de los algoritmos utilizados: árbol de decisión, regresión logística, SVM (con kernels PUK y RBF) y naïve Bayes (Tabla 5-30). Se ha podido comprobar en los modelos calculados por regresión logística y con árboles de decisión, que el atributo que más influencia tiene a la hora de predecir la tipología de un edificio es el material del tejado, seguido del número de alturas. El área tiene poca influencia, al igual que las coordenadas del centroide.

Esta clasificación de edificios en tipologías ha sido sometida a dos verificaciones. En primer lugar, se ha validado la clasificación con un dataset de muestras tomado de la base de datos del Ministerio (Tabla 5-31 y Tabla 5-32). El modelo calculado con SVM (PUK) es el que mejor equilibra los errores de clasificación, y aporta métricas del 88% en precisión global y 87% en F1-Score para el patrón Residencial. Ambas métricas son del 82% en áreas Urbanas Regulares, 80% en las Irregulares, 75% en las Rurales y 74% en los asentamientos Informales. El coeficiente kappa varía entre $0.48 \leq \kappa \leq 0.75$. La segunda comprobación se ha realizado con la base de datos de Corbane (Tabla 5-33), donde se puede ver que el modelo de SVM (PUK) presenta el mejor desempeño (la precisión global es 79% en las huellas regulares y 75% en las irregulares). Se aprecia que el error por falso positivo de mampostería es relativamente alto (23% en regulares y 24% en irregulares), al igual que ocurría en el escenario de Lorca, lo que introduce a la base de datos de vulnerabilidad más

edificios de mayor vulnerabilidad de que los que realmente hay. Habrá que comprobar si esto afecta al equilibrio del patrón de daños, aumentando los daños graves.

Se ha realizado, además un análisis visual comparativo entre la clasificación de edificios conseguida para las huellas irregulares con las regulares (Figura 5-48), donde se aprecia que las diferencias aumentan en los estratos Rural e Informal. Estos son, precisamente, los estratos que obtienen las métricas más bajas en todas las clasificaciones, lo que refleja la dificultad que supone para los modelos encontrar reglas o patrones entre los datos de los edificios de las zonas Rurales e Informales, ya que, en general, no siguen normas constructivas.

Es conveniente destacar que la base de datos creada por Corbane et al. (2011) con la que se ha hecho esta comparación es una fuente totalmente independiente y ajena a este estudio. Los datos fueron tomados por un equipo diferente de personas, en circunstancias y con objetivos particulares que no están relacionados con los de esta investigación, ni con el trabajo del Dr. D. Lallemant, cuyos datos son los utilizados para crear los modelos (base de datos del MTPTC). Poder obtener una verificación aceptable con una fuente de referencia independiente es muy importante, ya que es una forma de probar la robustez del modelo de clasificación y su capacidad de generalización y escalado.

4. La clasificación de edificios obtenida en las 21 áreas de muestra se ha extrapolado al resto de la ciudad de Puerto Príncipe. Para ello, se ha llevado a cabo un remuestreo por Bootstrap y se han calculado, para cada estrato, índices de densidad de edificios de cada clase por km² (Tabla 5-35). Por intersección espacial de los estratos con los distritos de la ciudad (Figura 5-51), se han creado 147 regiones en las que se han extrapolado estos índices multiplicando por el área de cada una, según el estrato en que se encontraban. Se ha conseguido con ello una base de datos de exposición y vulnerabilidad para Puerto Príncipe con un número de edificios mayor que la aportada por el Ministerio, que presentaba falta de completitud.

El tiempo estimado en realizar esta clasificación de tipologías constructivas para asignación de la vulnerabilidad sísmica a más de 6000 edificios, es de 11 días, para dos personas expertas en la materia y en la zona. Por comparación con el escenario de Lorca, es razonable plantear que se necesitarían unos 5 días para trabajar en 1000 edificios, por el mismo coste (4500 €).

Finalmente, se ha realizado un cálculo de daños por sismo en Puerto Príncipe, para la base de datos de Corbane y para las obtenidas en este estudio con las huellas regulares e irregulares. Se ha realizado en dos partes: por un lado, en las áreas de muestra donde hay edificios comunes con Corbane y, por otro, en toda la ciudad con los datos extrapolados. Se ha simulado un sismo de Mw 7.0 en un segmento de la falla de Enriquillo, a 20 km al este de la zona de aplicación, y se han comparado los resultados de daño obtenidos.

Cuando se comparan los daños desagregados por altura de las construcciones (Figura 5-55), se aprecian diferencias en los edificios más altos, cuyos tejados de hormigón son los que resultan más sobre-segmentados. Esto permite localizar dónde la sobre-segmentación causa un mayor impacto, y así poder acortarlo. En la comparación por tipologías con Corbane (Figura 5-54) se

observa que se sobrestima el número de edificios dañados de mampostería y madera con los dos datasets de huellas, pero también el de no dañados, lo que está en consonancia con el índice de falso positivo de mampostería de la clasificación de tipologías, que es del 23% para huellas regulares y del 24% para las irregulares. Es importante señalar que los patrones de daños obtenidos son iguales y las discrepancias más notables se obtienen en el grado de daño nulo. Esto permite concluir que ese error por falso positivo de la clase MAndW no afecta significativamente al cálculo del daño. Es decir, el hecho de clasificar el 23% o 24% de los edificios de hormigón (regulares e irregulares, respectivamente) como mampostería, que es más vulnerable, no aumenta los daños graves y no introduce un sesgo en el patrón de daños. De hecho, en el cómputo de daños totales, los daños calculados para las tres bases de datos de vulnerabilidad son prácticamente iguales (Figura 5-53). El error porcentual promediado (MAPE) para los cinco grados de daño es del 4% para los edificios digitalizados, y del 8% para los segmentados.

Estos resultados parecen indicar que las incertidumbres inherentes a las distintas fases del proceso de cálculo (clasificación del tipo de tejado, obtención de las alturas de los edificios y cálculo del número de plantas, principalmente), no tienen una influencia significativa en la determinación de la tipología constructiva y permiten obtener la vulnerabilidad con una precisión global del 79% y 75% (con huellas digitalizadas y segmentadas, respectivamente). Estos datos de vulnerabilidad dan lugar a un cómputo de daños que apenas difiere del obtenido con una base de datos de exposición y vulnerabilidad generada en campo, y serían un input fiable en el que basar la definición de políticas de prevención y mitigación ante el riesgo sísmico, que es el fin último de estos estudios.

Se ha realizado también un cálculo de daños en términos de área construida, en lugar de edificios (Figura 5-56). Esta sería una buena opción en aquellos casos en los que el número de huellas segmentadas no pudiera utilizarse para aproximar el número de edificios real con la precisión requerida, pero el área construida sí fuera precisa.

En cuanto a la comparación de los cálculos de daño realizados para toda el área de Puerto Príncipe con las huellas regulares y las irregularidades, se aprecia que las irregulares arrojan resultados que están siempre por encima de las regulares, como era de esperar por la sobre-segmentación (Figura 5-58). Cuando se separan los daños en edificios con diferente altura (Figura 5-59) se puede distinguir que hay mayor discrepancia en los edificios altos, de más de 4 plantas, donde los irregulares doblan en número a los regulares. Los edificios altos son, en su inmensa mayoría, de hormigón armado, y efectivamente son los tejados más sobre-segmentados. En el cómputo total de los daños (Figura 5-57) se puede ver cómo se ha igualado esa discrepancia y en todos los grados hay, aproximadamente, un 20% más de edificios irregulares. Esto concuerda con el coeficiente de comparación del número de huellas, que era de 1.25 (Tabla 5-25). Lo reseñable de este resultado es que el patrón de daños es el mismo para las dos bases de datos de vulnerabilidad, y no parece verse afectado por la sobre-segmentación.

De esa misma Tabla 5-25 se sabe que en el estrato Residencial es donde hay más segmentos que tejados, hasta el punto de que el número de huellas irregulares dobla (o casi triplica) el de

huellas regulares. La Figura 5-60 muestra el coeficiente de comparación del daño estimado para ambos datasets, por distritos y por estratos. Se aprecia claramente que hay dos distritos (23 y 29) con un coeficiente más alto, pero que no llega a 2. Parece que la agregación de los daños por distritos, que es como generalmente se presentan los resultados de riesgo a escala local, reduce las discrepancias puntuales encontradas a lo largo del proceso de cálculo. Si se analizan los dos distritos, se ve que en ellos hay sendas regiones residenciales con un número de edificios mayor en el dataset de huellas irregulares. Pero la diferencia en ambos distritos supone solo el 0.9% y el 0.5% del número total de edificios, respectivamente.

En cuanto al cálculo de edificios inhabitables, se puede concluir que ambos conjuntos de huellas arrojan resultados comparables, con distribuciones porcentuales prácticamente iguales y valores absolutos cercanos entre sí. La distribución espacial de los daños refleja la variabilidad del tipo y densidad de edificación presente en la ciudad, con distritos que registrarían apenas unas decenas de edificios inhabitables, mientras que otros superarían los 3000. La cifra total alcanza los 27000 edificios, que supone el 19% de las edificaciones de la ciudad.

6 Análisis comparativo de costes: estudio por teledetección Vs. campaña de campo

Una vez confirmada su viabilidad técnica, la segunda parte de la hipótesis general de esta tesis está dirigida a estudiar la viabilidad económica de la creación de bases de datos de exposición y vulnerabilidad con datos tomados por sensores remotos. Se consideraría viable económicamente si su coste es menor que el de una campaña de campo destinada a recoger información sobre edificios para generar el mismo producto. Para abordar esta cuestión, se ha realizado una comparación de los costes que supone realizar un estudio de teledetección con los de una campaña de campo. Se ha planteado como objetivo la caracterización de 1000 edificios en dos ciudades distintas, una en un país desarrollado y otra en un país en desarrollo (como en esta tesis), para valorar si existen diferencias significativas.

Coste de un estudio de exposición y vulnerabilidad por teledetección

Los costes de los estudios por teledetección se han obtenido a partir de la experiencia de desarrollar el presente estudio. En las secciones 5.1.3 y 5.2.3 se tiene la estimación de costes y tiempos para cada escenario lo más realista posible y adaptada a la posibilidad de realizar un estudio de nuevo.

Coste de un estudio de exposición y vulnerabilidad realizado en campo

Para la evaluación de costes de una campaña de campo, se ha contado con la colaboración, sin ánimo de lucro, de 12 expertos nacionales e internacionales, con experiencia acreditada en inspección de edificios in-situ. Los colaboradores han sido: Gaspar-Escribano, Jorge; Kalantoni, Dimitra; Lang, Dominik; Martínez, Paco; Martínez-Cuevas, Sandra; Meslem, Abdelghani; Molina, Sergio; Pittore, Massimiliano; Pomonis, Antonios; Quirós, Ligia; Spence, Robin; Wieland, Marc.

Todos ellos han contribuido rellenando un formulario preparado exprofeso para esta investigación, o bien aportando comentarios y sugerencias a partir de su experiencia en la materia. El modelo de formulario se adjunta en el Anexo VII, y contiene una estimación a modo de ejemplo (no se corresponde con ninguna de las contribuciones en particular). Para cada escenario tipo, han aportado datos sobre los recursos (humanos y materiales), tiempos y coste (unitario, parcial y global) que supondría la evaluación de 1000 edificios in-situ. En el formulario se han separado las tareas previas a la visita de campo, la propia visita (con viaje incluido) y los trabajos de post-proceso. A continuación, se presenta una lista confeccionada a partir de las tareas que han ido considerando los colaboradores:

1. *Preliminary tasks:*
 - 1.1. *Review literature and prepare a building classification scheme (MBT identification) based upon the review.*
 - 1.2. *Explore Google Earth and Google Street View.*
 - 1.3. *Implement GIS.*
 - 1.4. *Calculate (X,Y) coordinates of buildings to inspect.*
 - 1.5. *Introduce the coordinates into the GPS memory card.*

- 1.6. *Load building contours in portable system.*
- 1.7. *Print some screenshots for in-field support.*
- 1.8. *Workshop: Training of operators and testing of devices.*
2. *In-field survey:*
 - 2.1. *Meet local experts to confirm the preliminary identification of MBT*
 - 2.2. *Organize in groups of two people*
 - 2.3. *Design optimum routes (every day)*
 - 2.4. *Visit and inspect the buildings. Fill in the forms and take (linked) photos of each building.*
 - 2.5. *Depurate and load the forms from the tablets to a common server*
 - 2.6. *Exchange experiences in order to improve the daily work*
3. *Post-processing*
 - 3.1. *Review all the forms (or a representative sample) for quality and consistency assessment*
 - 3.2. *Merge all the forms into one single spatial database*
 - 3.3. *Load and store all the photos*
 - 3.4. *Validate data (spatial queries, statistics...)*
 - 3.5. *Assess vulnerability using all the data collected in the field.*

En la Tabla 6-1 se presenta un resumen de todos los formularios recibidos para el escenario en un país desarrollado. Se observa que hay dos grupos claramente diferenciados de estimaciones, siendo el coste total promedio de uno de ellos de algo más de 43500 €, y del otro, de casi 14000 €. Las principales diferencias entre un grupo y otro son el número de personas que participan en la recogida de datos (que varía entre 2 y 12) y los honorarios de estas. Estas diferencias se originan porque algunos colaboradores han enfocado el estudio logístico como una labor profesional, y otros, como una tarea de investigación con presupuesto limitado. En lugar de calcular un promedio del coste de los dos grupos, se ha optado por mantener ambas opciones, ya que son muy realistas.

Además de en los recursos humanos, existe una gran variabilidad en cuanto a tiempos. La estimación mínima de recursos plantea que 4 personas tardarían 5 días en recoger en campo toda la información necesaria sobre los 1000 edificios; mientras que la estimación más alta considera que 12 personas necesitarían 20 días. El resto de las estimaciones están entre estas dos. Una variabilidad similar se aprecia en los tiempos estimados para el trabajo preliminar y el post-proceso. Esto refleja otro cambio de enfoque: la actividad correspondiente a la estimación mínima tiene como único fin recoger la información de la manera más optimizada posible, contando con personas expertas; mientras que la segunda, tiene una componente de formación del equipo de personas que realiza la toma de datos. En cualquier caso, todos los autores están de acuerdo en que la inspección de los edificios se debe hacer en equipos de dos personas. Alguno ha reportado tiempos de observación de 5 minutos por edificio (unos 100 edificios al día), mientras que la mayoría afirma que se necesitan al menos 15 minutos para poder hacer mediciones, tomar la fotografía y rellenar todos los campos del formulario (unos 30 edificios al día). Un apunte interesante por parte de uno de los colaboradores, que hay que tener en cuenta, es que cuantos menos equipos participen en la toma de datos, más calidad tendrá en el producto generado (porque habrá una mayor consistencia y homogeneidad de criterios). Aunque el tamaño del área de estudio condiciona la cantidad de personal, y a veces hay que poner en marcha a un gran grupo para completar el trabajo en un tiempo razonable.

Tabla 6-1. Logística de una campaña de campo para toma de datos sobre edificios en un país desarrollado. Resumen de la información proporcionada por expertos.

	PRELIMINARY WORK				IN-FIELD SURVEY								POST-PROCCESING			TOTAL COST
	Human Res.	Material Res.	Time	Cost	Human Res.	Material Res.	Time	Cost				Human Res.	Time	Cost		
								Travel	Field work	Office work	Aggregated					
1	3	3 laptops 8 tablets	14	6900	10	4 transport 10 laptops	5	5200	6800	12000	24000	5	30	10700	41600	
2	3	3 laptops 2 GPS	10	18200	4	1 transport 2 tablets	7	3260	21720	4160	29140				47340	
3	3	Unspecified	10	7000	5	Unspecified	20	2800	4750	7700	15250	2	10	6000	42375	
4	3	3 laptops 8 GPS	25	9250	12	2 transport 2 laptops 8 tablets	20	3400	17680	2260	23340	6	30	12200	44790	
*5	4	4 laptops 1 SW lic.	5	11550	9	8 tablets	10	12800	15200	940	28940	4	2	1280	41770	
													Avg. Cost		43575 €	
**6	2	1 laptop	15	2370	2	2 GPS 2 tablets	12	1340	3500	200	5040	1	30	3100	10510	
7	3	3 laptops 4 GPS	10	4050	4	1 laptop 4 tablets	5	1000	5700	800	7500	3	10	2950	14500	
8	2	2 laptops 4 tablets	5	4350	4	2 transport	10	3200	5600	0	8800	5	5	1750	14900	
**9	2	1 laptop	3	1300	2	1 transport	17	3400	8450	700	12550	2	5	1900	15750	
													Avg. Cost		13915 €	

NOTE: time unit is working day (8hrs.); Cost unit is euro; GPS refers to handheld GPS; Transport refers to rented car.

COMMENTS:

* Eight people travelling. The author proposes the use of a dedicated SW for data collection, which reduces the office work during the field campaign and the post-processing.

** Examples taken from a PhD in-field research activity. Budget was limited.

Tabla 6-2. Logística de una campaña de campo para toma de datos sobre edificios en un país en desarrollo. Resumen de la información proporcionada por expertos.

	PRELIMINARY WORK				IN-FIELD SURVEY							POST-PROCCESING			TOTAL COST
	Human Res.	Material Res.	Time	Cost	Human Res.	Material Res.	Time	Cost				Human Res.	Time	Cost	
								Travel	Field work	Office work	Aggregated				
1	2	2 laptops 4 tablets	5	4350	4	2 transport	10	3600	3800	0	7400	4	5	1750	13500
2	3	3 laptops 3 tablets	14	5650	10	4 transport 5 tablets 7 laptops	10	7500	8300	10500	26300	5	30	10700	42650
3	3	3 laptops 2 GPS	10	18200	4	1 transport 2 GPS 2 tablets	7	3260	12760	4160	20180				38380
4	3	3 laptops 6 GPS	15	5975	6	3 transport 2 laptops 6 tablets	10	3000	10920	1400	15320	3	10	2950	24245
5	3	3 laptops 8 GPS	25	9250	12	2 transport 2 laptops 8 tablets	20	6200	17680	2260	26140	6	30	12200	47590
*6	3	3 laptops 8 GPS	21	8850	8	2 transport 2 laptops 8 tablets	20	15200	21800	7100	44100	5	30	10700	63650
													Avg. Cost		38336 €

NOTE: time unit is working day (8hrs.); Cost unit is euro; GPS refers to handheld GPS; Transport refers to rented car.

COMMENTS:

* Eight people travelling.

En cuanto a los perfiles profesionales, todos los colaboradores están de acuerdo en que se necesitan ingenieros civiles, ingenieros y científicos de datos, asistentes y expertos locales. Los ingenieros senior aparecen en los formularios con menor dedicación que los demás perfiles, porque sus labores principales son de diseño de la campaña, organización y toma de decisiones. Como asistentes suelen considerar a estudiantes de grado o a doctorandos, que realizan labores de oficina y también de campo. Todos los autores cuentan con expertos locales, especialmente para los trabajos preliminares, en los que se estudian las tipologías constructivas típicas de la zona de estudio, y para la toma de datos en campo. Donde hay grandes discrepancias es en las asignaciones económicas de estos perfiles profesionales. Algunos autores proponen honorarios para ingenieros de 120€/hora, mientras que otros consideran unos 30€/hora. La asignación a los estudiantes de doctorado va desde 80€/hora a 80€/día, similar a la de los expertos locales.

En los recursos materiales existe bastante homogeneidad. Todos proponen utilizar ordenadores y dispositivos GPS manuales y/o *tablets* (para la toma de datos), además de otros materiales de oficina (impresoras, material fungible, mapas, etc.). Estos costes también están incluidos en el monto total, si bien algunos investigadores proponen añadir solo la parte proporcional correspondiente a la amortización que de esos recursos materiales se imputaría a una campaña de campo.

En cuanto a las actividades, hay propuestas particulares que merece la pena destacar. Como se ha mencionado, algunos colaboradores proponen aprovechar la actividad para dar formación a ingenieros jóvenes. En cuanto a la duración, algunos dedican significativamente más tiempo a los trabajos preliminares y proponen estudiar la zona mediante exploración de Google Street View, e incluso, llevar cargados en el GPS manual las coordenadas de los edificios que se piensa visitar. Otros, en cambio, le dan más importancia a la fase de post-proceso para realizar trabajos de depuración y controles de calidad de los datos recogidos. Otras propuestas defienden que es más conveniente realizar parte de estas comprobaciones mientras se está en la ciudad de estudio (trabajo de oficina), para poder volver a tomar datos en caso de errores o incongruencias. Finalmente, un colaborador propone la utilización de una aplicación móvil para realizar una recogida de datos estandarizada y que almacene los registros directamente en un servidor común. Esto, asegura, permitiría ahorrar la mayor parte del post-proceso.

La logística propuesta para los países en desarrollo se presenta resumida en la Tabla 6-2. El coste promedio de las contribuciones es de algo más de 38000€, pero hay dos estimaciones *outliers* en esta distribución (una a la baja, de 13500 €, y otra al alza, de 63650 €). Las principales diferencias con las estimaciones hechas para países desarrollados están relacionadas, por un lado, con los honorarios profesionales, que para los expertos locales llegan a reducirse a la mitad, y para los estudiantes, algunos autores proponen cubrir solamente sus gastos de manutención diaria. Por otro lado, la mayoría de las estimaciones proponen aumentar el número de expertos locales, porque las tipologías constructivas suelen presentar particularidades relacionadas con la cultura y la tradición de construcción local; así como aumentar el tiempo en la toma de datos, por las dificultades e imprevistos que puedan surgir. Relacionado con esto, un colaborador ha considerado incluir un asistente local en campo para que dé apoyo logístico. Esto es importante, porque además de que habla la lengua local, puede

agilizar considerablemente los procesos de solicitud de permisos, algo que puede ser muy problemático en algunos países.

Figura 6-1. Desglose de costes de una campaña de campo.

La Figura 6-1 presenta un desglose de costes para cada tipo de escenario. En el caso de un país desarrollado, se aprecian claramente los dos grupos de estimaciones con promedios de coste diferentes. En general, el trabajo de toma de datos es el que conlleva mayor gasto, tanto por el tiempo, como por el personal, como por los desplazamientos a la zona de estudio. El coste del trabajo preliminar para estudio de tipologías constructivas locales y preparación de los materiales es comparable al destinado a la fase de post-proceso, en la que se depuran los datos y se analizan para extraer información relacionada con la clasificación de tipologías y su vulnerabilidad. En la estimación económica para un escenario en desarrollo, la distribución de costes sigue un esquema similar.

Comparación de costes

Los costes que conlleva una campaña de campo para la elaboración de una base de datos de exposición y vulnerabilidad son significativamente mayores que los que supone generarla con datos tomados por sensores remotos. Estos se han estimado en 4500 €, que son el 40% de la menor de todas las contribuciones (10510 €), un 30% del promedio del grupo de contribuciones más bajas (13915 €) y sólo un 10% si se compara con el promedio más alto (43575 €). En cuanto al escenario en un país en desarrollo, el coste de realizar un estudio por teledetección sería el mismo, 4500 €, y supone el 12% de la estimación hecha por los expertos para una campaña de campo (38336€).

Además de ser considerablemente más económico, el estudio por teledetección permite un escalado exponencial sin incurrir en gastos mucho mayores. Por ejemplo, Qi et al. (2017) reportan un tiempo de 10 días, que es comprable al estimado aquí para Puerto Príncipe, pero en analizar un área veinte veces más extensa (456 km²). Por el contrario, la campaña de campo

no permite un escalado de este tipo, ya que cuantos más edificios hay que inspeccionar, más tiempo (o personal) se necesita, y las posibilidades de optimización son muy reducidas. De hecho, si se comparan los tiempos estimados en realizar cada trabajo, se observa que el estudio por teledetección sólo consumiría entre $1/3$ y un $1/15$ del tiempo que proponen los expertos en su planificación logística.

Estos resultados permiten concluir que la generación de bases de datos de exposición y vulnerabilidad sísmica por teledetección es viable y rentable económicamente, tanto en países desarrollados como en países en desarrollo.

Bloque 4

7 Sumario, discusión y conclusiones generales

El propósito general de esta tesis es comprobar si es viable, técnica y económicamente, la incorporación de las bases de datos de exposición y vulnerabilidad generadas por teledetección en los estudios de riesgo sísmico. Es de especial interés cuantificar las incertidumbres y los costes asociados. Con base en una serie de objetivos planteados para dar cumplimiento a este propósito, se ha diseñado un procedimiento metodológico en tres fases, que se ha aplicado en dos escenarios urbanos muy distintos en extensión y en complejidad: Lorca y Puerto Príncipe. En ambas ciudades, se han analizado datos tomados por sensores remotos siguiendo las fases diseñadas: (1) se han identificado los patrones urbanos; (2) se han extraído las huellas de los edificios y los atributos que conforman la base de datos de exposición; y (3) se les ha asignado un modelo de vulnerabilidad empleando técnicas de aprendizaje automático. Finalmente, se han realizado sendos cálculos de daño por sismo en escenarios simulados. Todos los pasos han sido exhaustivamente verificados. Siempre que ha sido posible, la fuente de referencia ha sido una base de datos independiente.

El enfoque en cada escenario es diferente. En Lorca, se pretende diseñar un procedimiento que sea rápido de implementar y replicable en otras ciudades de España, de cara a los más de 700 planes municipales de prevención ante el riesgo sísmico que la Directriz Básica de Planificación de Protección Civil ordena que deben hacerse (DBPC, 2004). Esto motiva el uso de datos que ya estén disponibles y den cobertura a todo el territorio nacional, además de evitar intensivas tareas de refinamiento en post-proceso o manuales. En Puerto Príncipe, interesa diseñar un procedimiento que permita abarcar grandes extensiones y que sea replicable en otras ciudades cercanas a la falla de Enriquillo, dado que se ha demostrado que aún está cargada con potencial suficiente como para generar un terremoto similar al de 2010 (Symithe y Calais, 2016). Las dos ciudades elegidas son muy diferentes y plantean retos en todas las fases del procedimiento metodológico, que se han abordado de manera distinta en función de los datos disponibles y el enfoque. Por este motivo se propone que los procedimientos que se diseñen sean lo más flexibles posible para poder adaptarse a las particularidades de cada escenario (Torres et al., 2019b). Las diferentes técnicas de análisis empleadas en cada escenario se detallan en la Figura 5-4 y la Figura 5-26.

La revisión del estado del arte (sección 2 y Anexo II) mostraba que, a pesar de haberse desarrollado numerosos estudios en este campo en los últimos 10-15 años, aún quedan preguntas por responder. Estas preguntas están relacionadas con la forma de integrar los datos tomados por sensores remotos, con el proceso de muestreo, con la flexibilidad del procedimiento metodológico, con la aplicabilidad de los resultados y con la verificación de los mismos. En la presente tesis se ha tratado de dar respuesta a estas preguntas, contribuyendo así a la construcción del conocimiento en este campo.

A continuación, se presentan las principales conclusiones extraídas del desarrollo de esta tesis (marcadas en **negrita**), en relación con los objetivos planteados.

1. *Sobre el acceso a las fuentes de datos y su integración.*

En noviembre de 2010, el Dr. David Lallemant fue el ponente principal del *workshop* organizado por *The Information Products Laboratory for Emergency Response*, parte del *Chester F. Carlson Center for Imaging Science*. Lallemant habló de cómo la publicación de datos espaciales horas después del terremoto de Haití ocurrido el 10 de enero de 2010 estaba haciendo posible la estimación de los daños y la priorización de la distribución de la ayuda:

*"With all this data there was an outpouring of data products and a huge willingness of people to give their time and effort," Lallemant says. "And because the data was open, they could do that."*²⁰

"Con todos estos datos, hubo una efusión de productos derivados y una gran disposición de la gente a aportar su tiempo y su esfuerzo", dice Lallemant. "Y pudieron hacerlo porque los datos eran abiertos".

Pero sus palabras dicen más. Expresan la transformación social que se produce cuando las personas tienen acceso a datos abiertos: "están dispuestas a dar su tiempo y su esfuerzo". En particular para la ciencia, es de vital importancia contar con datos que hagan posible la investigación y permitan avanzar en el conocimiento.

En esta tesis, todas las fuentes de datos utilizadas son abiertas y, gracias a eso, ha sido posible realizar la investigación, contrastar las hipótesis y alcanzar los objetivos planteados. Aunque parte del software empleado es propietario, se ha ido creando en paralelo un código con Python que reproduce todo el proceso de análisis. Este código ha sido publicado en abierto en Github (Arredondo, 2019) y, junto con el uso de software libre de SIG, permite la reproducción del estudio sin coste adicional en programas informáticos.

En el escenario de Lorca, se han empleado parcelas del Catastro, LiDAR y ortofotos del PNOA, así como una imagen Landsat OLI8 del repositorio *EarthExplorer*. En cuanto a la evaluación de estas fuentes, no han supuesto ningún coste económico para el proyecto, han sido fácilmente accesibles (el Catastro, a través de INSPIRE, y el PNOA no requieren ni siquiera registro), e idóneas en relación con resoluciones, formato y extensión. Esto se ha conseguido probar con éxito un nuevo uso para los datos del PNOA. Además, el CNIG ha publicado recientemente un nuevo producto, el Mapa LiDAR, creado a partir de LiDAR clasificado con alta precisión para toda España, lo que refuerza el enfoque del escenario de Lorca de contar con datos actualizados con cobertura nacional. Este dataset podría ahorrar la parte de clasificación del LiDAR, lo que reduciría considerablemente el tiempo de análisis (1 día de 5. Sección 5.1.3). La fuente de referencia fue elaborada en una tesis doctoral previa (Martínez-Cuevas, 2014) y ha sido posible reutilizarla de nuevo en esta investigación. El reaprovechamiento de los datos abre las puertas a nuevos desarrollos. Esta opinión es compartida por otros autores (e.g. Geiss y Taubenböck, 2013) y por la comunidad científica, en general. **Con objeto de contribuir a esta corriente, todos los datos generados en esta tesis serán publicados en OpenStreetMap y Mendeley,**

²⁰ <https://www.rit.edu/news/helping-haiti-through-remote-sensing-world-bank-engineer-looks-lessons-learned>

y serán reutilizados en futuros proyectos y en docencia. De ahí la importancia del segundo objetivo que propone alinear el formato de las bases de datos de exposición aquí generadas con alguna taxonomía reciente para garantizar su interoperabilidad. En consecuencia, **los atributos generados siguen las indicaciones de la taxonomía de clasificación de edificios del proyecto RIESGOS, que es la más reciente y completa.**

En el escenario de Puerto Príncipe, se han empleado ortofotos y LiDAR del RIT, una imagen Landsat TM y dos bases de datos de referencia, una oficial (del MTPTC de Haití) y otra generada en un proyecto de investigación previo (Corbane et al., 2011). De nuevo, se ha contado con datos oficiales y de investigaciones previas, todos ellos publicados en abierto o aportados para esta tesis, que no han supuesto ningún coste económico. Además, han resultado idóneos en formato, resolución y extensión. Existe una limitación con la base de datos ministerial, y es que adolece de cierta falta de completitud y su acceso abierto no está garantizado.

En ambos escenarios se ha utilizado LiDAR como fuente de datos tridimensional para resolver el problema de la altura de las edificaciones. Si se revisa el estado del arte (Anexo II), se puede comprobar que muy pocos autores han recurrido a esta fuente, siendo más habitual el uso de DEM derivados de SAR. A pesar de que las resoluciones del Radar van mejorando, los estudios que se han realizado en este campo han trabajado con resoluciones de 12 m, 5 m y 1 m. El LiDAR supera estos valores, con densidades mínimas del orden de 1 pto/m² (de hecho, el LiDAR empleado en Haití tiene más de 3 pto/m²).

Finalmente, es importante remarcar el potencial que supone para un país tener fuentes de datos institucionales abiertas. Los casos del Catastro y el PNOA de España, que recibió un premio de la ONU en 2013, o el repositorio EarthExplorer del USGS, son claros ejemplos de ello. Hay que mostrar que, lejos de suponer un perjuicio para el país, contribuyen a su desarrollo y a su posicionamiento internacional.

2. Sobre el cálculo de atributos para confeccionar las bases de datos de exposición y vulnerabilidad.

Los procesos de generación de huellas de edificios son motivo de discusión en la comunidad científica (e.g. Qi et al., 2017). Algunos autores invierten tiempo en generar segmentaciones semi-automáticas anidadas o multi-nivel (Bruzzone y Carlin, 2006; Soille, 2009; Geiss et al., 2016; 2017) que conllevan un análisis más intensivo y un refinamiento en post-proceso; otros, en cambio, prefieren directamente digitalizar las huellas, reportando mejores precisiones y tiempos de pocas horas en conseguir varios miles de edificios (Su et al., 2015; Qi et al., 2017). En esta tesis se han aplicado ambos métodos en el escenario de Haití, y segmentación en el caso de Lorca, seguida de selección de segmentos con LiDAR. **Esta innovadora combinación tiene importantes ventajas respecto al OBIA, como que se consigue tener la vegetación clasificada en distintos rangos de altura, no se pierden segmentos por ser clasificados como sombra, ya que esta clase no existe cuando se trabaja con LiDAR, y se identifican segmentos correspondientes a objetos singulares, como puentes.**

A las huellas digitalizadas se les ha llamado regulares, e irregulares a las segmentadas. La conclusión que se extrae es que cada método presenta ventajas e inconvenientes y no es posible afirmar que uno es mejor que el otro. Hay que mantener el procedimiento lo suficientemente flexible como para poder elegir el más adecuado en cada caso de estudio.

- **La segmentación es más rápida y se pueden cubrir grandes extensiones sin incurrir apenas en más tiempo de procesado. Sin embargo, se ha podido comprobar en este estudio que da mejores resultados si se realiza de forma estratificada**, es decir, con diferentes parámetros en cada patrón urbano. Esto, que además de haberse comprobado aquí, resulta muy razonable, no es aplicado en general por otros autores. El gran inconveniente de la segmentación que no se ha resuelto aún, y está probablemente lejos de resolverse, es la sobre-segmentación de objetos. **Es la principal fuente de incertidumbre en este procedimiento, ya que la segmentación es el primer paso del cual parte el resto de la aplicación. Además, se ha comprobado que los resultados obtenidos con las huellas irregulares son siempre de peor calidad que los obtenidos con las huellas regulares (sin que por ello resulten inválidas)**. El problema de sobre-segmentación se da en todos los estudios citados en el estado del arte que han recurrido a esta técnica. Algunos autores lo resuelven incorporando segmentos mayores de un nivel superior (o dos) de la segmentación anidada, pero se corre el riesgo de agrupar segmentos pequeños en objetos a los que no pertenecen. En caso de utilizar una segmentación para obtener los edificios, en esta tesis se propone encontrar un tamaño de segmentos que permita un acuerdo entre el número de objetos que se generan y su tamaño. Es decir, si el número de segmentos supera al de edificios reales (que es lo habitual), al menos que no sean demasiado pequeños. Y es que, si el área de los segmentos es demasiado pequeña, se desvirtúa el concepto de edificio y no saldrían bien clasificados de los modelos de predicción de tipología constructiva que utilizan el área como atributo. **Conseguir una segmentación que cree una máscara de área construida precisa permitiría realizar el estudio de riesgo sísmico en términos de área construida, en lugar de en número de edificios. Esto es una solución innovadora al problema de la sobre-segmentación de edificios en este campo de aplicación**, que no había sido propuesta en la comunidad científica hasta ahora.

Es conveniente recordar que **en este estudio se ha comprobado que la inclusión del dato de altura en la segmentación mejora significativamente los resultados de la clasificación de objetos final**, con precisiones globales que pasan de estar por debajo del 70% a estar por encima del 90% (precisiones finales en la Figura 5-62).

- La digitalización, en cambio, proporciona resultados muy precisos para las huellas de los edificios, pero no es escalable a grandes extensiones de terreno en tiempos razonables. **Para cubrir grandes áreas, en esta tesis se propone la digitalización estratificada de huellas en áreas de muestra y la posterior extrapolación del número de edificios a través del valor de la densidad obtenido mediante un remuestreo con bootstrap**. En el escenario de Haití se han alcanzado resultados satisfactorios sobre la base de datos de referencia del ministerio, con errores

porcentuales en la estimación del número de edificios menores del 10% en los estratos que se pueden considerar más completos (el Residencial y los dos Urbanos). Un registro de tiempos de digitalización ha permitido demostrar que se pueden digitalizar entre 5 y 13 edificios por minuto (incluyendo el etiquetado del material del tejado), en línea con lo reportado por Qi et al. (2017).

El número de huellas generado por segmentación ha sido verificado por comparación con la base de datos de referencia en Lorca, y por comparación con las huellas digitalizadas, en Puerto Príncipe. **La segmentación de las zonas del Centro en Lorca no ha resultado muy satisfactoria por la complejidad en las formas de los tejados.** No se ha dedicado más tiempo a su refinamiento, ya que se cuenta con huellas catastrales para todo el territorio nacional y no es necesario generarlas automáticamente. En Puerto Príncipe, en cambio, donde no hay una base de datos de huellas de edificios oficial y abierta, la segmentación ha adquirido un papel más relevante al considerarse como única fuente probable en caso de replicar este estudio en otras ciudades del país. **En consecuencia, se ha obtenido un número de huellas más equiparable con el número de edificios real, al que supera en una media del 25% (Figura 5-62). Los edificios más sobre-segmentados son los de las zonas Residenciales. Este dato es relevante en sí mismo, pues permite tomar la decisión de descartar la segmentación en esas zonas y dirigir hacia ellas los trabajos de digitalización, que resultarían muy reducidos. O bien, establecer una regresión o un factor de conversión (de 1/1.25) para estimar el número de edificios a través del número de huellas segmentadas.**

En este paso de generación de huellas, así como en los siguientes, **se ha podido comprobar la eficacia de trabajar de manera estratificada en patrones urbanos, a pesar de no ser lo habitual.** Esto, que se ha mencionado para el caso de la segmentación, también se aplica a otros procesos como la clasificación del LiDAR o el cálculo de los modelos predictivos de tipologías. En Lorca se ha trabajado en zonas de muestra que están en 3 de los 5 patrones identificados en la ciudad: Ensanche, con edificios multifamiliares, principalmente; y dos zonas del centro histórico (Centro A, con edificios de mayor tamaño, y Centro B, con edificios más pequeños). En Puerto Príncipe, se ha trabajado en 5 patrones: Residencial, Urbano Regular (con calles más bien rectas y viviendas de tamaño medio), Urbano Irregular (con calles sinuosas y viviendas más pequeñas), Rural e Informal (infraviviendas). Los edificios, la red viaria y la vegetación tienen características distintas en cada estrato, y precisamente esas diferencias ayudan a mejorar los resultados de los procesos de análisis. En otras palabras, los parámetros máximos y mínimos (relativos a áreas, alturas, orientaciones, distancias, etc.) de los objetos (edificios, árboles, calles, etc.) son particulares en cada patrón urbano y es más acertado establecerlos por separado. Además, trabajar de manera estratificada permite extraer mejores conclusiones e identificar y localizar problemas, que pueden darse en un patrón urbano, pero en otro no.

Una de las preguntas que se han abordado en esta tesis tiene que ver con la forma de realizar el proceso de muestreo para las clasificaciones. La pregunta surge por las diferentes aproximaciones que se pueden encontrar en la literatura, donde la mayoría de los autores toman las muestras de manera aleatoria (e.g Geiss et al., 2016; Riedel et al., 2014; Mück et al., 2013) y muy pocos los hacen de forma manual (e.g Wieland et al, 2012a). Generalmente hacen

numerosas pruebas de entrenamiento de modelos con números de muestras que van desde varias decenas a cientos (Matsuka et al., 2012; Costanzo et al., 2016), pero rara vez optan por modelos entrenados con pocas muestras. En este estudio, se han probado ocho datasets con diferente configuración en cuanto al número y tipo de muestras, al proceso de selección de estas y al equilibrio entre clases para clasificar los patrones urbanos en la estratificación de Puerto Príncipe (sección 5.2.2.1). **El resultado permite concluir que se obtienen mejores precisiones con datasets equilibrados y con muestras del mismo tipo en las fases de entrenamiento y verificación. Esto plantea un reto cuando se entrenan modelos con una fuente de datos y después se verifican con otra, algo que se ha abordado también en esta tesis obteniendo resultados robustos (clasificación de MBT en Puerto Príncipe). Del experimento, se vio también que los resultados no varían mucho cuando las muestras se eligen aleatoria o manualmente, lo que abre la puerta al muestreo manual, siempre que se consiga recoger en la muestra toda la variabilidad presente en la población.** La preferencia por datasets equilibrados también se ha comprobado aun cuando la distribución real de las clases es muy dispar, como ocurre en las tipologías constructivas de Puerto Príncipe (sección 5.2.2.3).

El cálculo de los atributos es el paso siguiente a la creación de las huellas de los edificios. Un principio importante para esta tesis ha sido la separación del concepto de exposición y el de vulnerabilidad. Por ello, la base de datos de exposición es un resultado con valor por sí solo (Figura 5-10 y Figura 5-41). Hay autores que clasifican las tipologías constructivas directamente desde el análisis espectral de la imagen (Borfecchia et al., 2010; Geiss et al., 2015) empleando centenares de atributos. Esto da lugar a modelos que no son comprensibles y a resultados de clasificación de edificios en MBT cuyos atributos más relevantes han pasado a un segundo plano.

Los mayoría de los atributos incluidos en las bases de datos de esta tesis han sido adaptados a la taxonomía del proyecto RIESGOS. Algunos son sencillos de calcular, como los referentes a la posición y forma de las huellas (área, coordenadas planimétricas del centroide, grado de compacidad); otros han necesitado algún cálculo adicional (orientación, elongación, posición relativa en la manzana); y otros han requerido un proceso elaborado que desencadena, a su vez, otros procesos de análisis. Se trata de la altura, la inclinación del tejado (para obtener su forma) y los materiales de este. La altura se deriva del dato tridimensional, que es el LiDAR en esta tesis. El LiDAR ha de ser clasificado previamente, si no viene clasificado de origen. En este estudio se ha empleado para ese fin el algoritmo TOD que ha sido validado en un contexto internacional, quedando segundo en el ranking (una descripción más completa puede encontrarse en Torres et al., 2019a). La mediana de las alturas de los puntos LiDAR clasificados como edificio que caen dentro de cada huella ha sido el estadístico utilizado para aproximar la altura de los edificios. Para la pendiente, se ha utilizado la media de las pendientes de los triángulos del TIN creado dentro de cada huella con estos mismos puntos. En cuanto al material de las cubiertas, en Haití se ha obtenido por OBIA tras la segmentación y, en el caso de las huellas digitalizadas, etiquetándolo manualmente (metálico y de hormigón). **Una innovación destacable introducida aquí ha sido la inclusión de una banda ráster creada a partir de la intensidad del LiDAR para la clasificación OBIA de los tejados de Puerto Príncipe**

(además de la de alturas, que ya se viene utilizando). Con este dato, agregado a la composición de bandas RGB+SWIR+altura, se han podido distinguir los tejados por su material (hormigón y metal) en el 83% de las más de 6000 instancias. En Lorca, donde no se ha realizado un análisis OBIA ni se ha digitalizado, ha sido necesario calcular un modelo predictivo para obtener el material del tejado en función de los otros atributos (ordenados por su poder discriminante: número de plantas, coordenadas del centroide, área y pendiente del tejado).

En todos los casos posibles, los atributos se han verificado exhaustivamente. Las precisiones pueden consultarse en la Figura 5-20 y Figura 5-62. El resultado es una base de datos de exposición en cada escenario que recoge las características (atributos) de cada edificio, a nivel individual, y que pueden emplearse para analizar su vulnerabilidad sísmica, como en este caso, o como base para trabajar en su vulnerabilidad ante otros peligros naturales.

3. Sobre la clasificación de tipologías constructivas (MBT)

En Lorca se han identificado dos tipologías predominantes, de acuerdo con los autores de la base de datos de referencia (Martínez-Cuevas, 2014; Martínez-Cuevas et al., 2017): edificios de mampostería sin reforzar (M) y edificios de hormigón armado (RC). En Puerto Príncipe, también se han agrupado en dos, con base en estudios previos (Molina et al., 2014; Torres et al., 2016) y en línea con otros autores (Hancilar et al., 2013): edificios de mampostería y madera (MAndW) y edificios de hormigón armado (RC). Cada una de estas tipologías, ha sido subdividida en función del número de plantas.

La clasificación de los edificios en tipologías constructivas es el paso previo a la asignación de un modelo de vulnerabilidad a través de modelado predictivo usando técnicas de aprendizaje automático. De todos los atributos calculados aquí, sólo se han empleado los de mayor poder discriminante para la clasificación de MBT. Esta decisión está fundamentada en la premisa de crear modelos de predicción que sean sencillos y comprensibles, de manera que se puedan extraer conclusiones sobre qué atributos son los que más influyen en la clasificación y sopesar el coste-beneficio de obtenerlos. Así, **se ha podido comprobar que la clasificación de las MBT en Lorca está basada en los atributos altura y localización; mientras que en el escenario de Puerto Príncipe, son la altura y el material del tejado los atributos fundamentales.** La obtención de estos atributos no ha sido directa en este estudio, pero se pueden plantear a futuro nuevas vías más rápidas de obtenerlos. Por ejemplo, para la altura, el LiDAR ya se podrá descargar clasificado del CNIG para los estudios en España; y para el tejado, podría probarse si un OBIA es más rápido y eficaz, como ha ocurrido en Haití, que un modelo predictivo, como en Lorca.

Para la zona Centro de Lorca, se obtiene una precisión global (OA) en la clasificación de MBT del 82% con un modelo generado por regresión logística (Figura 5-20). **En Puerto Príncipe, los modelos creados con SVM y kernel PUK han permitido alcanzar una OA del 75% para huellas irregulares y del 79% para las regulares (Figura 5-62). En ambos escenarios, los errores provienen, principalmente, de falsos positivos de mampostería, que introducen más edificios de esta tipología en la base de datos de los que realmente hay. Los**

modelos aquí generados alcanzan precisiones del mismo orden, e incluso superiores, que las obtenidas por otros autores en estudios similares (ver tabla comparativa del Anexo II); **y presentan la ventaja añadida de utilizar un número muy reducido de atributos (2 en el caso de Lorca y 3-5 en el de Puerto Príncipe).**

En cuanto a las técnicas de aprendizaje automático utilizadas para crear los modelos predictivos, se han comparado diferentes algoritmos para tratar de responder a la pregunta habitual de cuál es el mejor clasificador. El hecho de utilizar relativamente pocos atributos ha permitido probar técnicas como la regresión logística o las redes Bayesianas (incluso naïve Bayes), que son más sencillas de calcular y, sobre todo, de interpretar, que los SVM. En la mayoría de los casos, el desempeño de las técnicas sencillas iguala, o incluso, supera al de las más sofisticadas. **La regresión logística ha resultado ser uno de los clasificadores más potentes**, además de ser muy transparente y de fácil interpretación. **Se ha comprobado que los SVM son muy sensibles al sobre-ajuste, pues los modelos predictivos para MBT creados con estos algoritmos son los que más precisión pierden en las fases de verificación en relación a las de entrenamiento.** Se han probado tres kernels diferentes para crear los modelos con SVM: polinómico, RBF y PUK. Este último, que ha sido introducido hace relativamente pocos años (Üstün et al., 2006), no había sido probado en este campo hasta ahora, lo que supone una innovación aportada en este estudio. De hecho, **el SVM creado con kernel PUK es el clasificador que mejores precisiones proporciona en casi la totalidad de los casos.**

4. Sobre la asignación de vulnerabilidad y cálculo de daños por sismo

A cada tipología constructiva identificada se le ha asignado el modelo de vulnerabilidad sísmica que mejor puede describir su comportamiento ante una sacudida sísmica. En Lorca, a la tipología M se le ha asignado el modelo M3, y a la tipología RC, el modelo RC1, ambos de Lagomarsino y Giovinazzi (2006). En Puerto Príncipe, se han tomado los modelos de vulnerabilidad del trabajo previo de Molina et al. (2014), asignando el modelo CM-UM a la tipología MAndW, y el RC-CB a la tipología RC. En cada modelo de vulnerabilidad se han considerado los distintos rangos de alturas, subdividiendo los edificios de cada tipología según el número de plantas. En total, han resultado 5 MBT para Lorca (2 de M y 3 de RC) y 4 MBT para Puerto Príncipe (2 de MAndW y 2 de RC).

En cada escenario, se ha simulado un sismo y se han estimado los daños utilizando el método analítico I-DCM (secciones 5.1.4 y 5.2.4). En Lorca, el evento simulado es de $M_w = 5.2$ y su epicentro se ha ubicado a unos 4 km al norte del área de estudio. En Puerto Príncipe, es de $M_w = 7.0$ y se ha localizado a unos 20 km al este de la ciudad. En los cálculos, se ha mantenido todo constante (parámetros del sismo, suelo, modelo de atenuación, modelos de vulnerabilidad, definiciones de grados de daño y metodología) y solo se ha variado el input de vulnerabilidad: (1) base de datos de referencia; (2) bases de datos estimadas en este estudio con huellas regulares y con irregulares, siempre que ha sido posible. Esto permite hacer una comparación precisa del impacto, únicamente, de las incertidumbres de la evaluación de la vulnerabilidad en el cálculo del daño. **En el escenario de Lorca, la comparación de daños totales arroja un**

MAPE del 4% para las huellas regulares de las zonas del Centro (Figura 5-20); en Puerto Príncipe, es del 4%, cuando se comparan las huellas regulares con la referencia, y del 8% en el caso de las irregulares (Figura 5-62). En un desglose detallado de los daños por tipologías y por alturas, se aprecia que las mayores diferencias se dan en el daño nulo y ligero. Estos grados de daños son los que menor influencia tienen en el cálculo de los indicadores más críticos de riesgo sísmico.

Estos valores bajos de error parecen demostrar que las incertidumbres encadenadas en los procesos previos de generación de huellas, cálculo de atributos y clasificación de tipologías tienen un impacto mínimo en la estimación de daños por sismo. Son resultados altamente satisfactorios, que permiten responder positivamente a la hipótesis de investigación sobre la viabilidad técnica del uso de las bases de datos creadas por teledetección en estudios de riesgo.

Si se analizan las tipologías por separado, se puede observar que los errores por falso positivo de mampostería se reflejan en un aumento del número de este tipo de edificios, como era de esperar. Sin embargo, hay más edificios de mampostería tanto dañados como no dañados, es decir, no se alteran los patrones de daños en ninguno de los escenarios. En cuanto a la separación de daños por altura de las edificaciones, no se aprecian diferencias reseñables.

Con las huellas irregulares de Puerto Príncipe, que presentaban un 25% de error porcentual en la estimación del número de edificios, pero un error menor (del 12%) en la estimación del área construida, se ha realizado un cálculo de daños en términos de área. Es la alternativa que se presenta para la estimación del riesgo sísmico en casos como este, o más extremos, en los que el número de huellas obtenido por procesos semi-automáticos no permita aproximar el número de edificios de forma precisa, pero sí el área construida.

Con la extrapolación de la distribución de edificios por tipologías a toda la ciudad de Puerto Príncipe, **se ha podido realizar otro cálculo de daños asociado al escenario sísmico. Las estimaciones de edificios inhabitables obtenidas con las huellas regulares e irregulares son comparables, con distribuciones espaciales prácticamente iguales. Se ha obtenido que algo de más de 27000 edificios quedarían inhabitables, lo que supone un 19% del total en la ciudad de Puerto Príncipe.**

5. Sobre la evaluación de costes

En la sección 6 se ha llevado a cabo un análisis comparativo del coste que conlleva la realización de una campaña de campo para recabar información sobre los edificios de cara a generar una base de datos de exposición y vulnerabilidad, y el coste que supone crearla por teledetección. Este último se ha derivado de la realización del presente estudio. El coste de una campaña de campo ha sido obtenido a partir de información aportada por 12 expertos internacionales en la materia. **De la comparación se deduce que el análisis de datos tomados por sensores remotos permite obtener un producto comparable al que se obtendría en campo, pero por un 10%-30% de su precio y dedicando, como máximo, un tercio del tiempo, lo que**

avala la viabilidad económica de estos estudios. Esto aplica tanto para ciudades de países en desarrollo como desarrollados.

La conclusión directa de esta comparación sería proponer la teledetección como la vía más adecuada para el análisis de vulnerabilidad en una ciudad, en lugar de las tradicionales campañas de campo. Pero la realidad es que lo óptimo (y lo que se propone en esta tesis) es una combinación de ambas, en la que colaborase un equipo interdisciplinar de expertos en Ingeniería civil/Arquitectura e Ingeniería/Ciencia de datos, como se reclama en la comunidad científica (e.g. Wieland, 2013; Geiss y Taubenböck, 2013; Taubenböck et al., 2009). Para ello, se propone un proceso de 5 fases generales:

1. Estudio cualitativo de la ciudad por medio del análisis de imágenes de satélite, aéreas o terrestres. Para esto hay diversas fuentes libres: Landsat, ASTER-GDEM, PNOA, Sentinel, Google Earth, Google Street View...
2. Estratificación de la ciudad en patrones urbanos homogéneos atendiendo al tamaño y la disposición de sus edificaciones, a la regularidad de la red viaria, a la presencia de vegetación, edad aproximada de las construcciones, etc.
3. Generación de huellas de los edificios. Para ello, se pueden consultar fuentes oficiales donde ya estén generadas (catastro, censo, instituto de estadística, etc.) u otras, como OSM. Si no las hay, se procedería a digitalizarlas o a segmentar las imágenes de alta resolución. Para esto es probable que sea necesario adquirir datos, tanto imágenes ópticas como información 3D (LiDAR o Radar).

En este paso se toma la decisión de si es necesario realizar una campaña de campo para toma de muestras.

4. En caso afirmativo, se diseñaría la campaña a partir de toda la información obtenida en los pasos previos. Con la ciudad estratificada y con las huellas generadas, se localizan aquellas áreas de la ciudad donde es indispensable ir a campo a recoger la información de los edificios. Esto vendría condicionado por la dificultad de identificar una tipología concreta, o de aplicar las técnicas de análisis en algún estrato o área específica dentro de un determinado estrato. Esto se ha visto en las zonas del Centro de Lorca y en el estrato Residencial de Haití, donde los casos de sobre-segmentación no permitían obtener huellas precisas.

Además, se podría realizar una toma de muestras en otras partes de la ciudad para poder entrenar nuevos modelos predictivos, en caso necesario, o para realizar un control de calidad de la base de datos que se genere finalmente.

En cualquier caso, la campaña de campo no estaría destinada a recoger información de todos los edificios (o de todas las tipologías), sino de una parte concreta, determinada de antemano. Esto permite abaratar los costes considerablemente y optimizar la planificación. Esta es la diferencia principal con los procedimientos actuales, según se desprende del estado de arte. Hoy en día, se va a campo a traer una muestra representativa de todas las tipologías. Con el enfoque que se propone aquí, sólo se iría a campo a tomar unos datos concretos. Un ejemplo puede ilustrar mejor esto: los edificios multifamiliares de hormigón armado, exentos, de grandes áreas y varias alturas, están presentes en muchas ciudades del mundo. Estos edificios son muy sencillos de identificar automáticamente en las imágenes y

sus atributos se calculan fácilmente. No sería necesario recoger ningún dato sobre esta tipología en campo. Si los expertos estiman que se pueden inspeccionar unos 4 edificios por hora, teniendo en cuenta la cantidad de edificios de este tipo que puede haber en una ciudad, se estarían ahorrando muchas horas de trabajo en campo y muchos desplazamientos, al no tener que visitar las zonas donde estos edificios predominan.

5. Cálculo de atributos dentro de las huellas y aplicación de los modelos predictivos para la clasificación de tipologías. Finalmente, asignación de vulnerabilidad a cada tipología.

El producto derivado de este procedimiento sería una base de datos de exposición y vulnerabilidad de la ciudad, de alta precisión, creada por combinación de análisis de teledetección y campaña de campo dirigida, que habría permitido optimizar costes y tiempo.

8 Summary, discussion and main conclusions

The main goal of this thesis is to design and apply a flexible procedure based on Earth Observation techniques to create exposure and seismic vulnerability databases, as well as to evaluate the technical and economic feasibility of their integration within seismic risk studies. The procedure has been applied in two cities that present significant differences regarding extension and complexity: Lorca (Spain) and Port Prince (Haiti). In both cities, remote sensing data have been analyzed following three major phases: (1) stratification of the city into homogeneous urban patterns; (2) building footprint extraction and attribute calculation to implement the exposure database; and (3) seismic vulnerability allocation using machine learning techniques. Finally, a damage scenario has been simulated in each city. All the steps have been exhaustively validated with reference databases.

The approach is adapted to each city. In Lorca, the procedure is meant to be fast to deploy and replicable in other Spanish cities, given that the General Direction of Civil Protection and Emergencies of Spain establishes that a list of more than 700 cities must develop earthquake risk studies at local scale in the years to come (DBPC, 2004). According to this, the data that is already available with national coverage are preferred. In this case, time-consuming post-processing for fine-tuning of partial results is avoided. In Port Prince, the procedure is expected to be applicable in large areas and replicable in other cities along the Enriquillo Fault-System, since it has been proved that the accumulated strain is still capable of generating another event similar to the 2010 Haiti earthquake (Smythe y Calais, 2016). Each scenario poses challenges in all the phases of the procedure, which have been addressed differently based on the specific objectives and data availability. For this reason, these procedures should be flexible and able to adapt to the particular features of different cities (Torres et al, 2019b). Figures 5-4 and 5-26 show the different techniques applied to each scenario.

The state-of-the-art (Section 2 and Appendix II) reveals that, despite the numerous studies that have been conducted in this application domain in the past 10-15 years, there are still several open questions. These questions are related to data fusion strategies, the sampling process, the flexibility of the procedure, or the applicability and validation of results. These questions have been addressed here with the aim to expand knowledge in this field.

In the following, the main conclusions drawn from this study are presented (in bold), linked to the objectives (Section 3).

1. On data access and data fusion strategies

In November 2010, Dr. David Lallemand was the keynote speaker at the workshop hosted by The Information Products Laboratory for Emergency Response, part of the Chester F. Carlson Center for Imaging Science. Dr. Lallemand explained how all the spatial data that were made public right after the 2010 Haiti earthquake enabled damage estimation and prioritization of the aid distribution.

*“With all this data there was an outpouring of data products and a huge willingness of people to give their time and effort,” Lallemand says. “And because the data was open, they could do that.”*²¹

But his words have a deeper meaning. They express the social transformation that occurs when people have access to open data: *“They are willing to give time and effort”*. Especially in science, open data are crucial for research and knowledge expansion.

All the data used in this thesis came from open sources. This has enabled to conduct the research, contrast the hypothesis and reach the pursued objectives. Despite proprietary software has been used, a code has been written in Python that reproduces the whole methodological procedure. This code has been made public through Github (Arredondo, 2019) with guaranteed open access. Along with the use of open-source GIS software, the code now allows the reproduction of the analysis with no additional costs of software licensing.

In Lorca, Cadastral parcels, orthophotos and LiDAR from PNOA, as well as one Landsat OLI8 image from the *EarthExplorer* repository have been used. Access to these sources was free and easy (Cadaster and PNOA do not even require a login), and the spatial resolutions, format, and extensions of the data were suitable for the study. The PNOA program was originally conceived as a source for continuous DTM and DSM generation to be applied in hydrology, urban change detection, rural land-use characterization, etc. This research has proved a new use of the PNOA data for earthquake risk analysis, an application domain that had not been even envisioned so far. In addition, the CNIG has recently published a new open product called LiDAR Map, which has been created from high quality classified LiDAR covering the whole Spanish territory. This supports the purpose of having updated data with national coverage for the application in Lorca and other Spanish cities. The already classified LiDAR Map dataset could save one working day in the analysis workflow (Section 5.1.3). The reference data for this study were created in a previous PhD. Thesis (Martínez-Cuevas, 2014) and it has been possible to reuse them in this research. Data recycle allows for new developments. This opinion is shared by other authors (e.g. Geiss and Taubenböck, 2013) and the general scientific community. **With the aim of contributing to this trend, all the data generated within this thesis will be made public in OpenStreetMap and Mendeley; and will be reused in research and teaching.** Data sharing motivated the second objective of this dissertation, according to which the exposure databases here implemented should be in compliance with any global building taxonomy in order to guarantee interoperability. Consequently, **the attributes fulfill the recommendations of the exposure model that is being designed within the RIESGOS project, as it is the most recent and comprehensive building taxonomy to date.**

For the Haitian scenario, orthophotos and LiDAR from RIT, along with one Landsat TM image have been integrated. In addition, two reference databases have been used: an official cadastral and census database (from the MTPTC of Haiti) and another building dataset compiled within a previous cooperative project (Corbane et al., 2011). All these data are either open data or have been contributed to this study; hence, data acquisition has not involved any additional cost.

²¹ <https://www.rit.edu/news/helping-haiti-through-remote-sensing-world-bank-engineer-looks-lessons-learned>

Besides, formats, spatial resolution, and extension were appropriate for this study, although the official cadastral database presents two limitations: lack of completeness and non-guaranteed open access.

In both scenarios, LiDAR has been incorporated as 3D data source to extract the building height. According to the state-of-the-art review (Appendix II), very few studies have resorted to this data source; instead the use of DEM derived from SAR is widespread. Despite Radar spatial resolutions are improving over time, the studies conducted in this field have worked with GSD of 12 m, 5 m, and 1 m. LiDAR provides better resolutions, with minimum point densities of 1 pts/m² (in fact, the LiDAR dataset for Haiti has a point density higher than 3 pts/m²).

Finally, it is worth remarking that it is important for a country to provide official open data. Exemplary cases are the Spanish Cadaster and PNOA (which received an important award from the United Nations in 2013), or the EarthExplorer repository. Far from being disadvantageous, sharing spatial data contributes to the country development and international positioning.

2. On attribute calculation to compile exposure and vulnerability databases.

Processes for building footprint extraction are a matter of discussion in the scientific community (e.g. Qi et al., 2017). Some authors invest time in computing semi-automatic multi-level image segmentations (Bruzzone y Carlin, 2006; Soille, 2009; Geiss et al., 2016; 2017) that entail an intensive analysis and a post-processing for refinement of results. In contrast, other authors prefer to manually digitize building contours, claiming to reach better accuracies and reporting digitization times of a few hours to obtain several thousands of footprints (Su et al., 2015; Qi et al., 2017). In this research, both approaches have been implemented in Port Prince, and only image segmentation has been applied in Lorca followed by a selection of segments with classified LiDAR. **This innovative combination presents advantages with respect to the OBIA: segments corresponding to vegetated objects are directly classified into height ranges (high, medium-sized, and low vegetation); segments corresponding to shadows are not discarded, since LiDAR points are not affected by shadows; other objects of interest are identified, such as bridges.**

The digitized footprints have been referred to as regular footprints, whereas the segmented ones have been called irregular footprints. The conclusion extracted from the application of both approaches is that each one presents advantages and drawbacks and it is not possible to choose one over the other. The procedure must be flexible enough to allow using the most appropriate in each particular scenario.

- **Semi-automated image segmentation is faster and able to cover large areas without substantially increasing the processing time. A stratified segmentation conducted separately in each urban pattern has proved to yield better results than one single segmentation of the whole study area.** The major weakness of automated image segmentation, which is still far from resolved, is object over-segmentation. **It is the main source of uncertainty in this procedure since it is the first step of the whole application. In fact, it has been observed that the results**

obtained with the irregular footprints are always poorer (while still usable) than those provided by the regular ones.

Over-segmentation is a common problem to all the studies analyzed in the state-of-the-art that have used this technique. Some authors try to solve it by incorporating segments from one or two higher levels of a multi-level segmentation. However, this entails the risk of merging segments to objects they do not belong to.

In case of using image segmentation to extract building footprints, the number of segments identified as buildings will likely be higher than the actual number of buildings. Nevertheless, **the built-up area estimation could be precise. In such a case, the risk assessment can be performed in terms of built-up area instead of number of buildings. This is a novel proposal to overcome the problem of building over-segmentation in this application domain.** To this end, it is essential to reach an agreement between the number of objects and their size, in order for the segment area values to be as realistic as possible.

It is worth recalling that, **when performing OBIA, the incorporation of elevation data in the segmentation process has yielded better classification results in this study**, with global accuracies increasing from below 70% to over 80% (figure 5-62).

- Footprint digitization provides an accurate estimation of the number of buildings. However, it is not scalable to large areas in a reasonable and affordable time. **To cover a wide zone in Port Prince, a stratified extrapolation of the building density computed in smaller, manageable sample areas has been performed in this study. The percentage error obtained in Residential and Urban areas with the reference database of the MTPTC is lower than 10%. The recorded digitization times showed that 3 to 15 building footprints can be digitized within one minute, including the labelling of roof material.** This is in line with the times reported by Qi et al. (2017).

The number of footprints obtained from image segmentation has been compared with the reference dataset in Lorca, and with the digitized footprints in Port Prince, for validation. **The segmentation of the zones Center A and B of Lorca has not been very satisfactory due to the complexity of the roof shapes.** They have not been fine-tuned since cadastral footprints are available for the whole national territory. Thus, it is not necessary to generate them automatically in case this study is replicated. On the contrary, in Port Prince, where there is no official footprint database available and open, the image segmentation played a more important role as it might be the only means for footprint generation in similar cities. **Consequently, the estimated number of buildings is close to the actual one, with an overestimation of 25% (Figure 5-62). The more noticeable over-segmentation occurs in Residential areas. This information is relevant for support decision making. For example, it can be decided to substitute the segmentation for manual digitization in Residential areas, where the scarcity of buildings could make this process faster and easier. Or a conversion factor (e.g. 1/1.25) could be applied to infer the actual number of buildings from the segmented footprints.**

When extracting the building footprints, as well as in other subsequent steps for LiDAR classification and predictive modelling, **it has been proved that following a stratified**

strategy is significantly more efficient than analyzing the entire city as a whole; despite it is not very common in this application domain. In Lorca, the analysis has been conducted in 3 sample areas that are located in 3 out of the 5 urban patterns identified in the city: Sprawl, mainly with multi-family buildings; and 2 areas of the historic center (Center A, with larger buildings; and Center B, with smaller dwellings). In Port Prince, sample areas have been selected in all 5 urban patterns: Residential, Urban Regular (with rather straight streets and medium-sized buildings), Urban Irregular (with narrow, winding streets and smaller dwellings), Rural, and Informal settlements. The buildings, street network, and vegetation present different features in each stratum, and precisely these differences help to improve the outcomes of the analysis. More specifically, the maximum and minimum values of areas, heights, orientations, distances, etc., of the objects (buildings, trees, streets, etc.) are distinctive of each urban pattern and it is more accurate to establish them separately. Furthermore, working in a stratified frame allows to extract better conclusions and to identify and locate problems that may appear in one stratum, but not in others.

A question addressed in this thesis is related to the sampling process for classification purposes. The question arises from the different approaches found in the literature, where most of the authors conduct random sampling (e.g. Geiss et al., 2016; Riedel et al., 2014; Mück et al., 2013), while a minority opt for a manual selection of samples (e.g. Wieland et al., 2012a). Authors usually compare the performance of several training datasets of different sizes (e.g. Matsuka et al., 2012; Costanzo et al., 2016), but small training sets are rarely preferred. In this study, eight datasets have been compared for the stratification of the study area in Port Prince (section 5.2.2.1). The datasets have different configurations regarding the number and type of samples, the approach followed to select them, and the balance of sample distributions. **The result shows that balanced distributions outperform unbalanced training sets, especially when working with samples of the same type in the training and testing phases. This poses a challenge when models are trained with a certain database but tested with an independent reference dataset of a different nature. This has also been addressed in this thesis, obtaining robust results (MBT classification in Port Prince). From the analysis, it has also been observed that the results did not vary significantly when using random or manually-created training datasets, which supports the preference for manual sampling, provided that the resulting sample is representative of the whole population.**

The attribute calculation is the following step once the building footprints have been created. An important principle of this dissertation has been the clear separation of the two concepts *exposure* and *vulnerability*. Consequently, the exposure database is a valuable result itself (figures 5-10 and 5-41). Other authors classify the building typologies directly from the image spectral analysis (Borfecchia et al., 2010; Geiss et al., 2015) using hundreds of attributes. This yields predictive models that are hardly understandable and building classifications in which the building main attributes are unnoticed.

Most of the attributes included in the exposure databases here compiled have been adapted to the taxonomy of the project RIESGOS. The calculation of attributes related to the location and geometric shape was straightforward (floor area, (X, Y) coordinates of the centroid, degree of

compactness); others involved additional calculations (orientation, elongation, position within the block); and others required a complete process (height, roof shape and roof material). The building height has been derived from the LiDAR, which was previously classified with the algorithm TOD. This algorithm has been validated within an international context, eventually ranking second (further details can be found in Torres et al., 2009a). The median of the height of all the points classified as building that fall inside one footprint has been chosen as the estimator of the building height. To compute the nr. of stories, the building height has been divided by the inter-story height. With the same LiDAR points, a TIN has been created inside each footprint to derive the building shape (flat or pitched) from its triangles' average slope.

The roof material in Port Prince (concrete slab or metal sheet) has been extracted from the OBIA carried out to generate the irregular footprints; while for the regular footprints, the roof was labeled as it was digitized. **An innovation introduced in this thesis is the integration of the LiDAR intensity for the OBIA (in addition to the spectral bands and the elevation data that are commonly used). LiDAR intensity helped separating the concrete roofs from the metal roofs in 83% of more than 6000 instances.** In Lorca, a predictive model has been created to infer the roof material using other attributes (ordered by decreasing discriminant power: nr. of stories, coordinates of the centroid, area, and roof slope).

The attributes have been exhaustively verified and the accuracy measures are shown in figures 5-20 and 5-62. As a result, an exposure database has been compiled in each scenario with attributes for all individual buildings. This database can be used to allocate the seismic vulnerability (as in this study) or as a basis to further analyze the vulnerability of the buildings to other natural hazards.

3. On the building typology (MBT) classification

In Lorca, two predominant typologies have been identified according to the authors of the reference database (Martínez-Cuevas, 2014; Martínez-Cuevas et al., 2017): unreinforced masonry buildings (M) and reinforced concrete buildings (RC). In Port Prince, the building stock has been grouped also into two MBT based on previous studies (Molina et al., 2014; Torres et al., 2016) and in line with other authors (Hancilar et al., 2013): masonry-and-wood buildings (MAndW) and reinforced concrete buildings (RC). Each typology has been further divided into sub-classes of different height range.

The building classification into MBT is a previous step to the vulnerability allocation through predictive modelling using machine learning techniques. Out of all the attributes here computed, only those with higher discriminant power have been used to learn the models. This decision is based on the purpose of keeping the models as simple and understandable as possible, in order to gain insight into the most important attributes for MBT separability. This would allow to establish a trade-off between the minimum number of attributes to be included in the models that provides an acceptable accuracy and the cost of achieving them. In this regard, **it has been proved that the MBT classification in Lorca is based mainly in the building height and location; whereas in Port Prince, the most relevant attributes are the building height and the roof material.** The calculation of these attributes is not a

straightforward process, but new, faster methods can be applied in future studies. For instance, the new LiDAR Map can be used directly to obtain the building height in other Spanish cities. And for the roof material, it would be worth evaluating whether an OBIA (as applied in Port Prince) is faster than predictive modelling.

For the zone Center of Lorca, the overall accuracy (OA) of the MBT classification with logistic regression is 82% (figure 5-20). In Port Prince, the models created with SVM and kernel PUK yield an OA of 75% for irregular (segmented) footprints and 79% for regular (digitized) footprints (figure 5-62). In both scenarios, the main error is a false positive of masonry buildings. The models created in this thesis provide comparable or even higher accuracies for MBT classification than those obtained in similar studies (reference is in the comparative table of similar studies in Appendix II); while using a rather reduced number of attributes (2 in Lorca and 3-5 in Port Prince).

With regard to the machine learning techniques used to create the predictive models, several algorithms have been tested to find the best classifier. The small number of attributes considered to learn the models has allowed using techniques such as logistic regression or Bayesian networks, which are simpler and easier to interpret than the commonly used SVM. In most of the cases, these simpler techniques have shown a comparable or even better performance than the most sophisticated ones. **In fact, logistic regression has proved to be one of the most powerful classifiers, while being easy to interpret. It has been observed that SVM are very susceptible to overfitting, as the accuracy of the models created with SVM in the testing phase was significantly lower than the training accuracy in most of the cases.** Three different kernels have been used to learn the models with SVM: linear, RBF and PUK. The latter has recently been introduced (Üstün et al., 2006) and had not been tested in this application domain yet. **The innovative use of a PUK kernel has actually provided the best accuracies of most of the models created with SVM in this study.**

4. On the vulnerability allocation and damage estimation

Each MBT has been assigned with the vulnerability model that better describes its performance under an earthquake shaking. In Lorca, the typology M has been assigned with the model M3, and the typology RC with the model RC1, both taken from Lagomarsino and Giovinazzi (2006). For Port Prince, the vulnerability models were defined in the study of Molina et al. (2014). The model CM-UM has been allocated to the typology MAndW and the model RC-CB to the typology RC. For each vulnerability model, different height ranges have been considered. A total of 5 MBT are present in Lorca (2 of M and 3 of RC) and 4 MBT in Port Prince (2 of MAndW and 2 of RC).

In each city, a damage scenario has been calculated using the analytical method I-DCM, after simulating a probable earthquake (Sections 5.1.4 and 5.2.4). In Lorca, the seism has a $M_w = 5.2$ and the epicenter is located 4 km north of the town. In Port Prince, a $M_w = 7.0$ event located 20 km east of the city has been simulated. With the aim of evaluating the impact of the MBT classification uncertainties in the results of damage, the other inputs remained constant: earthquake parameters, type of soil, ground motion prediction equations, vulnerability models, damage grades definition and methodology. Only the input of vulnerability has been changed

to calculate the damage for (1) the reference database and (2) the exposure database created in this study. **The comparison of both results yields a MAPE = 4% for the zone Center of Lorca (figure 5-20). In Port Prince, also a MAPE = 4% is obtained for the regular footprints and MAPE = 8% for the irregular footprints (figure 5-62). A more detailed comparison of damage disaggregated by typology and by height range shows that the differences mainly correspond to the damage grades none and slight. These damage grades have minimal influence on the estimation of crucial risk indicators.**

These low values of MAPE seem to prove that the uncertainties associated to the previous processes for footprint extraction, attribute calculation, and MBT classification have a negligible impact in the estimation of damage. These results are highly satisfactory as they provide a positive answer to the research hypothesis on the technical viability of using exposure and vulnerability databases created by means of EO techniques in seismic risk studies.

When analyzing the damage disaggregated by typology, it can be seen that the false positive errors of masonry buildings cause an increment in the number of buildings of this type in both scenarios, as expected. However, the increment is observed either in damaged or non-damaged masonry buildings, that is, the damage pattern is not biased. As for the damage disaggregation by height, no remarkable differences are observed.

The irregular footprints in Port Prince present a percentage error of 25% in the estimation of number of buildings, but a lower error (12%) in the built-up area estimation. Hence, the damage has also been calculated in terms of built-up area. Instead of the common damage calculation based on number of buildings, this is a suitable alternative in those cases where the number of buildings is not reliable, but the built-up mask is accurate.

After the extrapolation of the MBT distribution, **another damage scenario has been calculated in the whole city of Port Prince. The number of uninhabitable buildings estimated for both, regular and irregular footprints is comparable, with very similar spatial distributions. Overall, the simulated earthquake would leave about 27000 buildings uninhabitable in Port Prince (19% of the total building stock).**

5. On the cost evaluation

Section 6 presents a comparative analysis of the cost of an in-field survey to collect exposure and vulnerability data and the cost of creating the same data by means of Earth Observation techniques. The latter has been derived from this study, while the former has been provided by 12 international experts in this field from their own experience. **From the comparison, it can be deduced that the use of remote sensing allows for creating a high-quality exposure and vulnerability database at 10%-30% the cost of an in-field campaign in one third of time (as much). This seems to confirm the economic feasibility of these studies.**

The direct conclusion from this comparison would be to propose remote sensing as the most efficient means to analyze the earthquake vulnerability of a city, instead of the traditional in-field surveys. However, the optimum solution (and the proposal made in this thesis) is a

combination of both, in which an inter-disciplinary team of experts in Civil Engineering/Architecture and Data Engineering/Science might collaborate, as claimed by the scientific community (e.g. Wieland, 2013; Geiss y Taubenböck, 2013; Taubenböck et al., 2009). To this end, a general 5-stage procedure is proposed:

1. Preliminary analysis of the city under study from a qualitative perspective using aerial, terrestrial or satellite images. Several open sources are available to this end: Landsat, ASTER-GDEM, PNOA, Sentinel, Google Earth, Google Street View...
2. Stratification of the city into homogeneous urban patterns considering the building sizes, arrangement of urban blocks, street network regularity, presence of vegetation, approximate age of the buildings, etc.
3. Building footprint generation. Official sources can be explored to access the footprints (cadaster, census, statistical institutes, etc.) or other sources such as OSM. In case there are no footprints available, it should be necessary to digitize or segment the building contours using satellite/aerial images and LiDAR/Radar data.

At this point, a decision must be made on whether to carry out the in-field survey.

4. In case an in-field survey is needed, the first step is to design it using all the information collected in the previous stages. It is possible to identify and delimit the areas to be visited based on the city stratification and the footprints already generated. Areas susceptible to be visited are those in which remote sensing is not able to provide the target information, such as urban patterns where the segmentation does not yield satisfactory results (as in the city center of Lorca) or the LiDAR classification is problematic (as in the Informal stratum of Port Prince). Furthermore, other parts of the city could be visited for sampling purposes.

In any case, the field campaign would not be oriented to collect data of all the buildings or all the typologies but focused only on a certain amount of information which is determined beforehand. This allows to optimize the logistics and reduce the cost. This is the main difference with respect to the current procedures, since nowadays the in-field visit is commonly devoted to collect data of all the building types present in the study area.

An example can be illustrative: Multi-family reinforced concrete buildings, which are generally tall and isolated, are present in many cities around the world. These buildings are easily identifiable with remote sensing techniques; hence no field inspection is needed to collect their attributes. According to experts' estimation, it takes one hour to inspect four buildings. If all multi-family buildings can be excluded from an in-field survey, the result is a substantial reduction of time and cost.

5. Attribute calculation inside the footprints and predictive analysis for MBT classification. Finally, a vulnerability model is allocated to each MBT.

The final product is a high-quality exposure and vulnerability database of the city under study created by combination of remote sensing and a directed in-field survey that allows for optimization of resources and cost.

9 Líneas futuras

Al finalizar esta tesis quedan abiertas algunas cuestiones que surgen de la motivación por querer avanzar en los estudios realizados, pero también de la reflexión sobre algunos aspectos.

En cuanto a la continuidad de los estudios, sería interesante poder extender la estimación de vulnerabilidad del escenario de Lorca al resto de la ciudad, como se ha hecho en Haití. De esa manera, surge la oportunidad de analizar otros patrones urbanos, como el Rural, y otras MBT, sobre todo de mampostería, que son muy frecuentes en España y otros países del Mediterráneo. En el caso de Haití, sería conveniente validar la extrapolación de la distribución de MBT a toda la ciudad con una base de datos de edificación que esté completa, o al menos, que haya garantías de que lo está en algunas zonas de control. También sería interesante contrastar la estimación de daños por área construida y terminar el cálculo de riesgo sísmico con más indicadores.

Hay dos reflexiones importantes que surgen del desarrollo de esta investigación, que además están conectadas. Una tiene que ver con la colaboración entre los profesionales que elaboran las bases de datos de edificación y los que las utilizan en materia de gestión de riesgos naturales. Esta es una cuestión abierta, a la que se han referido ya otros autores como Dell'Acqua et al. (2013), Wieland (2013); Geiss y Taubenböck, (2013). En muchas ocasiones, aun teniendo acceso a bases de datos municipales o estatales de catastro o censo, en algunos países no sirve de mucho, pues la información allí contenida es muy limitada o está agregada, o simplemente no es útil para estimar la vulnerabilidad sísmica de un edificio. Dado que esa información se utiliza, y va a seguir utilizándose cada vez más, en este tipo de estudios, es conveniente que haya un intercambio de intereses entre ambas partes que den lugar a ideas y sugerencias que redirijan la forma de hacer las cosas hasta el momento. Un ejemplo claro se encuentra en la base de datos de la edificación del MTPTC de Haití, utilizada en este estudio, y que se recopiló entre los años 2010-2011. Es una de las bases de datos catastrales de más reciente creación en el mundo. Su concepción y la dirección de la toma de datos corrió a cargo del Dr. D. Lallemand, ingeniero civil experto en riesgo sísmico. Dado que el proyecto se llevó a cabo después del gran terremoto de Haití de 2010, y tratándose de una zona sísmica en la que se van a seguir produciendo terremotos en el futuro, Lallemand se propuso como objetivo que estos datos se utilizaran en estudios de riesgo. De ahí que la base de datos tenga tantos campos relacionados con el desempeño de un edificio ante una sacudida sísmica (Anexo I), que permiten la identificación de tipologías y su vulnerabilidad de manera directa y precisa (piso blando, columna corta, pendiente de la calle, efecto golpeteo, etc.).

En este sentido, los miembros del equipo de investigación en el que se ha desarrollado esta tesis ya han iniciado contactos con la Dirección General del Catastro de España y con el portal inmobiliario Idealista.com, para firmar sendos acuerdos de colaboración que permitan realizar intercambios de datos que sean de interés para ambas partes (geometría de los edificios, fecha de construcción, coste de reposición, etc.). Se espera con ello poder agilizar en el futuro muchos de los procesos aquí presentados para generación de huellas, cálculo de atributos y clasificación de MBT. Se prevé que el dato de coste o valor de mercado de los edificios sea un atributo con gran poder discriminante a la hora de identificar su tipología constructiva.

La otra reflexión importante está relacionada con la generación de huellas de los edificios, que sean precisas y útiles para obtener un número lo más exacto posible de edificios. Es el inicio de todo el proceso y condiciona la calidad de los resultados finales, como se ha podido comprobar en esta tesis por la comparación de las huellas regulares e irregulares. La comunidad científica lleva décadas haciendo segmentaciones para identificar edificios individuales, probando técnicas como la segmentación anidada, alineamientos con la calle, reclasificación con segmentos vecinos, fusión de imagen y LiDAR, etc. Aún no se han conseguido resultados aceptables, que no necesiten algún ajuste con edición manual de las geometrías, o con regresiones o factores de corrección. Incluso si se consigue un único polígono por edificio, es muy probable que no tenga ni la misma área, ni la misma forma, por lo que los atributos derivados de estas para cumplir con las taxonomías no serían correctos. Así como la obtención de la altura de los edificios por medio del análisis de sombras está siendo reemplazada por el uso, cada vez más extendido, de datos 3D, como Radar o LiDAR, la segmentación será sustituida por otra técnica. Parece claro que la digitalización manual no va a ser la solución, pues también tiene sus limitaciones en cuanto a su capacidad de escalado.

Lo que parece más conveniente para la obtención de huellas precisas es la apertura de bases de datos de edificios donde esas huellas ya están generadas. Hoy día, además de en el Catastro, existen multitud de fuentes donde se puede encontrar esta información en formato digital: bases cartográficas y topográficas a escalas grandes, ayuntamientos, inspecciones técnicas de edificios (ITEs), portales como OSM, proyectos de BIM, planos de estudios de arquitectura e ingeniería o colegios profesionales, etc. Y es que cuando se generan las huellas de edificios en una ciudad para un nuevo estudio de riesgo sísmico (u otro propósito), lo que se está haciendo es duplicar el trabajo, ya que, lo más probable, es que esas huellas ya hayan sido generadas en algún momento (al menos, en el momento de su construcción), y estén almacenadas en algún sitio. Hay que tener en cuenta que pueden existir restricciones por aspectos relacionados con la privacidad de los datos, que tendrían que ir resolviéndose. Pero al menos la información relacionada con edificios públicos y patrimoniales (no sólo la huella, sino también los atributos indicativos de su vulnerabilidad) podría estar ya puesta a disposición del público mediante políticas de acceso reglado.

Es más, está habiendo iniciativas mundiales que promueven y financian la generación de bases de datos de exposición de cara a realizar estudios de gestión de riesgos, cuando esas bases de datos ya están generadas. Solo hay que ponerlas a disposición de los analistas de riesgos. Por tanto, no solo es necesario el avance técnico para mejorar los algoritmos de segmentación y clasificación, sino que, en paralelo, es necesario que se establezcan políticas de distribución de este tipo de datos. La directiva INSPIRE, que promueve la compartición de datos espaciales, podría ser un buen mecanismo regulador en este sentido.

Bloque 5

10 Referencias

1. Alpaydin, E. (2009). Introduction to machine learning. MIT press.
2. Álvarez-Cabal, R., Díaz-Pavón Cuaresma, E., Rodríguez-Escribano, R. (2013). El terremoto de Lorca. Efectos en los edificios. INTEMAC, Consorcio de Compensación de Seguros, Madrid, 233 pp.
3. Arranz, J.J. (2013). Diseño, optimización y análisis de sistemas basados en técnicas láser, para el modelado geométrico, registro y documentación, aplicados a entidades de interés patrimonial, Tesis doctoral, E.T.S.I. Topografía, Geodesia y Cartografía (UPM). <http://oa.upm.es/15056/>. Último acceso 16/05/2019
4. Arredondo, A. (2019). GitHub repositorio, Haiti OBIA. <https://github.com/arredond/haitiobia>
5. ATC-40 (1996). "ATC-40 – Seismic Evaluation and Retrofit of Concrete Buildings," Vol. 1, Report No. SSC 96-01, Applied Technology Council, Redwood City, CA.
6. ASCE/SEI 31-03 (2003). Handbook for seismic evaluation of existing buildings. American Society of Civil Engineers.
7. Barbat, A. H., Eeri, M., Moya, F. Y., Canas, J. A. (1996). Damage Scenarios Simulation for Seismic Risk. Earthquake spectra, 12(3).
8. Bhattacharjee, A., Lossio, R. (2011). Evaluation of OCHA response to the Haiti earthquake. Final Report, January, 2011.
9. Banco Mundial - PIB2017 (2017). https://datos.bancomundial.org/indicador/NY.GDP.PCAP.PP.CD?most_recent_value_desc=false&view=chart. Último acceso 10/05/2019
10. Banco Mundial – TF99644 (2017). Implementation completion and results report (TF-99644) on a grant from the Haiti reconstruction fund in the amount of US\$65 million to the Republic of Haiti for a Port-au-Prince neighborhood housing reconstruction project (PREKAD). Report No: ICR00004162. October 31, 2017. Online: <http://documents.worldbank.org/curated/en/338511510321829472/pdf/Implementation-Completion-and-Results-Report-Port-au-Prince-Neighborhood-Housing-Reconstruction-P125805-11012017.pdf>. Último acceso: 18/06/2019.
11. Benedetti, D., Petrini, V. (1984). Sulla vulnerabilità sismica di edifici in muratura: un metodo di valutazione. A method for evaluating the seismic vulnerability of masonry buildings. L'industria delle Costruzioni, 149: 66-74.
12. Ben-Gal, I. (2007). Bayesian networks. in Ruggeri F., Faltin F. & Kenett R., Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons.
13. Benito, B., Capote, R., Murphy, P., Gaspar-Escribano, J.M., Martínez-Díaz, J.M., Tsige, M., Stich, D., García-Mayordomo, J., García-Rodríguez, M.J., Jiménez-Peña, M.E., Insua-Arévalo, J.M., Álvarez-Gómez, J.A., Canora, C. (2007). An Overview of the Damaging and Low Magnitude Mw 4.8 La Peca Earthquake on 29 January 2005: Context, Seismotectonics, and Seismic Risk Implications for Southeast Spain. Bull. Seismol. Soc. Am., 97: 671-690
14. Bilham, R. (2009). The seismic future of cities. Bulletin of Earthquake Engineering. 7(4): 839.
15. Bilham, R. (2010). Lessons from the Haiti earthquake. Nature, 463: 878-879.
16. Blaschke, T. (2010). Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 65(1): 2-16.

17. Boore D.M., Stewart J.P., Seyhan E. Atkinson G.M. (2013). NGA-West 2 Equations for predicting response spectral accelerations for shallow crustal earthquakes, PEER Rpt. 2013/05, Pacific Earthquake Engineering Research Center, Berkeley, California, 135 pp
18. Borfecchia, F., De Cecco, L., Pollino, M., La Porta, L., Lugari, A., Martini, S., L Porta, L., Ristorante, E., Pascale, C. (2010). Active and passive remote sensing for supporting the evaluation of the urban seismic vulnerability. *Italian Journal of Remote Sensing*, 42(3): 129-141.
19. Boser, B., Guyon, I., Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152.
20. Breiman, L. (2001). Random Forests. *Machine Learning*, 45: 5–32.
21. Bruzzone, L., Carlin, L. (2006). A multilevel context-based system for classification of very high spatial resolution images. *IEEE Trans Geosci Remote Sens* 44(9): 2587–2600
22. Brumana, R., Della Torre, S., Oreni, D., Cantini, L., Previtali, M., Barazzetti, L., Banfi, F. (2018). SCAN to HBIM-Post Earthquake Preservation: Informative Model as Sentinel at the Crossroads of Present, Past, and Future. In *Euro-Mediterranean Conference* (pp. 39-51). Springer, Cham.
23. Brzev, S., Scawthorn, C., Charleson, A.W., Allen, L., Greene, M., Jaiswal, K., Silva, V., (2013), GEM Building Taxonomy Version 2.0, GEM Technical Report 2013-02 V1.0.0, 188 pp., GEM Foundation, Pavia, Italy, DOI:10.13117/GEM.EXP-MOD.TR2013.02.
24. Cabañas, L., Carreño, E., Izquierdo, A., Martínez-Solares, J.M., Capote, R., Martínez-Díaz, J.J., Benito, B., Gaspar-Escribano, J.M., Rivas-Medina, A., García-Mayordomo, J., Pérez, R., Rodríguez-Pascua, J.M., Murphy, P. (2011). Informe del sismo de Lorca del 11 de mayo de 2011. Informe Técnico. [http://www.ign.es/web/resources/sismologia/www/lorca//Lorcainfo 2011.pdf](http://www.ign.es/web/resources/sismologia/www/lorca//Lorcainfo%202011.pdf). Último acceso 07/06/2019
25. Calais, E., Freed, A., Mattioli, G., Amelung, F., Jónsson, S., Jansma, P., Hong, S., Dixon, T., Prépetit, C., Momplaisir, R. (2010). Transpressional rupture of an unmapped fault during the 2010 Haiti earthquake, *Nature Geosciences*, 3: 794-799.
26. Calvi, G.M. (1999). A Displacement-Based Approach for Vulnerability Evaluation of Classes of Buildings, *Journal of Earthquake Engineering*, 3(3): 411-438.
27. Calvi, G. M., Pinho, R., Magenes, G., Bommer, J. J., Restrepo-Vélez, L. F., Crowley, H. (2006). Development of seismic vulnerability assessment methodologies over the past 30 years. *ISSET Journal of Earthquake Technology*, 43(3): 75-104.
28. Campbell, K. W., Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*, 30(3): 1087-1115.
29. Cerri, M. (2018). Multivariable flood damage modelling for residential buildings using open data with random forests. Technical University of Munich, Department of Civil, Geo and Environmental Engineering Master thesis in Environmental Engineering.
30. Chaudhuri, D., Kushwaha, N. K., Samal, A., Agarwal, R. C. (2015). Automatic building detection from high-resolution satellite images based on morphology and internal gray variance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5): 1767-1779.

31. Chiou, B. S. J., Youngs, R. R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 30(3): 1117-1153.
32. Coburn A.W., Spence, R. (2002). *Earthquake Protection*, Second Edition. Ed. John Wiley & Sons, Ltd. ISBN: 0-471-49614-6
33. Cooper, G. Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9: 309-347.
34. Corbane, C., Saito, K., Dell'Oro, L., Bjorgo, E., Gill, E., Piard, B.E., Huyck, C.K., Kemper, T., Lemoine, G., Spence, R.J.S., Shankar, R., Senegas, O., Ghesquiere, F., Lallemand, D., Evans, G.B., Gartley, R.A., Toro, J., Ghosh S., Svekla, W.D., Adams, B.J., Eguchi, R.T. (2011). A comprehensive analysis of building damage in the 12 January 2010 MW7 Haiti earthquake using high-resolution satellite and aerial imagery. *Photogrammetric Engineering & Remote Sensing*, 77(10): 997-1009.
35. Costanzo, A., Montuori, A., Silva, J., Silvestri, M., Musacchio, M., Doumaz, F., Stramondo, S., Buongiorno, M. (2016). The combined use of airborne remote sensing techniques within a GIS environment for the seismic vulnerability assessment of urban areas: an operational application. *Remote Sensing*, 8(2): 146.
36. Cox, B., Bachhuber, J., Rathje, E., Wood, C., Dulberg, R., Kottke, A., Green, R., Olson S. (2011). Shear Wave Velocity- and Geology-Based Seismic Microzonation of Port-au-Prince, Haiti. *Earthquake Spectra*, Vol. 27, No. S1, pag. S67-S92
37. CRED/UNISDR (2016). *Poverty and death: Disaster mortality 1996-2015*.
38. Crowley, H., Pinho, R., Bommer, J.J. (2004). A Probabilistic Displacement-Based Vulnerability Assessment Procedure for Earthquake Loss Estimation, *Bulletin of Earthquake Engineering*, 2 (2): 173-219.
39. Data-Landsat: <https://www.usgs.gov/land-resources/nli/landsat>. Último acceso: 09/06/2019.
40. Data-PNOA: <http://pnoa.ign.es/>. Último acceso 09/06/2019.
41. DBPC (2004). Directriz Básica de Planificación de Protección Civil ante el Riesgo Sísmico. Aprobación en el «BOE» núm. 124, de 25 de mayo de 1995, pp 15294-15304. Ref.: BOE-A-1995-12364. Modificación en el «BOE» núm. 238, de 2 de octubre de 2004, pp 33205-33208. Ref.: BOE-A-2004-17005.
42. Dell'Acqua F, Gamba P, Jaiswal K (2013) Spatial aspects of building and population exposure data and their implicaitons for global earthquake exposure modeling. *Nat Hazards*, 68(3):1291-1309.
43. DesRoches, R., Comerio, M., Eberhard, M., Mooney, W., Rix, G.J. (2011). Overview of the 2010 Haiti earthquake. *Earthquake Spectra*, 27(S1): S1-S21.
44. Dey, V., Zhang, Y., Zhong, M. (2010). A Review on Image Segmentation Techniques with Remote Sensing Perspective. In *ISPRS TC VII Symposium – 100 Years ISPRS*, edited by W. Wagner and B. Székely. Vienna: Institute of Photogrammetry and Remote Sensing, Vienna University of Technology.
45. Dolce, M., Masi, A., Marino, M., Vona, M. (2003). Earthquake Damage Scenarios of the Building Stock of Potenza (Southern Italy) Including Site Effects, *Bulletin of Earthquake Engineering*, 1 (1): 115-140.

Referencias

46. Dunbar, P.K., Bilham, R.G., Laituri, M.J. (2003). Earthquake loss estimation for India based on macroeconomic indicators. *Risk Science and Sustainability*, NATO Science 112: 163–180.
47. Duputel, Z., Rivera, L., Kanamori, H., Hayes, G. P., Hirshorn, B., Weinstein, S. (2011). Real-time W phase inversion during the 2011 off the Pacific coast of Tohoku Earthquake. *Earth, planets and space*, 63(7): 5.
48. Duputel, Z., Rivera, L., Kanamori, H., Hayes, G. (2012). W phase source inversion for moderate to large earthquakes (1990–2010). *Geophysical Journal International*, 189(2), 1125–1147. <https://doi.org/10.1111/j.1365-246X.2012.05419.x>
49. Eberhard, M.O., Balbridge, S., Marshall, J., Mooney, W., Glenn J.R. (2010). The Mw 7.0 Haiti Earthquake of January 12, 2010: USGS/EERI Advanced Reconnaissance Team Report. USGS Internal Report.
50. Ehrlich, D., Kemper, T., Blaes, X., Soille, P. (2013). Extracting building stock information from optical satellite imagery for mapping earthquake exposure and its vulnerability. *Natural hazards*, 68(1): 79-95.
51. Esch, T., Thiel, M., Bock, M., Roth, A., Dech, S. (2008). Improvement of image segmentation accuracy based on multiscale optimization procedure. *IEEE Geoscience and Remote Sensing Letters*, 5(3): 463-467.
52. Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters* 27 (8): 861–874.DOI:10.1016/j.patrec.2005.10.010.
53. Felzenszwalb, P., Huttenlocher, D. (2004). Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision* 59 (2): 167–181.DOI:10.1023/B:VISI.0000022288.19776.77.
54. FEMA-154 (1988). Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, FEMA 154, Federal Emergency Management Agency, Washington D.C., USA.
55. FEMA-154 (2002). Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook, FEMA 154, Second Edition, C. Rojahn, Principal Investigator, C. Scawthorn, co-Principal Investigator, prepared by the Applied Technology Council, Redwood City, CA, (ATC-21), for the Federal Emergency Management Agency, Washington, D.C., USA.
56. FEMA-178 (1992). NEHRP Handbook for the Seismic Evaluation of Existing Buildings, FEMA 178, Federal Emergency Management Agency, Washington, D.C., USA.
57. FEMA-440 (2005). Improvement of nonlinear static seismic analysis procedures. FEMA 440, Federal Emergency Management Agency, Washington D.C.
58. Feriche, M., Vidal, F., Alguacil, G. Navarro, M. (2012). Daños causados en construcciones de Lorca por la serie sísmica de 2011. Homenaje a la profesora Maria Dolores Romacho Romero. Editorial Universidad de Almeria, pp. 99-117.
59. Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. 4th Edition. SAGE publications. ISBN: 978-1-4462-4917-8. ISBN: 978-1-4462-4918-5 (pbk).
60. Fierro, E., Perry, C. (2010). Preliminary Reconnaissance Report – 12 January 2010 Haiti Earthquake. BFP Engineers. 12 pp.
61. Fleiss, J.L., Levin, B., Paik, M.C. (1981). *Statistical methods for rates and proportions*. Chapter 18. The measurement of interrater agreement. ISBN 0-471-52629-0.

62. Foody, G.M. (2004). Thematic map comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5): 627-633.
63. Freeman, S. A. (1998). Development and use of capacity spectrum method, *Proceedings of the 6th U.S. National Conference on Earthquake Engineering*, Seattle, Washington.
64. Friedman, N., Geiger, D., Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2-3):131-163.
65. GAR (2009). *Global Assessment Report on Disaster Risk Reduction*. United Nations International Strategy for Disaster Reduction Secretariat (UNISDR). ISBN/ISSN: 9789211320282. 207 pp.
66. Geiss, C., Taubenböck, H. (2013). Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. *Natural Hazards*, 68(1): 7-48.
67. Geiss, C., Pelizari, P.A., Marconcini, M., Sengara, W., Edwards, M., Lakes, T., Taubenböck, H. (2015). Estimation of seismic building structural types using multi-sensor remote sensing and machine learning techniques. *ISPRS journal of photogrammetry and remote sensing*, 104:175-188.
68. Geiss, C., Jilge, M., Lakes, T., Taubenböck, H. (2016). Estimation of seismic vulnerability levels of urban structures with multisensor remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(5): 1913-1936.
69. Geiss, C., Schauss, A., Riedlinger, T., Dech, S., Zelaya, C., Guzmán, N., Hube, M.A., Arsanjani, J.J., Taubenböck, H. (2017). Joint use of remote sensing data and volunteered geographic information for exposure estimation: evidence from Valparaíso, Chile. *Natural Hazards*, 86(1): 81-105.
70. Grünthal, G. (1998). European macroseismic scale 1998. *European Seismological Commission (ESC)*.
71. Gunel, M. H. Ilgin, H. E. (2007), A Proposal for the Classification of Structural Systems of Tall Buildings, *Building and Environment*, Elsevier, 42 (7): 2667-2675.
72. Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3): 389-422.
73. Hancilar, U., Taucer, F., Corbane, C. (2013). Empirical fragility functions based on remote sensing and field data after the 12 January 2010 Haiti earthquake. *Earthquake Spectra*, 29(4): 1275-1310.
74. Haralick, R. M., Shanmugam, K., Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* 3(6): 610-621. DOI:10.1109/TSMC.1973.4309314.
75. Harb, M. M., De Vecchi, D., Dell'Acqua, F. (2015). Physical vulnerability proxies from remote sensing: reviewing, implementing and disseminating selected techniques. *IEEE Geoscience and Remote Sensing Magazine*, 3(1): 20-33.
76. Hayes, G. P., Briggs, R. W., Sladen, A., Fielding, E. J., Prentice, C., Hudnut, K., Mann, P., Taylor, F. W., Crone, A. J., Gold, R., Ito, T., Simons, M. (2010). Complex rupture during the 12 January 2010 Haiti earthquake, *Nature Geosciences*, 3: 800-805.
77. Hayes, G.P., Myers, E.K., Dewey, J.W., Briggs, R.W., Earle, P.S., Benz, H.M., Smoczyk, G.M., Flamme, H.E., Barnhart, W.D., Gold, R.D., Furlong K.P., (2017). Tectonic summaries of magnitude 7 and greater earthquakes from 2000 to 2015: U.S. Geological Survey Open-File Report 2016-1192, 148 p., <https://doi.org/10.3133/ofr20161192>.

78. Hazus-MH (2003). Multi-hazard Loss Estimation Methodology. Technical manual. Federal Emergency Management Agency (FEMA), Washington DC, USA. 712 pp.
79. He, H., Zhou, J., Chen, M., Chen, T., Li, D., Cheng, P. (2019). Building Extraction from UAV Images Jointly Using 6D-SLIC and Multiscale Siamese Convolutional Networks. *Remote Sensing*, 11(9): 1040. <https://doi.org/10.3390/rs11091040>
80. Hermosilla, T. (2011) Tesis doctoral Detección automática de edificios y clasificación de usos del suelo en entornos urbanos con imágenes de alta resolución y datos LiDAR. Universidad Politécnica de Valencia, Departamento de Ingeniería Cartográfica, Geodesia y Fotogrametría. 187 pp.
81. Herold, M., Liu, X., Clarke, K. C. (2003). Spatial Metrics and Image Texture for Mapping Urban Land Use. *Photogrammetric Engineering & Remote Sensing* 69 (9): 991–1001.DOI:10.14358/PERS.69.9.991.
82. HFA-2005. Hyogo Framework for Action 2005-2015: Building the resilience of nations and communities to disasters. United Nations Office for Disaster Risk Reduction (UNDRR). A/CONF.206/6.44pp. <https://www.unisdr.org/we/coordinate/hfa>. Último acceso 14/05/2019
83. Hill, M., Rossetto, T. (2008). Comparison of building damage scales and damage descriptions for use in earthquake loss modeling in Europe. *Bull Earthquake Eng.* DOI 10.1007/s10518-007-9057-y
84. Huberty, C. J. (1994). *Applied Discriminant Analysis*. New York: Wiley.
85. IBM SPSS Regression 22 (2014). Manual version 22. Ref. Nr.: 7038407. 23-julio-2014.
86. Jaiswal, K.S., Wald, D.J. (2008). Creating a global building inventory for earthquake loss assessment and risk management: U.S. Geological Survey Open-File Report 2008-1160, 103 pp.
87. JBDPA (1990). Standard for seismic capacity assessment of existing reinforced concrete buildings. Tokyo: Japanese building Disaster Prevention Association, Ministry of Construction, 1990.
88. Jin, X. (2012). Segmentation-based image processing system. U.S. Patent 8,260,048, filed Nov. 14, 2007, and issued Sept. 4, 2012.
89. Johnson, B., Xie, Z. (2011). Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4): 473-483.
90. Kappos, A. J., Stylianidis, K. C., Pitilakis, K. (1998). Development of seismic risk scenarios based on a hybrid method of vulnerability assessment. *Natural Hazards*, 17(2): 177-192.
91. Kononenko, I. (1994). Estimating Attributes: Analysis and Extensions of Relief. *Lecture Notes in Computer Science* 784: 171–182.
92. Kuhn, M., Johnson, K. (2013). *Applied predictive modeling*. New York: Springer. Print ISBN 978-1-4614-6848-6; Online ISBN 978-1-4614-6849-3.
93. Lagomarsino, S., Giovinazzi, S. (2006). Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. *Bulletin of Earthquake Engineering*, 4(4): 415-443.
94. Lallemand, D., (2014). Waiting for the big one: the continued earthquake risk of Port-au-Prince, Haiti. *Proceedings of the 3rd International Conference on Urban Disaster Reduction*, Earthquake Engineering Research Institute, Boulder, CO, USA.

95. Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*. 33(1): 159-174.
96. Lang, D.H., Singh, Y., Prasad, J.S.R. (2012b). Comparing Empirical and Analytical Estimates of Earthquake Loss Assessment Studies for the City of Dehradun, India. *Earthquake Spectra*, 28 (2) : 595–619, May 2012
97. Lang, D.H. (2013). Earthquake Damage and Loss Assessment – Predicting the Unpredictable. Dissertation for the degree of Dr. Philos. University of Bergen. Bergen, Norway, 2013. 332 pp. ISBN: 978-82-308-2271-5
98. Lang, D.H., Kumar, A., Sulaymanov, S., Meslem, A. (2017). Building typology classification and earthquake vulnerability scale of Central and South Asian building stock. *Journal of Building Engineering* 15 (2018) 261–277. DOI: <https://doi.org/10.1016/j.jobeb.2017.11.022>
99. Lantada, M.N. (2007). Tesis doctoral Evaluación del riesgo sísmico mediante métodos avanzados y técnicas GIS. Aplicación a la ciudad de Barcelona. Departamento de Ingeniería del terreno, Cartográfica y Geofísica. Universidad Politécnica de Cataluña. 325 pp.
100. Lawson, R.S., Vance, V., Krawinkler, H. (1994). Nonlinear static pushover analysis—why, when and how?, *Proc., 5th U.S. Conf. Earthq. Engrg.* 1:283-292
101. Li, L., Wang, J., Leung, H. (2010) Using spatial analysis and bayesian network to model the vulnerability and make insurance pricing of catastrophic risk. *Int J Geogr Inf Sci* 24:1759–1784.
102. Ma, L., Sacks, R., Zeibak-Shini, R., Aryal, A., Filin, S. (2015). Preparation of synthetic as-damaged models for post-earthquake BIM reconstruction research. *Journal of computing in civil engineering*, 30(3), 04015032.
103. Manaker, M., Calais, E., Freed, A.M., Ali, S.T., Przybylski, P., Mattioli, G., Jansma, C., Prépétit de Chaballier, J. B. (2008). Interseismic Plate coupling and strain partitioning in the Northeastern Caribbean, *Geophysical Journal International*, 174, 889–903.
104. Martínez, P., Goula, X. (2019). Plateforme d'évaluation post-sismique de dommages de bâtiments. 10ème Colloque National AFPS 2019. Strasbourg (France).
105. Martínez-Cuevas, S. (2014). Evaluación de la vulnerabilidad sísmica urbana basada en tipologías constructivas y disposición urbana de la edificación. Aplicación en la ciudad de Lorca, región de Murcia. Tesis Doctoral, E.T.S. I. Topografía, Geodesia y Cartografía (UPM). <http://oa.upm.es/30447/>. Último acceso 16/05/2019.
106. Martínez-Cuevas, S., Gaspar-Escribano, J. M. (2016). Reassessment of intensity estimates from vulnerability and damage distributions: the 2011 Lorca earthquake. *Bull Earthquake Eng.* 14: 2679-2703. DOI:10.1007/s10518-016-9913-8.
107. Martínez-Cuevas, S., Benito, M.B., Cervera, J., Morillo M.C., Luna, M. (2017). Urban modifiers of seismic vulnerability aimed at Urban Zoning Regulations. *Bull Earthquake Eng* 15: 4719. DOI: 10.1007/s10518-017-0162-2
108. Martínez-Díaz, J.J., Álvarez-Gómez, J. A. (2011). Seismotectonic characterization of the seismic series of Lorca and geological analysis of the earthquake source. In Geological preliminary field report of the Lorca earthquake (5.1 Mw, 11th May 2011), Instituto Geológico y Minero de España. Internal Report. <http://www.igme.es>. Último acceso 18/06/2019.
109. Martínez-Díaz, J. J., Bejar-Pizarro, M., Álvarez-Gómez, J. A., de Lis Mancilla, F., Stich, D., Herrera, G., Morales, J. (2012). Tectonic and seismic implications of an intersegment

- rupture: The damaging May 11th 2011 Mw 5.2 Lorca, Spain, earthquake. *Tectonophysics*, 546: 28-37. <https://doi.org/10.1016/j.tecto.2012.04.010>
110. Matsuka P., Chanussot J., Erwan, P., Philippe, G. (2012). A support vector regression approach for building seismic vulnerability assessment and evaluation from remote sensing and in situ data. In: *Proceedings of the 2012 IEEE international geoscience and remote sensing symposium (IGARSS)*, Munich, Germany, 22-27 July 2012, pp. 7533-7536.
111. McGarigal, K., Cushman, SA., Ene, E. (2012). FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: <http://www.umass.edu/landeco/research/fragstats/fragstats.html>
112. Medvedev, S. V., Sponheuer, W., Kárník, V. (1967). Seismic intensity scale, version 1964 (MSK-64). Institute of Geodynamics, Jena, Germany, Publ, 48: 13.
113. Milutinovic, Z., Trendafiloski, G. S. (2003). An advanced approach to earthquake risk scenarios with applications to different European towns. Report to WP04: Vulnerability of current buildings. September, 2003. 111 pp.
114. Mix D., Kijewski-Correa, T., Taflanidis, A.A. (2011). Assessment of Residential Housing in Léogâne, Haiti, and Identification of Needs for Rebuilding after the January 2010 Earthquake, *Earthquake Spectra* 27, S1:299-322, DOI:10.1193/1.3637942
115. Molina, S., Torres, Y., Benito, B., Navarro, M., Belizaire, D. (2014). Using the damage from 2010 Haiti earthquake for calibrating vulnerability models of typical structures in Port-au-Prince (Haiti). *Bull Earthq Eng* 12(4):1459-1478
116. Mouroux, P., Bertrand, E., Bour, M., Le Brun, B., Depinois, S., Masure, P. (2004). The European RISK-UE project: an advanced approach to earthquake risk scenarios. In *Proc. of the 13th World Conference on Earthquake Engineering*.
117. Mouroux, P., Le Brun, B. (2006). Presentation of RISK-UE project. *Bulletin of Earthquake Engineering*, 4(4), 323-339.
118. Mück, M., Taubenböck, H., Post, J., Wegscheider, S., Strunz, G., Sumaryono, S., Ismail, F.A. (2013). Assessing building vulnerability to earthquake and tsunami hazard using remotely sensed data. *Nat. Hazards*, 68: 97-114
119. NatCatSERVICE (2019). <https://natcatservice.munichre.com/>. Último acceso 10/05/2019.
120. ODS-2015. Objetivos de Desarrollo Sostenible. Naciones Unidas. <https://www.un.org/sustainabledevelopment/es/>. Último acceso 14/05/2019.
121. Navarro, M., García-Jerez, A., Alcalá, F. J., Vidal, F., Enomoto, T. (2014). Local site effect microzonation of Lorca town (SE Spain). *Bulletin of Earthquake Engineering*, 12(5): 1933-1959.
122. Ozdemir, P., Boduroglu, M.H., Ilki, A. (2005). Seismic Safety Screening Method, *Proceedings of the International Workshop on Seismic Performance Assessment and Rehabilitation of Existing Buildings (SPEAR)*, Ispra, Italy, Paper No. 23 (on CD).
123. Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, 1985 (pp. 329-334).

124. Pittore, M., Wieland, M. (2013). Toward a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing. *Natural Hazards*, 68(1), 115-145.
125. Pittore, M., Wieland, M., Fleming, K. (2016). Perspectives on Global Dynamic Exposure Modelling for Geo-Risk Assessment. *Nat Hazards*, 86(Suppl 1): 7-30. ISSN: 0921-030X, 1573-0840. DOI: 10.1007/s11069-016-2437-3.
126. Poland, C. D., Mitchell, A. D. (2007). A New Seismic Rehabilitation Standard—ASCE/SEI 41-06. In *New Horizons and Better Practices* (pp. 1-2).
127. Polli, D., Dell'Acqua, F. (2011) Fusion of optical and SAR data for seismic vulnerability mapping of buildings. In: Prasad S, Bruce L, Chanussot J (eds) *Advances in signal processing and exploitation techniques*. Springer, Berlin Heidelberg, pp 329–341. DOI:10.1007/978-3-642-14212-3_15
128. Prentice, C. S., Mann, P., Crone, A. J., Gold, R. D., Hudnut, K. W., Briggs, R. W., Koehler, R. D., Jean, P. (2010). Seismic hazard of the Enriquillo–Plantain Garden fault in Haiti inferred from palaeoseismology, *Nature Geosciences*, 3: 789-793.
129. Qi, W., Su, G., Sun, L., Yang, F., Wu, Y. (2017). Internet+ approach to mapping exposure and seismic vulnerability of buildings in a context of rapid socioeconomic growth: a case study in Tangshan, China. *Natural Hazards*, 86(1): 107-139.
130. Quinlan, R. (1993). *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA
131. Quirós, L.E. (2017). Modelizaciones y análisis de sensibilidad en la evaluación integral del riesgo sísmico a escala urbana: aplicación a la ciudad de Lorca. Tesis Doctoral, E.T.S.I. Industriales (UPM). <https://doi.org/10.20868/UPM.thesis.48042>. Último acceso 16/05/2019.
132. Raccuglia, P., Elbert, K. C., Adler, P. D., Falk, C., Wenny, M. B., Mollo, A., Zeller, M., Friedler, S.A, Schrier, J., Norquist, A. J. (2016). Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601): 73.
133. RADIUS (1999). *Risk Assessment Tools for Diagnosis of Urban Areas Against Seismic Disasters*, United Nations International Strategy for Disaster Reduction Secretariat (UNISDR). <https://www.geohaz.org/risk-assessment-tools-for-diagnosis-of-u> ;Último acceso 09/04/2019
134. Riedel, I., Gueguen, P., Dunand, F., Cottaz, S. (2014). Macroscale vulnerability assessment of cities using association rule learning. *Seismological Research Letters*, 85(2): 295-305.
135. Riedel, I., Guéguen, P., Dalla Mura, M., Pathier, E., Leduc, T., Chanussot, J. (2015). Seismic vulnerability assessment of urban environments in moderate-to-low seismic hazard regions using association rule learning and support vector machine methods. *Natural hazards*, 76(2): 1111-1141.
136. Rojahn, C., Sharpe, R. L. (1985). *ATC-13: Earthquake damage evaluation data for California*. Applied technology council. Redwood City, California, U.S.A. 492 pp
137. Rossetto, T., Elnashai, A. (2003). Derivation of Vulnerability Functions for European-Type RC Structures Based on Observational Data, *Engineering Structures*, 25(10): 1241-1263.
138. Santos, I., Brezo, F., Ugarte-Pedrero, X., Bringas, P. G. (2013). Opcode sequences as representation of executables for data-mining-based unknown malware detection. *Information Sciences*, 231: 64-82.

Referencias

139. Sarabandi, P., Kiremidjian, A., Eguchi, R. T. and Adams, B.J. (2008). Building inventory compilation for disaster management: Application of remote sensing and statistical modeling. Technical Report Series MCEER-08-0025, MCEER: Buffalo.
140. Schröter, K., Lüdtke, S., Redweik, R., Meier, J., Bochow, M., Ross, L., Nagel, C., Kreibich, H. (2018). Flood loss estimation using 3D city models and remote sensing data. *Environmental Modelling & Software*, 105: 118-131.
141. Sendai-Framework-2015. Sendai Framework for Disaster Risk Reduction 2015-2030. United Nations Office for Disaster Risk Reduction (UNDRR). A/RES/69/283. 26 pp. <https://www.unisdr.org/we/coordinate/sendai-framework>. Último acceso 14/05/2019.
142. Silva, V., Henshaw, P., Huyck, C., O'Hara, M. (2018). GED4ALL: Global Exposure Database for Multi-Hazard Risk Analysis. D5 – Final Report. GEM Technical Report 2018-05, GEM Foundation, Pavia, Italy.
143. Sohn, G., Dowman, I. (2007). Data fusion of high-resolution satellite imagery and LiDAR data for automatic building extraction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(1): 43-63.
144. Soille, P. (2007). On genuine connectivity relations based on logical predicates. In: *Proceedings of 14th international conference on image analysis and processing*, Modena, Italy, pp 487-492. IEEE Computer Society Press, September 2007. doi:10.1109/ICIAP.2007.4362825
145. Soille, P. (2008) Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Trans Pattern Anal Mach Intell.* 30(7):1132-1145. doi:10.1109/TPAMI.2007.70817
146. Soille, P. (2009) Recent developments in morphological image processing for remote sensing. In: Bruzzone L, Benediktsson J, Serpico S (eds) *Image and signal processing for remote sensing*, volume SPIE-7477. SPIE, 2009. doi:10.1117/12.836024
147. Spence, R., Coburn, A.W., Pomonis, A. (1992). Correlation of Ground Motion with Building Damage: The Definition of a New Damage-Based Seismic Intensity Scale, *Proceedings of the Tenth World Conference on Earthquake Engineering*, Madrid, Spain, 1: 551-556.
148. Spence, R., E. So, S. Jenkins, A. Coburn, and S. Ruffle (2011). A global Earthquake building damage and casualty database, Chapter 5 in Spence R, So E and Scawthorn C (eds), *Human casualties in earthquakes: progress in modelling and mitigation*, Springer.
149. Stanimirova, I., Üstün, B., Cajka, T., Riddelova, K., Hajslova, J., Buydens, L. M. C. Walczak, B. (2010). Tracing the geographical origin of honeys based on volatile compounds profiles assessment using pattern recognition techniques. *Food Chemistry*, 118(1): 171-176.
150. Su, G.W., Qi, W.H., Zhang, S.L., Sim, T., Liu, X.S., Sun, R., Sun, L., Jin, YF. (2015) An integrated method combining remote sensing data and local knowledge for the large-scale estimation of seismic loss risks to buildings in the context of rapid socioeconomic growth: a case study in Tangshan, China. *Remote Sens* 7:2543-2601. DOI:10.3390/rs70302543
151. Symithe, S., Calais, E. (2016). Present-day shortening in Southern Haiti from GPS measurements and implications for seismic hazard. *Tectonophysics*, 679: 117-124.
152. SYNER-G D1.1 (2009). SYNER-G work plan. SYNER-G Project: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain. WP1-Project coordination and management. Online: <http://www.vce.at/SYNER-G/files/dissemination/deliverables.html> Último acceso 09/04/2019

153. SYNER-G D3.1 (2011). Fragility Functions for Common RC Building Types in Europe. SYNER-G Project: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain. WP3-Fragility Functions of Elements at Risk. Online: <http://www.vce.at/SYNER-G/files/dissemination/deliverables.html> Último acceso 09/04/2019
154. SYNER-G D3.2 (2011). Fragility Functions for Common Masonry Building Types in Europe. SYNER-G Project: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain. WP3-Fragility Functions of Elements at Risk. Online: <http://www.vce.at/SYNER-G/files/dissemination/deliverables.html> Último acceso 09/04/2019
155. SYNER-G D8.9 (2013). Development of inventory datasets through remote sensing and direct observation data for earthquake loss estimation. SYNER-G Project: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain. WP8-Guidelines, recommendations and dissemination. Online: <http://www.vce.at/SYNER-G/files/dissemination/deliverables.html> Último acceso 16/05/2019
156. Taft-Morales, M. (2017). Haiti's Political and Economic Conditions: In Brief. Congressional Research Service of the US Congress. Whashington, DC. December 1, 2017.
157. Taubenböck H, Roth A, Dech S, Mehl H, Muñich J, Stempniewski L, Zschau J. (2009). Assessing building vulnerability using synergistically remote sensing and civil engineering. In: Krek A, Rumor M, Zlatanova S, Fendel E (eds) Urban and regional data management. Taylor & Francis Group, London, pp 287–300. ISBN 978-0-415-55642-2
158. TerminologyDRR-2016. Terminology on Disaster Risk Reduction. Informe del grupo de trabajo intergubernamental de expertos de composición abierta sobre los indicadores y la terminología relacionados con la reducción del riesgo de desastres. Asamblea General de las Naciones Unidas. A/71/644. 39pp. <https://www.unisdr.org/we/inform/terminology>. Último acceso 16/05/2019.
159. Torres, Y., Molina, S., Martínez-Cuevas, S., Navarro, M., Martínez-Díaz, J. J., Benito, B., Galiana-Merino, J.J., Belizaire, D. (2016). A first approach to earthquake damage estimation in Haiti: advices to minimize the seismic risk. Bulletin of Earthquake Engineering, 14(1): 39-58.
160. Torres, Y., Arranz, J.J., Gaspar-Escribano, J.M., Haghi, A., Martínez-cuevas, S., Benito, B., Ojeda, J.C. (2019a). Integration of LiDAR and multispectral images for rapid exposure and earthquake vulnerability estimation. Application in Lorca, Spain. International Journal of Applied Earth Observation and Geoinformation. 81: 161-175. DOI: <https://doi.org/10.1016/j.jag.2019.05.015>
161. Torres, Y. Arranz, JJ., Martínez-Cuevas, S., Gaspar-Escribano, J.M., Benito, B., Staller, A., Gómez, I. (2019b). The new methods for rapid exposure and seismic vulnerability assessment. How do they adapt to different scenarios? Proceedings II Congreso Internacional en Ingeniería Geomática, Madrid, 26-27 Junio 2019.
162. UN (2011). Report of the United Nations in Haiti 2010. Situation, challenges and outlook.
163. UNDRO (1979). Natural Disasters and Vulnerability Analysis, Report of Experts group meeting, Genova, Julio 1979, 49pp.
164. UNDRR-AM (2017). Country document for disaster risk reduction: Haiti, 2016. United Nations Office for Disaster Risk Reduction – Regional Office for the Americas (UNDRR AM) and Haiti – government. 116pp.

165. Üstün, B., Melssen, W. J., Buydens, L. M. (2006). Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems*, 81(1): 29-40.
166. Vacareanu, R., Lungu, D., Aldea, A., Arion, C. (2004). RISK-UE, An advanced approach to earthquake risk scenarios with applications to different European towns. WP7 Report: *Seismic Risk Scenarios Handbook*.
167. Vapnik, V. (2000). *The Nature of Statistical Learning Theory*, 2nd ed.; Springer: New York, NY, USA, 2000. DOI 10.1007/978-1-4757-3264-1
168. Vincent, L., Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulation. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13: 583-598, Jun 1991.
169. Waters, T., (2010). The OpenStreetMap Project & Haiti Earthquake Case Study. SlideShare. Online: https://es.slideshare.net/chippy/openstreetmap-case-study-haiti-crisis-response?from_action=save. Último acceso 19/06/2019.
170. WHE, 2000: <http://www.world-housing.net/>. Último acceso 07/04/2019
171. Whitman, R.V., Reed, J.W., Hong, S.T. (1973). Earthquake Damage Probability Matrices, *Proceedings of the Fifth World Conference on Earthquake Engineering*, Rome, Italy, 2: 2531-2540.
172. Wieland, M., Pittore, M., Parolai, S., Zschau, J. (2012a). Exposure estimation from multi-resolution optical satellite imagery for seismic risk assessment. *ISPRS International Journal of Geo-Information*, 1(1): 69-88.
173. Wieland, M., Pittore, M., Parolai, S., Zschau, J., Moldobekov, B., Begaliev, U. (2012b). Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging. *Soil Dynamics and Earthquake Engineering*, 36: 70-83.
174. Wieland, M. (2013). Exposure estimation for rapid seismic vulnerability assessment. An integrated approach based on multi-source imaging. Thesis dissertation. Technische Universität Berlin, Fakultät VI - Planen Bauen Umwelt (English).
175. Wieland, M., Pittore, M. (2014). Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images. *Remote Sensing* 6 (4): 2912-2939. DOI:10.3390/rs6042912.
176. Wieland, M., Torres, Y., Pittore, M., Benito, B. (2016). Object-based urban structure type pattern recognition from Landsat TM with a Support Vector Machine. *International Journal of Remote Sensing*, 37(17): 4059-4083.
177. Wieland, M., Pittore, M. (2017). A Spatio-Temporal Building Exposure Database and Information Life-Cycle Management Solution. *ISPRS Int. J. Geo-Inf.* 2017, 6, 114; DOI:10.3390/ijgi6040114
178. Wu, H., Cheng, Z., Shi, W., Miao, Z., Xu, C. (2014). An object-based image analysis for building seismic vulnerability assessment using high-resolution remote sensing imagery. *Natural hazards*, 71(1): 151-174.
179. Yan, G., Mas, J.F., Maathuis, B.H.P., Xiangmin, Z., Van Dijk, P.M. (2006). Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China, *International Journal of Remote Sensing*, 27(18): 4039-4055, DOI: 10.1080/01431160600702632

180. Yepes-Estrada, C., Silva, V., Valcárcel, J., Acevedo, A.B., Tarque, N., Hube, M., Coronel, G., Santa María, H. (2017). Modeling the Residential Building Inventory in South America for Seismic Risk Assessment. *Earthquake Spectra*, 33(1): 299–322. DOI: 10.1193/101915EQS155DP
181. Zeibak-Shini, R., Sacks, R., Ma, L., Filin, S. (2016). Towards generation of as-damaged BIM models using laser-scanning and as-built BIM: First estimate of as-damaged locations of reinforced concrete frame members in masonry infill structures. *Advanced engineering informatics*, 30(3): 312-326.
182. Zhuo, X., Fraundorfer, F., Kurz, F., Reinartz, P. (2018). Optimization of OpenStreetMap building footprints based on semantic information of oblique UAV images. *Remote Sensing*, 10(4): 624. <https://doi.org/10.3390/rs10040624>

11 Glosario de términos y acrónimos

A continuación, se presenta un listado de los términos más relevantes que se han utilizado en esta tesis, acompañados de una breve definición.

- **Análisis predictivo:** técnica estadística que consiste en crear un modelo matemático para predecir el valor de una variable (dependiente) en función de los valores de otras variables (independientes o predictivas). En esta tesis, estos modelos se calculan con técnicas de aprendizaje automático (traducción del término inglés *Machine Learning*, ML).
- **Dataset equilibrado/desequilibrado (*balanced/unbalanced*):** base de datos de entrenamiento (*training dataset*) para una clasificación con igual/desigual número de muestras por clase.
- **Dataset de entrenamiento:** base de datos de entrenamiento para análisis predictivo. En inglés se denomina *training dataset*.
- **Dataset de verificación:** base de datos de verificación de los resultados de una clasificación realizada mediante análisis predictivo. En inglés se denomina *testing dataset*, *ground truth dataset*, *yet-to-be-seen data* o *unseen data*.
- **Elongación de un edificio:** relación entre el ancho y el largo del rectángulo envolvente de la huella de un edificio.
- **EO: *Earth Observation*** (Observación de la Tierra, OT). Recolección de información sobre los procesos que ocurren en la Tierra mediante sensores remotos.
- **F1-Score:** Métrica derivada de la matriz de confusión, que proporciona la media armónica de la precisión y el recall.
- **FP Rate: *False Positive Rate*.** Métrica derivada de la matriz de confusión. Es el índice de falso positivo, también denominado 1-especificidad. Junto con el TP-rate, se emplea para crear la curva ROC.
- **GLCM:** matriz de co-ocurrencia de niveles de gris, a partir de la cual se derivan las características de textura.
- **GMPE: *Ground Motion Prediction Equation*.** También llamados leyes de atenuación, son modelos predictivos del movimiento fuerte del suelo asociado a un terremoto.
- **Grado de compacidad de un edificio:** relación entre el área del edificio y el área de un círculo que tenga el mismo perímetro.
- **GSD: *Ground Sampling Distance*.** Tamaño del pixel, indicador de la resolución espacial de los datos tomados por sensores remotos.
- **HAZUS:** Metodología completa de estimación del riesgo sísmico de la FEMA con una clasificación de edificios propia para EE.UU. (aunque en la práctica, se aplica en otras muchas regiones del mundo).
- **Huella:** contorno de la forma en planta de un edificio (*building footprint*).
- **I-DCM: *Improved Displacement Coefficient Method*.** Método de cálculo del punto de desempeño de un edificio sometido a una sacudida sísmica empleado en esta tesis para calcular el daño por sismo. Se obtiene al cruzar el espectro de capacidad del edificio con

el espectro de demanda sísmica, y viene dado por el desplazamiento máximo espectral que sufre el edificio a causa del terremoto representado en el espectro de demanda.

- Kappa *coefficient of agreement*: métrica derivada de la matriz de confusión que compara la precisión observada con la precisión esperada.
- LiDAR: *Light Detection And Ranging*. Tecnología de registro basada en escaneo láser. Los datos tomados con LiDAR son nubes de puntos tridimensionales de alta densidad y precisión. Además de las coordenadas XYZ, cada punto almacena información de intensidad del pulso láser y el orden de registro. En esta tesis se utilizan datos LiDAR aéreos.
- M: *Masonry*: Tipología constructiva típica de Lorca, representativa de edificios contruidos con mampostería sin reforzar.
- MAndW: *Masonry and Wood*. Tipología constructiva típica de Puerto Príncipe, representativa de edificios contruidos con mampostería y madera.
- Matriz de confusión: estructura de comparación utilizada en esta tesis para verificar el resultado de una clasificación. La matriz se crea colocando por columnas (o filas) el número de instancias asignadas por el clasificador a cada clase, y por filas (o columnas), la instancias que realmente pertenecen a cada clase, según el dataset de verificación. De la comparación de filas y columnas se puede saber cuántas instancias han sido correctamente clasificadas, ya que son los valores de la diagonal principal de la matriz de confusión.
- Mw: Magnitud momento. Escala sismológica que informa acerca del tamaño de los terremotos y es función de la energía liberada al producirse estos.
- OA: *Overall Accuracy*. Precisión global de una clasificación. Métrica derivada de la matriz de confusión, que indica el número de instancias clasificadas correctamente respecto del total de instancias que han participado en la clasificación.
- OBIA: *Object-Based Image Analysis*. Técnica de análisis de imagen basada en objeto, destinada a detectar, delinear e identificar objetos del mundo real presentes en una imagen, como edificios, áreas vegetadas, masas arbóreas, cuerpos de agua, sombras, caminos, pavimento, etc. Se compone de dos fases: segmentación, por la que la imagen es subdividida en agrupaciones de píxeles homogéneas, llamadas segmentos u objetos; y clasificación, mediante la que se etiquetan esos objetos de manera automática, a través de algoritmos de aprendizaje automático, generalmente.
- Orientación de un edificio: acimut del lado más largo del rectángulo envolvente de la huella de un edificio.
- Precisión (*precision*): métrica derivada de la matriz de confusión que da una medida de los errores por comisión. También es conocida como precisión del usuario.
- Radar: *RAdio Detection And Ranging*. Tecnología de registro basada en la emisión y detección de ondas de radio. En esta tesis no se usan datos RADAR.
- RC: *Reinforce Concrete*. Tipología constructiva representativa de edificios contruidos con hormigón armado (nomenclatura usada tanto para Lorca como para Puerto Príncipe).

- RFE: *Recursive Feature Elimination*. Técnica estadística empleada para reducir la dimensión del vector de características.
- ROC: *Receiver Operating Characteristics*. Métrica de calidad de una clasificación que representa el TP Rate frente al FP rate.
- RVS: *Rapid Visual Screening*. Procedimiento de inspección de edificios in situ para la evaluación de su vulnerabilidad ante algún tipo de riesgo.
- Segmentos: polígonos resultado de una segmentación de imagen. También se llaman objetos en esta tesis. Las huellas de edificios obtenidas por segmentación se han denominado huellas irregulares (por contraposición a las huellas regulares provenientes de un proceso de digitalización manual o de una base de datos oficial).
- SGE 2.0: Software para Gestión de Emergencias. Herramienta que coordina a ciudadanos y a las instituciones que intervienen en la recuperación de una ciudad tras una catástrofe para compartir datos.
- Taxonomía de clasificación de edificios: código estandarizado para organizar la información relativa a los edificios en base a la cual se pueda establecer su vulnerabilidad ante algún riesgo natural.
- TP Rate: *True Positive Rate*. Métrica derivada de la matriz de confusión que da una medida de los errores por comisión. Es el índice de verdadero positivo, también conocido como sensibilidad, recall o precisión del productor. Junto con el FP-rate, se emplea para crear la curva ROC.
- Vector de características (*feature vector*): atributos propios de cada instancia en una clasificación. Las características también son llamadas atributos en esta tesis.

A continuación, se presenta un listado de acrónimos de interés que han sido mencionados en este documento.

- ASCE: American Society of Civil Engineers.
- ATC: Applied Technology Council.
- AUC: Area Under the Curve (ROC).
- BTEB: Bureau technique dévaluation de bâtiments d'Haïti.
- BIM: Building Information Modelling.
- CEQID: Cambridge Earthquake Impact Database.
- CNIG: Centro Nacional de Información Geográfica del IGN.
- CNIGS: Centre National de l'Information Géo-Spatiale d'Haïti.
- COPERNICUS: Programa de Observación de la Tierra de la Unión Europea.
- CSM: Capacity Spectrum Method.
- DRR: Disaster Risk Reduction.
- DBPC: Directriz Básica de Planificación de Protección Civil.
- EERI: Earthquake Engineering Research Institute.
- EMS: European Macroseismic Scale.
- FEMA: Federal Emergency Management Agency.
- FID: Identificador de las áreas de muestra de Puerto Príncipe.
- GEM: Global Earthquake Model.
- HFA: Hyogo Framework for Action.
- HOT: Humanitarian OpenStreetMap Team.
- HR, VHR: High Resolution, Very High Resolution (aplicadas a la resolución espacial de datos tomados por sensores remotos).
- IAEE: International Association for Earthquake Engineering.
- IDE: Infraestructura de datos espaciales.
- IGN: Instituto Geográfico Nacional de España.
- IGV: Información Geográfica Voluntaria.
- IMM: Intensidad de Mercalli Modificada.
- Iv: Índice de vulnerabilidad.
- JRC: European Commission's Joint Research Centre.
- MADRS: Modified Acceleration-Displacement Response Spectrum.
- MAPE: Mean Absolute Percentage Error.
- MBT: Model Building Type (tipología constructiva típica de una zona de estudio).
- MDT: Modelo Digital del Terreno.
- MDSn (nDSM): Modelo Digital de Superficie normalizado.
- MMS: Mobile Mapping System.
- MPD: Matrices de probabilidad de daño.
- MTPTC: Ministère des Travaux Publics, Transports et Communications d'Haïti.
- NIR: Near Infrared, banda del infrarrojo cercano de una imagen.
- ODS: Objetivos de Desarrollo Sostenible.
- OSM: OpenStreetMap.

- PAGER: Prompt Assessment of Global Earthquakes for Response.
- P.E.: Percentage Error.
- PNOA: Plan Nacional de Ortofotografía Aérea de España.
- PUK: Pearson VII Universal Kernel. Kernel utilizado con SVM para crear clasificadores en esta tesis.
- RADIUS Project: <<Risk Assessment Tools for Diagnosis of Urban Areas against Seismic Disasters.>>
- RBF: Radial Basis Function. Kernel utilizado con SVM para crear clasificadores en esta tesis.
- RGB: Red-Green-Blue: bandas del espectro visible de una imagen.
- RIESGOS Project: Multi-Risk Information SystEm for South America's Andes ReGiOnS.
- Risk-UE Project: An advanced approach to earthquake risk scenarios with applications to different European towns.
- RIT: Rochester Institute of Technology.
- RMSE: Root Mean Square Error.
- RPA: Remotely Piloted Aircraft.
- SAR: Synthetic Aperture RADAR.
- SARA Project: South America Risk Assessment.
- SBST: Seismic Building Structure Types (se puede considerar sinónimo de MBT en esta tesis).
- SVM: Support Vector Machine.
- SWIR: Short Wave Infrared: banda del infrarrojo de onda corta de una imagen.
- SYNER-G Project: European Collaborative Research Project focusing on systemic seismic vulnerability and risk analysis of buildings, lifelines and infrastructures
- TAN: Tree Augmented Naïve Bayes.
- TIN: Triangulated Irregular Network.
- TOD: TIN for Object Detection (algoritmo para clasificación de puntos LiDAR).
- UNDRO: United Nations Disaster Relief Organization.
- UNISDR: Actualmente es UNDRR, la oficina de Naciones Unidas para la Reducción del Riesgo de Desastres (UN Office for Disaster Risk Redution).
- USGS: U.S. Geological Survey.
- WHE: World Housing Encyclopedia.

Anexo I

Formularios de recogida de datos sobre edificación en campo.

Ejemplos del trabajo realizado por el Ministère des Travaux Publics, Transports et Communications de Haiti y por el equipo de SGE 2.0 en Lorca

MINISTÈRE DES TRAVAUX PUBLICS TRANSPORTS ET COMMUNICATIONS

Evaluation rapide de bâtiments

DTP-BE #1

Inspection

Id. Inspecteur : _____

Date et Heure d'inspection : _____ ☐ AM ☐ PM

Affiliation : _____

Zone inspectée: ☐ Extérieur uniquement ☐ Extérieur et intérieur

Description du bâtiment

Type de construction

Nom du bâtiment :

Structure

- ☐ Structure en béton
- ☐ Structure en béton armé
- ☐ Structure en acier / tôles
- ☐ Structure en bois / tôles
- ☐ Murs porteurs

Adresse :

Dalle

Plancher

- ☐ En béton
- ☐ En béton armé
- ☐ En bois

Murs

- ☐ Murs en béton armé
- ☐ Murs de blocs non armés
- ☐ Murs de blocs armés
- ☐ Maçonnerie de roches
- ☐ Murs de briques
- ☐ Bois + maçonnerie

Contact/Téléphone :

Toiture

- ☐ En béton armé
- ☐ En bois / tôle
- ☐ En acier / tôle

Coordonnées GPS:

Nombre d'étage :

Nombre de sous-sol :

Autre type:

Spécification du bâtiment

Type d'occupation

Superficie approximative (mètre carré) :

Nombre de résidences :

Nombre de locaux non habitables :

- ☐ Résidentiel- section unique
- ☐ Résidentiel- section multiple
- ☐ Assemblée publique
- ☐ Santé
- ☐ Hôtel

- ☐ Commerce
- ☐ Bureaux
- ☐ Industriel
- ☐ Services d'urgence
- ☐ Bâtiment religieux

- ☐ Gouvernement
- ☐ Historique
- ☐ Ecole
- ☐ Autre

Evaluation

Etudiez les bâtiments pour vérifier leur état et cochez la colonne appropriée ci-dessous

Etat observé

Bâtiment effondré, partiellement effondré ou déplacé Bâtiment ou étage penché

Murs intérieurs et extérieurs fissurés

Murs intérieurs et extérieurs effondrés

Colonnes, pilastres et corbeaux fissurés et émiétés

Dalles, poutres, solives fissurées et écaillées

Parapets, fermes, terrasses et escaliers endommagés

Détérioration du bâtiment

Fissures ou mouvement du sol

Autre (spécifier)

Mineur/Aucun

Modéré

Grave

Domage estimé

☐

☐

☐

- ☐ aucun
- ☐ 0-1%
- ☐ 1-10%
- ☐ 10-30%
- ☐ 30-60%
- ☐ 60-100%
- ☐ 100%

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

☐

Difficulté de la démolition :

Aucun/mineure

Simple

Moyenne

Complexe

☐

☐

☐

☐

Observations : _____

SIGNALISATION

Choisissez une signalisation basée sur l'évaluation et le jugement d'équipe. Les conditions graves mettant en danger tout le bâtiment sont des raisons pour en interdire l'accès par la signalisation ACCES NON AUTORISE. Les états modérés et peu graves peuvent permettre une signalisation Utilisation restreinte. La signalisation INSPECTÉ doit être affichée à l'entrée principale. Les signalisations ACCES LIMITE et ACCES NON AUTORISE doivent être placées à toutes les entrées.

☐ **Inspecté (fiche verte)** ☐ **accès limité (fiche jaune)** ☐ **accès non autorisé (fiche rouge)**

☐ **Le bâtiment est-il vert mais taggé jaune / rouge due au martellement du bâtiment adjacent**

Enregistrez toutes les restrictions telles qu'indiquées: _____

AUTRES VERIFICATIONS. Cochez les cases ci-dessous seulement si d'autres actions sont nécessaires

- ☐ Barricades nécessaires dans la zone: _____
- ☐ Expertise détaillée recommandée : ☐ Structurelle ☐ Géotechnique ☐ Autre _____
- ☐ Autres recommandations ou restrictions tel qu'inscrit sur la pancarte : _____

Observations: _____

Facteurs de vulnérabilité (cochez tout ce qui s'applique) :

Date de début de construction ☐ 1 – 10 années ☐ 11 – 25 années ☐ 26 – 50 années ☐ plus de 50 années

Date de travaux importants de rénovation ☐ 1 – 5 années ☐ 6 – 10 années ☐ plus de 10 années

Sols de l'emplacement :

☐Sable de plage ☐Sol mou ☐ Sol ferme ☐Roches

Inclinaison de construction :

☐Plate ☐ Modérée ☐Abrupte

Emplacement du bâtiment :

☐ Plage ☐Rivière ☐ Vallée ☐ Plaine ☐ Colline ☐Flan de coteau ☐Sommet

Fondation :

☐Dalle/pose sur Grade ☐ élevé sur Post ☐Piles profondes

Forme en plan :

☐ O ☐ E ☐ H ☐ L ☐ T ☐ U ☐Rectangulaires ☐Autres

Irrégularités verticales :

☐Etage mou ☐Dénivelés ☐Murs de cisaillement couplés ☐Colonnes raccourcies ☐Martèlement de bâtiment adjacent

FICHA DE EVALUACIÓN DE LA VULNERABILIDAD SÍSMICA DE LOS EDIFICIOS EN LORCA

DATOS

ID: Referencia Catastral: Año de Construcción:
 Nº de viviendas: Nº de plantas: Sobre rasante: Bajo rasante:
 Dirección postal: Uso actual:

Cargar foto perspectiva

Cargar foto alzado

Cargar foto detalles

CARACTERÍSTICAS CONSTRUCTIVAS DE LA EDIFICACIÓN

- ☐ **ADOBE** ☐ **TAPIAL**
- ☐ **MURO DE CARGA:** ☐ Cascote y piedras ☐ Piedra sin tallar ☐ Sillar ☐ Fábrica ladrillo ☐ Mixto
 Forjado: ☐ Madera ☐ Artesano ☐ Hormigón armado ☐ Otro
 Refuerzo: ☐ Contrafuertes ☐ Tirantes ☐ Otros
- ☐ **HORMIGÓN ARMADO:** ☐ Sin diseño sismorresistente ☐ Con diseño sismorresistente
 Forjado: ☐ Unidireccional ☐ Bidireccional
- CUBIERTA:** ☐ Pesada ☐ Ligera

NOTAS

DEGRADACIÓN O MAL ESTADO DE CONSERVACIÓN

- ☐ **MUY ALTA O RUINOSA**
- ☐ **ALTA** ☐ Desprendimientos ☐ Desplomes ☐ Grietas elementos estructurales ☐ Grietas en uniones
- ☐ **MEDIA** ☐ Grietas o Fisuras ☐ Humedad ☐ Mal mantenimiento
- ☐ **BAJA O NULA** (No se aprecia nada)

NOTAS

CARACTERÍSTICAS URBANÍSTICAS DE LA EDIFICACIÓN

TIPOLOGÍA EDIFICATORIA:

- ☐ Residencial multifamiliar: ☐ Entre medianeras: ☐ Manzana cerrada con patio central ☐ Manzana cerrada densa
☐ Bloque abierto
- ☐ Residencial unifamiliar: ☐ Aislada ☐ Adosada o pareada ☐ Entre medianeras ☐ Agrupada

Longitud de fachada: m

FORMA GEOMÉTRICA EN PLANTA: ☐ Regular ☐ ☐ Esquinas exteriores entrantes ☐ Esquinas interiores entrantes

FORMA GEOMÉTRICA EN ALZADO:

- ☐ Regular ☐ Irregular: ☐ Escalonamiento ☐ Base y torre ☐ Invertido ☐ Voladizos ☐ Mixta

ELEMENTOS EXTERIORES: ☐ Parapetos ☐ Chimeneas ☐ Molduras ☐ Otros

EFFECTO GOLPETEO: ☐ Golpeteo por la izquierda ☐ Golpeteo por la derecha

POSICIÓN DEL EDIFICIO EN LA MANZANA: ☐ Esquina ☐ Intermedio ☐ Terminal ☐ E ☐ I ☐ T

PISO BLANDO: ☐ Sin muros de cerramientos ☐ Parcialmente cerrado
☐ Uso diferente ☐ Distribución diferente ☐ Altura diferente

COLUMNA CORTA: ☐ Semisótano ☐ Por columnas cautivas

POSICIÓN RESPECTO A LAS ALINEACIONES: ☐ En línea ☐ Fuera de línea ☐ Retranqueada: m

ALTURA CON LOS COLINDANTES: Nº plantas Edificio Izquierda Edificio Derecha

PENDIENTE: ☐ Llana ☐ Media ☐ Abrupta

Notas aclaratorias:

DATOS	Marcar todos los campos
ID	Código identificador
Año de Construcción	Año de recepción final de obra
Nº de viviendas	No incluye locales comerciales ni trasteros
Nº de plantas	Distinguir <i>sobre rasante</i> (incl. planta baja) y <i>bajo rasante</i> (sótanos)
CARACTERÍSTICAS CONSTRUCTIVAS DE LA EDIFICACIÓN	
ADOBE, TAPIAL, MURO DE CARGA, HORMIGÓN ARMADO	Escoger una categoría, y una o varias subcategorías
CUBIERTA	Escoger una: <i>Pesada</i> cubierta de Hormigón Armado y <i>Ligera</i> cubierta de teja, madera ...
CARACTERÍSTICAS URBANÍSTICAS DE LA EDIFICACIÓN	
TIPOLOGÍA EDIFICATORIA	Escoger entre <i>Residencial multifamiliar</i> y <i>Residencial unifamiliar</i> , y una subcategoría de estas Si el edificio tiene más de una fachada a la calle, en el campo <i>Longitud de fachada</i> se pondrá la de mayor longitud
FORMA GEOMÉTRICA EN PLANTA	La parte gris de los dibujos representa espacio vacío
FORMA GEOMÉTRICA EN ALZADO	Hacer croquis en forma mixta
ELEMENTOS EXTERIORES	Cualquier elemento no estructural o arquitectónico que pueda caer a la vía pública en caso de un sismo
EFECTO GOLPETEO	Izquierda y derecha según se mira desde el acceso principal al edificio
POSICIÓN DEL EDIFICIO EN LA MANZANA	El edificio <i>terminal</i> solo tiene un edificio colindante.
PISO BLANDO	Se refiere a cerramientos externos <i>Sin nada de cerramiento</i> : Sin muros entre columnas, cerramientos totalmente de vidrio <i>Parcialmente cerrado</i> : Con muros parciales y vidrios <i>Totalmente cerrado</i> : Con muros de cerramiento
POSICIÓN RESPECTO A LAS ALINEACIONES	Indicar retranqueo en metros
DIFERENCIA DE ALTURA CON LOS COLINDANTES	Poner el número de plantas del edificio colindante a izquierda y derecha.
PENDIENTE	<i>Llana</i> : con pendiente de 0% o despreciable. <i>Media</i> : Pendiente menor a 10% (similar a una rampa de acceso garaje) <i>Abrupta</i> : Pendiente mayor a 10%. (Se podría desarrollar una aplicación para medir la pendiente con el móvil)

Anexo II

Tabla comparativa de estudios similares.

Características más relevantes de los estudios destinados a generar bases de datos de exposición y/o vulnerabilidad utilizando sensores remotos en los últimos 10 años.

Comparison of studies aimed at evaluating earthquake exposure and/or vulnerability based on in-field surveys, remote sensing and ancillary data over the past 10 years.

This table is compiled upon those from Su et al. (2015) and Qi et al. (2017).

Su, G.W., Qi, W.H., Zhang, S.L., Sim, T., Liu, X.S., Sun, R., Sun, L., Jin, Y.F. (2015) An integrated method combining remote sensing data and local knowledge for the large-scale estimation of seismic loss risks to buildings in the context of rapid socioeconomic growth: a case study in Tangshan, China. Remote Sens 7:2543–2601. DOI:10.3390/rs70302543

Qi, W., Su, G., Sun, L., Yang, F., Wu, Y. (2017). Internet+ approach to mapping exposure and seismic vulnerability of buildings in a context of rapid socioeconomic growth: a case study in Tangshan, China. Natural Hazards, 86(1): 107-139.

NOTES: Some statements are taken literally from the authors. In the column “Data”, the data geometric resolution is in brackets. Accuracy refers to global or overall accuracy. MAPE stands for Mean Absolute Percentage Error. MBT stand for Model Building Type, as a synonym of SBSTs (Seismic Building Structural Types). Abbreviations: attr. for attributes; bldgs. for buildings; DB for database.

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery					
Wieland et al. (2012a)	Time-series of Landsat TM (30 m) and MSS (60 m); QuickBird (0.61 - 2.4 m); manual digitization.	Date of buildings with change-detection of Landsat time series. OBIA of Landsat image for classification of composition of buildings, streets and open spaces. Footprints from QuickBird image analysis: segmentation + post-processing for over-segmentation solution + SVM for classification (building, other sealed surfaces, bare soil, vegetation and shadows).	Attr.: 26 features for OBIA of Landsat img. MBT: 3 masonry types of 1-2 stories (detached/attached 2 sides/attached 3 sides); 2 brick-concrete types (1-2 stories; 3-6 stories); 1 concrete frame (7-9 stories); 1 mixed.	City of Bishkek, Kyrgyzstan. 15098 buildings digitized for verification of the number of buildings.	Accuracy of OBIA of Landsat is 81%. Footprints from QuickBird (built-up mask): Accuracy = 90%.
Dell'Acqua et al. (2013)	QuickBird image.	Design of BREC-4-GEM software for extracting the number of buildings (footprints) from VHR satellite images for the GEM initiative.		City of Pylos, Greece	Buildings are detected with accuracy around 90%

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery (Cont.)					
Erlich et al. (2013)	QuickBird image (0.61 m); DB of damage assessment after the Haiti 2010 EQ; Ground truth footprint DB provided by the municipality.	Alpha-tree of multi-level segmentations (Soille et al. 2007, 2008, 2009). Building segments were selected using parameters related to shape, area, width, length, spectral rage and presence of shadow. Height is estimated from shadow, which was detected by thresholding the histogram of the panchromatic band. <u>NOTE:</u> The method for height estimation from shadow might have problems due to occlusions in densely populated areas and in case of uneven terrain.	Attr. in the exposure DB: footprint and height	Area of 6x8 km ² in Sana'a (Yemen). Ground truth footprint DB was not suitable for height verification. Sample area of 100x100 m ² in Port Prince (Haiti).	In Sana's, 78.86% of agreement between the estimated and reference built-up masks. The height estimation was not validated. In Port Prince, the correlation between the nr. bldgs. estimated with segmentation and that contained in the damage DB is R ² =0.64.

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery and 3D data					
Matsuka et al. (2012)	VHR panchromatic image (25 cm); DSM airborne acquisition (1 m); building DB from in-field survey; footprints available.	Roof type (flat/gable) using panchromatic band and DEM with already existing footprints. Image classification using characteristic scale and spectral features. Building height: using DSM and footprints already available. SVM for BMT classification.	Attr.: Roof type, height, characteristic scale. MBT: vulnerability classification is given in a quantitative scale.	Grenoble, France, 1875 buildings. 402 buildings for training the model	Correlation between estimated vulnerability and ground truth data is 75% for 200 samples in the training dataset.
Polli and Dell'Acqua (2011)	QuickBird panchromatic band; SAR (1 m)	W-filter (implemented in BREC software) for footprint extraction from image. Analysis of two SAR images of the same building for height extraction. SP-BELA SW for vulnerability analysis.	Attr.: Footprint and nr. stories	Some sample zones in Messina, Italy.	Not mentioned
Mück et al (2013)	QuickBird image (0.61 - 2.4 m); DEM (5 m) from Radar; building DB from in-field survey (500 bldgs.); manual digitization.	QuickBird for OBIA and building extraction; DEM for height. Building vulnerability to earthquake and tsunami hazards is considered. Buildings are classified into 4 classes according to the presence of relevant components related to stability, tsunami impact, and accessibility.	Attr.: size, height, shape, orientation, regularity, distance to road. Classes: A,B,C,VE.	Cilacap, Central Java, Indonesia. Digitized 520 bldgs. for footprint verification.	Accuracy of footprint comparison is 84 %. Building classification in testing phase: accuracy = 81%
Wieland et al. (2012b); Pittore and Wieland (2013)	Time-series of Landsat TM (30 m) and MSS (60 m) images; ground-based omnidirectional imaging; OSM footprints; building DB from in-field survey; manual digitization.	OBIA of images for extracting building data at both, block and building levels. Height is extracted from 360° images. Nr. buildings estimated through the building density per stratum.	Attr.: height, age, stratum from Wieland et al. (2012a). MBT: vulnerability classes from EMS-98: A, B, C, D, E, F.	City of Bishkek, Kyrgyzstan Around 900 buildings for MBT verification	Accuracy of OBIA of built-up mask detection ranges between 80%-90%. Validation of height not provided. Visual validation of MBT classification

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery and 3D data (Cont.)					
Geiss et al. (2015)	IKONOS VNIR (1–4 m); Time-series of Landsat TM and ETM+ (30 m); nDSM from Radar (5 m); vulnerability/damage DB from in-field survey of 3896 bldgs.	Calculation of features at both, building and block level. Multi-temporal analysis of Landsat images to extract the period of construction. The inventory DB contains information on geometry, material of structures and walls, foundations and local soil conditions, material of the roof, type of building, quality indicators, damage level due to previous earthquakes, etc. Also, a vulnerability allocation is done following a (1) scoring method (expert knowledge), (2) the EMS-98 scale, and (3) MBT. Damage calculation is performed using fragility functions derived with data from the 2009 earthquake.	Attr.: (1 st group) each bldg. has 132 attributes, whereby 73 are related to individual buildings, and 59 to the block. (2 nd group) each bldg. has 145 attributes, whereby 79 are related to individual buildings, and 66 to the block. MBT: (1) Scoring; (2) EMS-98 classes merged into 3; (3) SBSTs: CM, RC high, RC low, SF, TF res, TF non-res, URM.	Padang, Indonesia. (1 st group) 434 buildings. (2 nd group) Only 561 could be linked to a footprint due to GPS inaccuracies, which are used for verification.	MBT from 1 st group: -Compared with expert scoring: MAPE = 10.6%. -Compared with EMS-98, accuracy = 65.4% and kappa=0.36 MBT from 2 nd group: Only visual validation is done for the 561 bldgs. The remaining 2221 bldgs. are validated by searching within a radius of 50m, with accuracy =84%.
Riedel et al. (2014; 2015)	Aerial images (25 cm); DEM from airborne data (1 m); footprints from Cadaster; building DB from in-field survey; INSEE database.	Vulnerability proxy developed with the vulnerability DB. Calculation of 15 morphological indicators using cadastral data (footprints) and Remote Sensing. 3 indicators were identified as the most powerful: width of the enclosing rectangle, shared wall ratio and average distance to neighbours. SVM for creating 2 models: one with all attr., and the other one with only 2 attr., i.e., nr. stories and date, since these are the attr. present in the INSEE DB. The latter also created with association rule learning (ARL). Estimation of vulnerability is extended to the whole country using the INSEE French DB, with both models (SVM and ARL). A damage estimation is done and compared to a reference from the Risk-UE project.	Attr.: date, nr. stories, roof shape (flat or slope), and the 3 morphological indicators. MBT: EMS-98 vulnerability classes	Grenoble, France Vulnerability DB with 3860 buildings and INSEE DB with 6214 buildings. Verification in Nice with reference data from the Risk-UE project.	Accuracy in training phase is 62.4% when only date and nr. stories are used. Maximum accuracy in training phase is 94.5% when all attr. are included. and the 6 vulnerability classes are merged into 3. Accuracy is not provided for the MBT classification of the entire French territory. Accuracy of the damage comparison is not quantified.

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery and 3D data (Cont.)					
Costanzo et al. (2016)	LiDAR and IPERGEO 503-band VNIR images; cadastral data for verification.	DTM and DSM, together with image classification to characterize terrain, buildings and road network. <u>NOTE:</u> roads and concrete roofs are not distinguished from each other.	Attr.: Area, height and roof material.	City of Cosenza, Italy	Roof type classification in testing phase: accuracy = 95% and kappa = 0.92. The rest is not validated.
Geiss et al. (2016)	RapidEye RGB+NIR images (6.5 m); DEM from TanDEM-X (12 m); Time series of Landsat MSS (79 m), TM and ETM+ (both 30 m); Expected damage DB for the district Zeytinburnu; GPS-photos from Google Panoramio.	Multi-temporal analysis of Landsat images to extract the period of construction. Multi-scale segmentation to delineate the urban structures. Direct estimation of damage grades with SVR and training dataset sizes from 145 to 582. Elevation features do not allow estimating damage grades in a viable way. The inclusion of super-segments from the multi-scale segmentation improved the accuracy. Classification of urban structures into 2 classes first (industrial/commercial and detached residential bldgs.), and finally into 3 classes (adding another residential class). <u>NOTE:</u> the analysis is performed at an aggregated urban level, not at building level.	Attr.: spectral, textural, and spatial. 50 features used to learn the best model for damage estimation. All features used to learn the model for urban structure classification. MBT: urban structures	Istanbul, Turkey. District Zeytinburnu.	MAPE of damage assessment is 13%. Accuracy of 3 urban structures classification is 85.6%, and Kappa = 0.77.
Geiss et al. (2017)	WorldView 2 RGB+NIR (1.85 m); OSM complete DB; (census data for population, but it is beyond the interest of this summary table).	Multi-level segmentation and classification with 270 features and 600 samples per class to obtain LULC. Labeled training segments are obtained from the OSM data in conjunction with the satellite imagery. Post-classification refinement was needed. An empirical relation between the nr. buildings in OSM and the segments classified as buildings is calculated for nr. building estimation. A new segmentation is done which includes morphological operators of the panchromatic band.	Attr.: 270 features LULC Classes: buildings, roads, trees, meadow, water, bare soil.	Valparaiso, Chile. 1056 labelled segments to verify the LULC classification. DB of buildings counted manually to verify the nr. buildings.	LULC classification accuracy is 86.7% with kappa=0.81. MAPE of the best Nr. bldg. estimation is 26.21% (with support vector regression)

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery, 3D data and other useful information					
Sarabandi et al. (2008)	QuickBird images (0.72 m); Tax assessor database; building DBs from in-field surveys; manual digitization.	<p>Rational Polynomial Coefficients (RPC) are used to obtain the building height by measuring the image coordinates of the ground and roof points of a building corner.</p> <p><u>NOTE:</u> This method might have problems due to occlusions in densely populated areas.</p> <p>Footprint digitization for extraction of geometric and shape-related attributes.</p> <p>Multinomial logistic regression for MBT classification.</p> <p><u>NOTE:</u> occupancy and date of construction are powerful attributes that comes from the ancillary database (tax assessor).</p>	<p>Dataset A (tax assessor DB): structural type, occupancy type, date, height, area.</p> <p>Dataset B (tax assessor DB + attr. from remote sensing): structural type, occupancy type, date + height, area, roof type, plan configuration.</p> <p>MBT: C, C/S, RM, S, URM, W.</p>	<p>Sample of 23 buildings in London for height estimation.</p> <p>Two datasets collected for Southern California, USA (A: 38135 bldgs.; B: 1947 bldgs.) for MBT classification.</p>	<p>Accuracy of height extraction by comparison with independent data survey: MAPE = 1.42%.</p> <p>Accuracy in testing phase for 6 MBT classification in dataset A is 63%. In dataset B, it is 84%.</p>
Borfecchia et al. (2010)	LiDAR (4p/m ²); AISA-Eagle RGB-NIR images (2.3 m); Landsat ETM+ (30 m); QuickBird (2.8 m); building DB from in-field survey; census data.	<p>Footprints from local cartography 1:2000. LiDAR for nr. stories, geometric attributes of buildings and presence of vegetation. AISA-Eagle images RGB-NIR for spectral signatures of roof inside the footprints. Landsat ETM+ and QuickBird images to obtain the date of construction (> or < 2000) for RC buildings using change detection.</p> <p>Data mining: ANN, KNN and decision tree with all features (mean and sd of 31 bands of AISA img. + LiDAR-related attr. + age + census + etc.).</p> <p><u>NOTE:</u> data from statistic office (ISTAT).</p> <p>+Info: Cost of LiDAR and AISA airborne data: 1000€/km²</p>	<p>Attr: Footprint area and perimeter, height, volume, altitude, roof type and complexity, roof spectral signature, age, census data...</p> <p>MBT: RC-precode, RC-highcode, Masonry. Characterization through Iv.</p>	<p>Avellino municipality, Italy.</p> <p>In-field ground truth data with 233 labelled buildings: 38 of RC-precode, 189 of RC-highcode, 6 of Masonry. Predictive model created with 60 buildings for training and 173 for testing.</p>	<p>Accuracy of built-up areas: before 2000 is 84.5% for buildings and 15.5% for other classes; after 2000 is 70.5% for buildings and 29.5% for others.</p> <p>Accuracy in testing phase for 3 MBT classification with ANN is 85%. If RC classes are grouped, then accuracy for 2 MBT is 90%.</p>

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
Optical imagery, 3D data and other useful information (Cont.)					
Wu et al. (2014)	WorldView-2 images (0.5-2 m); building DB from in-field survey; vulnerability database.	OBIA for building detection (shape and shadow). Height extracted from shadow. Building structure parameters from field survey: use + structure type + geology + seismic intensity. SVM to classify buildings into damage grades using all previous attributes. <u>NOTE:</u> All 48 buildings are isolated → favorable for shadow detection. <u>NOTE:</u> Structure type from in-field survey.	Shape, height, use, structure type, geology, seismic intensity. Damage grades from DG1 to DG5	Wuhan Optics Valley, China. 48 bldgs. in total. 35 bldgs. for training the model.	Bldg. detection accuracy from OBIA is 81%. Height error < 2 m.
Su et al. (2015)	VHR Optical images; building-related local knowledge (Br-LK); building DB from in-field survey; manual digitization.	Scenario that follows clear rules. Digitization of footprints for area extraction. Algorithm for height extraction from optical images by measuring the building SIPD (shadow length). The structural type attribute comes from a simple rule based on height: if 1 story → S/B; if 2-6 stories → B/C; if +7 stories → R/C. <u>NOTE:</u> The method for height estimation from shadow might have problems due to occlusions in densely populated areas and in case of uneven terrain.	Attr.: Footprint area, nr. stories. MBT: S/B (1-story Bricks); B/C (1-4-story Bricks and Concrete); RC (2-7 story Reinforced Concrete)	Tangshan, China. In-field survey to create a DB of 954 bldgs.	MAPE of height is 3.23%. Estimated area is 2.99% lower than real. Accuracy of 3 MBT classification using rules is 96% and kappa = 0.88
Qi et al. (2017)	Google Earth images; Tencent/Baidu Street View; WeChat; VGI; Br-LK; Real state databases; institutional websites; local informants providing in-field data; manual digitization.	Improvement of previous study by Su et al (2015). Digitization of footprints using Google Earth images. Algorithm from Su et al. (2015) for height extraction from SIPD. The structural type attribute comes from a simple rule based on height: if +7 stories → R/C. For shorter buildings, authors incorporate a great amount of information from internet sources and in-field informants in order to improve their previous classification (Su et al., 2015). <u>NOTE:</u> 530 out of 532 blocks in residential areas had homogeneous structural type, provided by real state agency.	Attr.: Footprint area, nr. stories, structural type. MBT: S/B (1-story Bricks); B/C (1-4-story Bricks and Concrete); RC (2-7 story Reinforced Concrete).	Tangshan, China. Exposure test DB (to verify area and height): 971 bldgs. DB to verify structural type: 1016 buildings.	MAPE of area is 4.64%. Accuracy of nr. stories is 89% and kappa = 0.86 Accuracy of 3 MBT classification using rules is 98% and kappa = 0.94

Author	Data	Comments to the main topics addressed	Attributes used and MBT identified	Nr. Buildings or study area covered	Accuracy or reliability
		<p><u>NOTE:</u> the increment in accuracy and kappa is slight with respect to previous study (Su et al., 2015).</p> <p>+Info: digitization of 1200 building footprints in 4-5 hours, by 1 skilled person (12-15 sec. per building).</p> <p>+Info: characterization of buildings in 10 working days, by 2 skilled people, covering 456 km².</p> <p>+Info: in-field survey for collecting data from 115 buildings took 3 skilled people a total of 3 days.</p>	Division into urban and rural strata.		

Anexo III

Informes de los modelos predictivos generados para clasificar las MBT de Lorca.

- *Gráficos de los árboles de decisión.*
- *Gráficos de las redes Bayesianas (TAN).*
- *Variables y coeficientes de las regresiones logísticas.*

ÁRBOLES DE DECISIÓN

ESTRATO DE ENSANCHE

ESTRATOS DE CENTRO

REDES BAYESIANAS (TAN)

Se muestra el gráfico con las relaciones entre atributos, pero, por razones de espacio, no se incluyen las tablas de probabilidad condicionada en cada nodo, ya que son demasiado extensas.

ESTRATO DE ENSANCHE

ESTRATOS
DE CENTRO

Variables y coeficientes de las REGRESIONES LOGÍSTICAS calculadas
para predecir la MBT de los edificios de Lorca.

Estrato	Coeficientes				
	Nº Plantas	Tejado	X_Centroide	Y_Centroide	Cte.
Ensanche	49.926	-	-	-0.024	101130.986
Centro	1.305	1.184	0.005	-	-2845.747

Anexo IV

Comparación de los tejados obtenidos en Puerto Príncipe

Figuras de los tejados generados por digitalización (observados), con su verdadera forma, posición y magnitud, sobre los que se superponen los centroides de los tejados obtenidos en el proceso OBIA (estimados). Se compara tanto la posición como el número de tejados y el tipo de material de las cubiertas.

Comparación de los tejados obtenidos en Puerto Príncipe

ESTRATO RESIDENCIAL

Área de muestra 0

Área de muestra 6

Área de muestra 8

Área de muestra 13

Área de muestra 23

ESTRATO URBANO REGULAR

Área de muestra 1

Área de muestra 4

Área de muestra 7

Área de muestra 11

Área de muestra 12

Área de muestra 14

Área de muestra 20

ESTRATO IRREGULAR

Área de muestra 15

Área de muestra 16

Área de muestra 18

Área de muestra 21

ESTRATO RURAL

Área de muestra 5

Área de muestra 19

ESTRATO INFORMAL

Área de muestra 2

Área de muestra 3

Área de muestra 9

Anexo V

Informes de los modelos predictivos generados para clasificar las MBT de Puerto Príncipe.

- *Gráficos de los árboles de decisión.*
- *Variables y coeficientes de las regresiones logísticas.*

ÁRBOLES DE DECISIÓN

ESTRATO
INFORMAL

Variables y coeficientes de las REGRESIONES LOGÍSTICAS calculadas
para predecir la MBT de los edificios de Puerto Príncipe.

Estrato	Coeficientes			
	Tejado	Nº Plantas	Área	Cte.
Residencial	21.650	0.566	0.008	-2.085
Urb. Regular	21.953	0.309	0.001	-1.376
Urb. Irreg.	21.699	0.535	0.005	-1.694
Rural	21.516	0.977	0.002	-1.700
Informal	21.348	0.929	0.005	-1.693

Anexo VI

Comparación de las tipologías constructivas obtenidas en Puerto Príncipe

- *Figuras de las MBT obtenidas para las huellas regulares, con su verdadera forma, posición y magnitud, sobre las que se superponen los centroides de las huellas irregulares con su MBT. Se compara tanto la posición como el número de objetos y la tipología constructiva.*
- *Figuras de las MBT obtenidas en este estudio, tanto regulares como irregulares, sobre las que se superponen los centroides de los edificios recogidos en Corbane et al. (2011) con su tipología constructiva.*

Comparación de las tipologías constructivas identificadas en Puerto Príncipe

Comparación de las MBT obtenidas para huellas regulares e irregulares

ESTRATO RESIDENCIAL

Área de muestra 0

Área de muestra 6

Área de muestra 8

Área de muestra 13

Área de muestra 23

ESTRATO URBANO REGULAR

Área de muestra 1

Área de muestra 4

Área de muestra 7

Área de muestra 11

Área de muestra 12

Área de muestra 14

Área de muestra 20

ESTRATO IRREGULAR

Área de muestra 15

Área de muestra 16

Área de muestra 18

Área de muestra 21

ESTRATO RURAL

Área de muestra 5

Área de muestra 19

Área de muestra 2

Área de muestra 3

Área de muestra 9

Comparación de las MBT obtenidas en este estudio con las observadas en campo por
Corbane et al., (2011)¹

HUELLAS REGULARES

HUELLAS IRREGULARES

Área de muestra 12

Área de muestra 12

Área de muestra 16

Área de muestra 16

Área de muestra 19

Área de muestra 19

¹ Corbane, C., Saito, K., Dell'Oro, L., Bjorgo, E., Gill, E., Piard, B.E., Huyck, C.K., Kemper, T., Lemoine, G., Spence, R.J.S., Shankar, R., Senegas, O., Ghesquiere, F., Lallemand, D., Evans, G.B., Gartley, R.A., Toro, J., Ghosh S., Svekla, W.D., Adams, B.J., Eguchi, R.T. (2011). A comprehensive analysis of building damage in the 12 January 2010 MW7 Haiti earthquake using high-resolution satellite and aerial imagery. Photogrammetric Engineering & Remote Sensing, 77(10): 997-1009.

HUELLAS REGULARES

HUELLAS IRREGULARES

Área de muestra 20

Área de muestra 20

Área de muestra 21

Área de muestra 21

Área de muestra 23

Área de muestra 23

Anexo VII

Formulario de planificación logística de una campaña de campo para recogida de datos de vulnerabilidad sísmica.

Ejemplo de formulario aportado por expertos internacionales en el que se hace una estimación de los recursos materiales, el tiempo y el coste que conlleva la realización de una campaña de campo para recoger datos relacionados con la vulnerabilidad sísmica de 1000 edificios.

IN-FIELD SURVEY COST EVALUATION

PRELIMINARY WORK	OFFICE WORK	Human resources	Number	Time (days)	Unit cost (€)	Total cost	Observations
		Civil engineer	1	10	100	1000	
		Data Scientist	1	15	100	1500	
		Assistant	1	21	50	1050	
		Material resources	Number	-	Unit cost (€)	Total cost	Observations
		PC/laptop	3		600	1800	
		Printer	1		300	300	
		Office supplies	1		400	400	
		Handheld GPS	8		250	2000	
	PARTIAL COST					8050	
IN-FIELD SURVEY	TRAVEL + ACCOMM.	Concept	Number	Time (days)	Unit cost (€)	Total cost	Observations
		Transport	8		900	7200	
		Hotel	8	12	50	4800	
	FIELD WORK	Human resources	Number	Time (days)	Unit cost (€)	Total cost	Observations
		Civil engineer	1	10	200	2000	The professional fee is doubled when working out-of-the-office
		Architect	1	10	200	2000	
		Data Scientist	2	10	200	4000	
		Assistant	2	10	100	2000	
		Local Expert	2	10	80	1600	
		Material resources	Number	-	Unit cost (€)	Total cost	Observations
		Tablet	8		400	3200	
		Handheld GPS	8			0	GPS from preliminary work
		Office supplies	1		400	400	
		Transport mean	2	12	20	480	Car
	OFFICE WORK	Human resources	Number	Time (days)	Unit cost (€)	Total cost	Observations
		Civil engineer	1	2	200	400	
		Architect	1	2	200	400	
		Data Scientist	2	2	200	800	
		Assistant	2	2	100	400	
		Local Expert	2	2	80	320	
		Material resources	Number	-	Unit cost (€)	Total cost	Observations
		PC/laptop	2		600	1200	4 laptops from preliminary work
		Printer	1			0	1 printer from preliminary work
		Office supplies	1		100	100	
	PARTIAL COST					31300	
POST PROCESSING	OFFICE WORK	Human resources	Number	Time (days)	Unit cost (€)	Total cost	Observations
		Civil engineer	1	15	100	1500	
		Data Scientist	2	15	100	3000	
		Assistant	2	20	50	2000	
		Material resources	Number	-	Unit cost (€)	Total cost	Observations
		PC/laptop	5			0	5 laptops already counted
		Printer	1			0	1 printer already counted
		Office supplies	1		200	200	
	PARTIAL COST					6700	
TOTAL COST					46050		

CONCEPTS

Human resources	Material resources	Travel
Civil engineer (senior)	PC/laptop	Transport
Architect (senior)	Printer	Hotel
Data Scientist (senior)	Tablet	Other
Assistant/student (junior)	Handheld GPS	
Local Expert	Smart phone	
In-field assistant/logistics	Camera	
Other	Office supplies	
	Means of transportat.	
	Other	

ÍNDICE DE TABLAS

Tabla 2-1. Resumen comparativo de estudios destinados al análisis de vulnerabilidad sísmica con técnicas de teledetección en los últimos 10 años.....	42
Tabla 5-1. Distribución de MBT en las áreas de estudio. Datos tomados de la base de datos de referencia.....	80
Tabla 5-2. Combinación de parámetros del algoritmo TOD para clasificar el LiDAR en las distintas áreas de estudio.	87
Tabla 5-3. Valores umbrales de los indicadores de geometría utilizados para refinar la selección de huellas de edificios.	87
Tabla 5-4. Comparación del número de huellas obtenidas por procedimientos semiautomáticos usando datos tomados por sensores remotos, con el número de huellas reales contenidas en la base de datos de referencia.	88
Tabla 5-5. Número de muestras en los datasets de entrenamiento y verificación para la clasificación del tipo de tejado.....	90
Tabla 5-6. Ranking de los modelos predictivos para clasificar tejados que mayor precisión han alcanzado en la fase de entrenamiento.....	91
Tabla 5-7.Resultado de la verificación del atributo área.	93
Tabla 5-8. Matrices de confusión obtenidas en la verificación de la clasificación del tipo de tejado.....	96
Tabla 5-9. Ranking de los modelos predictivos creados para estimar el tipo de tejado en cada área de estudio.	96
Tabla 5-10. Número de muestras en los datasets de entrenamiento y verificación para la clasificación de la vulnerabilidad.	99
Tabla 5-11. Ranking de los modelos predictivos que mayor precisión han alcanzado en la fase de entrenamiento.	100
Tabla 5-12. Matrices de confusión obtenidas en la verificación de la clasificación de vulnerabilidad en la zona Centro, y métricas de precisión.....	101
Tabla 5-13. Estimación de tiempos para la realización de un análisis de vulnerabilidad por teledetección (a partir del estudio en Lorca)	106
Tabla 5-14. Estimación de costes para la realización de un análisis de vulnerabilidad por teledetección (a partir del estudio en Lorca)	106
Tabla 5-15. Parámetros del escenario sísmico simulado en Lorca.....	¡Error! Marcador no definido.
Tabla 5-16. Resultados del cálculo de daños en el escenario de Lorca.....	108
Tabla 5-17. Tipologías constructivas de Puerto Príncipe. Tomada de Torres et al. (2016).....	118
Tabla 5-18. Bases de datos de entrenamiento creadas para clasificar los patrones urbanos de Haití.	127
Tabla 5-19. Métricas de calidad de la estratificación del escenario de Haití para cada dataset.....	133
Tabla 5-20. Combinación de parámetros del algoritmo TOD para clasificar el LiDAR en los distintos estratos.....	135
Tabla 5-21. Media del coeficiente de silueta calculado para las distintas clasificaciones.....	136
Tabla 5-22. Informe de digitalización de huellas de edificios.....	139
Tabla 5-23. Parámetros de segmentación para P. Príncipe	140
Tabla 5-24. Datasets de entrenamiento y verificación para la detección y clasificación de huellas de edificios.....	141
Tabla 5-25. Comparación del número de huellas obtenidas por OBIA con el número de huellas digitalizadas	143
Tabla 5-26. Comparación del área construida (m ²) obtenida por OBIA con el número de huellas digitalizadas.....	147

Tabla 5-27. Matriz de confusión y métricas de calidad de la clasificación del tipo de tejado	149
Tabla 5-28. Número de muestras en los datasets de entrenamiento y verificación para la clasificación de vulnerabilidad.	152
Tabla 5-29. Precisiones para RC y MAndW obtenidas con los datasets desequilibrados	152
Tabla 5-30. Resumen de los modelos predictivos creados para clasificar los edificios en MBT.....	153
Tabla 5-31. Ranking de los modelos predictivos creados para cada estrato ordenados según las métricas obtenidas en fase de verificación.....	155
Tabla 5-32. Matrices de confusión de la fase de verificación de la clasificación de MBT en Puerto Príncipe	156
Tabla 5-33. Matrices de confusión y métricas de calidad de la verificación de las MBT	157
Tabla 5-34. Resumen de la extrapolación con Bootstrap para la B.D. del MTPTC	162
Tabla 5-35. Número de edificios de cada tipología estimado para el área total de los estratos	162
Tabla 5-36. Estimación de tiempos para la realización de un análisis de vulnerabilidad por teledetección (a partir del estudio en Puerto Príncipe).....	163
Tabla 5-37. Parámetros del escenario sísmico simulado en Puerto Príncipe	165
Tabla 5-38. Resultados de daño en las áreas de muestra comunes con Corbane et al (2011)	166
Tabla 5-39. Número de edificios por tipología en los distritos 23 y 29	171
Tabla 5-40. Número de edificios inhabitables calculado en el escenario de Puerto Príncipe con las huellas regulares (H. Digit.) e irregulares (H. Segm.).....	171
Tabla 6-1. Logística de una campaña de campo para toma de datos sobre edificios en un país desarrollado. Resumen de la información proporcionada por expertos	183
Tabla 6-2. Logística de una campaña de campo para toma de datos sobre edificios en un país en desarrollo. Resumen de la información proporcionada por expertos.	184

ÍNDICE DE FIGURAS

Figura 1-1. Número de terremotos, tormentas, temperaturas extremas e inundaciones ocurridos entre 1996 y 2015.	17
Figura 1-2. Definición de riesgo sísmico	21
Figura 1-3. Esquema del cálculo de los indicadores de riesgo sísmico	22
Figura 2-1. Comparación de varias taxonomías de clasificación de tipologías edificatorias.....	37
Figura 4-1. Esquema metodológico.....	56
Figura 4-2. Clasificador SVM (Support Vector Machine).	60
Figura 4-3. Esquema de un árbol de decisión (A) y de una red Bayesiana (B).....	71
Figura 5-1. (A) Derrumbamiento de la nave de crucero de la iglesia de Santiago. (B) Edificio colapsado en la calle Infante Juan Manuel.	77
Figura 5-2. Zonas de estudio en la ciudad de Lorca.	79
Figura 5-3. Construcciones típicas en Lorca.....	80
Figura 5-4. Proceso metodológico aplicado en el escenario de Lorca.....	83
Figura 5-5. Patrones urbanos identificados en la ciudad de Lorca.....	84
Figura 5-6. Resultado de la estratificación de Lorca en patrones urbanos.	85
Figura 5-7. Proceso para selección de segmentos correspondientes a edificios.	87
Figura 5-8. Ejemplos de segmentos seleccionados erróneamente como edificios.	88
Figura 5-9. Ejemplos de huellas de edificios obtenidas por segmentación.....	89
Figura 5-10. Ejemplo de edificio con los atributos calculados en el SIG.....	92
Figura 5-11. Huellas irregulares obtenidas para cada zona de estudio.....	94
Figura 5-12. Histograma de los residuos del número de plantas obtenidos para cada zona de estudio.	95
Figura 5-13. Resultados de la obtención del atributo “material del tejado” en comparación con la base de datos de referencia.	97
Figura 5-14. Resultados de la obtención de la tipología (MBT) en comparación con la base de datos de referencia.	101
Figura 5-15. Distribución de vulnerabilidad en Lorca	105
Figura 5-16. Escenario sísmico simulado en Lorca	108
Figura 5-17. Gráfica comparativa de los daños totales.....	109
Figura 5-18. Gráfica comparativa de los daños obtenidos desagregados por tipología.....	109
Figura 5-19. Gráfica comparativa de los daños obtenidos desagregados por número de plantas.....	110
Figura 5-20. Resumen de las métricas de calidad del procedimiento aplicado en Lorca	111
Figura 5-21. Ejemplos de prácticas constructivas típicas de Haití.	116
Figura 5-22. Construcciones típicas en Puerto Príncipe.	119
Figura 5-23. Áreas de estudio para la aplicación de Puerto Príncipe.	120
Figura 5-24. Evolución de la información geográfica digitalizada por HOT en Haití tras el terremoto (Waters, 2010)	121

Figura 5-25. Ubicación de los edificios (A) de la base de datos del MTPTC de Haití y (B) de Corbane et al. (2011).	123
Figura 5-26 Proceso metodológico aplicado en el escenario de Puerto Príncipe.	124
Figura 5-27. Patrones urbanos identificados en la ciudad de Puerto Príncipe.	126
Figura 5-28. Detalle de la segmentación en P. Príncipe.	127
Figura 5-29. Muestras cuadradas y segmentos en los datasets de entrenamiento.	128
Figura 5-30. Curvas ROC obtenidas para diferentes conjuntos de atributos, con muestras cuadradas (A) y con segmentos (B).	129
Figura 5-31. Curvas de aprendizaje del clasificador para diferentes tamaños muestrales.	130
Figura 5-32. Curvas ROC y métricas de precisión para 8 datasets distintos. Tomado de Wieland et al. (2016).	132
Figura 5-33. Resultado de la estratificación del área de estudio en patrones urbanos.	134
Figura 5-34. Ejemplos de LiDAR clasificado en segmentos de los diferentes patrones urbanos.	135
Figura 5-35. Métricas de paisaje obtenidas para los cinco patrones urbanos (1 a 5) con las seis clasificaciones calculadas y la base de datos de referencia (Ref label).	136
Figura 5-36. Patrones urbanos de Puerto Príncipe y áreas de muestra.	138
Figura 5-37. Ejemplos de resultados de clasificación con las 9 bandas.	142
Figura 5-38. Áreas de muestra excluidas del estudio.	142
Figura 5-39. Coeficiente de comparación del número de huellas regulares e irregulares.	144
Figura 5-40. Ejemplos de huellas obtenidas en diferentes patrones urbanos.	144
Figura 5-41. Atributos de la base de datos de exposición. Ejemplos de estrato Residencial e Informal.	145
Figura 5-42. Coeficiente de comparación del área construida de las huellas regulares e irregulares.	147
Figura 5-43. Principales fuentes de discrepancia en la estimación del área construida.	148
Figura 5-44. Comparación de tejados digitalizados (A) y segmentados (B).	150
Figura 5-45. Histograma de los residuos del número de plantas.	151
Figura 5-46. Edificios de tipología MAndW (A) y RC (B) con tejado metálico.	156
Figura 5-47. Fotografías de edificios con errores de clasificación.	158
Figura 5-48. Resultados de la clasificación de huellas regulares e irregulares en tipologías constructivas.	159
Figura 5-49. Distribución de vulnerabilidad en Puerto Príncipe.	160
Figura 5-50. Círculos creados para el Bootstrap.	161
Figura 5-51. Superposición de los estratos y los distritos de Puerto Príncipe.	165
Figura 5-52. Escenario sísmico simulado en Puerto Príncipe.	165
Figura 5-53. Gráfica comparativa de los daños totales.	166
Figura 5-54. Gráfica comparativa de los daños obtenidos por tipología.	167
Figura 5-55. Gráfica comparativa de los daños obtenidos desagregados por altura.	167
Figura 5-56. Estimación de daños empleando el área construida.	168
Figura 5-57. Gráficas comparativas de daños totales para toda la ciudad de Puerto Príncipe.	168
Figura 5-58. Gráficas comparativas de daños por tipología para toda la ciudad de Puerto Príncipe.	169

Figura 5-59. Gráficas comparativas de daños desagregados por altura para toda la ciudad de Puerto Príncipe	169
Figura 5-60. Coeficiente de comparación del daño obtenido para huellas regulares e irregulares por estrato (A) y por distrito (B).....	170
Figura 5-61. Mapas de edificios inhabitables.....	172
Figura 5-62. Resumen de las métricas de calidad del procedimiento aplicado en Puerto Príncipe.....	174
Figura 6-1. Desglose de costes de una campaña de campo.....	186