



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Estadística e Investigación Operativa  
Aplicadas y Calidad

Análisis y Modelización de la Demanda de Vehículos de  
Emergencia Sanitaria en la Provincia de Valencia utilizando  
Análisis de Procesos Puntuales

Trabajo Fin de Máster

Máster Universitario en Ingeniería de Análisis de Datos, Mejora de  
Procesos y Toma de Decisiones

AUTOR/A: Ródenas Moreno, Esther

Tutor/a: Villa Juliá, María Fulgencia

Cotutor/a: Sánchez Anguix, Víctor

Director/a Experimental: Karpova Krylova, Yulia

CURSO ACADÉMICO: 2023/2024



---

## Agradecimientos

En primer lugar, me gustaría agradecer a la fundación ValgrAI (Valencian Graduate School and Research Network of Artificial Intelligence) por concederme la beca que ha hecho posible mi participación en el Máster Universitario en Ingeniería de Análisis de Datos, Mejora de Procesos y Toma de Decisiones en la Universidad Politécnica de Valencia.

En segundo lugar, a todas aquellas personas que me han ayudado en la creación de este trabajo. Gracias por la paciencia al responder mis preguntas y por las maravillosas ideas que me habéis brindado.

En tercer lugar, a mi familia, por el apoyo incondicional que me dan. Gracias por acompañarme en los momentos difíciles y por permitirme siempre seguir mis sueños.

Finalmente pero no menos importante a mi tutora Ful, gracias por ser un modelo a seguir, por la confianza que depositas en mí y por haberme abierto las puertas a este camino tan maravilloso. También a mis tutores Víctor y Yulia, trabajar con vosotros ha sido muy enriquecedor y estoy expectante de ver lo que nos deparará el futuro.

---

## Resumen

Responder de forma rápida a llamadas de emergencia que requieran la asistencia de un Vehículo de Emergencia Sanitaria (VES) es crucial para la salud y el bienestar de la población. No obstante, la gestión y despliegue de los VES requiere de un proceso de toma de decisiones que no es nada trivial. Es, por tanto, que la anticipación y conocimiento del comportamiento de la demanda de VES puede ser interesante para mejorar la calidad del servicio prestado por parte de las entidades a su cargo. Se dispone de una base de datos con más de 98.000 llamadas de emergencia que necesitaron un VES en la provincia de Valencia durante el año 2019. El objetivo de este trabajo es focalizar el estudio en la información aportada por las coordenadas, longitud y latitud, donde ocurrieron esos eventos. Se desea analizar cómo se han distribuido esos puntos en el espacio utilizando las técnicas aportadas en el campo de análisis de procesos puntuales y se proponen distintos modelos para estimar su función de intensidad desde dos perspectivas diferentes: frecuentista y bayesiana. Este proceso se realiza con el software estadístico R.

**Palabras clave:** Estadística espacial, Demanda de VES, Análisis de procesos puntuales, Clustering, Patrones de puntos espaciales, Programa R.



---

## Resum

Respondre de manera ràpida a trucades d'emergència que requerisquen l'assistència d'un Vehicle d'Emergència Sanitària (VES) és crucial per a la salut i el benestar de la població. No obstant això, la gestió i desplegament dels VEUS requerix d'un procés de presa de decisions que no és gens trivial. És, per tant, que l'anticipació i coneixement del comportament de la demanda de VES pot ser interessant per a millorar la qualitat del servei prestat per part de les entitats al seu càrrec. Es disposa d'una base de dades amb més de 98.000 trucades d'emergència que van necessitar un VEUS a la província de València durant l'any 2019. L'objectiu d'este treball és focalitzar l'estudi en la informació aportada per les coordenades, longitud i latitud, on van ocórrer eixos esdeveniments. Es desitja analitzar com s'han distribuït eixos punts en l'espai utilitzant les tècniques aportades en el camp d'anàlisi de processos puntuals i es proposen diferents models per a estimar la seua funció d'intensitat des de dos perspectives diferents: frequentista i bayesiana. Este procés es realitza amb el programari estadístic R.

**Paraules clau:** Estadística espacial, Demanda de VES, Anàlisi de processos puntuals, Clustering, Patrons de punts espacials, Programa R.

---

## Abstract

Fast response to emergency calls requiring the assistance of Emergency Medical Services (EMS) is crucial for the health and well-being of the population. However, the management and deployment of EMS vehicles require a non-trivial decision-making process. It is, therefore, of interest to anticipate and understand the behaviour of the demand for EMS to improve the quality of the service provided by the entities in charge of them. A database with more than 98,000 emergency calls is available that required an EMS deployment in the province of Valencia during 2019. This work aims to focus the study on the information provided by the coordinates, longitude, and latitude at which these events occurred. The aim is to analyse how these points have been distributed in space using the techniques provided in point processes analysis and several models are proposed to estimate the intensity function from two perspectives: frequentist and bayesian. This project is developed using statistical analysis software R.

**Key words:** Spatial statistics; EMV Demand; Point Process Analysis; Clustering; Spatial Point Patterns; R Software.

# Índice general

<b>Agradecimientos</b>	<b>I</b>
<b>Resumen</b>	<b>II</b>
<b>Resum</b>	<b>III</b>
<b>Abstract</b>	<b>IV</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS) . . . . .	5
1.3. Objetivos . . . . .	5
1.4. Estructura del documento . . . . .	5
<b>2. Marco teórico</b>	<b>7</b>
2.1. Problemas de localización y cubrimiento . . . . .	7
2.2. Estadística Espacial . . . . .	7
2.2.1. Geoestadística . . . . .	8
2.2.2. Datos agregados o en redes de localizaciones . . . . .	8
2.2.3. Procesos puntuales . . . . .	8
2.3. Inferencia Bayesiana . . . . .	10
<b>3. Revisión Bibliográfica</b>	<b>11</b>
3.1. Estudio de la Demanda . . . . .	11
3.2. Predicción de la Demanda . . . . .	12
3.3. Análisis de Procesos Puntuales . . . . .	12
3.4. Aportación de este trabajo . . . . .	13
<b>4. Descripción del Problema</b>	<b>14</b>
4.1. Base de datos . . . . .	14
4.2. Localización y ventana $W$ . . . . .	17
4.3. Consideraciones iniciales . . . . .	18
4.4. Aleatoriedad Espacial Completa (CSR) . . . . .	19
4.5. Software utilizado . . . . .	20
<b>5. Perspectiva Frecuentista</b>	<b>22</b>
5.1. Análisis Exploratorio de los Datos . . . . .	22
5.1.1. Cuestiones Fundamentales . . . . .	23
5.1.2. Intensidad . . . . .	23
5.1.3. Correlación . . . . .	27
5.1.4. Espaciado . . . . .	28
5.2. Ajuste de Modelos . . . . .	31

---

5.2.1. Modelos de Poisson . . . . .	31
5.2.2. Validación del Modelo . . . . .	35
<b>6. Perspectiva Bayesiana</b>	<b>39</b>
6.1. Preparación de los datos . . . . .	39
6.2. Mallado . . . . .	40
6.3. Matriz Matérn . . . . .	40
6.4. Modelo y Resultados . . . . .	42
<b>7. Conclusiones y futuras líneas de investigación</b>	<b>45</b>
7.1. Conclusiones . . . . .	45
7.2. Futuras líneas de investigación . . . . .	46
<b>Bibliografía</b>	<b>47</b>
<b>A. Relación del Trabajo con los Objetivos de Desarrollo Sostenible de la Agenda 2030</b>	<b>52</b>
<b>B. Instrucciones de R que se han utilizado en el trabajo.</b>	<b>54</b>

# Índice de figuras

1.1. Evolución del número de demandas asistenciales de urgencias y emergencias 112/061 en España.	2
1.2. Estructura de iReves . . . . .	3
1.3. Recursos desplegados ante emergencias. . . . .	4
1.4. ODS a los que contribuye este trabajo . . . . .	5
2.1. Distribución espacial de tres especies de hongos alrededor de un árbol. . . . .	9
4.1. Histogramas para análisis exploratorio de las variables. . . . .	16
4.2. Recursos desplegados ante emergencias. . . . .	17
4.3. Ventanas separadas por territorio. . . . .	18
4.4. Ventana $W$ de la región de estudio. . . . .	18
4.5. Proceso homogéneo de Poisson generado artificialmente. . . . .	19
5.1. Esquema de la metodología seguida con la metodología de perspectiva frecuentista. . . . .	22
5.2. Comparación de aleatoriedad espacial completa (CSR). . . . .	24
5.3. Análisis de cuadrantes. . . . .	25
5.4. Funciones Kernel. . . . .	26
5.5. Densidad kernel adaptativa. . . . .	26
5.6. Representación de las funciones $K$ y $L$ obtenidas en R. . . . .	28
5.7. Representación de las funciones $G$ , $F$ y $J$ obtenidas en R. . . . .	30
5.8. Resultado del modelo $M_0$ . . . . .	31
5.9. Resultado del modelo $M_1$ . . . . .	32
5.10. Resultado del modelo $M_2$ . . . . .	32
5.11. Resultado del modelo $M_3$ . . . . .	33
5.12. Resultado del test ANOVA entre los modelos $M_0$ y $M_1$ . . . . .	33
5.13. Resultado del test ANOVA entre los modelos $M_0$ y $M_2$ . . . . .	34
5.14. Resultado del test ANOVA entre los modelos $M_0$ y $M_3$ . . . . .	34
5.15. Resultado del test ANOVA entre los modelos $M_1$ y $M_3$ . . . . .	34
5.16. Resultado del test ANOVA entre los modelos $M_2$ y $M_3$ . . . . .	34
5.17. Estimación de la función $\rho$ . . . . .	36
5.18. Gráfico de <i>lurking</i> para la covariable densidad de población. . . . .	37
5.19. Gráfico Q-Q de los residuos suavizados para el modelo $M_3$ . . . . .	37
6.1. Esquema de la metodología seguida con la metodología de perspectiva bayesiana. . . . .	39
6.2. Representación de las llamadas de emergencias en la Provincia de Valencia. . . . .	40
6.3. Representación de las llamadas de emergencias en la Provincia de Valencia junto con el mallado. . . . .	41
6.4. Predicción de las llamadas de emergencias en la Provincia de Valencia. . . . .	42
6.5. Representación de la mediana y superficies de densidad. . . . .	43
6.6. Funciones de densidad de probabilidad del rango, la desviación típica y el parámetro independiente. . . . .	43

# Índice de tablas

4.1. Variables disponibles en la base de datos de emergencias para el año 2019. . . . .	15
5.1. Resumen de los estadísticos para cada modelo. . . . .	35
6.1. Resumen de los valores seleccionados a priori. . . . .	41
6.2. Resumen de la predicción de llamadas de emergencia obtenidas. . . . .	44
A.1. Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS). . . . .	52
B.1. Instrucciones de R que se han utilizado en el trabajo. . . . .	55



# Capítulo 1

## Introducción

### 1.1. Introducción

La salud y el bienestar de la población son fundamentales para asegurar la prosperidad de una sociedad. Gracias a los avances médicos y tecnológicos, tanto la esperanza como la calidad de vida de las personas han ido aumentando. Sin embargo, existen desigualdades severas entre países que ponen en riesgo el acceso a un derecho humano fundamental como es el derecho a la salud (WHO 2023). Para cumplir con los objetivos de la Agenda 2030 para el Desarrollo Sostenible, la Organización Mundial de la Salud (WHO) propone un sistema de atención sanitaria universal (UHC) para que cualquier persona, sin importar donde viva o su condición, pueda recibir servicios sanitarios de calidad donde y cuando lo requiera sin tener que endeudarse.

En España se dispone de un Sistema Nacional de Salud (SNS) cuya gestión es responsabilidad del Ministerio de Sanidad y que tiene un gasto de 132.000 millones de euros al año (SNS 2024). Este servicio asegura una cobertura pública completa para todas las personas residentes en España. Este gasto es financiado gracias a los impuestos, por lo que su presupuesto es limitado y es de suma importancia que los departamentos que lo gestionan optimicen sus recursos y den el mejor servicio posible a la población.

La atención sanitaria del SNS español está dividida en diversos servicios: actividades preventivas, atención primaria, atención hospitalaria, atención urgente y morbilidad atendida. Este Trabajo de Fin de Máster pone su atención en la **atención urgente**, más concretamente, en **los servicios de urgencias y emergencias 112/061**, que alcanzaron una demanda de 8.670.565 llamadas en el año 2021 y tiene una tendencia al alza, tal y como se observa en la Figura 1.1. Según los últimos datos pertenecientes al año 2022, para atender a esta demanda se disponen de 3.086 ambulancias en toda España, lo que se traduce en 6,5 ambulancias por cada 1.000 habitantes. No obstante, estas ambulancias están repartidas por comunidades y son los gobiernos autonómicos los que se encargan de su gestión.

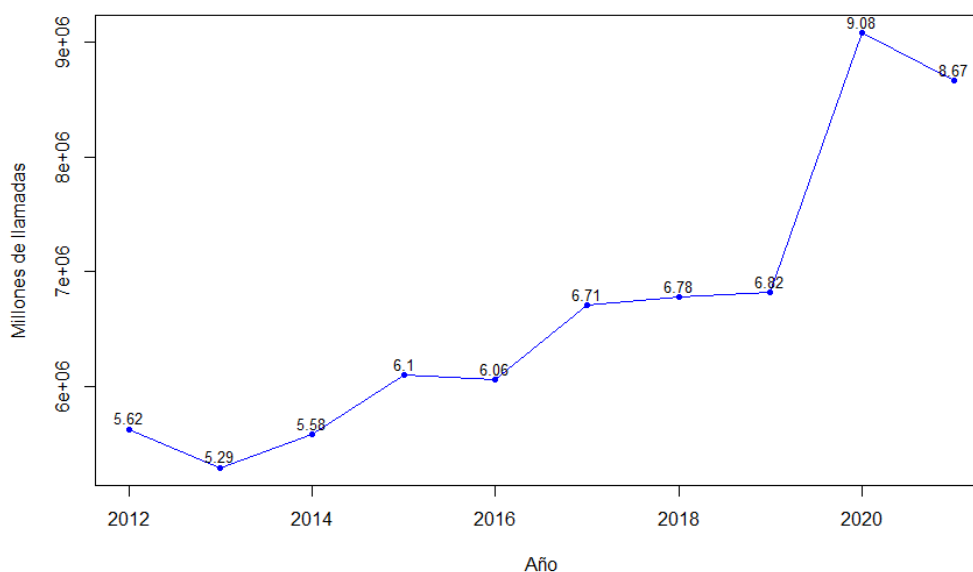
El organismo encargado de la gestión sanitaria en la Comunidad Valenciana es la *Conselleria de Sanitat Universal i Salut Pública* y dispone de 579 vehículos de emergencia sanitaria (VES) que corresponden a 11,7 VES por cada 1.000 habitantes. Para gestionar estos VES, se creó a principios de los 90 el Servicio de Emergencias Sanitarias de la Comunitat Valenciana (SESCV) que divide sus recursos en 10 categorías tal y como se observa en la Figura 1.3 (CV s.f.):

- **Centro de Información y Coordinación de Urgencias (CICU):** lugar donde se gestiona la demanda de llamadas de urgencias y emergencias realizadas al 112 CV y se coordinan los recursos disponibles para responder a las necesidades de los pacientes.
- **Helicópteros Medicalizados:** aeronaves utilizadas para transportar a pacientes que requieren soporte vital avanzado.
- **Servicio de Ayuda Médica Urgente (SAMU):** ambulancia medicalizada que dispone de personal médico,



de enfermería y técnico atendiendo emergencias sanitarias.

- **Soporte Vital Avanzado (SVA) de Enfermería:** ambulancia medicalizada que dispone de profesionales de la salud para atender a pacientes en estado crítico.
- **Soporte Vital Básico (SVB):** ambulancia medicalizada con menor equipamiento que un SVA, preparada para atender a pacientes con menor prioridad.
- **Ambulancia no asistida:** ambulancia destinada a transportar pacientes.
- **Vehículo de Intervención Rápida (VIR) o Unidades de Soporte y Coordinación (USC):** vehículo que permite prestar la asistencia necesaria en el lugar del suceso, pero que no está capacitada para el transporte de pacientes.
- **Emergencias Múltiples Víctimas (EMV):** vehículos especializados para prestar asistencia en situaciones donde hay múltiples víctimas.
- **Vehículo de Comunicaciones:** vehículos dotados para coordinar emergencias en el lugar del suceso.
- **Incubadora de transporte:** incubadora que se instala en los SAMU para transportar neonatos críticos acompañados por un especialista en pediatría.



**Figura 1.1:** Evolución del número de demandas asistenciales de urgencias y emergencias 112/061 en España.

**Fuente:** Elaboración propia con datos del Informe Anual del Sistema Nacional de Salud 2022

Es importante asegurar un servicio de emergencia sanitaria en un tiempo de respuesta corto, entendiendo como tiempo de respuesta el tiempo que tarda en llegar el VES desde que el usuario ha realizado la llamada, ya que las respuestas rápidas son clave para conseguir buenos resultados clínicos (O’Keeffe et al. 2011 y Blackwell y Kaufman 2002). Idealmente, el CICU dispondría de recursos suficientes para asegurar un tiempo de respuesta bajo en todo el territorio en cualquier momento del día, pero la realidad es que con una demanda con tendencia creciente, el CICU debe de gestionar, de la mejor manera posible, los recursos aquí expuestos para responder a las llamadas de emergencia de la forma más rápida.

Este Trabajo de Fin de Máster forma parte de un proyecto denominado **iReves** que ha sido financiado por la Agencia Valenciana de Investigación (AVI). El proyecto iReves busca desarrollar herramientas inteligentes que ayuden a los responsables encargados de la gestión de VES a ubicarlos en las mejores localizaciones con un tiempo de respuesta bajo. Es decir, busca mejorar la calidad del servicio consiguiendo responder a más emergencias en menos tiempo utilizando los recursos ya disponibles y no teniendo que depender de financiación extraordinaria.



**Figura 1.2:** Estructura de iReves

**Fuente:** <https://ireves.webs.upv.es/>

El proyecto iReves está integrado por dos piezas fundamentales. Por un lado está **reves data**, que se encarga de transformar los datos en información útil para que, por otro lado, **reves mind** utilice esa información para encontrar soluciones a los problemas planteados de reubicación de VES a través de las diferentes herramientas inteligentes que han sido creadas. Actualmente, estas herramientas suponen que existe una relación entre la demanda de emergencias y la población del territorio, por lo que su objetivo es cubrir la mejor cantidad de población posible. Para que estas herramientas proporcionen soluciones fiables y eficientes necesitan nutrirse de información de calidad. Este Trabajo de Fin de Máster forma parte de **reves data** y busca proporcionar modelos para predecir la demanda de emergencias teniendo en cuenta las características espaciales del problema. De esta forma, se proporcionará información de mayor calidad a las herramientas que buscan ayudar a la toma de decisiones de ubicación de VES. Dado que la flota de VES de la que dispone el SES es limitada, es interesante analizar qué zonas tienen mayor probabilidad de tener que responder a una emergencia y así tener un plan de despliegue que se anticipe a esa demanda de emergencias. El proyecto iReves trabaja sobre un caso real, la Comunidad Valenciana, por lo que se trabajará con una base de datos con información real de las llamadas de emergencia que se dieron en la provincia de Valencia en el año 2019.



(a) CICU.



(b) Helicóptero medicalizado.



(c) SAMU.



(d) SVA.



(e) SVB.



(f) Ambulancia no asistida.



(g) VIR.



(h) EMV.



(i) Vehículo de comunicaciones.



(j) Incubadora de transporte.

**Figura 1.3:** Recursos desplegados ante emergencias.

**Fuente:** Servei d'Emergències Sanitàries de la Comunitat Valenciana.

## 1.2. Relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS)



**Figura 1.4:** ODS a los que contribuye este trabajo

**Fuente:** Materiales de Comunicación de las Naciones Unidas

Cabe destacar que este trabajo colabora con la Agenda 2030 y los Objetivos de Desarrollo Sostenible (ODS). En concreto con los que se pueden observar en la Figura 1.4:

- **ODS 3. Garantizar una vida sana y promover el bienestar para todos en todas las edades:** En especial en la meta 3.d que busca reforzar la capacidad de los países en alerta temprana y gestión de riesgos.
- **ODS 9. Construir infraestructuras resilientes, promover la industrialización sostenible y fomentar la innovación:** más concretamente la meta 9.1 que busca desarrollar infraestructuras, en este caso de gestión de VES, que apoyen el bienestar humano.
- **ODS 10. Reducir la desigualdad en y entre los países y ODS 11. Lograr que las ciudades sean más inclusivas, seguras, resilientes y sostenibles:** asegurando que independientemente de la región en la que vivan, todas las personas tendrán una cobertura mínima de VES.

## 1.3. Objetivos

Este Trabajo de Fin de Máster tiene como meta colaborar en el proyecto de creación de herramientas inteligentes de iReves para asegurar que las personas encargadas de tomar las decisiones a la hora de gestionar la flota de VES en la provincia de Valencia dispongan de la mayor información posible presentada de forma sencilla e interpretable. Más concretamente, se centrará en estudiar las características espaciales de la demanda a través de dos objetivos:

1. En primer lugar, preparar y analizar la base de datos disponible de llamadas con despliegues de VES en la provincia de Valencia para el año 2019 utilizando análisis de procesos puntuales para evaluar si existen patrones espaciales.
2. En segundo lugar, proponer y evaluar distintos modelos basados en procesos puntuales para predecir de la demanda de VES y estudiar su validez.

## 1.4. Estructura del documento

A continuación se presenta la estructura de este Trabajo de Fin de Máster junto con una breve descripción de cada uno de los apartados:

- **Capítulo 1. Introducción:** Se describe brevemente el tema que envuelve al problema para facilitar el entendimiento del resto del trabajo y se detallan los objetivos a seguir.
- **Capítulo 2. Marco teórico:** Se destacan los conceptos teóricos más importantes explicando en qué consiste la estadística espacial haciendo hincapié en el análisis de procesos puntuales.
- **Capítulo 3. Revisión Bibliográfica:** Se realiza una breve exposición de la tendencia que han seguido los trabajos centrados en el estudio y predicción de la demanda de emergencias, especialmente aquellos que han analizado el aspecto espacial de los datos.
- **Capítulo 4. Descripción del Problema:** En este apartado se exponen los distintos elementos del problema, se explora la base de datos con la que se va a trabajar y se presenta la preparación que ha sido necesaria para realizar los análisis en R.
- **Capítulo 5. Perspectiva Frecuentista:** Se analizan los datos y se proponen modelos siguiendo la metodología aportada por el análisis de procesos puntuales desde la perspectiva de la estadística frecuentista.
- **Capítulo 6. Perspectiva Bayesiana:** Se plantea un modelos basado en un proceso cox log-gaussiano desde la perspectiva de la estadística bayesiana.
- **Capítulo 7. Conclusiones y Trabajo futuro:** Se expondrán las conclusiones a las que se ha llegado durante la realización del trabajo y se mencionarán las futuras ideas de mejora para continuar con el trabajo aquí iniciado.



## Capítulo 2

# Marco teórico

### 2.1. Problemas de localización y cubrimiento

Aunque los problemas de localización y cubrimiento no sean objeto de estudio de este Trabajo de Fin de Máster, el objetivo principal de este iReves es aumentar la cantidad de información que los encargados de tomar decisiones tienen a su alcance y, de esta manera, mejorar la calidad de las soluciones que proponen y del servicio con el que proveen a la sociedad. Actualmente el problema central consiste en buscar cómo repartir la flota de VES con el objetivo de cubrir la mayor cantidad de población.

Se entiende como base a las ubicaciones destinadas por los organismos públicos para situar a los VES mientras esperan a ser movilizados ante una emergencia. Cabe destacar la importancia que tiene en estas situaciones el tiempo de respuesta, ya que meros minutos pueden significar la supervivencia de una persona. Es, por tanto, que se establecen unos límites temporales que los VES más cercanos deben contemplar. Por ejemplo, si se establece el límite temporal en 15 minutos, se debe asegurar que toda la población tiene disponible al menos a uno de estos vehículos o, en otras palabras, cada individuo está cubierto por uno o más VES que cumplen esta característica. Los problemas que tienen esta restricción se denominan problemas de cubrimiento (García y Marín 2015).

Este problema es abordado en Vecina García (2022) donde se propone el uso de modelos de cubrimiento máximo. Estos modelos buscan maximizar la población cubierta en un espacio determinado manteniendo fijo el número de bases disponibles y permiten seleccionar la mejor combinación de bases que, aun sin proporcionar un cubrimiento total de la población, intentan cubrir la mayor parte de ella utilizando los recursos ya disponibles. Es decir, se busca sacar el máximo partido a los recursos revisando cuál sería la mejor localización para la flota de VES.

Los algoritmos propuestos para conseguir soluciones a estos problemas necesitan información, y, sobre todo, información de calidad. Esta información debe ser lo más real posible para que las soluciones propuestas sean eficientes. Este trabajo busca proporcionar información de calidad a los algoritmos propuestos por iReves proporcionando un mejor entendimiento de la demanda de VES en la Provincia de Valencia.

### 2.2. Estadística Espacial

Para poder proporcionar información de calidad, es importante realizar análisis adecuados al contexto del problema y a la tipología de los datos. Dado que se desea investigar el efecto de las localizaciones de las llamadas a las que ha tenido que acudir un VES, es necesario adentrarse en el campo de la estadística espacial.

Aunque no fue hasta 1970 cuando Waldo R. Tobler introdujo lo que se denominó la Primera Ley de la Geografía, que decía que sucesos espacialmente cercanos entre ellos son estadísticamente más dependientes

que aquellos que son lejanos, siempre ha existido el interés por parte de la humanidad de entender patrones espaciales. Desde hace siglos las personas han dibujado mapas y han intentado encontrar patrones subyacentes (Ripley 2005). Actualmente, gracias a la evolución digital y al poder computacional, existe la capacidad de analizar esos patrones espaciales y obtener información clara y útil que busca entender porqué se han originado esos patrones y cómo evolucionan.

La estadística espacial proporciona las herramientas estadísticas necesarias para responder a aquellas preguntas donde la localización y la información espacial sobre los datos es relevante. Básicamente, se busca modelar la incertidumbre expresada probabilísticamente como un proceso espacial estocástico tal y como recoge la expresión 2.1.

$$Y(\mathbf{s}) : \mathbf{s} \in D \quad (2.1)$$

donde  $Y(\mathbf{s})$  corresponde al valor aleatorio del atributo en la posición  $\mathbf{s}$ , y la  $D$  es un subconjunto del espacio de  $d$ -dimensiones que contiene todas las posibles localizaciones de interés (Cressie y Moores 2022).

Según la tipología de los datos, se distinguen tres áreas de estudio diferentes: la geoestadística, los datos agregados y los procesos puntuales (Gómez Rubio 2022).

### 2.2.1. Geoestadística

La geoestadística reúne diversas herramientas estadísticas para estudiar variables espaciales que cambian continuamente dentro de la región de estudio. Se dice que surgió en la década de 1960 con el objetivo de estimar reservas de minerales (Gómez Rubio 2022) ya que, en comparación con enfoques más tradicionales, era capaz de modelar tanto la tendencia espacial como la correlación espacial (Cressie 1993).

El objetivo principal de la geoestadística suele ser la predicción o estimación de una variable en cualquier punto de la región de estudio partiendo de un número limitado de localizaciones donde se ha medido esa variable, también en esa misma región. Una de las técnicas más populares de predicción espacial es la de *kriging*, denominada así por el ingeniero de minas D. G. Krige.

Algunos ejemplos prácticos de la aplicación de geoestadística son estudios de contaminación ambiental en determinadas zonas o distintos análisis de epidemiología para evaluar el riesgo de padecer una determinada enfermedad (Gómez Rubio 2022).

### 2.2.2. Datos agregados o en redes de localizaciones

Cabe la posibilidad de que los datos estén presentados de forma agregada y no se puedan separar para conocer la localización exacta de los sucesos o que se den únicamente en las localizaciones establecidas de una red. Un ejemplo puede ser los datos socioeconómicos que el Instituto Nacional de Estadística proporciona de manera agregada.

Este tipo de análisis busca modelos estadísticos que expresen que las observaciones cercanas tienen a parecerse entre ellas, por lo que se basan en modelos de regresión que incluyen la correlación existente entre regiones vecinas.

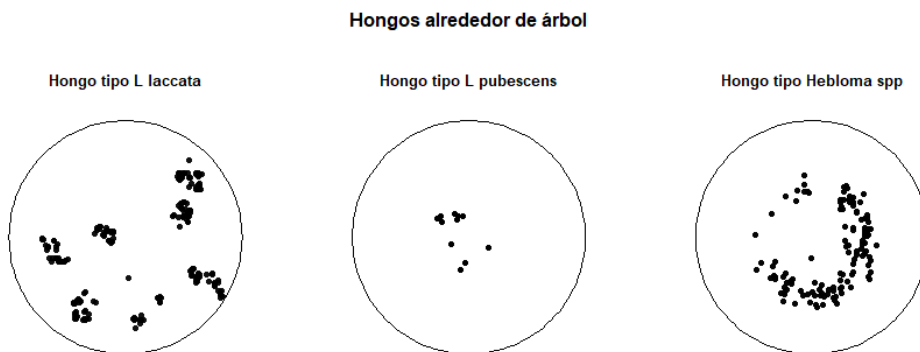
Algunos ejemplos de aplicación de los modelos de datos agregados son la econometría espacial o los atlas de mortalidad (Gómez Rubio 2022).

### 2.2.3. Procesos puntuales

El análisis de procesos puntuales busca estudiar la importancia de la localización de determinados eventos o elementos en una región concreta. Un patrón de puntos espaciales en un conjunto de datos que contiene la

localización de dichos eventos o elementos.

Un posible ejemplo de proceso puntual es el proporcionado en la base de datos *sporophores* (Ford, Mason y Pelham 1980) de la librería *spatstat* (Baddeley, Rubak y Turner 2015a). Se pueden ver representados en la Figura 2.1 tres especies de hongo diferentes que han crecido alrededor de un árbol situado en el centro de la región de estudio. Se aprecia que la distribución de los hongos no es homogénea y que, además, se rige por un patrón distinto para cada tipo de hongo.



**Figura 2.1:** Distribución espacial de tres especies de hongos alrededor de un árbol.

**Fuente:** Elaboración propia en R.

Una de las tareas más importantes a la hora de estudiar procesos puntuales es identificar si existen tendencias espaciales que afecten a la densidad de los puntos, como ha ocurrido en el caso de los hongos. Este tipo de análisis es clave ya que puede proporcionar información muy útil como, por ejemplo, ayudar a identificar un punto de contaminación que esté provocando contagios en la población o analizar qué zonas de una ciudad son más comunes para diferentes tipos de crímenes. Es posible que los puntos no sigan un patrón homogéneo, sino que estén formando distintos clusters o que estén más alejados de lo esperado inicialmente.

Partiendo del caso más sencillo, aquél en el que se asume que los eventos son independientes unos de otros y se sigue una distribución uniforme a lo largo de toda la región  $D$ , se puede decir que el número de eventos viene definido por una distribución de Poisson de media  $\lambda$ . Se denomina  $\lambda$  a la intensidad del proceso puntual, es decir, el número de casos por unidad de área (Gómez Rubio 2022). Este caso, también conocido como proceso de Poisson homogéneo, es la base de la teoría de procesos puntuales espaciales y representa aleatoriedad espacial completa o *complete spatial randomness* (CSR) en inglés. Un proceso puntual  $\mathbf{X}$  tendrá intensidad homogénea si, para cualquier subregión  $B$  del espacio inicial, el número de puntos *esperado* para  $B$  es proporcional al área de  $B$ :

$$\mathbb{E}[n(\mathbf{X} \cap B)] = \lambda|B| \tag{2.2}$$

Se puede representar la intensidad empírica de los puntos como:

$$\bar{\lambda} = \frac{n(\mathbf{X})}{|W|} \tag{2.3}$$

donde  $\bar{\lambda}$  es un estimador insesgado de la verdadera intensidad  $\lambda$ , asumiendo un proceso puntual homogéneo.  $\mathbf{X}$  representa al conjunto de datos del patrón puntual, observado en la ventana  $W$  con un número de puntos  $n(\mathbf{X})$  (Baddeley, Rubak y Turner 2015b).

Por otro lado, los eventos en el patrón de puntos pueden llevar asociados dos tipos distintos de datos (Baddeley, Rubak y Turner 2015b):



- **Marcas:** información auxiliar sobre el punto. Por ejemplo la especie de un hongo, el diámetro de un árbol o el tipo de VES que ha respondido a la llamada.
- **Covariables:** datos exploratorios que no se incluyen en la respuesta. Por ejemplo la altitud en un determinado punto, la acidez del terreno o funciones de distancia.

Los modelos utilizados para ajustar procesos puntuales están muy relacionados con los modelos de regresión comunes (Renner et al. 2015), no obstante, los modelos de procesos puntuales tienen un planteamiento diferente. En un modelo de regresión típico el foco está en la variable aleatoria  $y_i$ , de la que se desea modelar la media estimada  $\mu_i$  dependiendo de una serie de variables  $x_i$ . Eso no ocurre en los modelos de procesos puntuales, cuyo foco está en las localizaciones espaciales de los eventos que han sucedido, es decir, el estudio se centra en **dónde** han ocurrido los eventos. La similitud con los modelos de regresión proviene del modelado de la intensidad  $\lambda(s)$ , el número de eventos esperado por unidad de área, ya que se modela como una función de covariables  $x(s)$  medidas en toda la región de estudio  $A$ .

Hasta ahora, las herramientas disponibles en iReves están resolviendo un problema de cubrimiento de población tal y como se observa en Vecina García (2022). Este trabajo busca que esas herramientas puedan adaptarse a cubrir la demanda de emergencias, una característica más volátil que la población y que requiere, debido a la tipología de sus datos, requiere un tratamiento distinto. Para ello se analizará la base de datos de emergencias reales dadas en la provincia de Valencia para el año 2019 utilizando técnicas de análisis de procesos puntuales para determinar si la demanda de emergencias en la provincia de Valencia sigue un proceso de Poisson homogéneo o si, por el contrario, existe algún tipo de tendencia espacial entre los puntos.

### 2.3. Inferencia Bayesiana

La inferencia bayesiana es el proceso de ajustar un modelo probabilístico a un conjunto de datos y resumir su resultado mediante una distribución de probabilidad sobre los parámetros del modelo y sobre cantidades no observadas de las que se quiere aprender información (Rubin et al. 2004). Los métodos bayesianos destacan por el uso explícito de la probabilidad para cuantificar la incertidumbre en la inferencia de manera que se puede incluir en los modelos información conocida por la persona experta. Esta información se aporta en los modelos asignando distribuciones de probabilidad a los parámetros del modelo (Martínez-Minaya 2015). Esta metodología permite hacer inferencia en modelos complejos, como son los modelos espaciales, gracias a modelos jerárquicos bayesianos que descomponen la información conocida a priori en tres niveles: el primer nivel corresponde a la verosimilitud, que contiene la información aportada por los datos; en el segundo nivel están los parámetros  $\theta$  de la distribución que se le ha asignado previamente al modelo; finalmente, cabe la posibilidad de que los parámetros del segundo nivel dependan de unos hiperparámetros que formarían el tercer nivel y el final de la jerarquía (Martínez-Minaya 2015). Además, la inferencia bayesiana es capaz de adaptarse muy bien a los problemas espaciales por lo que se ha decidido incluir esta metodología como alternativa para el estudio de la demanda de llamadas de emergencia de la Provincia de Valencia.

Uno de los problemas que existe a la hora de resolver problemas espaciales es la alta complejidad que hay a la hora de encontrar expresiones analíticas para las distribuciones posteriores de parámetros e hiperparámetros (Martínez-Minaya 2015). El método numérico de aproximación anidada integrada de Laplace (INLA) es una propuesta para resolver este problema cuya sencilla implementación y velocidad lo hace más interesante como alternativa a otros métodos como pudieran ser los métodos de Montecarlo basados en cadenas de Markov. La librería `inlabru` es una interfaz que permite realizar modelado espacial utilizando internamente el paquete INLA, que estudia los patrones puntuales y estima su función de intensidad  $\lambda(u)$  haciendo uso de procesos cox log-gaussianos (LGCP) usando ecuaciones diferenciales estocásticas (SPDE) (Gómez-Rubio 2020). Se propone en este trabajo explorar la aplicación de esta metodología a la demanda del caso de estudio real de la demanda de VES en la Provincia de Valencia como posible alternativa al análisis de procesos puntuales frecuentista.

## Capítulo 3

# Revisión Bibliográfica

El análisis de la demanda de VES es clave para prepararse a las necesidades sanitarias de la sociedad en momentos donde el tiempo de respuesta es crítico para la supervivencia del paciente. En la mayor parte de países, el despliegue y la gestión de los VES viene dada por organizaciones gubernamentales cuyo presupuesto no es ilimitado, por lo que deben trabajar en optimizar el uso de sus recursos disponibles y asegurar un servicio de calidad. No obstante, no es tarea sencilla encontrar soluciones que satisfagan la demanda que a su vez cumplan otra serie de restricciones, por lo que muchos equipos deciden aplicar técnicas de investigación operativa para cumplir sus objetivos, como puedan ser bajos costes, cobertura máxima o la planificación de personal necesario, entre otros (Vile et al. 2012). Pero para que estas técnicas sean efectivas se deben utilizar predicciones de la demanda precisas, especialmente para modelos que utilizan despliegues dinámicos, es decir, aquellos que utilizan la predicción del número de llamadas para organizar los nuevos despliegues cada cierto número de horas (Setzler, Saydam y Park 2009). Es, por tanto, que disponer de una buena predicción de la demanda de emergencias es importante para obtener el máximo partido de los modelos de investigación operativa diseñados para cada caso. En 2010, Sasaki et al. (2010) utilizaron un algoritmo genético de agrupación modificada para comparar las localizaciones óptimas de los VES actuales y futuras que incluía las predicciones de los casos de emergencia dados anteriormente en la región de estudio en Niigata, Japón. Obtuvieron las predicciones en intervalos de 5 años desde 2020 hasta 2050 y analizaron la correlación existente entre diversos factores demográficos y las predicciones de los posibles cambios en la población japonesa.

### 3.1. Estudio de la Demanda

Para poder realizar predicciones de la demanda precisas, es necesario analizar primero su naturaleza y entender qué factores están afectando a esa demanda de VES.

En Aldrich, Hisserich y Lave (1971) utilizan modelos de regresión de mínimos cuadrados para analizar y estimar la demanda de VES en una zona urbana, más concretamente, en la ciudad de Los Ángeles, Estados Unidos, concluyendo que la demanda es altamente predecible utilizando modelos lineales simples. Años más tarde, Siler (1975) utiliza modelos de regresión múltiple y *stepwise* para predecir la misma demanda utilizando variables de tres tipos: socioeconómicas, sobre uso del terreno y relacionadas con los servicios de emergencias sanitarias. También Kvålseth y Deems (1979) propone utilizar modelos de regresión de mínimos cuadrados de primer y segundo orden para predecir la demanda de servicios de emergencias sanitarias en Atlanta, Estados Unidos. Cabe destacar que, aunque predominen los casos estudiados en áreas urbanas, también existe interés por las zonas rurales, como en Wrigley et al. (2002), donde realizan un análisis retrospectivo de los despliegues de VES en Wiltshire, Inglaterra.

Por otro lado, aparece otra línea de investigación que se centra en analizar modelos de simulación de sistemas de emergencias sanitarias y qué afecta a los tiempos de respuesta a las emergencias (Ziad Nehme y Smith 2016). En Hall (1971) se propusieron diversos modelos para evaluar el número y las localizaciones de VES necesarias

para reducir el tiempo de respuesta de las emergencias en la ciudad de Detroit, Estados Unidos. Aboueljinane, Sahin y Jemai (2013) revisan los modelos de simulación que se aplican a las operaciones de servicios de emergencias sanitarias y destaca la dificultad de utilizar la localización de las llamadas de emergencias y cómo se suele agregar los datos en zonas para reducir tiempos de computación. En Eastwood et al. (2015), se analizan los despliegues de VES en Australia y se propone un método de triaje suplementario que permita responder a las necesidades de los pacientes que tienen baja prioridad sin la necesidad de desplegar un VES con el objetivo de reducir la carga que los sistemas de emergencias sanitarias sufren ya que la demanda no para de crecer.

### 3.2. Predicción de la Demanda

A lo largo de los años, se han ido proponiendo diferentes métodos para realizar modelos que predigan la demanda de VES con la mayor precisión posible, como se exponía en Aboueljinane, Sahin y Jemai (2013).

En Baker y Fitzpatrick (1986) se propone construir un modelo de suavizado exponencial para predecir la demanda de VES en Carolina del Sur, Estados Unidos. Además, proponen utilizar programación por objetivos para encontrar los parámetros óptimos del modelo. En la misma línea, existen diversos estudios que deciden utilizar las técnicas propuestas por el campo de las series temporales para realizar las predicciones, como Channouf et al. (2007) que utiliza modelos autorregresivos de medias móviles o Matteson et al. (2011) que utiliza una combinación entre modelos dinámicos de factores latentes con modelos de series temporales. Vile et al. (2012) propone utilizar una técnica no paramétrica para el análisis de series temporales conocida como análisis de espectros singulares (SSA), que proporciona predicciones mejores a largo plazo en comparación con otros métodos más tradicionales.

Las técnicas de minería de datos han resultado ser muy útiles en diversos campos de estudio y se ha popularizado su uso en estos últimos años. No es de extrañar que se haya decidido emplear este tipo de técnicas para intentar realizar predicciones de la demanda de VES. En Setzler, Saydam y Park (2009), Grekousis y Liu (2019), Martin, Mousavi y Saydam (2021) y Wang et al. (2021) se proponen diversas versiones de modelos de redes neuronales para predecir las demandas. Li, Vanberkel y Zhong (2022) propone utilizar un árbol de decisiones híbrido que utiliza un clasificador Naive Bayes para predecir el retraso en los tiempos de despliegue de VES en Canadá. Otro ejemplo, es el de Blanco Prieto et al. (2024), donde utilizan técnicas de minería de datos para determinar cuáles son las variables determinantes a la hora de predecir la demanda de VES.

No obstante, los estudios mencionados hasta ahora tienen en común que se centran en la parte temporal del problema pero simplifican o descartan directamente los datos aportados por las localizaciones de las llamadas por tener demasiado coste computacional (Aboueljinane, Sahin y Jemai 2013).

### 3.3. Análisis de Procesos Puntuales

Según la literatura estadística, se conoce como un proceso puntual a un conjunto de eventos puntuales cuya localización y número de ocurrencias es aleatorio (Renner et al. 2015). La inquietud de analizar los mapas de patrones puntuales espaciales no es algo reciente, desde la década de los 70, investigadores como Lotwick (1984) o Ripley (2005), proponen diferentes herramientas para analizar si la intensidad  $\lambda$  (ver Ecuación 2.3), sigue un proceso de puntos estacionario o no.

Las técnicas de análisis de procesos puntuales se han aplicado a estudios de temáticas muy diversas, desde el estudio de ocurrencia de actividades sísmicas (Ogata 1999) al análisis de la presencia de ciertas especies de animales (Renner et al. 2015). Hay muchos ámbitos que utilizan este enfoque para analizar sus datos y es fácilmente extrapolable al ámbito de las emergencias sanitarias. Esto es debido a que los datos asociados a las llamadas de emergencias de VES suelen ser patrones de puntos espaciales (Bayisa et al. 2020) y es importante no descartar la información que la localización de los eventos aporta.

En Ong et al. (2009), Swalehe y Aktas (2016) y Chen et al. (2016) introducen el uso de sistemas de infor-

mación geográfica (GIS) para incluir los datos geográficos en sus problemas de localización de ambulancias. Un GIS es un entorno que permite analizar datos espaciales y obtener información relacionada con patrones o relaciones espaciales subyacentes. Además, suele ofrecer la representación en mapas de los datos para ayudar al usuario a obtener información útil de sus *datasets* y que aporte valor a los problemas con los que está trabajando. También Hassler y Ceccato (2023) utiliza un GIS para evaluar las variaciones espacio temporales en la demanda de VES en Suecia con el objetivo de estudiar como las diferencias en el uso del terreno y en las características socioeconómicas afectan a las variaciones en la demanda en el país.

Tanto Zhou y Matteson (2015) como Nilsang y Yuangyai (2021) proponen el uso de una estimación espacio temporal de densidad *kernel* por su rapidez y sencillez a la hora de ser interpretado por los profesionales que se encargan de gestionar la flota de VES. Este método busca evaluar las características de las distintas demandas históricas que se poseen y les asocia un peso que establece su capacidad predictiva. Se construye un estimador espacial de densidad *kernel* utilizando esos pesos y se toman las estimaciones obtenidas como predicciones (Zhou y Matteson 2015). Un año más tarde, (Zhou y Matteson 2016) proponen utilizar un método de deformación *kernel* (*kernel warping*), que incluya también el efecto de la complejidad de las estructuras urbanas y las características geográficas del terreno. Este método selecciona los incidentes más relevantes para la predicción que desea realizar y les ajusta un estimador de densidad *kernel* pero deforma el *kernel* utilizando un grupo de datos históricos más grande independientemente de si tienen o no relevancia con la predicción que se desea realizar.

No obstante, todos los métodos expuestos hasta ahora no consideran la tipología de los datos como patrones de puntos espaciales y existen, actualmente, pocos estudios que utilicen un modelado basado en patrones espaciales para predecir la demanda de VES. En Zhou et al. (2015) se propone utilizar un *Gaussian mixture model* (GMM) de  $m$ -componentes utilizando procesos de puntos espaciales para modelar la demanda en Toronto, Canadá. De manera parecida, Yang et al. (2020) utiliza GMM clustering para incluir cuantitativamente la incertidumbre de la demanda en su algoritmo para el problema de localización de VES en el distrito de Songjian en Shanghai, China.

En Nicoletta et al. (2021) proponen el uso de modelos lineales generalizados mixtos, pero, en lugar de utilizar las coordenadas puras de las localizaciones, clasifican las llamadas en zonas  $z$ . Utilizan la distribución de Poisson porque es la que mejor se ajusta a las distribuciones empíricas de los datos de los que disponen para la ciudad de Montreal y un suburbio cercano a Laval, Canadá.

Finalmente, en Nabarro, Fletcher y Shawe-Taylor (2018) y Bayisa et al. (2020) proponen el uso de procesos espacio temporales cox log-gaussianos (LGCP) para modelar la demanda. Los LGCP son procesos puntuales doblemente estocásticos que siguen un proceso de Poisson no homogéneo y cuyo logaritmo de la intensidad estocástica viene definida por un proceso Gaussiano. Además, en Bayisa et al. (2020) se propone utilizar un LGCP con tres componentes: uno espacial, uno temporal y uno estocástico que modele la dependencia y la variación espacio temporal de los eventos.

### 3.4. Aportación de este trabajo

Como se ha expuesto anteriormente, son escasos los estudios donde se propongan modelos basados en procesos puntuales a pesar de que, generalmente, los datos de llamadas de VES son patrones puntuales ya que vienen recogidos en un conjunto de datos que contiene las coordenadas donde han sucedido los eventos. La tendencia ha sido descartar o simplificar en gran manera la parte espacial de los distintos casos de estudio debido a la complejidad computacional que aportan a problemas que ya son bastante complejos de por sí. Además, se observa en los trabajos analizados que no existe un método general, si no que cada servicio propone una solución diferente debido a que este tipo de problema tiene la suficiente envergadura para ser tratado como un problema independiente. El estudio de la demanda de emergencias tiene dos grandes pilares: la parte temporal, donde se ha puesto el foco históricamente, y la parte espacial. Este trabajo busca centrarse en la parte espacial, analizando los patrones espaciales subyacentes en la base de datos del caso real en el que *iReves* está trabajando.

## Capítulo 4

# Descripción del Problema

En este capítulo se van a describir los elementos que forman parte del problema y que son necesarios para su entendimiento y resolución. Como se ha comentado anteriormente, este TFM surge de la necesidad de proporcionar a las herramientas de iReves mind la predicción de la demanda de emergencias para mejorar la calidad de sus soluciones, ya que hasta este momento se utiliza únicamente la población de las distintas zonas al tener un coeficiente de correlación muy elevado con el número de emergencias (Vecina García 2022). Se va a describir la base de datos con la que se ha trabajado, el área de estudio, las consideraciones iniciales, la hipótesis inicial de la que se parte y las librerías de R que se han utilizado.

### 4.1. Base de datos

Como se ha visto en el capítulo 3, las características de este tipo de problemas implican trabajar con datos asociados a casos reales de ciudades o regiones concretas. En este TFM se va a trabajar con una base de datos que recoge las llamadas de emergencias recibidas donde fue necesario el despliegue de un VES para el año 2019 en la provincia de Valencia. Esta base de datos ha sido cedida a iReves por el Servicio de Emergencias Sanitarias de la Comunidad Valenciana. Se tienen datos de 98.594 llamadas de emergencias desde el 1 de enero hasta el 31 de diciembre de 2019 y 39 variables asociadas a estas. No obstante, la mayoría de esas variables son repetidas, están vacías o son redundantes ya que siempre tienen el mismo valor. Al descartarlas, quedan las 19 variables descritas en la tabla 4.1.

Los procesos puntuales se caracterizan por tener la forma de una serie de eventos que han ocurrido en una región  $W$  y de los cuales se conoce su posición geográfica concreta. Del mismo modo, en la base de datos proporcionada por el SES de la Comunidad Valenciana se tienen las diferentes ocasiones en las que se ha producido una llamada de emergencia y su posición geográfica en la provincia de Valencia. Esta posición geográfica queda recogida en las variables  $x$ , que contiene la información de la longitud, y la  $y$ , que contiene la información de la latitud. Por esta razón se van a aplicar técnicas de análisis de procesos puntuales para analizar la base de datos y obtener predicciones de la demanda de emergencias sanitarias.

Uno de los aspectos más complicados de este TFM ha sido trabajar con la base de datos. Inicialmente proporcionada en archivo de extensión hoja de cálculo de Microsoft Excel (*.xlsx*), daba problemas al intentar cargarla en el software R debido a la gran cantidad de información que había en su interior. Por lo tanto, hubo que transformar el archivo a un formato más sencillo, en este caso se eligió el tipo texto (*.txt*) para que este pudiera ser procesado. Además, fue necesario cambiar la codificación de lectura ya que el idioma español, idioma utilizado por el sistema de registro de emergencias sanitarias del SES, dispone de caracteres específicos no reconocidos por la codificación por defecto *UTF-8*. Se seleccionó la codificación *Latin1* para evitar que la base de datos se cargara de forma incompleta. Del mismo modo, la gran cantidad de datos ha provocado que los tiempos de computación de las operaciones fueran muy elevados y, en ocasiones, que no se pudieran realizar algunas operaciones.

Variable	Descripción
<b>Incidente</b>	Número de identificación del incidente.
<b>Día</b>	Día en el que se ha recibido la llamada (número).
<b>Mes</b>	Mes en el que se ha recibido la llamada (número).
<b>Día semana</b>	Día de la semana en el que se ha recibido la llamada (nombre).
<b>Fiesta</b>	Variable binaria que vale 1 si el día ha sido festivo y 0 en caso contrario.
<b>Hora</b>	Hora del día en la que se ha recibido la llamada.
<b>Minuto</b>	Minuto en el que se ha recibido la llamada.
<b>Segundo</b>	Segundo en el que se ha recibido la llamada.
<b>Período</b>	Fase del día en el que se ha recibido la llamada, hay 24 períodos.
<b>Municipio</b>	Municipio en el que ha ocurrido la emergencia.
<b>Dirección</b>	Dirección de la emergencia.
<b>Recurso</b>	Código del recurso que se despliega
<b>Edad</b>	Edad de la persona que ha tenido la emergencia.
<b>Sexo</b>	Sexo de la persona que ha tenido la emergencia.
<b>SVB o SAMU</b>	Tipo de VES que se despliega.
<b>x</b>	Longitud de la emergencia.
<b>y</b>	Latitud de la emergencia.
<b>Motivo</b>	Razón de la emergencia.
<b>Hospital</b>	Centro al que el VES llevará al paciente.

**Tabla 4.1:** Variables disponibles en la base de datos de emergencias para el año 2019.

**Fuente:** Elaboración propia.

Una vez ya con los datos cargados en el software elegido, especificado al final de este capítulo, se ha decidido descartar una serie de variables ya que no aportaban información útil al problema. Más concretamente, las variables que se han descartado son de tipo identificativas como puedan ser el número de incidente, el municipio, la dirección escrita de la localización de la emergencia, el código del recurso que se despliega para responder a la llamada, el motivo de la emergencia y el hospital de destino. El resto de variables, sin contar la latitud ni la longitud, se han tomado como marcas, es decir, información auxiliar sobre el evento como se describía en el capítulo 2. Dado que aportan información extra, se ha decidido representar algunas de estas variables utilizando diagramas de barras con fines meramente exploratorios.

En primer lugar, algunas marcas asociadas a la parte temporal del problema son las vistas en las Figuras 4.1a y 4.1b. A priori, analizando la Figura 4.1a, podría parecer que hay un menor número de llamadas en los meses asociados a la estación de verano, de junio a septiembre, y un mayor número en los meses de otoño-invierno, de octubre a enero, aunque es difícil ver los efectos que podrían tener las estaciones disponiendo datos de un único año. Por otro lado, en la Figura 4.1b parece que los días correspondientes al fin de semana, sábado y domingo, tienen un menor número de emergencias en comparación con los días laborables. Estas características son interesantes a la hora de realizar el estudio temporal de la demanda de emergencias y, aunque se escapa del alcance de este TFM, se tendrán en cuenta en las líneas futuras de trabajo que se describen en el capítulo 7.

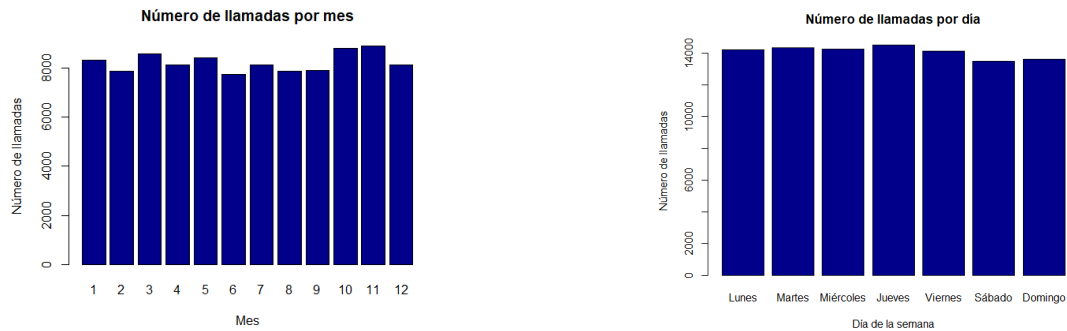
En segundo lugar, se realiza un diagrama de barras representando en intervalos de 5 años la edad del paciente que ha tenido la emergencia, tal y como se ve en la Figura 4.1c. Destacar que las emergencias aumentan considerablemente con la edad del paciente, especialmente en los pacientes de 70 a 90 años. Esto puede ser muy útil, ya que se podría tener en cuenta el envejecimiento de las diferentes zonas de la provincia y asignar más VES en bases cuya población esté muy envejecida.

En tercer lugar, tal y como se aprecia en la Figura 4.1d, se observa que el número de llamadas de pacientes hombres es ligeramente superior al número de llamadas de pacientes mujeres. Mencionar que en este apartado había 1.700 llamadas que no disponían de información relativa al sexo del paciente, pero solo suponen un 1.7 %

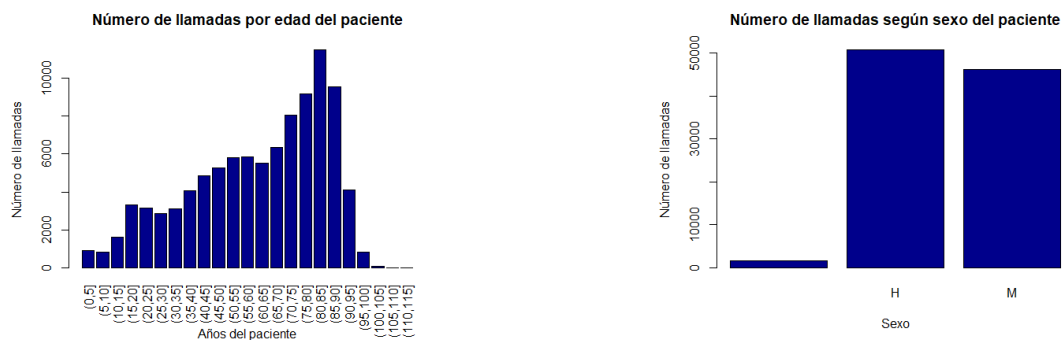


de los datos.

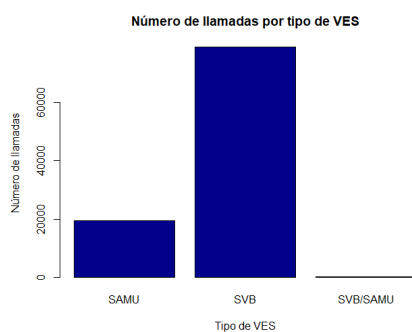
En último lugar, en la Figura 4.1e se observa claramente como el recurso tipo SVB es desplegado con bastante mayor frecuencia que el recurso tipo SAMU. Cabe destacar, que el SES de la Comunidad Valenciana no hace distinción en sus registros entre SVA y SAMU, vehículos que disponen de recursos para atender a situaciones más críticas en comparación con los vehículos SVB. Se pueden ver imágenes de estos recursos en la Figura 4.2.



(a) Número de llamadas de emergencia recibidas por mes en el año 2019. (b) Número de llamadas de emergencia recibidas por día en el año 2019.



(c) Número de llamadas de emergencia recibidas por edad del paciente en el año 2019. (d) Número de llamadas de emergencia recibidas según sexo en el año 2019.



(e) Número de llamadas según tipo de VES que se desplegó en el año 2019.

**Figura 4.1:** Histogramas para análisis exploratorio de las variables.

**Fuente:** Elaboración propia en R.

Aunque este estudio exploratorio no se pueda aplicar directamente en la modelización de los procesos puntuales, la información que recogen las marcas es muy útil y puede ayudar en gran manera a los organismos encargados del despliegue de VES. Al final, las personas encargadas de tomar las decisiones no solo utilizan las

herramientas inteligentes que equipos como iReves están creando, sino que también aplican sus conocimientos y su experiencia previa en el sector. Los modelos y programas no siempre son capaces de cuantificar todos los aspectos asociados a un problema, por lo que no se debe descartar la perspectiva que un análisis exploratorio puede aportar a la persona que debe tomar la decisión final y que sí es capaz de tener en cuenta el contexto en el que se debe tomar la decisión.



(a) SAMU.



(b) SVB.

**Figura 4.2:** Recursos desplegados ante emergencias.

**Fuente:** Servici d'Emergències Sanitàries de la Comunitat Valenciana.

## 4.2. Localización y ventana $W$

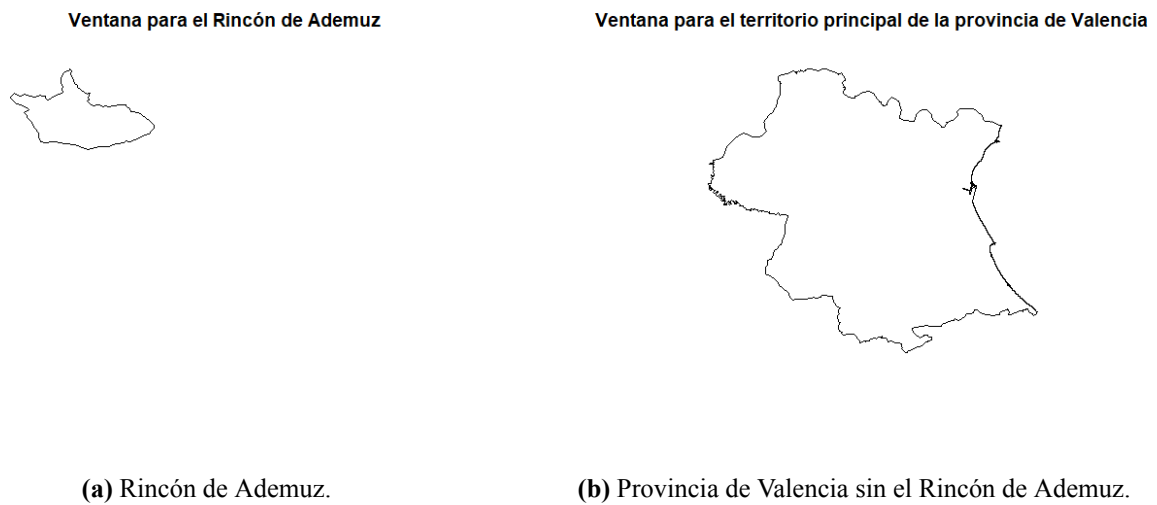
Uno de los aspectos más importantes del análisis de procesos puntuales es definir correctamente la región o ventana de estudio  $W$  que aparece en la Ecuación 2.3 y que contiene los eventos que han ocurrido recogidos en la base de datos. Este caso de estudio contiene casos reales que se dieron en el año 2019 en la provincia de Valencia perteneciente a la Comunidad Valenciana, España. Para poder crear la ventana  $W$  fue necesario descargar un archivo tipo *shapefile* que contenía el mapa geográfico de España con la posibilidad de dividirlo en provincias. Este mapa fue obtenido de la página web GADM y utiliza como sistema geográfico de coordenadas la longitud/latitud y el sistema geodésico mundial WGS 84. No obstante, la creación de la ventana no fue para nada trivial y hubo que realizar diferentes operaciones para conseguir una utilizable:

1. En primer lugar, hubo que cargar el archivo *shapefile* y seleccionar la provincia de Valencia dentro del mapa de España.
2. En segundo lugar, el objeto obtenido no era compatible con la instrucción de creación de ventanas, por lo que se tuvieron que extraer las coordenadas de forma manual y comprobar que estaban ordenadas en sentido de las agujas del reloj, ya que de otra forma no se reconoce como polígono.
3. En tercer lugar, al crear la ventana con las coordenadas extraídas del punto 2, se incurrió en un error provocado por la geografía de la provincia de Valencia. A diferencia de otras provincias, la provincia de Valencia tiene un territorio que no es colindante con el resto de municipios, el Rincón de Ademuz. Esto provocaba que el polígono creado tuviera un hueco en medio de la ventana y que eliminara la mayoría de eventos al no caer dentro del polígono. Por lo tanto, fue necesario crear dos ventanas independientes: una para el Rincón de Ademuz y otra para el territorio principal de la provincia que se pueden apreciar en la Figura 4.3a y a Figura 4.3b, respectivamente.
4. Finalmente, una vez creadas las dos ventanas para los dos territorios de forma independiente, se unieron para formar una única ventana cuya representación se puede ver en la Figura 4.4a.

Una vez que ya se dispone de la ventana  $W$ , se pueden representar los puntos en el mapa tal y como se observa en la Figura 4.4b. Se puede ver que la mayor parte de los puntos están concentrados cerca de las grandes

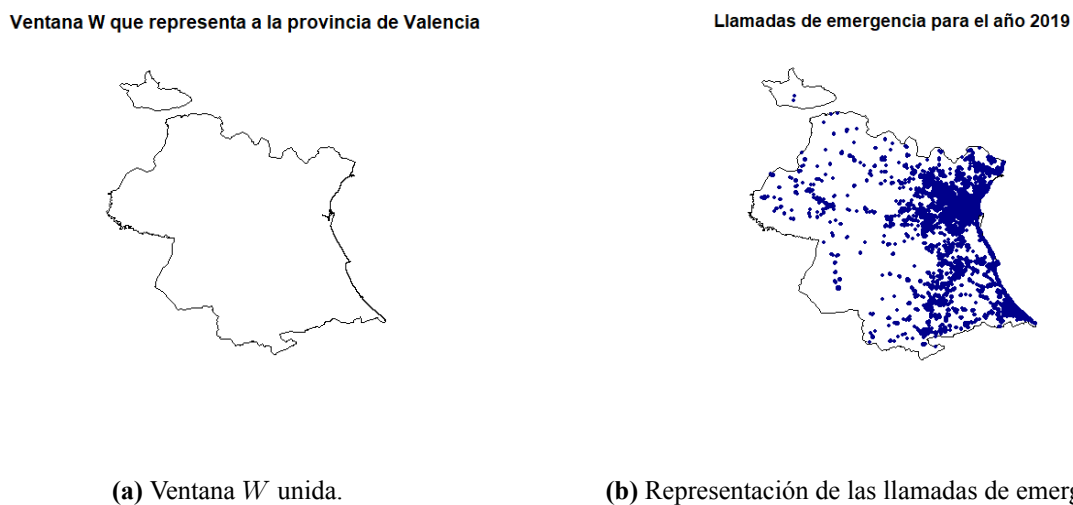


ciudades, especialmente de la ciudad de Valencia, y de la costa. Destacar que se han eliminado 88 puntos que caían fuera de la ventana al tener localizaciones muy cercanas a la frontera, dado que representan tan solo un 0,09% de los datos se decide descartarlos.



**Figura 4.3:** Ventanas separadas por territorio.

**Fuente:** Elaboración propia en R.



**Figura 4.4:** Ventana  $W$  de la región de estudio.

**Fuente:** Elaboración propia en R.

### 4.3. Consideraciones iniciales

Hay que considerar dos cuestiones fundamentales cuando se quiere implementar un análisis de procesos puntuales: constatar si corresponde aplicar técnicas de procesos puntuales y verificar que se cumplen una serie de suposiciones iniciales (Baddeley, Rubak y Turner 2015b). Primeramente, se debe revisar si aplicar las técnicas de procesos puntuales a los datos disponibles se ajusta al contexto científico del problema en el que se está trabajando. Para explicar un patrón de puntos espaciales como resultado de un proceso de puntos espaciales es necesario suponer que el patrón es aleatorio, es decir, que la localización y el número de puntos no está

previamente fijado y que el patrón que siguen los puntos es el verdadero interés del problema. Una vez verificado que el contexto es el correcto, se debe considerar la segunda cuestión, que busca constatar que el conjunto de puntos con el que se está trabajando cumple las suposiciones estándares para su análisis (Baddeley, Rubak y Turner 2015b):

1. La localización de los puntos se ha medido de manera exacta.
2. No existen dos puntos que caigan en la misma localización.
3. El estudio de la región  $W$  ha sido exhaustiva.
4. Los puntos podrían haber aparecido en cualquier punto de la región  $W$ .

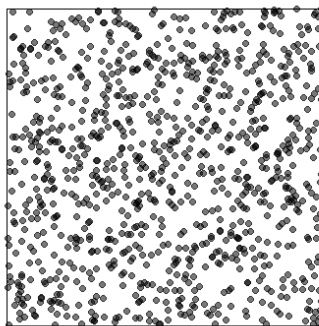
En el Capítulo 5 se consideran estas dos cuestiones y se comprueba el cumplimiento de estas suposiciones para establecer que el análisis de procesos puntuales es correcto y pertinente en este problema.

#### 4.4. Aleatoriedad Espacial Completa (CSR)

Tal y como se comentaba en el Capítulo 2, la base de la teoría de procesos puntuales espaciales radica en la aleatoriedad espacial completa (CSR) cuya intensidad viene definida en la Ecuación 2.1. Este proceso tiene dos características clave: la homogeneidad y la independencia. La homogeneidad establece que los puntos no tienen preferencia por ninguna localización y pueden aparecer aleatoriamente en cualquier lugar de la región  $W$ . La independencia establece que los resultados en una región del espacio no afectan a los resultados de otras regiones. Esta distribución se puede ver en la naturaleza en eventos como la radiación y sirve como modelo estándar de referencia para comparar con otros posibles patrones que estén subyacentes en los datos a estudiar (Baddeley, Rubak y Turner 2015b). Por lo tanto, la hipótesis nula con la que se va a iniciar el estudio de la base de datos de la demanda de VES es la descrita a continuación:

$H_0$ : La base de datos sigue un proceso homogéneo de Poisson (CSR) $H_\mu$ : La base de datos <b>no</b> sigue un proceso homogéneo de Poisson (No CSR)
---

Se puede ver un ejemplo gráfico de la distribución espacial de un proceso homogéneo de Poisson generado artificialmente en un cuadrado de lado 1 con una intensidad de mil puntos en la Figura 4.5.



**Figura 4.5:** Proceso homogéneo de Poisson generado artificialmente.

**Fuente:** Elaboración propia en R.

## 4.5. Software utilizado

Se dispone de una base de datos con llamadas de emergencia donde ha sido necesario el despliegue de un VES, pero para poder obtener conclusiones de estos datos es necesario transformarlos en información. Para trabajar con la base de datos y que se pueda aplicar a las operaciones correspondientes en el entorno de trabajo se han preparado los datos tal y como se ha explicado a lo largo de este capítulo. Para conseguir esta transformación se ha utilizado principalmente dos programas informáticos: Microsoft Excel y el software estadístico R a través de la interfaz RStudio.

Para poder abrir la base de datos cedida por el SESCOV ha sido necesario el uso del programa de cálculo Microsoft Excel. Inicialmente se intentó cargar la base de datos directamente con la extensión de Excel *.xlsx*, no obstante, debido a la magnitud de los datos, fue imposible utilizar esa extensión ya que saltaban errores al intentar procesar la base en el software R. Para evitar este problema, se decidió modificar el tipo de archivo a uno de tipo texto de extensión *.txt* ya que su codificación es mucho más sencilla y su lectura mucho más rápida siempre y cuando se hayan seleccionado los parámetros de lectura correctamente.

R es un entorno de software libre especializado en cálculos estadísticos y gráficos, que dispone de un conjunto de herramientas muy amplio gracias a las librerías y a la libertad de poder crear funciones propias (R Core Team 2023). La versión utilizada para este TFM ha sido la R-4.3.1 y, para facilitar su uso, todo el código y las operaciones realizadas se han desarrollado en RStudio. Rstudio es un entorno de desarrollo integrado que ayuda al usuario a ser más eficiente al utilizar el programa R y que, además, permite generar documentos utilizando la extensión R Markdown.

Para tratar datos espaciales existen dos enfoques: el enfoque frecuentista y el enfoque bayesiano. En un principio, este TFM iba a seguir exclusivamente el enfoque frecuentista utilizando el manual escrito por Bivand, Pebesma y Gómez Rubio (2013) que detalla con mucha precisión cómo se debe realizar análisis de procesos puntuales en el entorno R. En este manual se utilizan una serie de librerías que se detallan a continuación:

- **sp**: librería que contiene clases y métodos para datos de tipo espacial y que permite representarlos, dividirlos o analizarlos entre otros (Pebesma y Bivand 2005).
- **rgdal**: librería que permite obtener datos geoespaciales de la biblioteca GDAL y que trabaja junto con otras muchas librerías, más concretamente con la **sp** (Bivand, Keitt y Rowlingson 2023).
- **rgeos**: proporciona una interfaz para utilizar GEOS, una librería de geometría computacional centrada en los sistemas de información geográfica (GIS) (Bivand y Rundel 2023).
- **mapprools**: librería con herramientas para manipular datos de tipo geográfico (Bivand y Lewin-Koh 2023).

Lamentablemente, debido a la jubilación del investigador encargado del mantenimiento de las librerías **rgdal**, **rgeos** y **mapprools**, Roger Bivand (Pebesma y Bivand 2022), han dejado de recibir soporte y sus creadores han aconsejado la transición a otras librerías. Dado que el objetivo de este TFM continúa a largo plazo, se tomó la decisión de cambiar de manual de referencia para no tener que realizar el cambio de sistema una vez que estuviera más avanzado el proyecto. El nuevo manual escogido es el escrito por Baddeley, Rubak y Turner (2015) y que, del mismo modo que el manual de Bivand, Pebesma y Gómez Rubio (2013), proporciona una serie de librerías para poder trabajar cómodamente con datos de puntos espaciales. Las librerías que se han utilizado, incluyendo las recomendadas por Baddeley, Rubak y Turner (2015) son:

- **spatstat**: librería especializada en estadística espacial enfocada en el análisis de patrones de puntos espaciales en dos dimensiones (Baddeley y Turner 2005).
- **sf**: paquete que permite codificar datos de tipo vector espacial y la sustitución propuesta del paquete **rgdal** (Pebesma 2018).

- **dplyr**: librería que facilita la selección de datos y su manipulación (Wickham et al. 2023).
- **raster**: librería que permite trabajar con rásteres de datos, necesarios para incluir las covariables en los modelos (Hijmans 2023).
- **stars**: paquete para transformar y manipular diferentes tipos de datos espaciales y rásteres (Pebesma y Bivand 2023).

Finalmente, debido a que la metodología de análisis de procesos puntuales daba problemas de computación al tratar una base de datos tan grande, se ha decidido probar otra metodología basada en modelización espacial de enfoque bayesiano con aproximación anidada integrada de Laplace. Para ello ha sido necesario utilizar las siguientes librerías:

- **inlabru**: librería que permite hacer modelización espacial utilizando aproximación anidada de Laplace utilizando el paquete INLA (Bachl et al. 2019).
- **INLA**: paquete que permite realizar inferencia bayesiana con modelos latentes gaussianos (Rue et al. 2017).
- **mgcv**: para realizar inferencia bayesiana con modelos aditivos generalizados (GAM) mixtos (Wood 2017).

Se han utilizado también las librerías básicas de R (R Core Team 2023) para la realización de gráficos, instrucciones simples y lectura de la base de datos. Mencionar que ha sido necesario la utilización de algunas de las librerías descatalogadas ya que no se han encontrado métodos alternativos para algunas de sus instrucciones.

# Capítulo 5

## Perspectiva Frecuentista

En este capítulo se va a realizar un análisis exploratorio de los datos en el contexto de los procesos puntuales analizando si la intensidad, la correlación y el espaciado de los puntos corresponden con un proceso puntual CSR o, si por el contrario, existe algún tipo de patrón en la demanda de emergencias de la provincia de Valencia. Una vez determinado que los datos no siguen un proceso homogéneo de Poisson, se procede a ajustar diferentes modelos basados en la distribución de Poisson y se evalúan con diferentes estadísticos seleccionando aquél que sea significativamente mejor a la hora de estimar la función de intensidad de la demanda de VES. Una vez seleccionado el modelo se valida si realmente hacen una representación fidedigna de los datos disponibles y, en caso contrario, se proponen alternativas. Todo este proceso queda resumido en el esquema de la Figura 5.1.

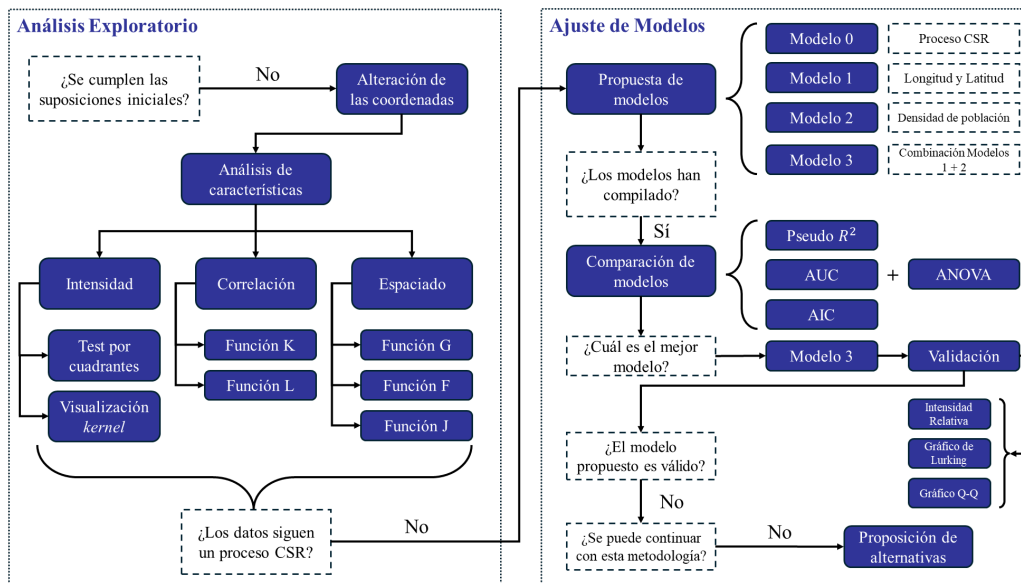


Figura 5.1: Esquema de la metodología seguida con la metodología de perspectiva frecuentista.

Fuente: Elaboración propia.

### 5.1. Análisis Exploratorio de los Datos

### 5.1.1. Cuestiones Fundamentales

Es necesario comprobar las dos cuestiones fundamentales que se establecen en la Sección 4.3. En este trabajo, donde se estudia la demanda de VES en la Provincia de Valencia, el interés recae en analizar el patrón que provoca la aparición de las emergencias que requieren el desplazamiento de un VES y, en caso de ser posible, encontrar el proceso puntual que sigue para así poder predecir esa demanda, por lo que se puede asumir que el contexto del problema es el correcto. También se debe comprobar si se cumplen las suposiciones iniciales que se detallan a continuación:

1. **La localización de los puntos se ha medido de manera exacta:** se supone que las herramientas de las que dispone el SES de la Comunidad Valenciana intentan medir de la forma más exacta la localización de las emergencias teniendo en cuenta la infraestructura disponible actual.
2. **No existen dos puntos que caigan en la misma localización:** esta suposición no se cumple de manera inicial ya que hay puntos duplicados en la base de datos. Existen dos enfoques diferentes para resolver este problema: el descarte y la alteración. El primero de ellos es descartar aquellos puntos que están duplicados en la base de datos, al intentar aplicar este método se eliminaban 50.670 llamadas de la base de datos, aproximadamente un 51,44 % de los datos. Se descartó rápidamente esta opción y se optó por aplicar el enfoque alternativo que consiste en introducir un pequeño ruido para alterar ligeramente las coordenadas de los puntos y evitar que se repitan, teniendo en cuenta la frontera de la ventana  $W$  para que estos no se salgan de la región de estudio.
3. **El estudio de la región  $W$  ha sido exhaustiva:** es decir, que no ha habido errores en la detección de las llamadas de emergencia. Dada la importancia del servicio de emergencias, se asume que los procesos dispuestos por el SES aseguran la detección de todas las llamadas que se realizaron en el año 2019.
4. **Los puntos podrían haber aparecido en cualquier punto de la región  $W$ :** se asume que las llamadas podrían haber ocurrido en cualquier lugar dentro de la provincia de Valencia.

### 5.1.2. Intensidad

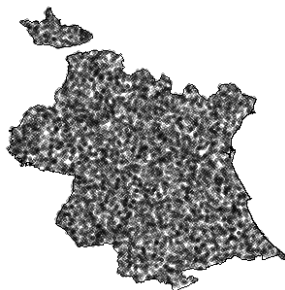
El estudio de la intensidad de un patrón de puntos es uno de los aspectos más importantes en el análisis de procesos puntuales y suele ser de los primeros elementos en ser investigados. Esto es debido a que la intensidad es una característica descriptiva básica de un proceso puntual y, además, no requiere de demasiadas suposiciones (Baddeley, Rubak y Turner 2015b). Las herramientas que se van a utilizar para el estudio de la intensidad van a ser de tipo no paramétricas.

Por lo general, tal y como se describía en los Capítulos 2 y 4, se parte de la hipótesis que establece que la intensidad es de tipo homogénea y no hay una preferencia espacial a la hora de que ocurran los eventos. Sin embargo, en muchos casos, esa hipótesis es inapropiada ya que la intensidad sí que varía en el espacio. Algunos ejemplos son la propagación de ciertos árboles en bosques o la preferencia de algunas especies animales por hábitats concretos. Cuando esto ocurre se pueden utilizar técnicas estadísticas para estimar la intensidad a partir de los datos.

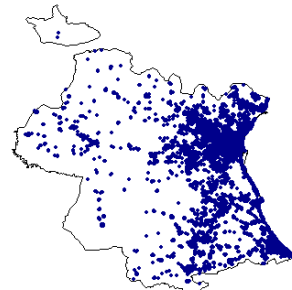
Como exploración inicial se ha decidido comparar los puntos disponibles actualmente correspondientes a la base de datos descrita con un CSR generado artificialmente con la misma intensidad media y en la misma región  $W$ . Para calcular esta intensidad se ha utilizado la Ecuación 2.3 obteniendo una media de  $8.717 \text{ puntos}/\text{km}^2$ . En la Figura 5.2 se presentan dos casos. Por un lado vemos que en la Figura 5.2a se muestra la imagen de cómo se debería detectar el mapa de la Provincia de Valencia para que tuviera aleatoriedad espacial completa. Por otro lado, en la Figura 5.2b están representadas las llamadas de la base de datos de la que se dispone. Llama la atención que en la Figura 5.2a, correspondiente al CSR, los puntos están distribuidos por toda la ventana  $W$  mientras que en la Figura 5.2b los puntos están muchos más agrupados quedando incluso zonas en blanco. A priori, no parece demasiado probable que la base de datos del caso de la provincia de Valencia vaya a seguir

un proceso homogéneo de Poisson. Para realizar esta comprobación se van a utilizar dos métodos: el test por cuadrantes y la visualización *kernel*.

CSR con la misma intensidad que la de la base de datos



Llamadas de emergencia para el año 2019



- (a) Representación de las llamadas de emergencia de un CSR artificial. (b) Representación de las llamadas de emergencia de la base de datos real.

**Figura 5.2:** Comparación de aleatoriedad espacial completa (CSR).

**Fuente:** Elaboración propia en R.

### Test por Cuadrantes

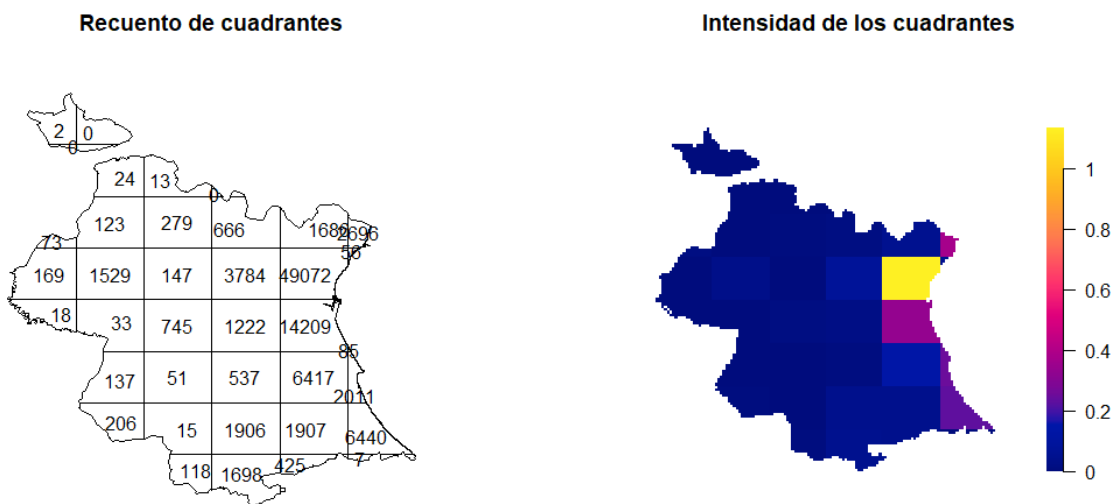
Se comienza realizando el recuento por cuadrantes con la instrucción `quadratcount` en R. Se puede observar en la Figura 5.3a cómo el número de llamadas de emergencia en la ventana de región  $W$  varía bastante entre regiones. Otra forma de analizar los cuadrantes de cada ventana es representando la intensidad de cada cuadrante, tal y como se ve en la Figura 5.3b, donde colores fríos como el azul indican intensidades próximas a cero y colores más cálidos como el amarillo indican una intensidad muy elevada. Atendiendo al gráfico de la Figura 5.3b, la intensidad es más elevada en la zona de la costa y especialmente en el cuadrante que contiene a la ciudad de Valencia. Es importante encontrar un buen equilibrio a la hora de seleccionar el tamaño de los cuadrantes, ya que cuadrantes más grandes reducen el error relativo del conteo de puntos en ese área, pero elimina el efecto que produce la variación espacial dentro del propio cuadrante (Baddeley, Rubak y Turner 2015b). Cabe destacar que debido a que la ventana correspondiente a la provincia de Valencia no es regular, las áreas de los cuadrantes no son exactamente iguales por lo que este análisis no es suficiente para determinar la inhomogeneidad de los datos. Otra forma de verificar la existencia de inhomogeneidad es realizar un test estadístico donde la hipótesis nula sea que la intensidad es homogénea y la hipótesis alternativa sea que la intensidad es inhomogénea pero sin especificar de qué forma. Asumiendo que el proceso de puntos espaciales sigue una distribución de Poisson, la hipótesis nula es la descrita en la Sección 4.4. El test estadístico a aplicar en este caso es el denominado test de uniformidad  $\chi^2$  que se observa en la Ecuación 5.1.

$$\chi^2 = \sum_j \frac{(n_j - e_j)^2}{e_j} = \sum_j \frac{(n_j - \bar{\lambda}a_j)^2}{\bar{\lambda}a_j} \quad (5.1)$$

donde  $n_j$  es el número de puntos en un cuadrante  $j$ ,  $a_j$  es el área del cuadrante  $j$ ,  $e_j$  es el número estimado de puntos en el cuadrante  $j$  y  $\bar{\lambda}a_j$  es la intensidad estimada en el cuadrante  $j$ .

Se debe destacar que la selección del número de cuadrantes es bastante importante ya que puede poner en peligro la potencia del test  $\chi^2$  si hay más demasiados cuadrantes con menos de 5 puntos (Baddeley, Rubak y Turner 2015b). Al realizar el test con la instrucción `quadrat.test` de la librería `spatstat` (Baddeley y Turner 2005), se obtiene un p-valor de 2,2e-16 por lo que se puede descartar la hipótesis nula y confirmar que la

intensidad no es homogénea. Además, la instrucción `quadrat.test` también permite realizar el test con una simulación Monte Carlo, cuyo p-valor obtenido es de 0,001 y confirma los resultados obtenidos con el test  $\chi^2$ .



(a) Recuento de los puntos en diferentes cuadrantes. (b) Representación de intensidad de cada cuadrante.

**Figura 5.3:** Análisis de cuadrantes.

**Fuente:** Elaboración propia en R.

### Visualización Kernel

Dado que la intensidad de los datos no es homogénea, se puede estimar de forma visual la función de intensidad  $\lambda(u)$  de forma no paramétrica utilizando un estimador *kernel*. En R se puede realizar esta estimación *kernel* con el función `density.ppp` de la librería `spatstat` (Baddeley y Turner 2005). Se ha seleccionado el estimador corregido de Diggle (Diggle 1985), cuya expresión matemática se describe en la Ecuación 5.2, ya que reduce el sesgo introducido por los efectos de la frontera. Además, esta instrucción utiliza suavizado *kernel* de densidad Gaussiana isotrópica.

$$\lambda(u) = \sum_{i=1}^n \frac{1}{e(x_i)} K(u - x_i) \tag{5.2}$$

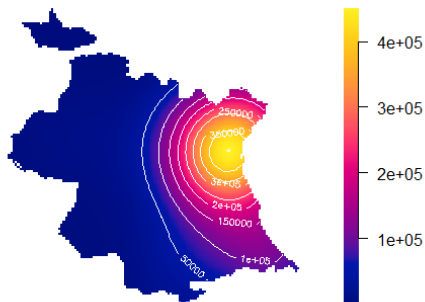
donde  $x_i$  es la localización del punto  $i$ ,  $n$  es el total de puntos,  $K(u)$  es la función *kernel* y  $e(x_i)$  es la corrección del sesgo provocado por los efectos de frontera.

Utilizando la instrucción `density.ppp` con los datos de la provincia de Valencia se obtienen las estimaciones de la Figura 5.4. En la Figura 5.4a se deja el ancho de banda por defecto y se obtiene una estimación poco precisa que centra el foco en el área de la ciudad de Valencia. Como ésta no es una buena estimación, ya que queda sesgada debido a la alta concentración de eventos en la ciudad de Valencia, se decide utilizar la instrucción `bw.diggle` para calcular el ancho de banda óptimo correspondiente a los datos del caso de estudio y se obtiene la función de densidad representada en la Figura 5.4b, que es mucho más precisa. Aunque el foco sigue estando cercano a la ciudad de Valencia, aparecen otros puntos de interés cercanos a la costa y por el sur de la provincia que corresponden con núcleos urbanos. Esto tiene bastante sentido ya que se había visto anteriormente que existía una correlación muy elevada entre la demanda de emergencias y la población. Destacar que este tipo de estimador *kernel* cuyo ancho de banda es fijo para toda la región puede no ser el mejor método

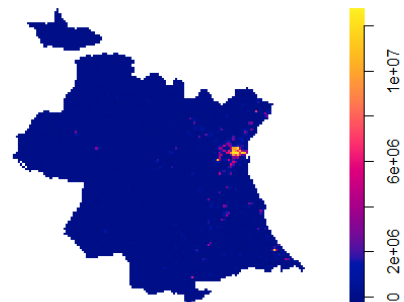


para datos cuya intensidad varía en gran manera a lo largo del territorio, ya que probablemente esté suavizando en exceso las zonas con intensidades altas y suavizando de manera insuficiente las zonas con intensidades bajas. En la Figura 5.5, se puede observar que utilizando una densidad kernel adaptativa a la intensidad, el foco de la ciudad de Valencia está mucho menos extendido y lo que anteriormente parecían intensidades más altas por la costa de la provincia ahora apenas se diferencian.

Densidad Kernel con ancho de banda por defecto



Densidad Kernel con ancho de banda óptimo



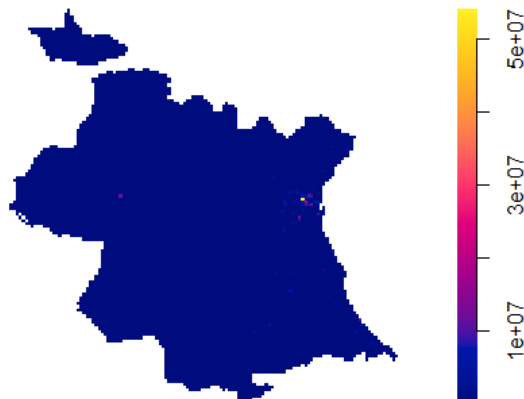
(a) Densidad kernel con ancho de banda por defecto.

(b) Densidad kernel con ancho de banda óptimo.

**Figura 5.4:** Funciones Kernel.

**Fuente:** Elaboración propia en R.

Densidad Kernel adaptativa



**Figura 5.5:** Densidad kernel adaptativa.

**Fuente:** Elaboración propia en R.

### 5.1.3. Correlación

Una de las razones más comunes por las que se estudian datos de patrones puntuales es para determinar si los puntos han aparecido de forma independiente entre ellos o si existe algún tipo de dependencia interpuntual. Esta dependencia entre puntos se puede analizar con la correlación o, en términos más generales, la covarianza. Una covarianza positiva indicaría que hay *clustering* o que los puntos tienden a aparecer cerca de otros puntos, una covarianza cercana a cero implicaría un patrón completamente aleatorio (CSR) y una covarianza negativa señalaría que hay *regularidad* o que los puntos tienden a aparecer lejos los unos de los otros. Se debe tener un cierto conocimiento del proceso puntual antes de estudiar la posible correlación existente, pero este aspecto ya ha sido estudiado en este TFM. Además, cabe recordar que la correlación no tiene porqué indicar causalidad y es insuficiente para determinar las posibles causas que estén provocando los patrones en los puntos (Baddeley, Rubak y Turner 2015b). Destacar que el tamaño de la base de datos con la que se está trabajando en este TFM ha provocado muchos problemas en este apartado y ha impedido utilizar varias herramientas de estudio de la correlación.

#### Función K de Ripley

La función  $K$  de Ripley (Ripley 1977) es una de las técnicas más populares para estudiar la correlación espacial en patrones puntuales y se define como el número de puntos medio acumulado situados a una distancia  $r$  de un punto típico. Además, la función  $K$  de Ripley introduce una corrección que tiene en cuenta los efectos provocados en la frontera de la ventana  $W$  y normaliza el resultado dividiéndolo por la intensidad con el objetivo de poder comparar patrones de puntos espaciales diferentes. La ecuación empírica de la función  $K$  se describe en 5.3.

$$K(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n 1\{d_{ij} \leq r\} e_{ij}(r) \quad (5.3)$$

donde  $r$  es el radio,  $W$  es el área de la ventana,  $n$  es el número total de puntos,  $d_{ij}$  es la distancia entre el punto  $i$  y el punto  $j$  y  $e_{ij}(r)$  es la corrección del efecto frontera.

#### Función L

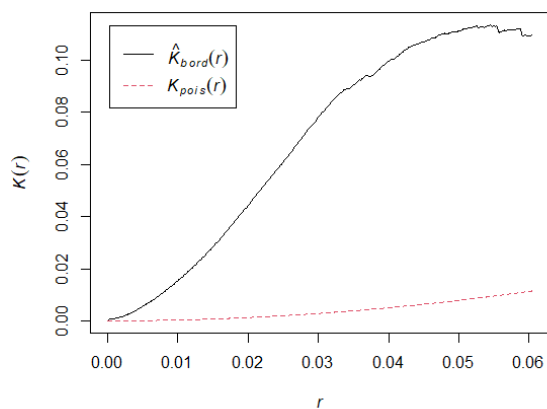
La función  $L$  propuesta por Besag (Besag 1977) transforma la función  $K$  en una función lineal para facilitar su comprensión. Su expresión empírica está descrita en la Ecuación 5.4.

$$L(r) = \sqrt{\frac{K(r)}{\pi}} \quad (5.4)$$

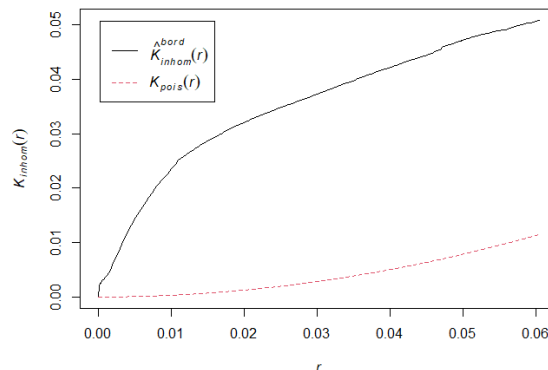
Estas funciones se pueden estimar en R con las instrucciones `Kest` y `Lest` de la librería `spatstat` (Baddeley y Turner 2005). Se ha seleccionado la corrección de tipo *border* ya que era la única que compilaba debido a la magnitud de la base de datos. En la Figura 5.6a se observa el resultado de aplicar la función  $K$  empírica a los datos, mientras que en la Figura 5.6c corresponde a la aplicación de la función  $L$ . En este tipo de gráficas la línea roja de puntos indica el valor de las funciones si los datos siguieran un proceso de Poisson homogéneo, mientras que la línea negra representa el valor de las funciones para la base de datos que se está analizando. Tanto para el caso de la función  $K$  como el de la función  $L$ , la correlación supera lo esperado por un proceso CSR, ya que la línea roja  $K_{pois}(r)$  queda siempre por debajo de la línea negra, lo que significa que podría existir *clustering* en los datos. Cabe mencionar que no ha sido posible representar intervalos de confianza por problemas computacionales.

Dado que se ha demostrado anteriormente la inhomogeneidad del patrón de puntos, se debería tener en cuenta esta característica a la hora de calcular la correlación con las funciones  $K$  y  $L$ . Para ello, se deben utilizar las funciones  $K$  y  $L$  inhomogéneas disponibles en R `Kinhom` y `Linhom` respectivamente, pertenecientes a la librería `spatstat` (Baddeley y Turner 2005). Se puede observar el resultado en las Figura 5.6b y Figura

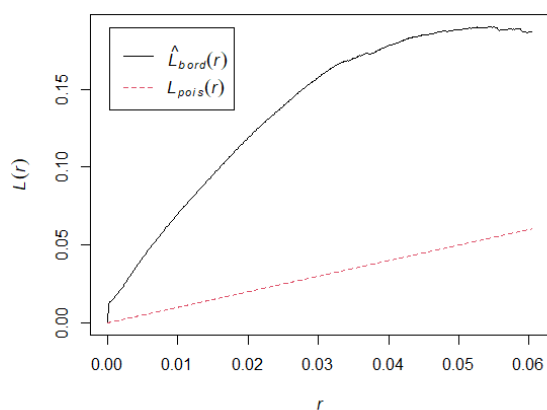
5.6d, respectivamente, donde ambas funciones llevan a la misma conclusión que la obtenida con las funciones homogéneas, existe *clustering* en el proceso puntual que se está estudiando.



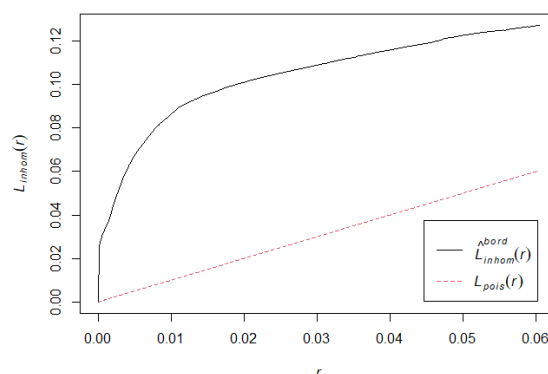
(a) Función  $K$  empírica.



(b) Función  $K$  inhomogénea.



(c) Función  $L$  empírica.



(d) Función  $L$  inhomogénea.

**Figura 5.6:** Representación de las funciones  $K$  y  $L$  obtenidas en R.

**Fuente:** Elaboración propia en R.

### 5.1.4. Espaciado

A pesar de su popularidad, la función  $K$  de Ripley descrita anteriormente puede obviar algunas características de la dependencia entre puntos. Por tanto, el análisis de los espaciados o las distancias cortas entre puntos puede proveer mayor información sobre el patrón puntual, aunque también obvia características que solo se aprecian analizando la correlación. Esto es debido a que el estudio de la correlación y el espaciado son complementarios uno del otro, por lo que un buen análisis debería estudiar ambas características.

#### Función $G$

La función  $G$ , también denominada la función de distribución de distancia al vecino más próximo, permite obtener mayor información de un patrón puntual en comparación con el estudio de la media de distancias al vecino más próximo. Se puede ver su expresión matemática en la Ecuación 5.5. Se puede calcular en R gracias a la instrucción `Gest` de la librería `spatstat` (Baddeley y Turner 2005). Se ha seleccionado la corrección de

tipo frontera para minimizar los efectos de tipo borde y se ha calculado el intervalo de confianza con la función *envelope*, realizando 99 simulaciones. El resultado se puede apreciar en la Figura 5.7a y se observa que, de interpretación similar a la función  $K$  de Ripley, la curva negra de probabilidad de la distribución empírica de distancia al vecino más próximo es superior a la curva roja de un proceso CSR. Es decir, las distancias al vecino más próximo son inferiores a lo esperado por un proceso CSR con la misma intensidad, lo que podría corresponder con un patrón de tipo *clustering* (Baddeley, Rubak y Turner 2015b).

$$G(r) = \mathbb{P}\{d(u, \mathbf{X} \setminus u) \leq r \mid \mathbf{X} \text{ tiene un punto en } u\} \quad (5.5)$$

donde  $r$  es el radio,  $u$  es la localización,  $\mathbf{X}$  es el proceso puntual y  $d(u, \mathbf{X} \setminus u)$  es la distancia al vecino más próximo.

### **Función F**

Por otro lado, la función  $F$ , utiliza la distancia de espacio vacío para su función de distribución. La distancia de espacio vacío es la distancia entre una localización  $u$  fija, que se encuentre dentro de la región de estudio, y el punto del proceso más cercano. Su expresión matemática es la descrita en la Ecuación 5.6. Se puede calcular en R gracias a la instrucción `Fest` de la librería `spatstat` (Baddeley y Turner 2005). Se han seleccionado las mismas características que para la función  $G$ . Cabe destacar que en el caso de la función  $F$  el efecto de tipo *clustering* aparece de forma totalmente opuesta a la función  $K$  de Ripley. Tal y como se observa en la Figura 5.7c, la curva negra para la función de distribución empírica de distancia de espacio vacío está por debajo de un patrón CSR o, lo que es lo mismo, las distancias de vacío son mayores que un proceso CSR con la misma intensidad, lo que es consistente con un posible patrón de tipo *clustering* (Baddeley, Rubak y Turner 2015b).

$$F(r) = \mathbb{P}\{d(u, \mathbf{X}) \leq r\} \quad (5.6)$$

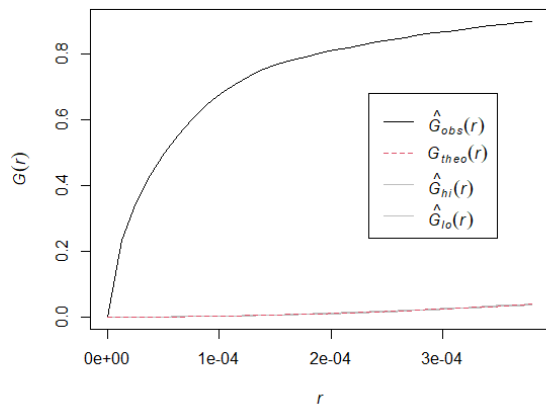
donde  $r$  es el radio,  $u$  es la localización,  $\mathbf{X}$  es el proceso puntual y  $d(u, \mathbf{X})$  es la distancia de espacio vacío.

### **Función J**

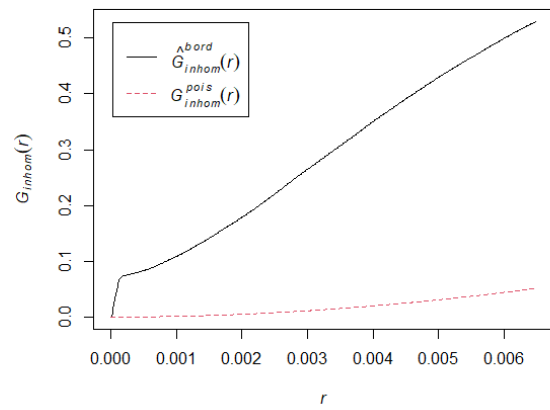
Ambas funciones,  $G$  y  $F$ , se pueden combinar utilizando la función  $J$  cuya expresión se puede ver en la Ecuación 5.7. Cuando los valores de  $J(r) > 1$  se tendrá un patrón de tipo *regular* y cuando  $J(r) < 1$  se tendrá un patrón de tipo *clustering*. Tal y como se observa en la Figura 5.7e, el valor de  $J(r) < 1$  correspondiente a los datos, la línea negra, es consistente con las conclusiones previas. Este gráfico se ha obtenido con la instrucción `Jest` y se han seleccionado las mismas características que en los casos anteriores para su cálculo y el de los intervalos de confianza.

$$J(r) = \frac{1 - G(r)}{1 - F(r)} \quad (5.7)$$

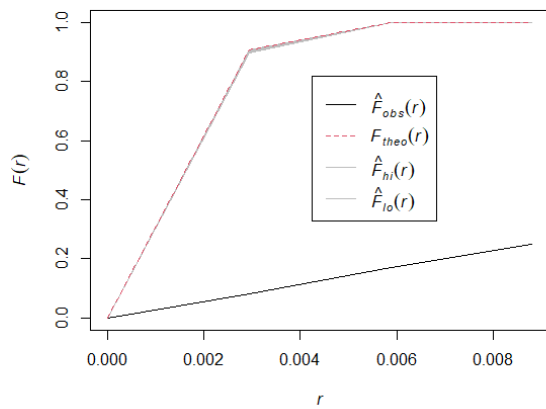
Tal y como ocurría con el estudio de la correlación, se debería tener en cuenta la inhomogeneidad del proceso puntual para el cálculo de estas funciones  $G$ ,  $F$  y  $J$ . Por lo tanto se va a proceder al cálculo de las funciones  $G$ ,  $F$  y  $J$  inhomogéneas en R utilizando las instrucciones `Ginhom`, `Finhom` y `Jinhom`, respectivamente. De nuevo se se ha seleccionado la corrección de tipo frontera para corregir los efectos de los bordes. En la Figura 5.7b se puede ver como la curva negra de la función  $G$  inhomogénea vuelve a superar la curva roja del proceso CSR, en la Figura 5.7d la curva negra de la función  $F$  inhomogénea vuelve a ser inferior a la curva roja del proceso CSR y en la Figura 5.7f como la  $J(r) < 1$ . De nuevo se confirma que el patrón de puntos espaciales de la base de datos de emergencias de la provincia de Valencia sigue un patrón de tipo *clustering*. También se puede confirmar que los resultados obtenidos analizando la correlación y el espaciado son consecuentes entre ellos.



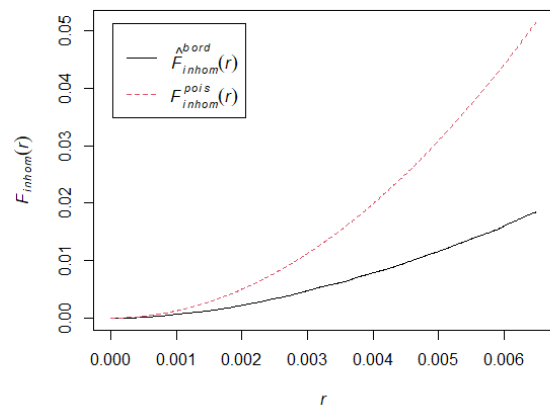
(a) Función  $G$  empírica.



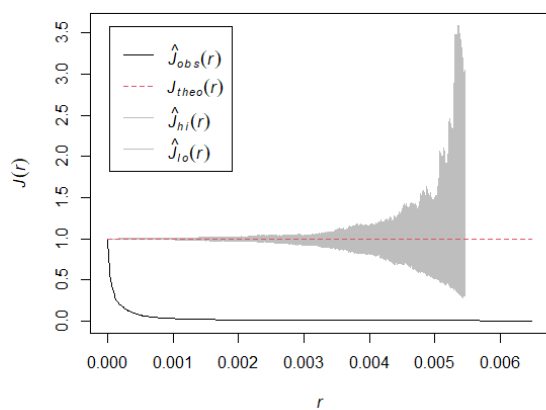
(b) Función  $G$  inhomogénea.



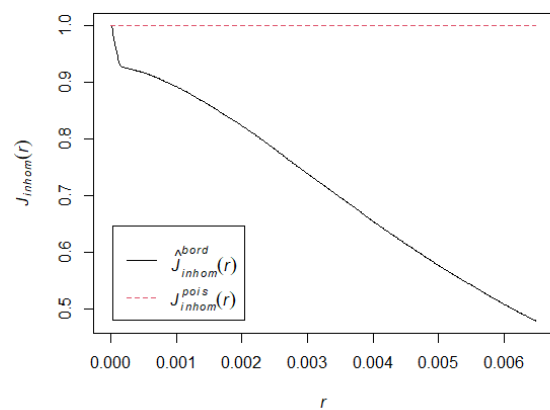
(c) Función  $F$  empírica.



(d) Función  $F$  inhomogénea.



(e) Función  $J$  empírica.



(f) Función  $J$  inhomogénea.

**Figura 5.7:** Representación de las funciones  $G$ ,  $F$  y  $J$  obtenidas en R.

**Fuente:** Elaboración propia en R.

## 5.2. Ajuste de Modelos

A lo largo de este TFM se han ido analizando diferentes aspectos de los patrones puntuales y su posible aplicación en la demanda de emergencias. El objetivo principal de este TFM es el de proporcionar a las herramientas de iReves con información de mayor calidad para que éstas puedan mejorar sus soluciones, por lo tanto, es necesario construir un modelo que permita realizar esa inferencia. A lo largo de este trabajo se ha podido exponer las particularidades de la base de datos de llamadas de emergencia en la provincia de Valencia para el año 2019 y sus características en términos de patrones puntuales: los datos siguen un patrón de puntos con intensidad inhomogénea y sus puntos tienden a aparecer más cercanos los unos de los otros en comparación con un proceso de CSR. Se ha decidido utilizar un modelo basado en la distribución de Poisson ya que se trabaja con la suposición de que los puntos son independientes los unos de los otros, lo que permite simplificar el análisis estadístico y permite aplicar más técnicas que otros procesos. Los procesos de Poisson son descritos completamente por su función de intensidad  $\lambda(u)$ , por lo que los modelos buscarán describir la ecuación que da forma a esta intensidad (Baddeley, Rubak y Turner 2015b). Además, ha sido imposible aplicar otras distribuciones ya que la gran cantidad de puntos en la base de datos provocaba que no hubiera suficiente espacio en la memoria del ordenador y las instrucciones no compilaran. Esta limitación de hardware ha estado muy presente a lo largo de todo el desarrollo de este trabajo ya que varias instrucciones limitaban muchas de sus opciones a bases de datos con tan solo 1.000 puntos, aproximadamente cien veces menos de los que se disponen en este caso de estudio.

### 5.2.1. Modelos de Poisson

Una vez demostrado que los datos no siguen un proceso de Poisson homogéneo (CSR) se puede proceder a modelar la función de intensidad de la demanda de VES. Esto se realiza utilizando la instrucción `ppm` de la librería `spatstat` (Baddeley y Turner 2005) y su construcción es análoga al ajuste de modelos estándar en R con `lm` para modelos lineales y `glm` para modelos lineales generalizados. En primer lugar, se crea un esquema de cuadratura con la instrucción `quadscheme`, que introduce una serie de puntos dummy al set de datos y permite realizar el ajuste de los modelos con `ppm`.

#### Modelo 0

El modelo cero (Ecuación 5.8) corresponde a un proceso de Poisson homogéneo o CSR. Se observa en resultado expresado la Figura 5.8 que el término de intersección  $\beta_0$  es significativo. Aunque se ha demostrado anteriormente que los datos no siguen un proceso de Poisson homogéneo, se construye este modelo con el objetivo de comparar si los modelos con intensidades inhomogéneas son significativamente mejores a la hora de estimar la función de intensidad de la demanda de VES que este modelo que considera la intensidad como homogénea.

$$\log(\lambda) = \beta_0 \quad (5.8)$$

```
Stationary Poisson process
Fitted to point pattern dataset 'Qe'
Intensity: 87159.25
      Estimate      S.E. CI95.lo CI95.hi Ztest   Zval
log(lambda) 11.37549 0.003186168 11.36925 11.38174 *** 3570.274
```

**Figura 5.8:** Resultado del modelo  $M_0$ .

**Fuente:** Elaboración propia en R.

**Modelo 1**

El primer modelo busca analizar el posible efecto espacial a través de las coordenadas de los puntos, donde  $X$  es la latitud e  $Y$  es la longitud, y su interacción (Ecuación 5.9). Tal y como se ve en la Figura 5.9, todos sus parámetros son significativos.

$$\log(\lambda(u)) = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 XY \tag{5.9}$$

```

Nonstationary Poisson process
Fitted to point pattern dataset 'qe'

Log intensity: ~x + y + x * y

Fitted trend coefficients:
(Intercept)      x          y          x:y
-91.833059 -98.215978  2.701378  2.629779

(Intercept) Estimate S.E. CI95.lo CI95.hi Ztest Zval
x -98.215978 1.94618610 -102.030433 -94.401524 *** -50.46587
y 2.701378 0.02271379 2.656860 2.745897 *** 118.93123
x:y 2.629779 0.04943989 2.532879 2.726680 *** 53.19145
    
```

**Figura 5.9:** Resultado del modelo  $M_1$ .

**Fuente:** Software R.

**Modelo 2**

El segundo modelo introduce como covariable la densidad de población ya que tiene una correlación muy elevada con la demanda de llamadas de emergencia (Vecina García 2022). En un principio, esta covariable se introdujo como función de distancia hasta los núcleos urbanos de la provincia tal y como se ha hecho con otras bases de datos como el estudio del contagio de cólera con los datos del Doctor John Snow (Snow 1988). Finalmente se descartó esta idea al conseguir un mapa ráster con la densidad de población en España (Copernicus 2022), seleccionado la zona de la Provincia de Valencia y convirtiéndolo en una imagen pixelada, con la instrucción `as.im.RasterLayer` debido a las condiciones impuestas por la instrucción `ppm`. Esta imagen pixelada que contiene la densidad de población se puede ver como covariable en la Ecuación 5.10. De nuevo se observa en la Figura 5.10 que todos los parámetros son significativos.

$$\log(\lambda(u)) = \beta_0 + \beta_1 Densidad \tag{5.10}$$

```

Nonstationary Poisson process
Fitted to point pattern dataset 'qe'

Log intensity: ~densidad

Fitted trend coefficients:
(Intercept) densidad
10.890368897 0.004717174

(Intercept) Estimate S.E. CI95.lo CI95.hi Ztest Zval
densidad 0.004717174 5.087938e-06 0.004707202 0.004727146 *** 927.1289
    
```

**Figura 5.10:** Resultado del modelo  $M_2$ .

**Fuente:** Software R.

**Modelo 3**

El último modelo propone unificar los modelos  $M_1$  y  $M_2$ , tal y como se observa en la Ecuación 5.11. Finalmente, se vuelve a obtener en el resultado del modelo de la Figura 5.11 que todos los parámetros son significativos y no se descarta ninguno.

$$\log(\lambda(u)) = \beta_0 + \beta_1 \text{Densidad} + \beta_2 X + \beta_3 Y + \beta_4 XY \tag{5.11}$$

```

Nonstationary Poisson process
Fitted to point pattern dataset 'qe'

Log intensity: ~densidad + x + y + x * y

Fitted trend coefficients:
(Intercept)  densidad      x          y      x:y
-59.401926727  0.003976727 -72.337579960  1.852137499  1.950460506

(Intercept)  Estimate      S.E.      CI95.lo      CI95.hi  Ztest  Zval
densidad    -59.401926727  1.077296e+00 -61.513387790 -57.290465663 *** -55.13985
densidad     0.003976727  5.949227e-06  0.003965067  0.003988387 *** 668.44428
x            -72.337579960  2.212562e+00 -76.674122400 -68.001037520 *** -32.69403
y            1.852137499  2.740385e-02  1.798426948  1.905848050 *** 67.58677
x:y         1.950460506  5.624335e-02  1.840225573  2.060695438 *** 34.67896
    
```

**Figura 5.11:** Resultado del modelo  $M_3$ .

**Fuente:** Elaboración propia en R.

**ANOVA**

Una vez se han propuesto los modelos y se ha comprobado que todos sus parámetros son significativos, se debe comprobar también que éstos son significativamente mejores a la hora de estimar la función de intensidad que un proceso homogéneo de Poisson, que correspondería con el modelo  $M_0$ . Para ello se van a realizar tres test ANOVA por pares, comparando cada uno de los modelos ( $M_1$ ,  $M_2$  y  $M_3$ ) con el modelo  $M_0$  utilizando la instrucción `anova.ppm` y seleccionando el test de razón de verosimilitud (LRT). Tal y como se observa en las Figuras 5.12, 5.13 y 5.14, para los tres test ANOVA se obtienen p-valores muy bajos e inferiores al nivel de significación, por lo que se puede decir que los tres modelos propuestos son significativamente mejores a la hora de estimar la función de intensidad de la demanda de llamadas de emergencia en comparación con un proceso de Poisson homogéneo. Dado que el test `anova.ppm` compara modelos anidados, también se comprueba si el modelo  $M_3$  es significativamente mejor que los modelos  $M_1$  y  $M_2$ . Se puede ver en las Figuras 5.15 y 5.16, que se obtienen p-valores muy bajos y que el  $M_3$  se posiciona como significativamente superior a los otros dos.

```

Analysis of Deviance Table

Model 1: ~1      Poisson
Model 2: ~x + y + x * y      Poisson
  Npar Df Deviance Pr(>Chi)
1     1
2     4 3  180014 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
    
```

**Figura 5.12:** Resultado del test ANOVA entre los modelos  $M_0$  y  $M_1$ .

**Fuente:** Elaboración propia en R.



```

Analysis of Deviance Table

Model 1: ~1      Poisson
Model 2: ~densidad      Poisson
  Npar Df Deviance Pr(>Chi)
1     7
2     8 1  308996 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Figura 5.13:** Resultado del test ANOVA entre los modelos  $M_0$  y  $M_2$ .

**Fuente:** Elaboración propia en R.

```

Analysis of Deviance Table

Model 1: ~1      Poisson
Model 2: ~densidad + x + y + x * y      Poisson
  Npar Df Deviance Pr(>Chi)
1     7
2    11 4  394881 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Figura 5.14:** Resultado del test ANOVA entre los modelos  $M_0$  y  $M_3$ .

**Fuente:** Elaboración propia en R.

```

Analysis of Deviance Table

Model 1: ~x + y + x * y      Poisson
Model 2: ~densidad + x + y + x * y      Poisson
  Npar Df Deviance Pr(>Chi)
1    10
2    11 1  214668 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Figura 5.15:** Resultado del test ANOVA entre los modelos  $M_1$  y  $M_3$ .

**Fuente:** Elaboración propia en R.

```

Analysis of Deviance Table

Model 1: ~densidad      Poisson
Model 2: ~densidad + x + y + x * y      Poisson
  Npar Df Deviance Pr(>Chi)
1     2
2     5 3   85885 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

**Figura 5.16:** Resultado del test ANOVA entre los modelos  $M_2$  y  $M_3$ .

**Fuente:** Elaboración propia en R.

Aunque se haya demostrado con los test ANOVA que el modelo  $M_3$  es superior en términos de estimación de la función de intensidad, en este trabajo se va a estudiar otra forma de comparar y evaluar los distintos modelos: haciendo uso de estadísticos. Se puede ver un resumen de los estadísticos para cada modelo en la Tabla 5.1. Los datos del modelo  $M_0$  se presentan con el objetivo de ser informativos.

**Pseudo  $R^2$**

En primer lugar, está el valor  $pseudoR^2$  que mide el porcentaje de la variabilidad de los datos que es capaz de explicar el modelo propuesto teniendo en cuenta que la variable respuesta no es continua. Dado que el  $M_0$  no tiene ninguna variable explicativa tiene sentido que su valor de  $pseudoR^2$  sea cero, del mismo modo que se observa que conforme se añaden más variables explicativas al modelo, mayor valor de  $pseudoR^2$  se obtiene.

**Área Bajo la Curva (AUC)**

En procesos puntuales, el área bajo la curva (AUC) mide la capacidad que tienen las covariables para separar el dominio espacial en áreas de baja y alta densidad de puntos (Baddeley y Turner 2005). El modelo  $M_0$  con un AUC de 0,5 indica que está seleccionando aleatoriamente estas áreas de densidad, sería como lanzar una moneda al aire y asignar el resultado según si se ha obtenido cara o cruz. En los modelos más sofisticados se obtienen mejores valores de AUC, especialmente en los modelos  $M_2$  y  $M_3$ .

**Criterio de Información de Akaike (AIC)**

Finalmente, si se estudia el valor del criterio de información de Akaike (AIC), que es un estimador que mide a su vez la complejidad del modelo propuesto y su bondad de ajuste, se observa que el modelo  $M_3$  es el mejor. Por lo tanto, se llega a la misma conclusión que con el test ANOVA y se selecciona el modelo  $M_3$  como el mejor a la hora de estimar la función de intensidad de demanda de llamadas de emergencia.

	$M_0$	$M_1$	$M_2$	$M_3$
$PseudoR^2$	0	0,248	0,426	0,544
AUC	0,5	0,864	0,966	0,923
AIC	-2044e3	-2225e3	-2354e3	-2440e3

**Tabla 5.1:** Resumen de los estadísticos para cada modelo.

**Fuente:** Elaboración propia.

**5.2.2. Validación del Modelo**

Tras haber ajustado el modelo de procesos puntuales es necesario verificar que este modelo es capaz de representar de manera fiel a los datos debido a que cabe la posibilidad de que éste no sea capaz de recoger de manera fidedigna el patrón subyacente de los datos. Haber obtenido buenos valores en los estadísticos puede hacer creer que se tiene un buen modelo cuando la realidad es muy distinta. Por esta razón se debe validar el modelo  $M_3$  (Ecuación 5.11) realizando una serie de tests que aseguren el buen planteamiento y ajuste del modelo a los datos.

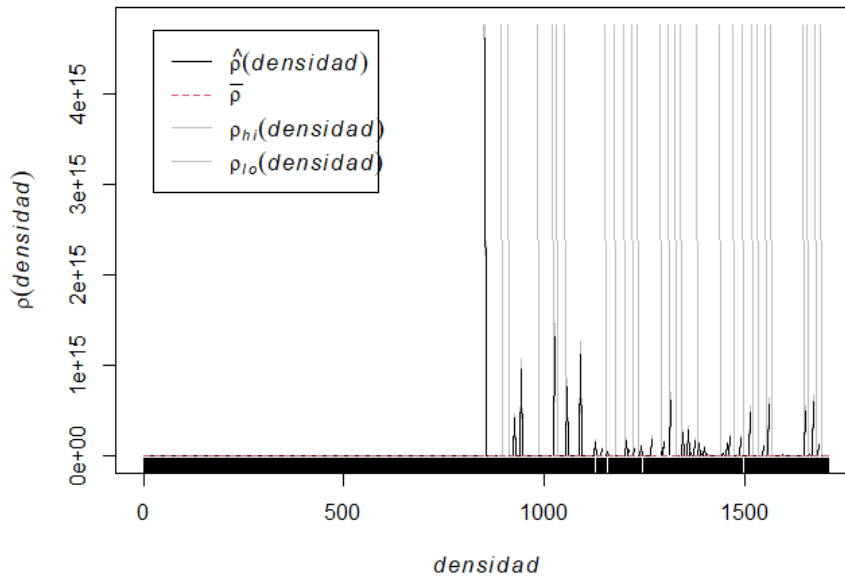
**Intensidad Relativa**

Se entiende como intensidad relativa  $r(u)$  a la relación entre la verdadera función de intensidad  $\lambda(u)$  y su función de intensidad modelada  $\lambda_0(u)$  de un proceso puntual  $X$ . Se puede ver la expresión en la Ecuación 5.12. Valores de  $r(u)$  cercanos a 1 indicarán un buen ajuste, mientras que valores superiores a 1 indican que se está infravalorando la función de intensidad (Baddeley, Rubak y Turner 2015a). Utilizando la instrucción `density`, se puede calcular una estimación del valor de la intensidad relativa  $\hat{r}(u)$ . Se obtiene un rango de  $\hat{r}(u)$  entre 0,093 y 2,25, lo que indica que se está infravalorando la verdadera función de intensidad y el modelo propuesto no es un buen ajuste. Por otro lado, se puede comprobar cómo se ajusta la distribución observada de los valores de una covariable espacial con la distribución esperada, si se asume que la verdadera intensidad de un proceso

viene dada por la Ecuación 5.13, donde  $\rho$  es una función desconocida (Baddeley, Rubak y Turner 2015b). Esta función  $\rho$  se puede estimar con la instrucción `rhohat.ppp` y su resultado se ve en la Figura 5.17. Se observa que para densidades de población muy altas la intensidad está muy infravalorada.

$$ru(u) = \frac{\lambda(u)}{\lambda_0(u)} \tag{5.12}$$

$$\lambda(u) = \rho(Z(u))\lambda_0(u) \tag{5.13}$$

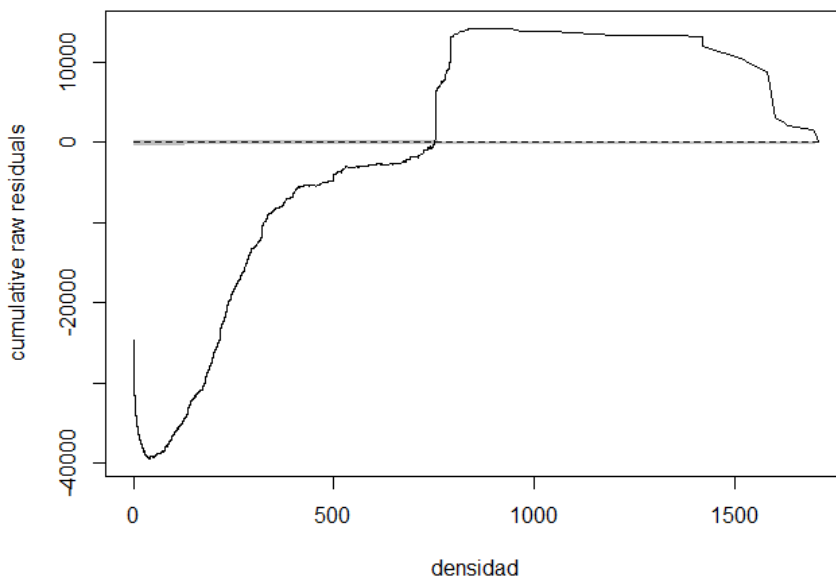


**Figura 5.17:** Estimación de la función  $\rho$ .

**Fuente:** Elaboración propia en R.

### Gráfico de lurking

El gráfico de *lurking* ayuda a identificar la importancia de una covariable del modelo y se puede obtener en R con la instrucción `lurking`. Se observa en la Figura 5.18, que el modelo  $M_3$  está sobrestimando la intensidad en densidades de población bajas y subestimando en altas, lo que indica que la densidad de población es una covariable importante a la hora de estimar la función de intensidad, o tal vez lo sea también otra covariable relacionada con ésta, y el modelo no está recogiendo bien su efecto.

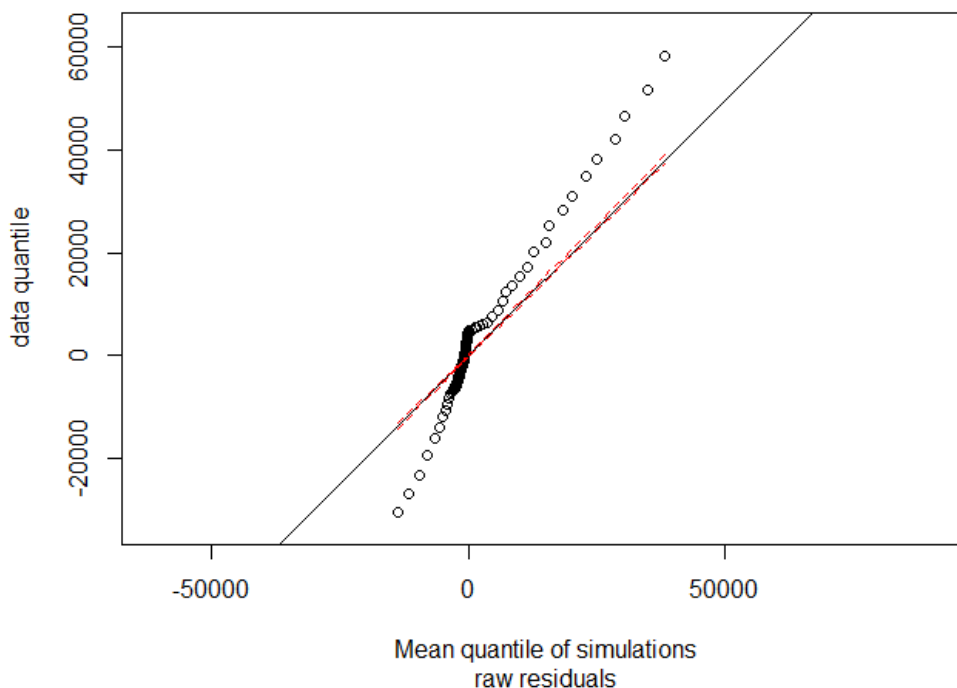


**Figura 5.18:** Gráfico de *lurking* para la covariable densidad de población.

**Fuente:** Elaboración propia en R.

**Gráfico Q-Q**

Finalmente, si se estudia el gráfico Q-Q de la Figura 5.19, se llega a la misma conclusión que el modelo planteado no presenta un buen ajuste a los datos y hay efectos que el modelo no está recogiendo.



**Figura 5.19:** Gráfico Q-Q de los residuos suavizados para el modelo  $M_3$ .

**Fuente:** Elaboración propia en R.

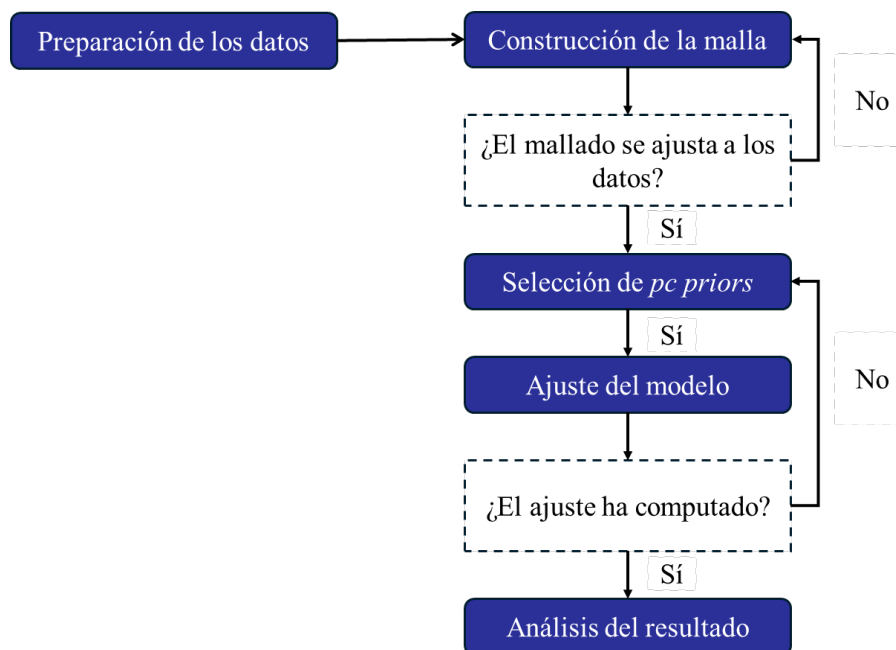
En conclusión, el modelo  $M_3$  que se ha propuesto no proporciona un buen ajuste a los datos por lo que no se puede utilizar para predecir la demanda de llamadas de emergencia y debe descartarse. En este momento se decide plantear modelos más sofisticados ya que, tal y como se ha presentado en el capítulo 3, varios estudios utilizan procesos log cox-gaussianos (LGCP) o modelos de clustering para sus casos de estudio. En este momento se presentó un problema, a pesar de que la librería `spatstat` proporciona la instrucción `kppm` para el planteamiento de estos modelos, no era posible hacer uso de ella ya que la magnitud de los datos implicaba que no había suficiente memoria en el ordenador para ejecutar la instrucción. Se plantearon diversas soluciones:

- En primer lugar, dividir la base de datos en secciones más reducidas, por mes o por semana, y plantear modelos para cada uno de esos subconjuntos de datos siguiendo la misma metodología que aquí se ha presentado. Esta opción se descartó debido a que se desea continuar con este estudio en el futuro y analizar también la componente temporal para proponer soluciones más detalladas. Las limitaciones de hardware impiden que se pueda seguir con esta metodología.
- En segundo lugar, se planteó realizar agrupaciones de puntos ya sea por distrito médico, por proximidad o por diferentes propuestas. Aunque no se descarta esta propuesta de manera permanente, esto implicaría que el problema dejaría de estar recogido en procesos puntuales y pasaría a ser un problema de datos agregados tal y como se presentaba en el Capítulo 2 con una metodología completamente diferente.
- En tercer lugar, se plantea proponer un modelo basado en inferencia bayesiana que permite realizar modelado espacial utilizando aproximaciones integradas anidadas de Laplace con el paquete de R `inlabru` (Bachl et al. 2019) que permite de forma sencilla aplicar diferentes efectos al modelo. Se decide adoptar esta solución para la continuación de este trabajo.

## Capítulo 6

# Perspectiva Bayesiana

En esta sección se va a presentar la metodología de inferencia bayesiana como alternativa a la propuesta frecuentista. En primer lugar se va a preparar la base de datos para poder utilizar las instrucciones de la librería `inlabru`, interfaz que permite el ajuste de modelos espaciales aplicando técnicas basadas en la inferencia bayesiana. Tras preparar los elementos necesarios para ejecutar el ajuste, la malla y la Matérn, se ajusta el modelado basado en un proceso de Cox log-gaussiano y se analizan los resultados obtenidos. El proceso de trabajo que se ha seguido queda resumido en el esquema de la Figura 6.1.



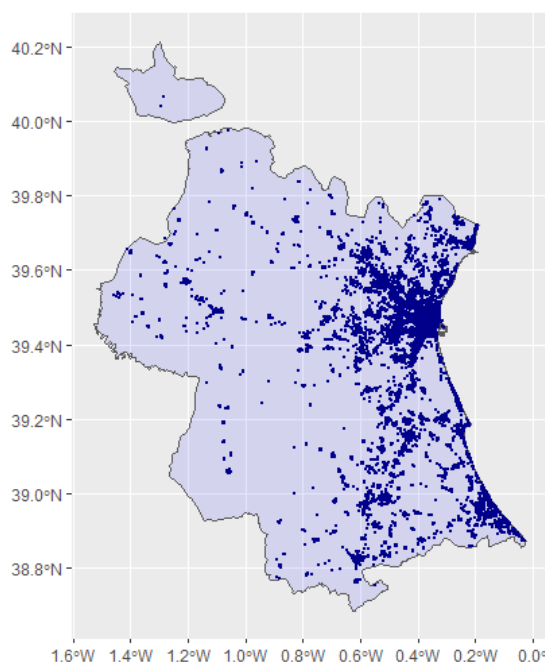
**Figura 6.1:** Esquema de la metodología seguida con la metodología de perspectiva bayesiana.

**Fuente:** Elaboración propia.

### 6.1. Preparación de los datos

Ha sido necesario realizar un nuevo tratamiento de los datos para poder utilizar las instrucciones disponibles en la librería `inlabru` debido a que ésta requiere que sus entradas sean de tipo *Simple Features* (`sf`). Para ello se ha construido un objeto de tipo `sf` (Pebesma 2018) que contuviera la información relativa a las llamadas de

emergencia y otro objeto `sf` que contuviera la frontera donde se posicionan los datos, que en este caso es la delimitación de la Provincia de Valencia. Se ha tenido que seleccionar un sistema de referencia de coordenadas (CRS) que pudiera identificar correctamente las coordenadas de los puntos y de la frontera. El CRS seleccionado es el *WGS 84 (World Geodetic System 1984)* y ha sido aplicado a todos aquellos elementos que lo han necesitado: el objeto de datos, el mapa de la Provincia de Valencia y la malla. Se puede ver una representación del objeto de datos y de la frontera en la Figura 6.2.



**Figura 6.2:** Representación de las llamadas de emergencias en la Provincia de Valencia.

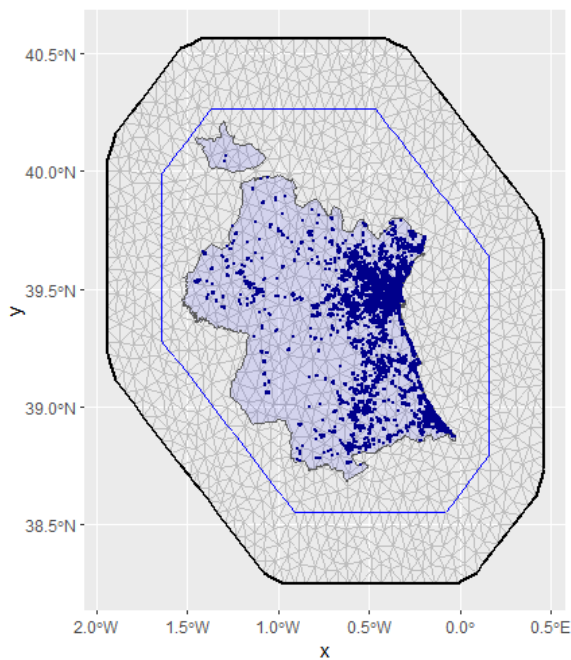
**Fuente:** Elaboración propia en R.

## 6.2. Mallado

Una vez ya se han preparado los datos de forma correcta es indispensable construir una malla que cubra la región de estudio para poder plantear el modelo LGCP con `inlabru`. Con este fin, se ha seguido el procedimiento indicado en la guía escrita por Krainski et al. (2018). Existen contables alternativas a la hora de construir el mallado (Belmont 2022), pero se ha decidido optar por un mallado con una capa exterior para evitar el posible efecto frontera provocado por todos los puntos que están cercanos a las zonas limítrofes de la región de estudio. Se puede ver el resultado del mallado junto con el conjunto de datos y el mapa que delimita región de estudio en la Figura 6.3.

## 6.3. Matriz Matérn

El siguiente paso consiste en introducir los valores a priori a la matriz Matérn. La función de covarianza Matérn es muy utilizada en análisis de datos espaciales (Porcu et al. 2024). Sirve para modelar la correlación entre medidas en localizaciones geoespaciales y permite hacer predicciones de localizaciones que no forman parte del set de datos pero sí están en la región de estudio.



**Figura 6.3:** Representación de las llamadas de emergencias en la Provincia de Valencia junto con el mallado.

**Fuente:** Elaboración propia en R.

Para la construcción de la Matérn, es necesario seleccionar dos valores a priori: el rango (*range*) y la desviación típica (*sigma*). El rango hace referencia a la distancia límite donde se puede considerar que dos puntos están relacionados espacialmente. La desviación típica está directamente relacionada con el predictor local y delimita entre que intervalo puede variar el número medio de ocurrencias. Es importante realizar una buena elección de los valores a priori (*pc priors*) ya que una mala elección puede provocar que el LGCP no llegue a ningún resultado o que éste no sea válido. En este trabajo se han seleccionado valores similares a los propuestos en la guía de Krainski et al. (2018), aunque se desea en el futuro desarrollo del proyecto ahondar más en este aspecto. Los valores seleccionados se pueden consultar en la Tabla 6.1.

<b>prior.range</b>	$\rho_0 = 0,05$	$\rho = 0,01$
<b>prior.sigma</b>	$\sigma_0 = 0,111$	$\sigma = 0,01$

**Tabla 6.1:** Resumen de los valores seleccionados a priori.

**Fuente:** Elaboración propia.

El `prior.range` es un vector de dos elementos que especifica los valores de la Ecuación 6.1, donde  $\rho$  es el rango espacial,  $\rho_0$  es el rango seleccionado a priori y  $\lambda_1$  es la probabilidad. Según los valores de la Tabla 6.1, se está indicando que la probabilidad de que el rango sea menor de 5,5 kilómetros es del 1%. Nótese que el rango está expresado en grados y es necesario multiplicar su valor por 110 para hacer una conversión aproximada de grados a distancia ya que 1 grado equivale aproximadamente a 110 km.

$$P(\rho < \rho_0) = \lambda_1 \tag{6.1}$$

El `prior.sigma` es también un vector de dos elementos que especifica los valores de la Ecuación 6.2, donde  $\sigma$  es la desviación típica,  $\sigma_0$  es la desviación típica seleccionada a priori y  $\lambda_2$ . Según los valores de la Tabla 6.1, se está indicando que la probabilidad de que la desviación típica sea menor de aproximadamente 11.000



llamadas es del 1 %. Para obtener el número de llamadas se ha multiplicado el número de llamadas totales por el valor seleccionado de  $\sigma_0$ .

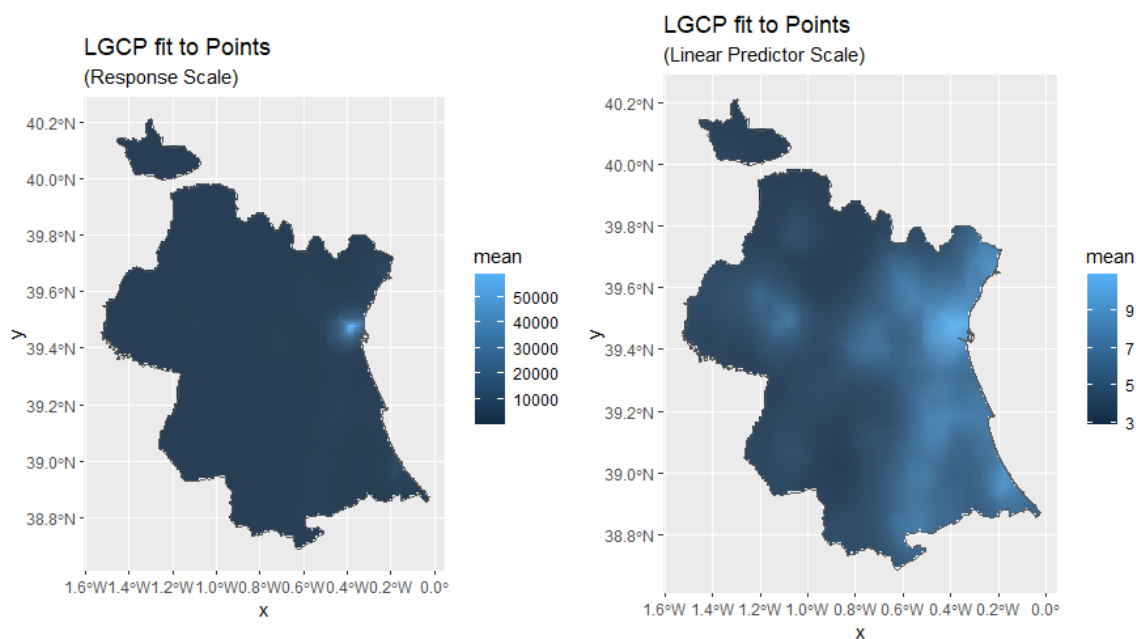
$$P(\sigma < \sigma_0) = \lambda_2 \quad (6.2)$$

## 6.4. Modelo y Resultados

Tras haber construido todos los objetos necesarios se procede a ajustar un modelo a la demanda de llamadas de emergencia en la Provincia de Valencia con la instrucción `lgcp` de la librería `inlabru`. El modelo propuesto viene definido en la Ecuación 6.3, donde  $\eta_i$  es el predictor lineal,  $\beta_0$  es el parámetro independiente,  $\omega_i$  es el efecto espacial.

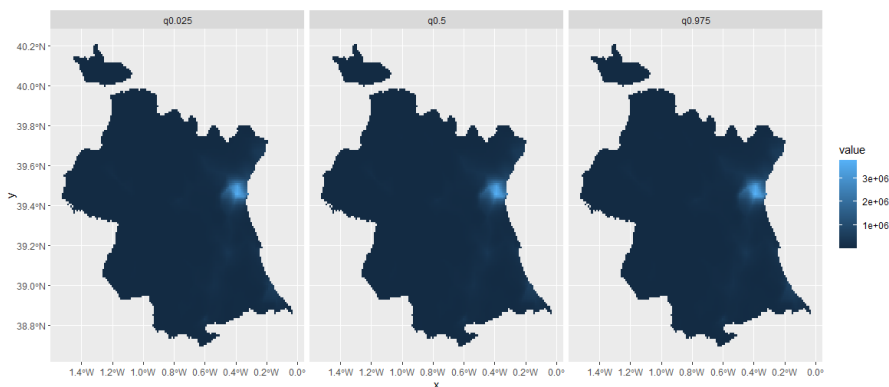
$$\eta_i = \beta_0 + \omega_i \quad (6.3)$$

Una vez introducida la ecuación del modelo en R, se ejecuta la instrucción `lgcp` para el ajuste y se comprueba que el método ha proporcionado una solución. Tras este paso, es posible realizar una predicción de la intensidad utilizando el modelo LGCP que ha sido ajustado a los datos. Esta predicción se puede visualizar en el gráfico de la izquierda de la Figura 6.4, donde se observa que la gran mayoría de llamadas se concentra en la zona urbana de Valencia y, aunque apenas perceptible, cerca de la costa. En la parte derecha de la Figura 6.4 se puede observar la evolución del valor medio del predictor lineal a lo largo de todo el territorio. También es posible representar la mediana y las superficies de densidad inferior al 95 % y superior al 95 % que aparecen en la Figura 6.5 aunque la diferencia es muy pequeña debido a la alta concentración de llamadas en Valencia. Para el futuro sería interesante plantear la utilización de una escala variable para mostrar las diferencias de forma más clara.



**Figura 6.4:** Predicción de las llamadas de emergencias en la Provincia de Valencia.

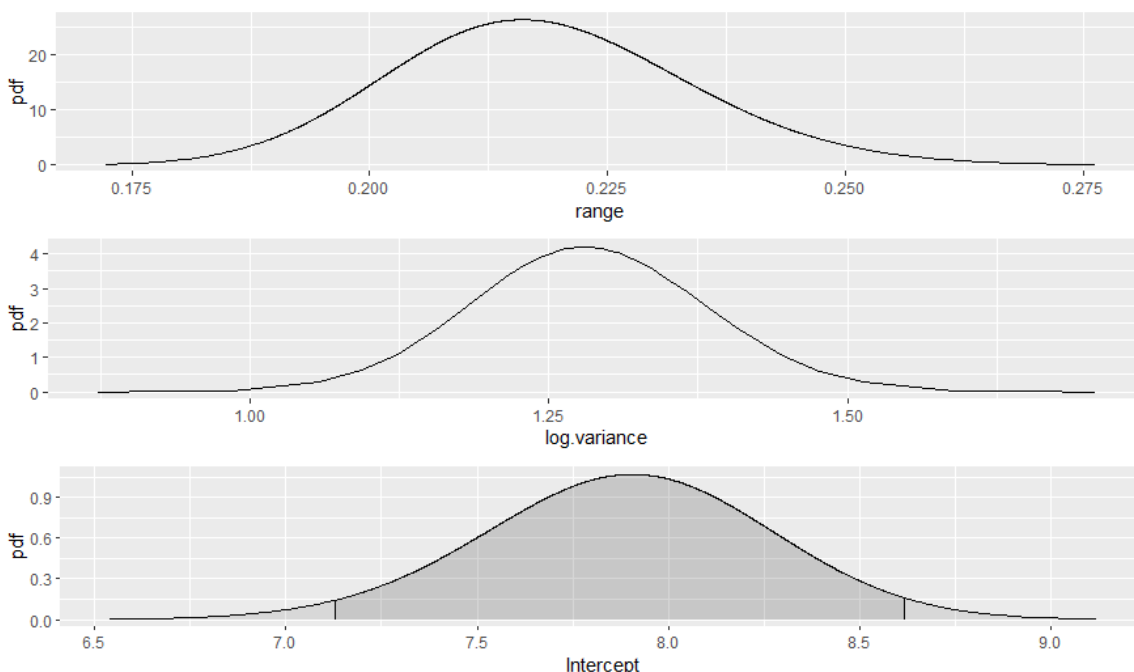
**Fuente:** Elaboración propia en R.



**Figura 6.5:** Representación de la mediana y superficies de densidad.

**Fuente:** Elaboración propia en R.

Se puede representar las funciones de densidad de probabilidad del rango, la desviación típica y el parámetro independiente en un gráfico como el de la Figura 6.6. El valor medio del rango está cercano a 0,21 lo que estaría diciendo que dos puntos en un rango de aproximadamente 22 kilómetros estarían relacionados espacialmente, un valor bastante elevado teniendo en cuenta que el mapa de la Provincia de Valencia abarca una distancia de entre 150 y 170 kilómetros. Con respecto a la desviación típica y al parámetro independiente se obtiene unos valores medios de aproximadamente 1,3 y 7,8 aproximadamente, lo que implicaría que la media de llamadas de emergencia variaría entre los valores 6,5 y 9,1, teniendo en cuenta la naturaleza de los datos se podría decir que la media de llamadas varía entre 7 y 9.



**Figura 6.6:** Funciones de densidad de probabilidad del rango, la desviación típica y el parámetro independiente.

**Fuente:** Elaboración propia en R.

Es posible estimar el número de llamadas de emergencia en la Provincia de Valencia utilizando la función `predict`. En la Tabla 6.2 se puede ver un resumen de los valores predichos obtenidos. Destacar que, mientras que el valor real de las llamadas de emergencia es de 98.594, la media obtenida por el modelo es de 98.775 llamadas, un valor muy cercano al real.

Media	$\sigma$	Q1	Q2	Q3	Mediana	Error Media	Error $\sigma$
98.775,08	352.14	98.090,91	98.770,01	99.452,43	98.770,01	35,21	26,23

**Tabla 6.2:** Resumen de la predicción de llamadas de emergencia obtenidas.

**Fuente:** Elaboración propia.

Ha sido muy interesante estudiar las técnicas disponibles en la librería `inlabru` y la perspectiva bayesiana a la hora de resolver problemas tan complejos como los espaciales. Es una metodología que ha funcionado muy bien a pesar de disponer de una base de datos de llamadas de emergencias relativamente grande. Por todo esto, se ha decidido mantener el interés por esta metodología aplicando alternativas al modelado aquí propuesto e introduciendo nuevos datos a los que se ha tenido acceso. Finalmente, cabe destacar que esta metodología todavía no ha sido aplicada a los problemas espaciales de demanda de VES por lo que sería una aportación interesante al estado del arte.

## Capítulo 7

# Conclusiones y futuras líneas de investigación

### 7.1. Conclusiones

Es importante que los organismos responsables de los servicios de urgencias y emergencias dispongan de herramientas que aseguren una respuesta rápida y eficaz ante eventos que requieran un despliegue de Vehículos de Emergencias Sanitarias (VES). Tener un servicio de calidad no solo repercutirá en el estado de bienestar de su población, sino que además es un derecho fundamental al que todo el mundo debería tener acceso. Aportar información sobre la demanda de llamadas de emergencia a los algoritmos de ubicación/reubicación de VES podría implicar mejores soluciones y una mejor preparación ante la incertidumbre de lo que podría pasar.

Tal y como se ha visto, estudiar la demanda de llamadas de emergencia en un territorio es un problema completo de por sí. No existe un método único que resuelva este tipo de problema ya que cada región espacial tiene una casuística distinta que requiere de un tratamiento particular. En este trabajo se ha estudiado el caso concreto de la Provincia de Valencia, territorio que no había sido analizado anteriormente, por lo que se ha podido proporcionar una visión introductoria de dónde se concentran esas emergencias: cerca del núcleo urbano de Valencia.

La base de datos proporcionada por el Servicio de Emergencias Sanitarias de la Comunidad Valenciana es muy interesante y contiene datos que, analizados de forma paralela podrían aportar información que después se podría trasladar a los modelos.

Una parte bastante importante del desarrollo de este trabajo se ha empleado en la preparación de los datos en el software R para poder cumplir los requerimientos de las diferentes instrucciones y en buscar librerías que estuvieran disponibles para realizar los análisis. No ha sido posible el análisis de procesos puntuales sin utilizar librerías descatalogadas, lo que ha supuesto un problema debido a que, en ocasiones, ha sido complicado encontrar un lugar donde obtener los paquetes para después instalarlos de forma manual en R.

Se ha podido demostrar de diversas formas que la función de intensidad de la demanda de llamadas de emergencia en la Provincia de Valencia no sigue un proceso de Poisson homogéneo y tiene algún tipo de patrón de datos subyacentes utilizando técnicas de análisis de procesos puntuales.

Se han propuesto diversos modelos basados en distribuciones de Poisson para intentar estimar la función de intensidad de la demanda pero, a pesar de que eran significativamente mejores a la hora de estimar que un proceso de Poisson homogéneo, no ha sido posible utilizarlos ya que no han superado el proceso de validación.

La combinación de la magnitud de los datos con la limitación de hardware ha sido un problema significativo en el desarrollo de este trabajo. Intentar proponer modelos más sofisticados u otras instrucciones con la metodología descrita en el manual de Baddeley, Rubak y Turner (Baddeley, Rubak y Turner ) ha sido imposible ya que el ordenador no tenía la memoria suficiente para tratar con la gran cantidad de datos. A pesar de que se buscaron alternativas para no cambiar la metodología, al final se decidió descartarla.

Los resultados obtenidos con la librería `inlabru` y el método de inferencia bayesiana con aproximación integrada anidada de Laplace (INLA) han resultado muy interesantes y prometedores. A pesar de que inicialmente era un campo completamente desconocido ha sido posible plantear un modelo basado en patrones log cox-gaussianos (LGCP) que ha estimado una demanda muy cercana a la demanda real.

Finalmente, destacar que todavía quedan posibles metodologías para aplicar al estudio de la demanda de llamadas de emergencia de la Provincia de Valencia por lo que sería interesante analizar estas metodologías en conjunto con el equipo encargado de los algoritmos de ubicación/reubicación y con las personas encargadas de tomar las decisiones relacionadas con la flota de VES para seleccionar aquellas que se acomoden mejor a sus necesidades.

## **7.2. Futuras líneas de investigación**

En este trabajo ha sido posible presentar un análisis introductorio de la demanda de VES de la Provincia de Valencia, pero como bien indica la palabra, queda todavía mucho por estudiar y datos que explotar.

La metodología de inferencia bayesiana tiene mucho potencial, especialmente debido a la facilidad que tiene a la hora de modelar efectos complejos de los datos. Dado que el Servicio de Emergencias Sanitarias de la Comunidad Valenciana ha podido proporcionar datos relativos a más años, es interesante proponer una línea de investigación que analice y construya modelos más sofisticados que incluyan posibles efectos temporales y de covariables.

Tal y como se presentaba, sería posible aplicar otras técnicas de estudio espacial como la metodología de datos agregados entre otros y comprobar cómo varían los resultados y cuáles se adecuarían mejor a las necesidades de los organismos decisores.

# Bibliografía

- WHO (2023). “Tracking Universal Health Coverage: 2023 Global monitoring report: Executive summary”. En: SNS (2024). *Principales Datos del Sistema Nacional de Salud*.
- CV, SES (s.f.). Acceso: 20-02-2024. url: <https://ses.san.gva.es/ca/descripcio-de-recursos>.
- O’Keeffe, Colin, Jon Nicholl, Janette Turner y Steve Goodacre (2011). “Role of ambulance response times in the survival of patients with out-of-hospital cardiac arrest”. En: *Emergency medicine journal* 28.8, págs. 703-706.
- Blackwell, Thomas H y Jay S Kaufman (2002). “Response time effectiveness: comparison of response time and survival in an urban emergency medical services system”. En: *Academic Emergency Medicine* 9.4, págs. 288-295.
- García, Sergio y Alfredo Marín (2015). “Covering Location Problems”. En: *Location Science*. Ed. por Gilbert Laporte, Stefan Nickel y Francisco Saldanha da Gama. Cham: Springer International Publishing, págs. 93-114. isbn: 978-3-319-13111-5. doi: 10.1007/978-3-319-13111-5\_5. url: [https://doi.org/10.1007/978-3-319-13111-5\\_5](https://doi.org/10.1007/978-3-319-13111-5_5).
- Vecina García, M.A. et al. (2022). “A mixed integer linear program for a real relocation problem of emergency medical vehicles in the province of Valencia”. En: *J Appl Res Eng Technol and Engineering* 3(2), págs. 85-92. doi: 10.1007/978-3-319-13111-5\_5.
- Ripley, Brian D (2005). *Spatial statistics*. John Wiley y Sons.
- Cressie, Noel y Matthew T Moores (2022). “Spatial statistics”. En: *Encyclopedia of Mathematical Geosciences*. Springer, págs. 1-11.
- Gómez Rubio, Virgilio (2022). “Una introducción a la estadística espacial”. En: *Boletín de Estadística e Investigación Operativa* 38(1). url: <https://www.seio.es/beio/una-introduccion-a-la-estadistica-espacial/>.
- Cressie, Noel (1993). *Statistics for Spatial Data*. John Wiley y Sons, Ltd. Cap. 1, págs. 1-26. isbn: 9781119115151. doi: <https://doi.org/10.1002/9781119115151.ch1>.
- Ford, E.D., P.A. Mason y J. Pelham (1980). “Spatial patterns of sporophore distribution around a young birch tree in three successive years”. En: *Transactions of the British Mycological Society* 75.2, págs. 287-296. issn: 0007-1536. doi: [https://doi.org/10.1016/S0007-1536\(80\)80091-X](https://doi.org/10.1016/S0007-1536(80)80091-X). url: <https://www.sciencedirect.com/science/article/pii/S000715368080091X>.
- Baddeley, Adrian, Ege Rubak y Rolf Turner (2015a). *Spatial Point Patterns: Methodology and Applications with R*. London. url: <http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/>.
- (2015b). *Spatial point patterns: methodology and applications with R*. CRC press.
- Renner, Ian W., Jane Elith, Adrian Baddeley, William Fithian, Trevor Hastie, Steven J. Phillips, Gordana Popovic y David I. Warton (2015). “Point process models for presence-only analysis”. En: *Methods in Ecology and Evolution* 6.4, págs. 366-379. doi: <https://doi.org/10.1111/2041-210X.12352>. eprint: <https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12352>. url: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12352>.
- Rubin, Donald B., Andrew Gelman, John B. Carlin y Hal S. Stern (2004). *Bayesian data analysis*. eng. 2nd ed. Texts in statistical science. Boca Raton: Chapman Hall. isbn: 158488388X.

- Martínez-Minaya, Joaquín (2015). “Factores Climáticos y Espaciales Asociados con la Enfermedad ‘Citrus Black Spot’”. Análisis Bayesiano de la Dispersión de la Enfermedad en Sudáfrica”. Tesis de máster. Universidad de Valencia.
- Gómez-Rubio, V. (2020). *Bayesian Inference with INLA*. A Chapman & Hall book. CRC Press. isbn: 9781138039872. url: <https://books.google.es/books?id=DvTLyGEACAAJ>.
- Vile, Julie Leanne, Jonathan W Gillard, Paul R Harper y Vincent A Knight (2012). “Predicting ambulance demand using singular spectrum analysis”. En: *J Oper Res* 63, págs. 1556-1565. doi: 10.1057/jors.2011.160.
- Setzler, Hubert, Cem Saydam y Sungjune Park (2009). “EMS call volume predictions: A comparative study”. En: *Computers and Operations Research* 36.6, págs. 1843-1851. issn: 0305-0548. doi: <https://doi.org/10.1016/j.cor.2008.05.010>. url: <https://www.sciencedirect.com/science/article/pii/S0305054808001068>.
- Sasaki, Satoshi, Alexis J Comber, Hiroshi Suzuki y Chris Brunson (2010). “Using genetic algorithms to optimise current and future health planning - the example of ambulance locations”. En: *International Journal of Health Geographics* 9.4, págs. 1-10. doi: 10.1186/1476-072X-9-4.
- Aldrich, Carole A, John C Hisserich y Lester B Lave (1971). “An analysis of the demand for emergency ambulance service in an urban area”. En: *American journal of Public Health* 61.6, págs. 1156-1169. doi: 10.2105/ajph.61.6.1156.
- Siler, K. F. (1975). “Predicting demand for publicly dispatched ambulances in a metropolitan area”. En: *Health services research* 10.3, págs. 254-263.
- Kvålseth, T O y J M Deems (1979). “Statistical models of the demand for emergency medical services in an urban area.” En: *American Journal of Public Health* 69.3. PMID: 420371, págs. 250-255. doi: 10.2105/AJPH.69.3.250. eprint: <https://doi.org/10.2105/AJPH.69.3.250>. url: <https://doi.org/10.2105/AJPH.69.3.250>.
- Wrigley, Hannah, Steve George, Helen Smith, Helen Snooks, Alan Glasper y Eileen Thomas (2002). “Trends in demand for emergency ambulance services in Wiltshire over nine years: observational study”. En: *BMJ* 324.7338, págs. 646-647. issn: 0959-8138. doi: 10.1136/bmj.324.7338.646. eprint: <https://www.bmj.com/content/324/7338/646.full.pdf>. url: <https://www.bmj.com/content/324/7338/646>.
- Ziad Nehme, Emily Andrew y Karen Smith (2016). “Factors Influencing the Timeliness of Emergency Medical Service Response to Time Critical Emergencies”. En: *Prehospital Emergency Care* 20.6. PMID: 27487018, págs. 783-791. doi: 10.3109/10903127.2016.1164776. eprint: <https://doi.org/10.3109/10903127.2016.1164776>. url: <https://doi.org/10.3109/10903127.2016.1164776>.
- Hall, William K. (1971). “Management science approaches to the determination of urban ambulance requirements”. En: *Socio-Economic Planning Sciences* 5.5, págs. 491-499. issn: 0038-0121. doi: [https://doi.org/10.1016/0038-0121\(71\)90007-3](https://doi.org/10.1016/0038-0121(71)90007-3). url: <https://www.sciencedirect.com/science/article/pii/0038012171900073>.
- Aboueljjanane, L., E. Sahin y Z. Jemai (2013). “A review on simulation models applied to emergency medical service operations”. En: *Computers and Industrial Engineering* 66.4, págs. 734-750. issn: 0360-8352. doi: <https://doi.org/10.1016/j.cie.2013.09.017>. url: <https://www.sciencedirect.com/science/article/pii/S0360835213003100>.
- Eastwood, Kathryn, Ameer Morgans, Karen Smith, Angela Hodgkinson, Gareth Becker y Johannes Stoelwinder (2015). “A novel approach for managing the growing demand for ambulance services by low-acuity patients”. En: *Australian Health Review* 40.4, págs. 378-384.
- Baker, J. R. y K. E. Fitzpatrick (1986). “Determination of an Optimal Forecast Model for Ambulance Demand Using Goal Programming”. En: *Journal of the Operational Research Society* 37.11, págs. 1047-1059. doi: 10.1057/jors.1986.182. eprint: <https://doi.org/10.1057/jors.1986.182>. url: <https://doi.org/10.1057/jors.1986.182>.
- Channouf, Nabil, Pierre L'Ecuyer, Armann Ingolfsson y Athanassios N Avramidis (2007). “The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta”. En: *Health Care Management Science* 10.1, págs. 25-45.



- Matteson, David S, Mathew W McLean, Dawn B Woodard y Shane G Henderson (2011). “Forecasting emergency medical service call arrival rates”. En: *Annals of Applied Statistics* 5.2B, págs. 1379-1406.
- Grekousis, George y Ye Liu (2019). “Where will the next emergency event occur? Predicting ambulance demand in emergency medical services using artificial intelligence”. En: *Computers, Environment and Urban Systems* 76, págs. 110-122. issn: 0198-9715. doi: <https://doi.org/10.1016/j.compenvurbsys.2019.04.006>. url: <https://www.sciencedirect.com/science/article/pii/S0198971519300146>.
- Martin, R. Justin, Reza Mousavi y Cem Saydam (2021). “Predicting emergency medical service call demand: A modern spatiotemporal machine learning approach”. En: *Operations Research for Health Care* 28, pág. 100285. issn: 2211-6923. doi: <https://doi.org/10.1016/j.orhc.2021.100285>. url: <https://www.sciencedirect.com/science/article/pii/S2211692321000011>.
- Wang, Zhaonan, Tianqi Xia, Renhe Jiang, Xin Liu, Kyoung-Sook Kim, Xuan Song y Ryosuke Shibasaki (2021). “Forecasting Ambulance Demand with Profiled Human Mobility via Heterogeneous Multi-Graph Neural Networks”. En: *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, págs. 1751-1762. doi: 10.1109/ICDE51399.2021.00154.
- Li, Mengyu, Peter Vanberkel y Xiang Zhong (2022). “Predicting ambulance offload delay using a hybrid decision tree model”. En: *Socio-Economic Planning Sciences* 80, pág. 101146. issn: 0038-0121. doi: <https://doi.org/10.1016/j.seps.2021.101146>. url: <https://www.sciencedirect.com/science/article/pii/S0038012121001385>.
- Blanco Prieto, Jorge, Marina Ferreras González, Steven Van Vaerenbergh y Oscar Jesús Cosido Cobos (2024). “A Data Mining Approach for Health Transport Demand”. En: *Machine Learning and Knowledge Extraction* 6.1, págs. 78-97. issn: 2504-4990. doi: 10.3390/make6010005. url: <https://www.mdpi.com/2504-4990/6/1/5>.
- Lotwick, H. W. (1984). “Some Models for Multitype Spatial Point Processes with Remarks on Analysing Multitype Patterns”. En: *Journal of Applied Probability* 21.3, págs. 575-582. issn: 00219002. url: <http://www.jstor.org/stable/3213619> (visitado 27-02-2024).
- Ogata, Yoshihiko (1999). “Seismicity Analysis through Point-process Modeling: A Review”. En: *Seismicity Patterns, their Statistical Significance and Physical Meaning*. Ed. por Max Wyss, Kunihiko Shimazaki y Akihiro Ito. Basel: Birkhäuser Basel, págs. 471-507. isbn: 978-3-0348-8677-2. doi: 10.1007/978-3-0348-8677-2\_14. url: [https://doi.org/10.1007/978-3-0348-8677-2\\_14](https://doi.org/10.1007/978-3-0348-8677-2_14).
- Bayisa, Fekadu L., Markus Ådahl, Patrik Rydén y Ottmar Cronie (2020). “Large-scale modelling and forecasting of ambulance calls in northern Sweden using spatio-temporal log-Gaussian Cox processes”. En: *Spatial Statistics* 39, pág. 100471. issn: 2211-6753. doi: <https://doi.org/10.1016/j.spasta.2020.100471>. url: <https://www.sciencedirect.com/science/article/pii/S2211675320300658>.
- Ong, Marcus EH, Faith SP Ng, Jerry Overton, Susan Yap, Derek Andresen, David KL Yong, Swee Han Lim y V Anantharaman (2009). “Geographic-time distribution of ambulance calls in Singapore: utility of geographic information system in ambulance deployment (CARE 3)”. En: *Annals Academy of Medicine Singapore* 38.3, pág. 184.
- Swalehe, Masoud y Semra Gunay Aktas (2016). “Dynamic ambulance deployment to reduce ambulance response times using geographic information systems: A case study of Odunpazari District of Eskisehir Province, Turkey”. En: *Procedia Environmental Sciences* 36, págs. 199-206.
- Chen, Albert Y., Tsung-Yu Lu, Matthew Huei-Ming Ma y Wei-Zen Sun (2016). “Demand Forecast Using Data Analytics for the Preallocation of Ambulances”. En: *IEEE Journal of Biomedical and Health Informatics* 20.4, págs. 1178-1187. doi: 10.1109/JBHI.2015.2443799.
- Hassler, Jacob y Vania Ceccato (2023). “Spatiotemporal variations in ambulance demand: towards equitable emergency services in Sweden”. En: *Geografiska Annaler: Series B, Human Geography* 0.0, págs. 1-21. doi: 10.1080/04353684.2023.2208591. eprint: <https://doi.org/10.1080/04353684.2023.2208591>. url: <https://doi.org/10.1080/04353684.2023.2208591>.
- Zhou, Zhengyi y David S. Matteson (2015). “Predicting Ambulance Demand: a Spatio-Temporal Kernel Approach”. En: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. url: <https://api.semanticscholar.org/CorpusID:13912724>.



- Nilsang, Suriyaphong y Chumpol Yuangyai (2021). “Activity detection for multi-factors of ambulance demand areas: A case study in Bangkok”. En: *AIP Conference Proceedings*. Vol. 2397. 1. AIP Publishing.
- Zhou, Zhengyi y David S. Matteson (2016). “PREDICTING MELBOURNE AMBULANCE DEMAND USING KERNEL WARPING”. En: *The Annals of Applied Statistics* 10.4, págs. 1977-1996. issn: 19326157. url: <http://www.jstor.org/stable/44252223> (visitado 28-02-2024).
- Zhou, Zhengyi, David S. Matteson, Dawn B. Woodard, Shane G. Henderson y Athanasios C. Micheas (2015). “A Spatio-Temporal Point Process Model for Ambulance Demand”. En: *Journal of the American Statistical Association* 110.509, págs. 6-15. issn: 01621459. url: <http://www.jstor.org/stable/24739283> (visitado 28-02-2024).
- Yang, Wei, Qiang Su, Ming Zhou y Xinghong Qin (2020). “Ambulance allocation considering the spatial randomness of demand”. En: *Computers and Industrial Engineering* 139, pág. 106202. issn: 0360-8352. doi: <https://doi.org/10.1016/j.cie.2019.106202>. url: <https://www.sciencedirect.com/science/article/pii/S0360835219306710>.
- Nicoletta, Vittorio, Alessandra Guglielmi, Angel Ruiz, Valérie Bélanger y Ettore Lanzarone (ene. de 2021). “Bayesian spatio-temporal modelling and prediction of areal demands for ambulance services”. En: *IMA Journal of Management Mathematics* 33.1, págs. 101-121. issn: 1471-6798. doi: 10.1093/imaman/dpaa028. eprint: <https://academic.oup.com/imaman/article-pdf/33/1/101/42114404/dpaa028.pdf>. url: <https://doi.org/10.1093/imaman/dpaa028>.
- Nabarro, Seth, Tristan Fletcher y John Shawe-Taylor (2018). “Spatiotemporal Prediction of Ambulance Demand using Gaussian Process Regression”. En: *arXiv preprint arXiv:1806.10873*.
- GADM (s.f.). Acceso: 04-04-2023. url: [https://gadm.org/download\\_country.html](https://gadm.org/download_country.html).
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. url: <https://www.R-project.org/>.
- Bivand, Roger, Edzer Pebesma y Virgilio Gómez Rubio (ago. de 2013). *Applied Spatial Data Analysis With R*. isbn: 0387781706. doi: 10.1007/978-1-4614-7618-4.
- Pebesma, Edzer J. y Roger Bivand (2005). “Classes and methods for spatial data in R”. En: *R News* 5.2, págs. 9-13. url: <https://CRAN.R-project.org/doc/Rnews/>.
- Bivand, Roger, Tim Keitt y Barry Rowlingson (2023). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. <http://rgdal.r-forge.r-project.org>, <https://gdal.org>, <https://proj.org>, <https://r-forge.r-project.org/projects/rgdal/>.
- Bivand, Roger y Colin Rundel (2023). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. <https://r-forge.r-project.org/projects/rgeos/> <https://libgeos.org> <http://rgeos.r-forge.r-project.org/index.html>.
- Bivand, Roger y Nicholas Lewin-Koh (2023). *maptools: Tools for Handling Spatial Objects*. R package version 1.1-9/r424. url: <https://R-Forge.R-project.org/projects/maptools/>.
- Pebesma, Edzer y Roger Bivand (2022). *R-spatial evolution: retirement of rgdal, rgeos and maptools*. url: <https://r-spatial.org/r/2022/04/12/evolution.html>.
- Baddeley, Adrian y Rolf Turner (2005). “spatstat: An R Package for Analyzing Spatial Point Patterns”. En: *Journal of Statistical Software* 12.6, págs. 1-42. doi: 10.18637/jss.v012.i06.
- Pebesma, Edzer (2018). “Simple Features for R: Standardized Support for Spatial Vector Data”. En: *The R Journal* 10.1, págs. 439-446. doi: 10.32614/RJ-2018-009. url: <https://doi.org/10.32614/RJ-2018-009>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller y Davis Vaughan (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, <https://github.com/tidyverse/dplyr>. url: <https://dplyr.tidyverse.org>.
- Hijmans, Robert J. (2023). *raster: Geographic Data Analysis and Modeling*. R package version 3.6-26. url: <https://CRAN.R-project.org/package=raster>.
- Pebesma, Edzer y Roger Bivand (2023). *Spatial Data Science: With applications in R*. London: Chapman y Hall/CRC, pág. 352. doi: 10.1201/9780429459016. url: <https://r-spatial.org/book/>.
- Bachl, Fabian E., Finn Lindgren, David L. Borchers y Janine B. Illian (2019). “inlabru: an R package for Bayesian spatial modelling from ecological survey data”. En: *Methods in Ecology and Evolution* 10, págs. 760-766. doi: 10.1111/2041-210X.13168.

- Rue, Håvard, Andrea I. Riebler, Sigrunn H. Sørbye, Janine B. Illian, Daniel P. Simpson y Finn K. Lindgren (2017). “Bayesian computing with INLA: A review”. En: *Annual Reviews of Statistics and Its Applications* 4.March, págs. 395-421. url: <http://arxiv.org/abs/1604.00860>.
- Wood, S.N (2017). *Generalized Additive Models: An Introduction with R*. 2.<sup>a</sup> ed. Chapman y Hall/CRC.
- Diggle, Peter (1985). “A Kernel Method for Smoothing Point Process Data”. En: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 34.2, págs. 138-147. issn: 00359254, 14679876. url: <http://www.jstor.org/stable/2347366> (visitado 11-03-2024).
- Ripley, B. D. (1977). “Modelling Spatial Patterns”. En: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.2, págs. 172-212. issn: 00359246. url: <http://www.jstor.org/stable/2984796> (visitado 12-03-2024).
- Besag, J. (1977). “Contribution to the Discussion on Dr. Ripley’s Paper”. En: *Journal of the Royal Statistical Society* B39, págs. 193-195. issn: 0035-9246. doi: 10.1111/j.2517-6161.1977.tb01616.x.
- Snow, John (1988). *On the mode of communication of cholera*. isbn: 9275115052.
- Copernicus (2022). *GHSL - Global Human Settlement Layer*. <https://human-settlement.emergency.copernicus.eu/download.php?ds=pop> [Acceso: 22/04/2024].
- Krainski, Elias, Virgilio Gómez-Rubio, Haakon Bakka, Amanda Lenzi, Daniela Castro-Camilo, Daniel Simpson, Finn Lindgren y Håvard Rue (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman y Hall/CRC.
- Belmont, Jafet (2022). *Building the Mesh*. <https://rpubs.com/jafet089/886687> [Acceso: 20/05/2024].
- Porcu, Emilio, Moreno Bevilacqua, Robert Schaback y Chris J Oates (2024). “The Matérn model: A journey through statistics, numerical analysis and machine learning”. En: *Statistical Science* 39.3, págs. 469-492.

## Anexo A

# Relación del Trabajo con los Objetivos de Desarrollo Sostenible de la Agenda 2030

Anexo al Trabajo de Fin de Grado y Trabajo de Fin de Máster: Relación del trabajo con los Objetivos de Desarrollo Sostenible de la agenda 2030.

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No Procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.	X			
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.				X
ODS 9. Industria, innovación e infraestructuras.			X	
ODS 10. Reducción de las desigualdades.			X	
ODS 11. Ciudades y comunidades sostenibles.			X	
ODS 12. Producción y consumo responsables.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz Justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos				X

**Tabla A.1:** Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Es responsabilidad de todos realizar esfuerzos para asegurar un futuro justo y sostenible para el planeta y solo será posible con la colaboración activa entre países y, en relación con este trabajo, con profesionales de todas las disciplinas. En 2015 la Organización de Naciones Unidas propuso la Agenda 2030 como marco de referencia para unificar los objetivos a los que aspirar ya que es imprescindible trabajar en un marco común para conseguir que las acciones tengan un verdadero impacto. Esta Agenda 2030 contiene 17 Objetivos de Desarrollo Sostenible (ODS) que contienen diferentes aspectos críticos para el planeta y la humanidad. Es importante que en las investigaciones científicas que se realicen se tengan en cuenta estos ODS y se busquen soluciones para los problemas que estos abordan. La Universidad Politécnica de Valencia realiza de forma activa acciones para colaborar con la Agenda 2030 como campañas de concienciación y promoción de trabajos académicos, entre otros.

Este trabajo está altamente relacionado con el ODS número 3 *Salud y Bienestar*. El derecho a la salud es fundamental y así es recogido en la lista de derechos humanos. Es indispensable asegurar un acceso libre y de calidad a servicios sanitarios para toda la población independientemente de sus características. Mejorando la calidad de la respuesta a las emergencias que requieren de un vehículo de emergencia sanitaria, se consiguen mejores resultados clínicos lo que implica una mejoría en el bienestar de la sociedad. Esto queda recogido en la Meta 3.D Refuerzo en la gestión de riesgos sanitarios. Existe también una ligera relación con los ODS 9 *Industria, Innovación e Infraestructura*, ODS 10 *Reducción de las Desigualdades*, y ODS 11 *Ciudades y Comunidades Sostenibles*, ya que mejorando la infraestructura del servicio de atención a emergencias se obtiene un acceso eficaz, sostenible e igualitario a servicios médicos de urgencia independientemente de la localización o las características socioeconómicas del paciente.

## Anexo B

# Instrucciones de R que se han utilizado en el trabajo.

Instrucción	Librería	Uso
read.delim	utils	Lectura de la base de datos.
readOGR	rgdal	Carga de archivo shapefile.
owin	spatstat	Creación de ventana.
union.owin	spatstat	Unión de ventanas.
rjitter	spatstat	Alteración de coordenadas.
rpoispp	spatstat	Simulación de un CSR.
ppp	spatstat	Creación de objeto de tipo patrón de datos (ppp).
intensity	spatstat	Cálculo de la intensidad de un objeto ppp.
quadratcount	spatstat	Conteo de cuadrantes de un objeto ppp.
quadrat.test	spatstat	Test de cuadrantes para un objeto ppp.
density.ppp	spatstat	Cálculo de la función de intensidad suavizada por kernel.
bw.diggle	spatstat	Cálculo del ancho de banda óptimo para la estimación kernel.
Kest	spatstat	Cálculo de la función K de Ripley.
Lest	spatstat	Cálculo de la función L.
Gest	spatstat	Cálculo de la función G.
Fest	spatstat	Cálculo de la función F.
Jest	spatstat	Cálculo de la función J.
Kinhom	spatstat	Cálculo de la función K de Ripley inhomogénea.
Linhom	spatstat	Cálculo de la función L inhomogénea.
Ginhom	spatstat	Cálculo de la función G inhomogénea.
Finhom	spatstat	Cálculo de la función F inhomogénea.
Jinhom	spatstat	Cálculo de la función J inhomogénea.
envelope	spatstat	Simulación de envolventes.
quadscheme	spatstat	Creación de un esquema de cuadratura.
ppm	spatstat	Ajuste de un modelo de procesos puntuales.
pseudoR2	spatstat	Cálculo del valor pseudoR2 para un modelo de procesos puntuales.
auc.ppm	spatstat	Cálculo del valor AUC para un modelo de procesos puntuales.
rho.hat	spatstat	Cálculo de la función de intensidad relativa.
lurking	spatstat	Creación del gráfico de lurking.
qqplot.ppm	spatstat	Creación del gráfico Q-Q.
st_as_sf	sf	Creación de un objeto de tipo sf.
st_crs	sf	Selección de un sistema de referencia de coordenadas (CRS).

inla.mesh.2d	INLA	Creación de una malla.
inla.spde2.pcmatern	INLA	Creación de la matrén.
lgcp	inlabru	Inferencia en un proceso log cox-gaussiano.
predict	stats	Predicción de la media del número de llamadas.

**Tabla B.1:** Instrucciones de R que se han utilizado en el trabajo.

**Fuente:** Elaboración propia.