



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**ESTADÍSTICA ESPACIAL: MODELOS DE  
PROCESOS PUNTUALES APLICADO AL ANÁLISIS  
DE INCIDENCIA DE INCENDIOS FORESTALES**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**A C T U A R I O**

**P R E S E N T A:**

**SERVANDO VALDES CRUZ**



**DIRECTOR DE TESIS:  
DR. CARLOS DIAZ AVALOS  
2017**

**CIUDAD UNIVERSITARIA, CDMX**



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno  
Apellido paterno  
apellido materno  
Nombre (s)  
Teléfono  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Carrera  
Número de cuenta

2. Datos del tutor  
Grado  
Nombre(s)  
Apellido paterno  
Apellido materno

3. Datos del sinodal 1  
Grado  
Nombre(s)  
Apellido paterno  
Apellido materno

4. Datos del sinodal 2  
Grado  
Nombre(s)  
Apellido paterno  
Apellido materno

5. Datos del sinodal 3  
Grado  
Nombre(s)  
Apellido paterno  
Apellido materno

6. Datos del sinodal 4  
Grado  
Nombre(s)  
Apellido paterno  
Apellido materno

7. Datos del sinodal 5  
Grado  
Nombre(s)  
Apellido paterno  
Apellido materno

8. Datos del trabajo escrito  
Título  
Subtítulo  
Número de páginas  
Año

1. Datos del alumno  
Valdés  
Cruz  
Servando  
58 41 54 36  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Actuaría  
307313311

2. Datos del tutor  
Dr  
Carlos  
Díaz  
Ávalos

3. Datos del sinodal 1  
Dr  
Ramsés Humberto  
Mena  
Chávez

4. Datos del sinodal 2  
Dr  
Gonzalo  
Pérez  
De la Cruz

5. Datos del sinodal 3  
Dr  
Carlos  
Díaz  
Ávalos

6. Datos del sinodal 4  
Dra  
Nancy Raquel  
Mejía  
Domínguez

7. Datos del sinodal 5  
Dr  
Delfino  
Vargas  
Chanes

8. Datos del trabajo escrito  
Estadística Espacial: Modelos de Procesos Puntuales  
aplicado al análisis de incidencia de incendios forestales  
108  
2017



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

FACULTAD DE CIENCIAS  
Secretaría General  
División de Estudios Profesionales

Votos Aprobatorios

LIC. IVONNE RAMÍREZ WENCE  
Directora General  
Dirección General de Administración Escolar  
Presente

Por este medio hacemos de su conocimiento que hemos revisado el trabajo escrito titulado:

**Estadística Espacial: Modelos de Procesos Puntuales aplicado al análisis de  
incidencia de incendios forestales**

realizado por **Servando Valdés Cruz** con número de cuenta **307313311** quien ha decidido titularse mediante la opción de tesis en la licenciatura en **Actuaría**. Dicho trabajo cuenta con nuestro voto aprobatorio.

Propietario Dr. Ramsés Humberto Mena Chávez

Propietario Dr. Gonzalo Pérez de la Cruz

Propietario Dr. Carlos Díaz Ávalos  
Tutor

Suplente Dra. Nancy Raquel Mejía Domínguez

Suplente Dr. Delfino Vargas Chanes

Atentamente

"POR MI RAZA HABLARÁ EL ESPÍRITU"

CIUDAD UNIVERSITARIA, CD. MX., A 15 DE DICIEMBRE DE 2016

JEFE DE LA DIVISIÓN DE ESTUDIOS PROFESIONALES

ACT. MAURICIO AGUILAR GONZÁLEZ

Señor sinodal: antes de firmar este documento, solicite al estudiante que le muestre la versión digital de su trabajo y verifique que la misma incluya todas las observaciones y correcciones que usted hizo sobre el mismo.

A tu memoria

Hoy donde quiera que estés, me doy cuenta de que tus cuidados y cariños cuando fui niño han sido parte de mi formación, de mi manera de ver la vida, gracias por compartir conmigo ese amor único, tus palabras siempre vivirán en mi alma, gracias por enseñarme que nada en la vida es imposible. Hoy sé que tengo allá arriba entre las estrellas a alguien que nunca olvidaré.

Te quiere y te extraña.  
Flaco.

## RESUMEN

En el presente trabajo se muestra el enfoque de los procesos puntuales, una de las 3 ramas de la estadística espacial se aplica a un ejemplo práctico como lo es el modelado de incidencia de incendios forestales, a lo largo del trabajo se da una breve introducción a la estadística espacial así mismo se profundiza sobre la teoría de los procesos puntuales.

Para la discusión de las técnicas y validación de los supuestos e hipótesis plateados se utiliza una base de datos asociados a los incendios forestales de la provincia de Castellón en España, aunque no son datos de México, debido a las restricciones de acceso a la información, la metodología y el comportamiento de los incendios es muy similar que si el análisis se hubiera realizado con datos de la República Mexicana.

El principal objetivo del análisis con este tipo de datos es poder encontrar el mecanismo estocástico generador que puede reproducir el patrón observado o algo similar, mediante un modelo el cual puede tener información adicional asociada a cada evento. Los resultados obtenidos dan continuidad a profundizar en estudios más complejos así como muestran mapas que ubican las zonas de mayor probabilidad para la incidencia de incendios. La separación entre incendios por tipo de causa, dan una visión más amplia y completa del análisis para ubicar comportamiento de negligencia y pirómanos.

# ABSTRACT

In the present work the approach of the point processes is shown, one of the three branches of spatial statistics is applied to a practical example such as the modeling of forest fire incidence, throughout the work is given a brief introduction to The spatial statistic also goes deeper into the theory of point processes.

For the discussion of the techniques and validation of the assumptions and silver hypotheses, a database is used associated with the forest fires of the province of Castellón in Spain, although they are not data of Mexico, due to the restrictions of access to the information, The methodology and the behavior of the fires is very similar that if the analysis had been done with data of the Mexican Republic.

The main objective of the analysis with this type of data is to be able to find the generator stochastic mechanism that can reproduce the observed pattern or something similar, through a model which may have additional information associated with each event. The results obtained give continuity to deepen in more complex studies as well as they show maps that locate the zones of greater probability for the incidence of fires. The separation between fires by type of cause, give a broader and complete view of the analysis to locate negligence and pyromaniac behavior.

# Índice general

Índice de Tablas	xii
Índice de Figuras	xx
<b>1. Presentación del estudio</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Objetivo general . . . . .	4
1.3. Objetivos específicos . . . . .	4
1.4. Hipótesis . . . . .	5
<b>2. Conceptos básicos e introducción a la Estadística Espacial</b>	<b>6</b>
2.1. Fundamentos teóricos preliminares . . . . .	6
2.1.1. Definición de espacio y medida de probabilidad . . . . .	6
2.1.2. $\sigma$ -álgebra, espacio medible y evento . . . . .	7
2.1.3. Medida de probabilidad . . . . .	7
2.1.4. Conjuntos de Borel . . . . .	8
2.1.5. Variable aleatoria . . . . .	8
2.1.6. Procesos estocásticos . . . . .	8
2.2. Modelo espacial general . . . . .	9
2.3. Principales enfoques del análisis espacial . . . . .	9
2.3.1. Datos lattices . . . . .	10



2.3.2.	Datos geoestadísticos . . . . .	10
2.3.3.	Patrones puntuales . . . . .	11
2.4.	Dependencia espacial . . . . .	11
2.4.1.	Autocorrelación espacial . . . . .	11
2.4.2.	Efectos de la correlación en inferencia estadística . . . . .	12
<b>3.</b>	<b>Fundamentos teóricos de Procesos Puntuales</b>	<b>14</b>
3.1.	Introducción a los procesos puntuales . . . . .	14
3.2.	Proceso puntual espacial en $\mathbb{R}^d$ . . . . .	15
3.2.1.	Patrón puntual . . . . .	15
3.2.2.	Espacio exponencial . . . . .	16
3.2.3.	Distribución del número de puntos de un proceso puntual . . . . .	16
3.2.4.	Distribución de un proceso puntual . . . . .	17
3.2.5.	Estacionariedad e isotropía . . . . .	17
3.2.6.	Proceso ordenable . . . . .	17
3.2.7.	Proceso puntual marcado . . . . .	17
3.3.	Momentos de un proceso puntual . . . . .	18
3.3.1.	Propiedades de primer y segundo orden de un proceso puntual . . . . .	19
3.3.2.	Propiedad de primer orden de un proceso puntual . . . . .	20
3.3.3.	Propiedad de segundo orden de un proceso puntual . . . . .	20
3.3.4.	Estructura de covarianza . . . . .	21
3.3.5.	Intensidad de primer y segundo orden para un proceso multivariante o multitypo . . . . .	22
3.3.6.	Independencia y etiquetado aleatorio . . . . .	23
3.3.7.	Tipos de patrones . . . . .	23
3.4.	Intensidad condicional . . . . .	24
3.5.	Teorema de Campbell . . . . .	25

3.6.	Análisis preliminar de un patrón puntual . . . . .	25
3.7.	Prueba de CSR (aleatoriedad espacial completa) . . . . .	26
3.7.1.	Índice de dispersión . . . . .	26
3.7.2.	Método de cuadrantes . . . . .	27
3.7.3.	Estimación por Kernel . . . . .	29
3.8.	Método basado en distancias . . . . .	30
3.8.1.	Función G (distribución del vecino más cercano) . . . . .	30
3.8.2.	Función F (distribución del espacio vacío) . . . . .	31
3.8.3.	Función K de Ripley . . . . .	32
3.9.	Modelos de procesos puntuales . . . . .	36
3.9.1.	Proceso puntual de Poisson . . . . .	36
3.9.2.	Proceso de Poisson no homogéneo . . . . .	37
3.9.3.	Proceso de Cox . . . . .	37
3.9.4.	Modelo de Cluster . . . . .	38
3.9.5.	Modelo Hard Core . . . . .	38
3.9.6.	Modelo de Neymann-Scott . . . . .	39
3.9.7.	Proceso de Gibbs . . . . .	39
<b>4.</b>	<b>Aplicación de la teoría de procesos puntuales al análisis de incidencia de incendios forestales</b>	<b>40</b>
4.1.	Descripción de la base de datos . . . . .	40
4.2.	Descripción general del análisis . . . . .	42
4.3.	Análisis exploratorio de datos . . . . .	44
4.3.1.	Análisis de la relación entre las variables . . . . .	44
4.3.2.	Análisis del número de incendios forestales por causa y área quemada . . . . .	45
4.3.3.	Dispersión y distribución de las principales variables de interés	49
4.4.	Prueba de aleatoriedad espacial completa (CSR) . . . . .	49

4.4.1.	Método de Kernel . . . . .	50
4.4.2.	Método de cuadrantes . . . . .	54
4.5.	Método de distancias: funciones de segundo orden . . . . .	55
4.6.	Ajuste de modelos para los datos de incendios forestales . . . . .	58
4.6.1.	Modelo ajustado para cada uno de los patrones de estudio . . . . .	60
4.6.2.	Análisis de Bondad de Ajuste . . . . .	60
4.6.3.	Mapas de predicción para los diferentes patrones ajustados . . . . .	63
4.7.	Análisis de interacción mediante la k-cruzada . . . . .	63
4.8.	Análisis de dependencia entre tamaño y localización . . . . .	65
<b>5.</b>	<b>Conclusiones</b>	<b>70</b>
5.1.	Objetivo general . . . . .	70
5.2.	Objetivos específicos . . . . .	70
5.3.	Hipótesis . . . . .	70
5.4.	Contraste entre las conclusiones los objetivos y las hipótesis . . . . .	71
<b>6.</b>	<b>Apéndice A.</b>	<b>73</b>
6.1.	Funciones L no homogéneas. . . . .	73
6.1.1.	Modelos de Poisson no homogéneos. . . . .	73
6.1.2.	Modelos de cluster. . . . .	77
6.1.3.	Modelos de Gibbs. . . . .	87
6.1.4.	Análisis de interacción mediante la k-cruzada. . . . .	91
6.1.5.	Análisis de dependencia entre tamaño y localización. . . . .	94
6.1.6.	Mapas de predicción para los diferentes patrones ajustados. . . . .	98
<b>7.</b>	<b>Apéndice B.</b>	<b>101</b>
7.1.	Código en R. . . . .	101



# Índice de tablas

4.1. Número de incendios por el tamaño del mismo y área quemada por cada año. . . . .	45
4.2. Número de incendios por causa de origen y el área quemada por cada año de estudio. . . . .	46
4.3. Área quemada por año y causa de incendio ( $m^2$ ). . . . .	47
4.4. Razón de área quemada entre número de incendios según tipo de causas para el año 2001 a 2006. . . . .	48
4.5. Valor de los valores de $p$ para contrastar la hipótesis de aleatoriedad espacial completa mediante el método de cuadrantes . . . . .	54

# Índice de figuras

3.1. Ejemplo de un proceso puntual marcado. . . . .	18
3.2. Simulación de los tres diferentes tipos de patrones puntuales: aleatorio panel (A); agregado panel (B) y regular panel (C). . . . .	24
3.3. Ejemplo del método de cuadrantes para un patrón puntual homogéneo de poisson con 100 puntos: patrón puntual panel (A); método de cuadrantes con división de 3x3 panel (B). . . . .	28
3.4. Estimación de Kernel gaussiano para los diferentes tipos de patrones puntuales: patrón aleatorio panel (A); patrón no homogéneo panel (B); patrón agregado panel (C) y patrón regular panel (D). . . . .	29
3.5. Funciones de distancias con intervalos de confianza construidos mediante simulaciones de Montecarlo correspondientes a un patrón aleatorio: función G panel (A); función F panel (B); función K panel (C) y función L panel (D). . . . .	34
3.6. Funciones de distancias con intervalos de confianza construidos mediante simulaciones de Montecarlo correspondientes a un patrón cluster: función G panel (A); función F panel (B); función K panel (C) y función L panel (D). . . . .	34
3.7. Funciones de distancias con intervalos de confianza construidos mediante simulaciones de Montecarlo correspondientes a un patrón regular: función G panel (A); función F panel (B); función K panel (C) y función L panel (D). . . . .	35
4.1. Localización de la provincia de Castellón en España. . . . .	41

4.2. Patrones de interés para el desarrollo del estudio originados por diversas causas durante el periodo 2001-2006: patrón general de los incendios ocurridos en la provincia de Castellón panel (A); incendios originados por causas naturales panel (B); incendios originados por negligencia o accidente panel (C); incendios causados de forma intencional panel (D) incendios menores a una hectárea panel (E); incendios mayores a una hectárea panel (F). . . . .	42
4.3. Covariables utilizadas como información adicional para el ajuste de los modelos: Ubicación de los incendios forestales en el periodo 2001-2006 panel (A); Densidad de población panel (B); Elevación panel (C); Orientación panel (D); Pendiente panel (E); Uso de suelo panel (F). . . . .	43
4.4. Grafica de correlación entre las variables asociadas a los datos de los incendios forestales. . . . .	44
4.5. Número de incendios con respecto al área quemada para los años 2001 a 2006. . . . .	45
4.6. Número de incendios por causa de origen en relación con el área quemada.	47
4.7. Trayectorias longitudinales del promedio de área quemada por número de incendios según tipo de causas para el año 2001 a 2006. . . . .	48
4.8. Graficas de caja para la variable pendiente y altura de cada registro de incendio dividido por año y tipo de causa. . . . .	49
4.9. Graficas de verosimilitud de validación cruzada para el parámetro $\sigma$ que maximiza la función de cada patrón de estudio: Total de los incendios de 2001 a 2006 panel (A); Naturales panel (B); Negligencia panel (C). Estimación del Kernel para la intensidad para cada patrón de estudio: Total de los incendios de 2001 a 2006 panel (D); Naturales panel (E); Negligencia panel (F). . . . .	51
4.10. Graficas de verosimilitud de validación cruzada para el parámetro $\sigma$ que maximiza la función de cada patron de estudio: Intencionados panel (A); Conatos (B); Grandes incendios panel (C). Estimación del Kernel para la intensidad para cada patrón de estudio: Intencionados panel (D); Conatos panel (E); Grandes incendios panel (F). . . . .	52
4.11. Estimación de Kernel para el total de incendios forestales correspondientes a cada uno de los años del estudio: 2001 panel (A); 2002 panel (B); 2003 panel (C); 2004 panel (D); 2005 panel (E); 2006 panel (F).	53
4.12. Estimación de Kernel por causas y tamaño de incendio: Conatos humanos panel (A); Conatos naturales panel (B); Incendios humanos panel (C); Incendios Naturales panel (D). . . . .	53

4.13. Estimación de Kernel por causas: Humanas panel (A); Naturales panel (B). . . . .	54
4.14. Función de correlación para los diferentes tipos de patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); Incendios por negligencia panel (C); Incendios intencionados panel (D); conatos (E); Grandes incendios (F). . . . .	55
4.15. Función F o distribución del espacio vacío para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F). . . . .	56
4.16. Función G o distribución de vecino más cercano para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F). . . . .	57
4.17. Función K de Ripley no homogénea para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F). . . . .	58
4.18. Función L no homogénea para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F). . . . .	59
4.19. Función L no homogénea para el total de incendios con un modelo de Cauchy con las coordenadas geográficas: Funcion L no homogénea con bandas de confianza panel (A); Transformación de la función L no homogénea con bandas de confianza panel (B). . . . .	61
4.20. Función L no homogénea para los incendios naturales con un modelo de Thomas con las coordenadas geográficas: Funcion L no homogénea con bandas de confianza panel (A); Transformación de la función L no homogénea con bandas de confianza panel (B). . . . .	61
4.21. Gráficos de diagnóstico para medir la bondad de ajuste en el modelo de Hard Core con un polinomio de grado 2: Grafico de marcas y grafico de lurking panel (A); grafica de cuantil-cuantil panel(B). . . . .	62
4.22. Gráficos de diagnóstico para medir la bondad de ajuste en el modelo de Strauss con un polinomio de grado 4: Gráfica de marcas y gráfica de lurking panel (A); Gráfica de cuantil-cuantil panel(B). . . . .	63



4.23. Mapa de predicciones ajustado con un modelo de Cauchy con coordenadas geográficas: Mapa de predicción para el total de incendios forestales panel (A); Mapa de predicción para el total de incendios forestales con la ubicación de los eventos ocurridos en periodo 2001-2006 panel (B).	64
4.24. Mapa de predicciones ajustado con un modelo de Thomas con coordenadas geográficas: Mapa de predicción para los incendios naturales panel (A); Mapa de predicción para los incendios naturales con la ubicación de los eventos ocurridos en periodo 2001-2006 panel (B).	64
4.25. Función de k cruzada para los casos relevantes: Interacción entre los patrones de los años 2002 y 2003 panel (A); Interacción entre los patrones de los años 2002 y 2006 panel (B); Interacción entre los patrones de causas desconocidas y los incendios intencionados panel (C); Interacción entre los patrones de incendios por negligencia e intencionados panel (D).	65
4.26. Función de k cruzada para la división de los incendios: Interacción entre conatos e incendios panel (C) y panel (B); Función k no homogénea para conatos panel(A); Función k no homogénea para incendios panel(D).	66
4.27. Gráficas correspondientes al análisis de marcas del año 2001: Histograma panel (A); Estimación de Kernel panel (B); función de correlación panel (C); Esperanza condicional inversa panel (D).	67
4.28. Graficas correspondientes al análisis de marcas para las causas humanas: Histograma panel (A); Estimación de Kernel panel (B); función de correlación panel (C); Esperanza condicional inversa panel (D).	68
4.29. Graficas correspondientes al análisis de marcas para las causas naturales: Histograma panel (A); Estimación de Kernel panel (B); función de correlación panel (C); Esperanza condicional inversa panel (D).	69
6.1. Funciones L no homogénea para un modelo de Poisson ajustado a los diferentes patrones de estudio.	73
6.2. Funciones L no homogénea para un modelo de Poisson ajustado con las coordenadas geográficas para los diferentes patrones de estudio.	74
6.3. Funciones L para un modelo no homogéneo de Poisson ajustado con las covariables para los diferentes patrones de estudio.	74
6.4. Funciones L para un modelo no homogéneo de Poisson ajustado con las covariables seleccionadas mediante el método de backward para los diferentes patrones de estudio.	75

6.5. Funciones L para un modelo no homogéneo de Poisson ajustado con las covariables y coordenadas geográficas para los diferentes patrones de estudio. . . . .	75
6.6. Funciones L para un modelo no homogéneo de Poisson ajustado con las covariables y coordenadas geográficas seleccionadas mediante el método backward para los diferentes patrones de estudio. . . . .	76
6.7. Funciones L para un modelo no homogéneo de Poisson ajustado con un polinomio de grado 4 para los diferentes patrones de estudio. . . . .	76
6.8. Funciones L para un modelo de Matern homogéneo para los diferentes patrones de estudio. . . . .	77
6.9. Funciones L para un modelo de Matern ajustado con las coordenadas geográficas para los diferentes patrones de estudio. . . . .	77
6.10. Funciones L para un modelo de Matern ajustado con las covariables para los diferentes patrones de estudio. . . . .	78
6.11. Funciones L para un modelo de Matern ajustado con las covariables seleccionadas por el método backward para los diferentes patrones de estudio. . . . .	78
6.12. Funciones L para un modelo de Matern ajustado con las covariables y las coordenadas geográficas para los diferentes patrones de estudio. . . . .	79
6.13. Funciones L para un modelo de Matern ajustado con las covariables y las coordenadas geográficas seleccionadas por el método de backward para los diferentes patrones de estudio. . . . .	79
6.14. Funciones L para un modelo de Matern ajustado con un polinomio de grado 2 para los diferentes patrones de estudio. . . . .	80
6.15. Funciones L para un modelo de Matern ajustado con un polinomio de grado 4 para los diferentes patrones de estudio. . . . .	80
6.16. Funciones L para un modelo de Thomas homogéneo para los diferentes patrones de estudio. . . . .	81
6.17. Funciones L para un modelo de Thomas ajustado con las coordenadas geográficas para los diferentes patrones de estudio. . . . .	81
6.18. Funciones L para un modelo de Thomas ajustado con las covariables para los diferentes patrones de estudio. . . . .	82
6.19. Funciones L para un modelo de Thomas ajustado con las covariables seleccionadas por el método backward para los diferentes patrones de estudio. . . . .	82

6.20. Funciones L para un modelo de Thomas ajustado con las covariables y las coordenadas geográficas para los diferentes patrones de estudio. . . . .	83
6.21. Funciones L para un modelo de Thomas ajustado con las covariables y las coordenadas geográficas seleccionadas por el método de backward para los diferentes patrones de estudio. . . . .	83
6.22. Funciones L para un modelo de Thomas ajustado con un polinomio de grado 2 para los diferentes patrones de estudio. . . . .	84
6.23. Funciones L para un modelo de Thomas ajustado con un polinomio de grado 4 para los diferentes patrones de estudio. . . . .	84
6.24. Funciones L para un modelo de Cauchy ajustado con las coordenadas geográficas para los diferentes patrones de estudio. . . . .	85
6.25. Funciones L para un modelo de Cauchy ajustado con un polinomio de grado 2 para los diferentes patrones de estudio. . . . .	85
6.26. Funciones L para un modelo de Cauchy ajustado con un polinomio de grado 4 para los diferentes patrones de estudio. . . . .	86
6.27. Funciones L para un modelo de Cauchy ajustado con las coordenadas geográficas y las covariables para los diferentes patrones de estudio. . . . .	86
6.28. Funciones L para un modelo de Cauchy ajustado con las coordenadas geográficas y las covariables seleccionadas por el método de backward para los diferentes patrones de estudio. . . . .	87
6.29. Funciones L para un modelo de Strauss ajustado con un polinomio de grado 2 para los diferentes patrones de estudio. . . . .	87
6.30. Funciones L para un modelo de Strauss ajustado con un polinomio de grado 4 para los diferentes patrones de estudio. . . . .	88
6.31. Funciones L para un modelo de Strauss ajustado con las coordenadas geográficas y las covariables para los diferentes patrones de estudio. . . . .	88
6.32. Funciones L para un modelo de Strauss ajustado con las covariables para los diferentes patrones de estudio. . . . .	89
6.33. Funciones L para un modelo de Hard Core ajustado con las coordenadas geográficas para los diferentes patrones de estudio. . . . .	89
6.34. Funciones L para un modelo de Hard Core ajustado con un polinomio de grado 2 para los diferentes patrones de estudio. . . . .	90
6.35. Funciones L para un modelo de Hard Core ajustado con un polinomio de grado 4 para los diferentes patrones de estudio. . . . .	90

6.36. Función de k cruzada para los años 2001-2006. . . . .	91
6.37. Función de k cruzada para las causas de origen de cada incendio. . . . .	91
6.38. Función de k cruzada para los incendios mayores y menores a 1 ha para las causas humanas y naturales. . . . .	92
6.39. Función de k cruzada para las causas naturales y humanas. . . . .	92
6.40. Función de k cruzada para los diferentes tamaños de incendios. . . . .	93
6.41. Función de k cruzada para los diferentes tamaños de incendios. . . . .	93
6.42. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2001. . . . .	94
6.43. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2002. . . . .	94
6.44. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2003. . . . .	95
6.45. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2004. . . . .	95
6.46. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2005. . . . .	96
6.47. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2006. . . . .	96
6.48. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas por causas humanas. . . . .	97
6.49. Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas por causas naturales. . . . .	97
6.50. Mapa de predicciones ajustado con un modelo de Cauchy con coordenadas geográficas. . . . .	98

6.51. Mapa de predicciones ajustado con un modelo de Thomas con coordenadas geográficas. . . . .	98
6.52. Mapa de predicciones ajustado con un modelo de Hard Core con un polinomio de grado 2. . . . .	99
6.53. Mapa de predicciones ajustado con un modelo de Thomas con covariables seleccionadas. . . . .	99
6.54. Mapa de predicciones ajustado con un modelo de Thomas con un polinomio de grado 2. . . . .	100

# Capítulo 1

## Presentación del estudio

### 1.1. Introducción

El impacto de fenómenos naturales en la época contemporánea requiere del análisis de datos de forma sistemática que ayude a la comprensión y el estudio de los mismos de una forma sencilla y eficaz. Es por ello que se hace uso de herramientas estadísticas más avanzadas, las cuales gracias al avance de la ciencia, así como el surgimiento de nuevas áreas y metodologías, ha motivado la creación de técnicas más robustas y sofisticadas.

Las técnicas estadísticas clásicas sirven para la construcción de modelos, además de estimar los parámetros, a partir de datos que provienen de una muestra aleatoria (Cressie, 1993). Esto significa que las observaciones del fenómeno de interés se toman bajo las mismas condiciones y de forma independiente; sin embargo, hay que recordar que en la vida cotidiana los modelos estadísticos son más reales si se incluye la dependencia, es decir, la relación entre los elementos de la muestra.

En la actualidad existen diversas formas de representar a los fenómenos naturales, una de ellas es la representación gráfica y mejor aún una representación espacial que aporta una perspectiva de cómo se distribuyen geográficamente los elementos de estudio en una región determinada.

La estadística espacial es la rama de la estadística que reúne un conjunto de metodologías cuyo objetivo es el estudio y análisis de datos que corresponden a mediciones de variables aleatorias en diversos puntos de una región de interés (Abellan et al., 2002). En una definición más formal decimos que la estadística espacial estudia el análisis de realizaciones de un proceso estocástico  $Z(s) : s \in D \subseteq \mathbb{R}^d$  en el cual  $s \in \mathbb{R}^d$  representa una ubicación en el espacio euclidiano  $d$ -dimensional,  $Z(s)$  representa una variable aleatoria de interés para la ubicación  $s$  la cual varía en un conjunto de índices  $D \subseteq \mathbb{R}^d$ .

La estadística espacial se conforma por tres diferentes grupos de datos, uno de

ellos y sobre el cual se desarrolló el siguiente trabajo, son los procesos puntuales, donde los eventos de interés se presentan en un conjunto finito de puntos dentro de un dominio  $D$  el cual recibe el nombre de patrón puntual. Un patrón puntual se define heurísticamente como un conjunto de datos los cuales están distribuidos sobre una región determinada y cuya ubicación se debe principalmente a un mecanismo estocástico, además de traer asociada una variable o característica de interés junto con sus coordenadas geográficas.

El principal objetivo de trabajar con patrones puntuales es comprender e intentar reproducir el mecanismo generador de dicho patrón de interés mediante la postulación de un modelo, además determinar si los eventos bajo estudio tienen cierta atracción, o suceden unos cerca de otros (patrón de agregación), o si por el contrario, presentan cierta repulsión o diseminación (procesos de inhibición o regulares), o su distribución es aleatoria.

También es de interés saber si los eventos ocurren con la misma probabilidad sobre un espacio. La teoría para el desarrollo del análisis en patrones puntuales engloba una serie de técnicas estadísticas cuya principal función es el estudio del mecanismo estocástico por el cual se genera el patrón de estudio.

En el presente trabajo se muestran y se estudian algunas de las técnicas de análisis espacial enfocadas a los procesos puntuales y se presenta una aplicación orientada al análisis de la incidencia de incendios forestales. Se centra en conocer con precisión la distribución espacial de los focos de incendios registrados en la región de estudio para después modelar y obtener un mapa de riesgo de la incidencia de los incendios forestales.

Estos eventos son interesantes de modelar debido a que encontramos diferentes tipos de comportamientos asociados, muchas veces, a factores externos que influyen sobre su incidencia. Sin embargo, mediante el uso de un modelo podemos dar una información más clara y precisa del comportamiento de los mismos, de esta manera se pueden recrear escenarios similares mediante simulaciones apoyadas por agentes externos, que muchas veces aportan información al modelo (covariables), y sirven para tomar acciones y evitar catástrofes.

Los incendios forestales tienen un comportamiento bastante interesante ya que se pueden originar de forma natural y su importancia en la dinámica de los ecosistemas es un asunto complejo de comprender. Por un lado, estos fenómenos se asocian con destrucción y daño, lo cierto es que el fuego tiene una influencia positiva en la naturaleza, pues ayuda a mantener ese equilibrio en los ecosistemas (Moritz 2003).

Estos eventos influyen en la regeneración y reproducción de ciertas especies forestales, propician el rebrote de vegetación y pastizal que sirve para la alimentación del ganado y de la fauna silvestre, también se facilita la germinación de las semillas de algunas especies de árboles, además de que sirve para controlar y erradicar ciertas plagas, enfermedades y vegetación indeseada, lo cual mantiene activa la dinámica forestal de una zona.

Sin embargo también son considerados como uno de los fenómenos más destructivos de la naturaleza (Oliver & Larson, 1996), ya que mediante la contribución al producir gases de efecto invernadero, estimulan y generan cambios dentro de la dinámica forestal que asociados con algunas agentes climatológicos pueden ser uno de los fenómenos más devastadores y en cuyo caso las áreas afectadas pueden tardar mucho tiempo en recuperarse.

Según la Comisión Nacional Forestal (CONAFOR, 2016) los incendios forestales se pueden catalogar dependiendo del origen por el cual fue causado de la siguiente manera:

- Incendios superficiales: son el tipo de incendio más común no solo en México sino en el resto del mundo estos se propagan sobre la superficie del terreno, se localizan desde la superficie del suelo hasta los 1.5 metros de altura.
- Incendios de copa o aéreos: consumen la totalidad de la vegetación los más peligrosos y difíciles de controlar.
- Incendios subterráneos: inician de forma superficial, bajo el suelo mineral debido a la acumulación y compactación de los combustibles. Por lo general, no producen llama y emiten poco humo.

Otra clasificación es atribuida a sus causas:

- Accidentales: Relacionados con accidentes automovilísticos, ferroviarios, aéreos y con ruptura de líneas eléctricas.
- Negligencias: Las quemas agropecuarias no controladas, fogatas de excursionistas, fumadores, quema de basura, limpieza de vías en carreteras y uso del fuego en otras actividades productivas dentro de las áreas forestales.
- Intencionales: Se refiere a las quemas que se realizan en el contexto de conflictos entre personas o comunidades, tala ilegal o litigios, entre otros.
- Naturales: Caída de rayos o erupciones volcánicas.

Otro tipo de clasificación propuesta es la que considera la normativa europea (PLADIGA) por área quemada:

- Conato: Incendios cuya superficie quemada es menor a 1 ha.
- Incendio: Cuando la superficie quemada es mayor a 1 ha.

La intervención de los humanos ha contribuido al desarrollo de los incendios forestales, un ejemplo de ello son las actividades agrícolas, donde los agricultores utilizan



el fuego, como herramienta para poder desarrollar técnicas de cultivo y de limpieza de terrenos (Rodríguez-Trejo & Fule, 2003). Además de causas accidentales se tiene también la generación de incendios por conducta de negligencia ya que según CONAFOR (2016), 9 de cada 10 incendios forestales son causados por seres humanos, de los cuales cada vez son mayores los incendios registrados cerca de núcleos habitacionales y tienden a seguir el patrón de estacionalidad, lo cual hace más viable la premisa del aumento de incendios forestales atribuibles a causas humanas (Martell et al., 1987).

La base de datos utilizada para este trabajo es un registro de los incendios forestales ocurridos en la provincia de Castellón en España durante los años de 2001 a 2006. En ella se observan diferentes características de interés, en particular, las coordenadas espaciales, la ubicación los centroides de cada registro de incendio, el área quemada así como la clave de causa u origen del mismo. Con esta información se puede realizar un análisis a través del espacio y del tiempo, así como un análisis sobre las causas que dan origen a los incendios forestales.

De igual forma, se crea un mapa de riesgos que ayuda a determinar las zonas de mayor riesgo de incidencia de incendios forestales con la finalidad de poder ayudar a la prevención y control de los mismos.

## 1.2. Objetivo general

- Aplicar los métodos desarrollados en procesos puntuales como una alternativa estadística para analizar la incidencia y el comportamiento del fenómeno de incendios forestales.

## 1.3. Objetivos específicos

- Construir modelos de la distribución geográfica de los incendios forestales y concluir si existe agrupamiento, repulsión o interacción entre ellos, esto se realiza en un contexto determinado por un espacio-tiempo, dimensión del área quemada y causas.
- Detectar las áreas con mayor probabilidad de ocurrencia de incendios forestales mediante un mapa de riesgo, tanto para causas naturales como humanas (negligencia o accidente e intencionado).
- Analizar el efecto de las covariables seleccionadas sobre la intensidad de los incendios forestales.
- Localizar tendencia de incendios forestales sobre la región de estudio.
- Identificar comportamientos reiterativos en la propagación de los incendios.

## 1.4. Hipótesis

- El total de incendios forestales para el periodo 2001-2006 en la región de estudio forman conglomerados ya sea por comportamientos de reincidencia, por factores externos como el tipo de suelo, altitud, clima o por la misma dinámica forestal.
- Los incendios por causas humanas (negligencia accidentes o intencionados) tienden a estar más concentrados cerca de áreas densamente pobladas.
- Los incendios por causas naturales tienden a ocurrir de forma aleatoria y no dependen de la información adicional que aportan las covariables.
- Existe una relación de dependencia por marcas entre el área quemada y por cada una de las causas mencionadas previamente.

# Capítulo 2

## Conceptos básicos e introducción a la Estadística Espacial

En este capítulo se hace una breve introducción, de las características, aplicaciones y diferentes enfoques de la estadística espacial. Dentro de la bibliografía principal que se utilizó para el desarrollo de esta sección está Cressie (1993), Rincón (2007) así como otras fuentes de las cuales se hará mención más adelante. Al inicio de esta sección se presentan conceptos básicos de probabilidad y procesos estocásticos que permitirán posteriormente entender las áreas que componen al estudio de la estadística espacial.

### 2.1. Fundamentos teóricos preliminares

La rama de las matemáticas que se encarga del estudio, la observación y descripción de los fenómenos o experimentos aleatorios, es la probabilidad. Un fenómeno o experimento aleatorio será aquel que al realizarse desde un principio bajo condiciones similares, los resultados no siempre sean los mismos (Rincón, 2007). Los conceptos probabilísticos tales como sigma algebra, espacio medible, evento y variable aleatoria definen a los procesos estocásticos, los cuales son sucesiones de variables aleatorias.

#### 2.1.1. Definición de espacio y medida de probabilidad

Se define un espacio de probabilidad como una terna  $(\Omega, \mathcal{F}, \mathbb{P})$  donde  $\Omega$  representa un conjunto arbitrario,  $\mathcal{F}$  es una  $\sigma$ -álgebra de conjuntos que pertenecen a  $\Omega$  y  $\mathbb{P}$  es una medida de probabilidad que está definida sobre  $\mathcal{F}$ .

El espacio de probabilidad es el modelo matemático creado para estudiar los experimentos aleatorios, dicho modelo considera una terna ordenada compuesta por  $(\Omega, \mathcal{F}, \mathbb{P})$  donde definiremos a  $\Omega$  como el espacio muestral, dicho conjunto no es vacío

y contiene todos los posibles resultados de interés del fenómeno o experimento aleatorio.  $\mathcal{F}$  es una sigma algebra, la cual es una colección o clase de subconjuntos no vacía del espacio muestral. Los elementos que conforman dicho subconjunto se le denominan eventos, sucesos o conjuntos medibles.  $\mathbb{P}$  es la medida de probabilidad la cual es una función definida sobre una sigma algebra y con valores en el intervalo  $(0,1)$ .

### 2.1.2. $\sigma$ -álgebra, espacio medible y evento

Se Define a una  $\sigma$  - álgebra como una colección  $\mathcal{F}$  de subconjuntos de  $\Omega$  que cumple las siguientes condiciones:

- $\Omega \in \mathcal{F}$ .
- si  $\mathcal{A} \in \mathcal{F}$  entonces  $\mathcal{A}^c \in \mathcal{F}$ .
- si  $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \dots \in \mathcal{F}$  entonces  $\bigcup_{n=1}^{\infty} \mathcal{A}_n \in \mathcal{F}$ .

A la pareja  $(\Omega, \mathcal{F})$  se le llama espacio medible y a los elementos de  $\mathcal{F}$  se les llama eventos o conjuntos medibles. Con las condiciones anteriores podemos observar que la  $\sigma$  -álgebra está definida bajo la operación de complemento y de uniones infinitas lo cual sustenta que pueda ser aplicable las operaciones de conjuntos.

### 2.1.3. Medida de probabilidad

Sea  $(\Omega, \mathcal{F})$  un espacio medible. Definiremos a una medida de probabilidad como una función  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  la cual satisface las siguientes condiciones:

- $\mathbb{P}(\Omega) = 1$ .
- $\mathbb{P}(\mathcal{A}) \geq 0$  para cualquier  $\mathcal{A} \in \mathcal{F}$ .
- si  $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \dots \in \mathcal{F}$  son ajenos 2 a 2 es decir  $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$  para  $i \neq j$  entonces
 
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \mathcal{A}_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(\mathcal{A}_n).$$

Entonces toda función  $\mathbb{P}$  definida sobre una  $\sigma$  -álgebra  $\mathcal{F}$  con valores en el intervalo  $[0, 1]$  y que cumplan con los axiomas de Kolmogorov, se le conoce como medida de probabilidad o probabilidad axiomática.

### 2.1.4. Conjuntos de Borel

Considere la colección de todos los intervalos abierto  $(a, b)$  de  $\mathbb{P}$  donde  $a \leq b$ . La mínima  $\sigma$ -álgebra generada por esta colección se le llama  $\sigma$ -álgebra de Borel de  $\mathbb{R}$  y se le denota por  $B(\mathbb{R})$ .

Definición: ( $\sigma$ -álgebra de Borel de  $\mathbb{R}$ )

$$B(\mathbb{R}) = \sigma(a, b) \subseteq \mathbb{R} : a \leq b.$$

A los elementos de  $B(\mathbb{R})$  se les llama conjuntos de Borel, borelianos o conjuntos de Borel medibles. De esta forma se pueden asociar la sigma algebra  $B(\mathbb{R})$  al conjunto de números reales y así obtener el espacio medible  $(\mathbb{R}, B(\mathbb{R}))$ .

### 2.1.5. Variable aleatoria

Hablaremos de un concepto fundamental en la teoría de la probabilidad, el de variable aleatoria que se define como una función  $\mathcal{X} : \Omega \rightarrow \mathbb{R}$  tal que para cualquier conjunto boreliano  $B$ , se cumple que el conjunto  $\mathcal{X}^{-1}$  es un elemento de  $\mathcal{F}$ .

Una variable aleatoria es una función del espacio muestral en el conjunto de números reales que satisface cierta condición de medibilidad (Rincón, 2007), donde es representado cada uno de los resultados del espacio muestral, formalmente hablando es una función de  $\Omega$  en los reales  $\mathbb{R}$  tal que la imagen inversa de cualquier boreliano es un elemento de la  $\sigma$ -álgebra del espacio de probabilidad.

### 2.1.6. Procesos estocásticos

Definiremos un proceso estocástico tomando en cuenta la definición de Rincón (2012), donde habla de un proceso estocástico como un mecanismo azaroso de una sucesión de variables aleatorias que varían a través de un espacio de estados. Formalmente se define enseguida:

Sea una familia de variables aleatorias  $Z(s) : s \in D \subseteq \mathbb{R}^p$  definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . El conjunto  $D$  de índices del proceso estocástico se denomina espacio de parámetros. Los valores que  $Z(s)$  toma se llaman estados y el conjunto de todos los posibles valores de  $Z(S)$  se llama espacio de estados.

Una vez tomando los conceptos básicos de probabilidad y el concepto de proceso estocástico se define la terminología de estadística espacial, tomando en cuenta que se trata de un proceso estocástico cuyo espacio de parámetros puede ser continuo o discreto, y que es el que define el tipo de datos espaciales que son de interés en este trabajo. La estadística espacial se trata en el próximo capítulo así como el tema de los procesos puntuales que es el eje central del trabajo.

## 2.2. Modelo espacial general

Definimos el modelo espacial general bajo los siguientes términos (Cressie, 1993): Sea  $s \in \mathbb{R}^d$  una localización en el espacio euclidiano  $d$ -dimensional y  $Z(s)$  una medición observada en una localización  $s$ . Al variar  $s$  sobre un conjunto de índices  $D \subseteq \mathbb{R}^d$  tenemos el proceso aleatorio:

$$Z(s) : s \in D \subseteq \mathbb{R}^d \quad (2.1)$$

Cuyas realizaciones son

$$z(s) : s \in D \subseteq \mathbb{R}^d \quad (2.2)$$

Algunas observaciones sobre el modelo espacial general:  $Z(s)$  representa la variable aleatoria indexada al dominio  $D$ . Dicha variable es la de interés en el sitio  $s$ .  $D$  representa el conjunto de los sitios de donde fueron obtenidos los datos, en dicho dominio está bien definido el proceso.

Los datos espaciales se pueden clasificar en tres grupos dependiendo del dominio asociado del cual se trata:

- Datos Geoestadísticos: Si  $D$  es un dominio continuo (fijo). El principal objetivo de este tipo de modelos es poder hacer predicciones sobre los sitios no medidos. Le corresponde al investigador decidir los sitios de interés de la región donde se llevarán a cabo las mediciones.
- Datos en retícula (Lattices): Si  $D$  es un conjunto numerable (fijo). Al igual que los datos anteriores, le corresponde al investigador realizar la selección.
- Procesos Puntuales: Si  $D$  es un dominio aleatorio. A diferencia de los datos geoestadísticos y en retícula (Lattices), los sitios de interés no pueden ser definidos por el investigador. El propósito de este análisis a diferencia de los otros radica en que nos interesa conocer acerca de la distribución de los fenómenos de interés.

## 2.3. Principales enfoques del análisis espacial

Cuando se habla de datos espaciales, nos referimos a aquellos que traen consigo asociadas aparte de las variables de interés, sus localizaciones georeferenciadas en un área de estudio. Dicha localización puede ser regular o irregular, comportamiento que es más frecuente ver en la naturaleza, ya que es muy poco probable que un fenómeno suceda o se obtengan de él, mediciones uniformemente distribuidas a través del espacio.

Se mencionó con anterioridad que la estadística espacial se divide en tres grandes grupos de estudio: geoestadística, datos lattices o de área y procesos puntuales; dicha división se da dependiendo del tipo de datos que se está estudiando y estos se diferencian a su vez dependiendo de las características asociadas al conjunto de índices  $D$  del proceso estocástico de interés.

### 2.3.1. Datos lattices

Para los datos lattices el procesos estocástico tiene un espacio de parámetros  $D$  que es una colección fija (regular o irregular) de muchos puntos contables de  $\mathbb{R}^d$ , se presentan en regiones o lattices de estudio que son continuas. Sobre regiones espaciales por lo cual la información adicional asociada a los eventos estará directamente relacionada al centroide de la región, suponiendo una distribución uniforme de la característica de interés sobre toda la región de estudio. Este tipo de datos se complementa adicionalmente con la información de las regiones o sitios adyacentes a la estudiada, originando la estructura de vecindad.

El investigador, experto en el área es quien puede decidir en donde realizar la medición de la característica de interés que se quiera estudiar. Una de las desventajas es que si se desconoce el tipo de investigación, la interpolación espacial puede carecer de sentido o no puede ser tan sencillo de darle una interpretación.

### 2.3.2. Datos geoestadísticos

El origen de estudio con este tipo de datos se originó en el campo de la minería y geología en los años 60, la aportación de esta gran rama de la estadística espacial se debe al enfoque más clásico que se le ha dado, donde se reconoce variabilidad espacial, lo cual en términos más estadísticos podemos decir que el modelo presenta tendencia espacial y correlación.

Para este tipo de datos nos referiremos a un campo aleatorio, el cual es un conjunto de datos espacialmente continuos, cuyo dominio  $D$  elemento de  $\mathbb{R}^d$  es no numerable y cuyas mediciones son tomadas en puntos fijos con localizaciones continuas en el espacio. Los datos geoestadísticos los podemos representar como  $z(s_1), z(s_2), \dots, z(s_n)$ . Un modelo simple para datos geoestadísticos es  $z(s) = \mu + \varepsilon(s)$ , donde se asume que el error tiene esperanza cero, además de asumir homocedasticidad.

$$\mathbb{E}[z(s)] = \mu \tag{2.3}$$

$$Var[z(s)] = \sigma^2. \tag{2.4}$$

Esto es para todos los puntos  $s$  del dominio  $D$ . En el estudio de series de tiempo la función de autocovarianza para una serie estacionaria solo dependía de la diferencia entre los puntos de la serie, este mismo principio será similar en Geoestadística ya que

consideraremos que la función de covarianza para 2 diferentes puntos, solo dependerá de la distancia y la dirección entre ambos, es decir:

$$\text{cov}[z(s_1), z(s_2)] = c(s_1 - s_2),$$

para alguna función  $c(\cdot)$ . Los datos geoestadísticos que satisfacen las condiciones anteriores son llamados estacionarios de segundo orden.

El objetivo de trabajar con dicho tipo de datos es poder tomar un muestreo de una cantidad finita de puntos en una región determinada y de esta forma extender la distribución espacial de los valores de un atributo sobre el total de la región de estudio.

Dentro de este análisis podemos contemplar la modelización del patrón de variabilidad, así como la obtención de una predicción de la variable en puntos que no se hayan muestreado, dando origen a la interpolación espacial. Los estudios de tipo Geoestadísticos son distinguibles de los datos de tipo lattices y patrones puntuales por la habilidad del índice espacial  $s$  para variar continuamente sobre el subconjunto de  $\mathbb{R}^d$ .

### 2.3.3. Patrones puntuales

Estos serán nuestros datos de interés en este trabajo. Los patrones puntuales se diferenciarán de los otros tipos de datos por una simple razón las localizaciones y no las mediciones son las variables de interés. Dicho patrón consiste en un número finito y determinado de localizaciones observadas en una región de estudio.

En este tipo de datos se trata de valorar si existe una tendencia de los eventos a exhibir un patrón sistemático, especialmente de alguna forma de regularidad o de agregación. El objetivo principal de los procesos puntuales es conocer la variación de la intensidad de los eventos sobre la región de estudio y buscar modelos que ayuden a explicar y reproducir el comportamiento del fenómeno.

## 2.4. Dependencia espacial

### 2.4.1. Autocorrelación espacial

Uno de los supuestos más fuertes dentro de la estadística clásica es la independencia, y violar este supuesto estima de forma incorrecta las varianzas de los estimadores de interés, distorsionando la distribución de los errores, obteniendo así resultados inválidos, y es que en la vida real la mayoría de los sucesos presentan una estrecha relación entre ellos, esto es que la realización de un suceso a veces depende o tiene influencia



en gran medida de otros (Diggle, 2003) dando origen al concepto estadístico de correlación. En la estadística espacial dicho concepto hace referencia a la estructura que presentan las variables del modelo espacial general  $\mathcal{Z}(s) : s \in D \subseteq \mathbb{R}^d$ .

¿Por qué y cómo detectar autocorrelación espacial? La importancia de detectar la autocorrelación espacial se debe principalmente a que de no hacerlo tiene consecuencias de gravedad a la hora de realizar procesos inferenciales, ya que se pueden obtener estimadores ineficientes e inconsistentes, produciendo una estimación sesgada. La dependencia, según Cressie (1993); se observa en la distancia entre dos eventos donde al estar cerca unos de otros existirá más similitud entre ellos, es decir, la noción de que los datos más cercanos, tienden a estar más correlacionados, en otras palabras la existencia de dependencia espacial se resume en encontrar similitud entre los individuos que se encuentran cercanos unos con otros.

En la estadística espacial al igual que en la clásica se tiene numerosas herramientas que se pueden utilizar para medir la estrecha relación espacial entre las observaciones. Dichos estadísticos evalúan o miden la dependencia espacial a través de la comparación de las distancias entre los puntos de la muestra y sus vecinos.

Una de las principales medidas para obtener autocorrelación es el Test de Mantel, que es un estadístico que se utilizaba principalmente en procesos puntuales, dicho estadístico, no es muy común utilizarlo en la actualidad debido al avance de herramientas de computo que lograron desarrollar métodos más robustos. El otro es el Test de Moran, el cual es utilizado para datos lattices o de retícula, su objetivo principal es determinar si regiones adyacentes tienen valores similares en la variable de interés, y el variograma que es la medida utilizada para datos geoestadísticos.

## 2.4.2. Efectos de la correlación en inferencia estadística

Hablando desde el enfoque clásico, la dependencia entre las observaciones causan que las estimaciones del modelo propuesto no sean tan eficientes además de que presenten subestimación en las varianzas, caso contrario al considerar muestras independientes. Este problema no cambia al tratar con datos espaciales, de hecho como se ha planteado la dependencia espacial será uno de los supuestos más importantes dentro del análisis, y de igual forma no se puede plantear los mismos supuestos que en la estadística clásica a la hora de hacer inferencias ya que la dependencia espacial nos lleva a una pérdida de poder explicativo (Diggle, 2003).

Es por ello que se debe considerar al conjunto de datos espaciales como una realización de un proceso estocástico y no como un conjunto de muestras. En el caso de un proceso estocástico todas las observaciones se utilizan de manera combinada para describir el patrón espacial del fenómeno estudiado.

Existen otro tipo de modelos como los espacio-temporales que incorporan la correlación y corrigen el problema de subestimación de la varianza. Dicha corrección es

de suma importancia si los resultados obtenidos del análisis se usan para la toma de decisiones. El objetivo principal del modelo general será resumir la información sobre el fenómeno de interés, ya sea alguna característica o buscar patrones acerca de las ubicaciones donde sucedió dicho fenómeno, para que una vez teniendo un planteamiento adecuado del modelo se pueda predecir el comportamiento en sitios donde no observamos la variable de interés o para poder modelar un comportamiento similar al observado (Cressie, 1993).

# Capítulo 3

## Fundamentos teóricos de Procesos Puntuales

En este capítulo se desarrolla el marco teórico de los procesos puntuales. Se toma como referencia bibliográfica para escribir este capítulo: Moller & Waagepetersen (2004), Cressie (1993) y Gaetan & Guyon (2010). Recordar que bajo el modelo espacial general  $Z(s) : s \in D \subseteq \mathbb{R}^d$ , el dominio  $D$  es aleatorio, y el principal objetivo del análisis de este tipo de puntos es poder determinar si existe un patrón sistemático para las localizaciones de los eventos y si se puede explicar con un modelo dicho comportamiento.

### 3.1. Introducción a los procesos puntuales

En primer lugar, se define heurísticamente a un patrón puntual como un conjunto de puntos que se encuentran en una cierta región llamada área de estudio. Dichos puntos representan eventos de interés de un fenómeno distribuido espacialmente y tienen asociados sus coordenadas  $(x, y)$ . Por lo que un proceso puntual se define como un arreglo o patrón de puntos distribuidos en un cierto conjunto. En general un patrón puntual no es más que un proceso estocástico donde se observa la localización de eventos de interés dentro de una región acotada  $A$ .

La definición de un proceso puntual según Diggle (2003) es aquel mecanismo estocástico que genera un conjunto numerable de eventos. En tanto que Arunachalam (2005) define a un proceso estocástico puntual como el que corresponde a la abstracción matemática de un fenómeno localizado aleatoriamente en un espacio de parámetros o como secuencia de eventos en el tiempo. El proceso puntual puede ser especificado en términos de sus propiedades de conteo.

Se define a los procesos puntuales de acuerdo a Moller & Waagepetersen (2003); este autor es una referencia sustantiva para el presente capítulo para desarrollar la

teoría de procesos puntuales. Dentro de los objetivos de los procesos puntuales está el responder a diversas cuestiones acerca del patrón y en general del fenómeno entre las cuales queremos observar si:

- Los eventos se distribuyen al azar o existe alguna dependencia entre cada uno de ellos.
- Se tiene alguna configuración espacial, es decir de clúster o regular.
- Que proceso ha generado ciertos patrones y si es posible reproducirlos mediante un modelo propuesto.

## 3.2. Proceso puntual espacial en $\mathbb{R}^d$

Los procesos puntuales son modelos estocásticos de patrones de puntos irregulares. En la teoría matemática un proceso puntual puede ser definido y estudiado en un espacio  $\mathbb{R}^d$ . Sin embargo, la mayoría de las veces solo consideraremos un espacio  $\mathbb{R}^2$  o  $\mathbb{R}^3$ . Un proceso puntual espacial es un proceso estocástico  $X$  cuyas realizaciones son un conjunto numerable y finito de eventos  $x_1, \dots, x_n$  localizados aleatoriamente en un espacio  $S$ . Sea  $S$  un subconjunto cerrado en  $\mathbb{R}^d$ , se define a  $B$  como el conjunto de todos los conjuntos de Borel de  $S$  y  $\nu$  la medida de Lebesgue sobre  $B$ .

Se define a un proceso puntual como sigue: Un proceso puntual sobre un espacio  $S$  es un mapeo  $X$  de un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  al conjunto  $N$  de configuraciones localmente finitas de tal manera que para cada conjunto acotado de Borel  $B$ ,  $N(B) = N_x(B)$  es una variable aleatoria que toma valores no negativos en los enteros  $\mathbb{Z}^+$ .

Se define a la variable aleatoria  $N(B)$  que representa el número de puntos en una región plana  $B$ . Se dice que un proceso puntual es simple si con probabilidad uno todos sus puntos  $x$  se producen en ubicaciones distintas en  $S$ . Para consultar literatura más profunda acerca de la teoría fundamental de procesos puntuales se encuentran en: Cox & Tsham (1980), Daley & Vere-Jones (2003) y Moller & Waagepetersen (2003).

### 3.2.1. Patrón puntual

En seguida se presenta una definición formal sobre un patrón puntual. Un patrón puntual en  $S$  es la salida  $x$  de un proceso puntual  $X$  en un conjunto localmente finito de puntos de  $S$ ,  $x = x_1, x_2, \dots, x_n \in S$ .

### 3.2.2. Espacio exponencial

Cuando  $S$  es acotado el espacio de configuraciones  $E$  de un proceso puntual  $X$  sobre  $S$  es equivalente a la unión de espacios  $E_n$  de configuraciones de  $n$  puntos de  $S$ , para toda  $n \geq 0$

$$\bigcup_{n \geq 0} E_n.$$

Esta unión se llama espacio exponencial y esta asociado con la  $\sigma$ -álgebra  $\varepsilon$  para lo cual todas las variables contables estarán definidas de la siguiente forma ya que la función que va del espacio exponencial al conjunto de los números naturales:

$$N(A) : E \longrightarrow \mathbb{N}$$

donde

$$N(A) = \#(x \cap A)$$

Y además  $A \in B_b(S)$  es medible. Es decir  $N(A)$  representa el número aleatorio de puntos en  $A$ . Es importante mencionar que la distribución de un proceso puntual  $X$  estará inducida por la probabilidad  $\mathbb{P}$  sobre el espacio denotado por  $(E, \varepsilon)$  de  $\mathbb{P}$ . La distribución de dimensiones finitas de un proceso puntual  $X$  está definida por la distribución de  $(N(A_1), N(A_2), \dots, N(A_m))$  sobre  $\mathbb{N}^m$ , para cada  $m \geq 1$  y  $m$ -tuple  $(A_1, A_2, \dots, A_m)$  de  $B_b(S)$ .

### 3.2.3. Distribución del número de puntos de un proceso puntual

En la estadística clásica usualmente se utiliza una distribución simple que defina o describa un fenómeno en específico, dicha distribución puede ser continua o discreta. En particular en un proceso puntual la distribución del número de puntos la definimos como:

$$\mathbb{P}(N(B) = n).$$

Para todo  $n = 0, 1, 2, \dots$ , la cual representa la probabilidad de que haya exactamente  $n$  puntos en el conjunto  $B$ , además de que también representa una distribución univariada. Un tipo específico de probabilidades son las vacías o también conocidas como probabilidades nulas, cuya expresión describe la probabilidad de que no haya o existan puntos en un subconjunto específico de  $B$  es decir:

$$\mathbb{P}(N(B) = 0).$$

Este concepto es de utilidad para definir algunas medidas de distancia que se utilizaran más adelante. De manera similar se define la distribución de puntos para un proceso puntual multivariado  $(X_1, \dots, X_n)$  con  $x_1, \dots, x_n$  definidos sobre  $S \subseteq \mathbb{R}^d$  y con su correspondiente función contable  $N_1, \dots, N_n$  la determinada por las probabilidades vacías  $P(N_1(B_1) = 0, \dots, N_n(B_n) = 0)$ .

### 3.2.4. Distribución de un proceso puntual

Usando la representación del espacio exponencial y la propiedad de probabilidad nula se describe la probabilidad de un proceso puntual  $N$  la cual es denotada por un conjunto  $A$

$$P(N \in A) = P(N(B) = 0)$$

Donde  $A$  es un conjunto de patrones puntuales que cumplen con cierta propiedad específica.

### 3.2.5. Estacionariedad e isotropía

Los siguientes dos conceptos importantes para nuestro análisis son la estacionariedad y la isotropía.

Un proceso puntual  $X$  sobre  $\mathbb{R}^d$  es estacionario si para cada  $\zeta \in \mathbb{R}^d$  la distribución del proceso trasladado  $X_\zeta = X + \zeta$  es la misma que la del proceso  $X$ , en otras palabras decimos que es invariante ante translaciones.

Un proceso puntual  $X$  sobre  $\mathbb{R}^d$  es isotrópico si la distribución de  $\rho X$  obtenida de la rotación de  $X$  por cualquier  $\rho$  es la misma que la del proceso  $X$ , con lo cual decimos que es invariante ante rotaciones.

### 3.2.6. Proceso ordenable

Un proceso puntual es ordenable si para alguna localización  $x \in S$

$$\lim_{|\delta| \rightarrow 0} \frac{\mathbb{P}(N[B(x, \delta)] > 1)}{|\delta|} = 0 \quad (3.1)$$

Se dice que un proceso es ordenable cuando no hay puntos coincidentes, en otros términos matemáticos también se define como:

$$\lim_{|d_x| \rightarrow 0} \frac{\mathbb{P}(N(d_x) > 1)}{|d_x|} = 0 \quad (3.2)$$

La ordenabilidad se puede generalizar a un segundo orden cuando para cualquier par de eventos  $x$  e  $y$  tenemos que:

$$\lim_{|d_x| \rightarrow 0 |d_y| \rightarrow 0} \frac{\mathbb{P}(N(d_x) > 1) \mathbb{P}(N(d_y) > 1)}{|d_x| |d_y|} = 0 \quad (3.3)$$

### 3.2.7. Proceso puntual marcado

Un proceso puntual marcado es un proceso puntual con una particularidad muy especial, además de sus coordenadas espaciales para su ubicación geográfica, lleva

asociada consigo una marca, que incluye una característica de interés del fenómeno y aunque al principio se mencionaba que el principal objetivo de este tipo de datos era su distribución, es una característica que ayuda a responder algunas preguntas bastante interesantes.

Se define un proceso puntual estocástico marcado de la siguiente manera: Sea  $K$  un espacio métrico (usualmente  $K \subseteq \mathbb{R}^m$ ), un proceso puntual marcado  $(X, M)$  sobre  $S \times K$  tal que  $X$  es un proceso puntual sobre  $S : (x, m) = (x_1, m_1), (x_2, m_2), \dots$  donde  $m_i \in K$  es una marca asociada con el sitio  $x_i$ . En la figura 3.1 se observa un ejemplo

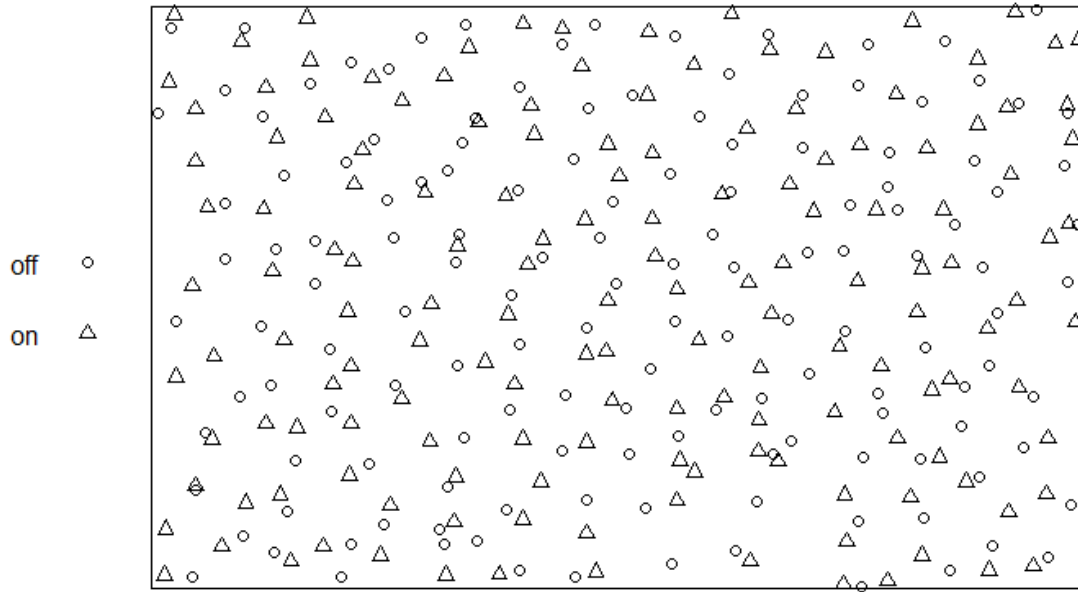


Figura 3.1: Ejemplo de un proceso puntual marcado.

de un patrón puntual marcado, la principal característica de este tipo de patrones es que además de su ubicación tiene asociado una característica de interés, la cual marca la diferencia entre cada tipo de patron.

Para definir de forma correcta a los procesos puntuales, falta considerar las propiedades necesarias que son de vital importancia para la construcción de modelos, además de la estimación de la intensidad y para realizar adecuadamente procesos de inferencia que se definen a continuación.

### 3.3. Momentos de un proceso puntual

En la estadística clásica los conceptos de esperanza y varianza son fundamentales y es que a partir de estas medidas no solo se tiene una idea acerca de los datos, sino

que también ayudan en procesos posteriores de inferencia. En el contexto de procesos puntuales la idea es muy similar, se usan dichos conceptos para definir algunas medidas importantes, ya que la estructura del proceso puntual está descrita por dichas medidas de momentos que aportan información acerca de la estructura del proceso puntual, principalmente la existencia de homogeneidad y la de interacción entre los puntos observados.

Las medidas de momentos son análogas a los momentos de las variables aleatorias en estadística clásica y son un pilar importante en el modelado y la inferencia de los procesos puntuales espaciales.

Se definen las propiedades de primer y segundo orden tanto para el caso univariado como para el multivariado y cuya utilidad es necesaria cuando se hace el análisis de los patrones de interés y su estimación para la construcción de los modelos más adecuados.

### 3.3.1. Propiedades de primer y segundo orden de un proceso puntual

La función de intensidad, es el concepto más importante en la teoría de procesos puntuales, su interpretación usa muchos supuestos tales como la homogeneidad e isotropía. Se definen los momentos de orden  $K$  de la siguiente manera:

$$\mu^{(k)} = (B_1 \times \dots \times B_k) = \mathbb{E}[N(B_1), \dots, N(B_k)]$$

Al tomar los casos especiales cuando  $k=1$  se tiene la función de primer orden que se define más adelante. Para el caso de  $k=2$  tenemos la función de segundo orden, la descomposición es la siguiente:

Para regiones acotadas:  $B_1 \subseteq \mathbb{R}^2$  y  $B_2 \subseteq \mathbb{R}^2$

$$\begin{aligned} \mu^{(k)} = (B_1 \times B_2) &= \mathbb{E} \left[ \sum_{\xi \in X} I(\xi \in B_1) \times \sum_{\eta \in X} I(\eta \in B_2) \right] \\ &= \mathbb{E} \left[ \sum_{\xi, \eta \in X} I((\xi, \eta) \in (B_1, B_2)) \right] \\ &= \mathbb{E} \left[ \sum_{\xi \in X} I(\xi \in (B_1 \cap B_2)) \right] + \mathbb{E} \left[ \sum_{\xi \neq \eta \in X} I((\xi, \eta) \in (B_1, B_2)) \right] \end{aligned}$$

donde la covarianza:

$$Cov[N(B_1)N(B_2)] = \mu^{(2)} = (B_1 \times B_2) + \mu(B_1 \cap B_2) - \mu(B_1)\mu(B_2)$$

La estructura de covarianza de la variable contable  $N(B)$  es más conveniente definirla en términos del momento de segundo orden, y se define más adelante. Las características de primer y segundo orden están definidas para la estructura de la media de una



variable contable  $N(B)$ , donde  $B \subseteq \mathbb{R}^2$ . La intensidad  $\lambda$  se interpreta como el número esperado de eventos por unidad de área. A partir de ahora, y a menos de que se especifique lo contrario, tomaremos que todos los procesos en este trabajo serán estacionarios e isotrópicos, es un supuesto muy fuerte pero tendrá que ser considerado, ya que bajo estas condiciones la interpretación de  $\lambda$  será más sencilla de realizar.

### 3.3.2. Propiedad de primer orden de un proceso puntual

La propiedad de primer orden se define:

$$\mu^{(1)}(B) = \mathbb{E}[N(B)] = \sum_{x \in X} I(x \in B)$$

La cual se define como el número promedio de puntos para el área  $B$ . Claramente observamos que la media depende del conjunto  $B$ , por lo cual es una función que opera sobre conjuntos. Se define a la función de intensidad como:

$$\lambda(B) = \mathbb{E}[N(B)]$$

si la medida de  $\mu^{(1)}(\cdot)$  es absolutamente continua se define a la función de intensidad como:

$$\mu^{(1)}(B) = \int \lambda(u) du$$

Otra definición que se puede tomar en cuenta para las propiedades de primer y segundo orden son las que se dan a partir de  $\mathbb{E}[N(x)]$ , que es el número promedio de puntos. La función de intensidad de primer orden es:

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \frac{\mathbb{E}[N(dx)]}{|dx|}$$

Donde  $dx$  será una pequeña región que contiene el punto  $x$ .

### 3.3.3. Propiedad de segundo orden de un proceso puntual

Suponiendo estacionariedad e isotropía definiremos a la propiedad de segundo Momento como:

$$K(h) = \lambda(x)^{-1} \mathbb{E}[N_0(h)]$$

donde  $N_0(h)$  denota el número de eventos que están situados a una distancia menor que  $h$  de un evento arbitrario. La interpretación de esta función no es tan arbitraria como la de primer orden. Sin embargo, su importancia practica sí, ya que esta definición permite expresarla como la media de una cantidad observada, con lo cual facilita

la estimación a partir de datos que ya se conocen. De esta manera se establece una relación entre  $K(h)$  y  $\lambda_2(h)$  y se tiene que:

$$K(h) = 2\pi\lambda^{-1} \int_0^h \lambda_2(x) dx.$$

Dicha relación es válida si se supone ordenabilidad en el proceso estudiado, es decir, que  $\mathbb{E}[N(dx)][N(dy)] \sim \mathbb{P}[N(dx)][N(dy) = 1]$  al cumplir esta condición se espera que el número promedio de eventos menor a una distancia  $h$  se puede calcular obteniendo la integral de la intensidad condicional, que es la integral de la intensidad de segundo orden. Algunas observaciones importantes a mencionar es que es más sencillo interpretar  $K(h)$  sin embargo teóricamente es más sencillo trabajar con  $\lambda_2(h)$ :

$$\lambda_2(h) = \lambda^{(2)}(2\pi h)^{-1}K(h).$$

Tomamos  $l^{(2)}$  como la densidad de producto de segundo orden la función de densidad asociada a la medida del momento K-esimo factorial será:

$$\alpha^{(2)}(B_1 \times B_2) = \int \int l^{(2)}(x, y) dx dy.$$

$l^{(2)}(x, y) dx dy$  puede ser interpretado como la probabilidad de que exista un punto del proceso puntual en cada dos grupos de datos especificados infinitesimalmente cuyas áreas son  $dx dy$  respectivamente.

La función de intensidad de segundo orden también se puede definir como:

$$\lambda_2(x) = \lim_{|dx||dy| \rightarrow 0} \frac{\mathbb{E}[N(dx)]\mathbb{E}[N(dy)]}{|dx||dy|} \quad (3.4)$$

### 3.3.4. Estructura de covarianza

La estructura de covarianza de la variable contable  $N(B)$  es más conveniente definirla en términos del momento de segundo orden  $\alpha^{(2)}$  como sigue: Para un proceso puntual  $X \in S$  el momento de segundo orden factorial se define como:

$$\alpha^{(2)}(B_1 \times B_2) = \mathbb{E} \sum_{\xi \neq \eta \in X} I((\xi, \eta) \in (B_1, B_2)) = \mu^{(2)}(B_1 \times B_2) - \lambda(B_1 \cap B_2)$$

Se tiene al igual que la función de intensidad el caso cuando  $\alpha^{(2)}$  es absolutamente continua con respecto de la medida de Lebesgue sobre  $(S, B(S))^2$  y  $\lambda^{(2)}(\xi, \eta)$  es la densidad de la intensidad de orden 2 de  $X$  entonces:

$$\mu^{(2)}(B_1 \times B_2) = \int_{-a}^b I((\xi, \eta) \in (B_1, B_2)) \lambda^{(2)}(\xi, \eta) d\xi d\eta$$

donde  $\lambda^{(2)}(\xi, \eta)$  puede ser interpretado como la probabilidad de observar un punto en cada una de las dos regiones en áreas infinitesimalmente muy pequeñas.

Para hacer una caracterización acerca de la tendencia de un patrón, es decir, observar si los puntos siguen un comportamiento de atracción o de repulsión entre ellos se puede auxiliar de la función de correlación:

$$g(u, v) = \frac{\lambda^{(2)}(u, v)}{\lambda(u)\lambda(v)}; \lambda(u), \lambda(v) > 0.$$

Si los puntos aparecen de forma independiente uno del otro tenemos que  $\lambda^{(2)}(u, v) = \lambda(u)\lambda(v)$  es decir  $g(u, v) = 1$  tiene un patrón completamente aleatorio, si  $g(u, v) > 1$  representa atracción entre los puntos, mientras que si  $g(u, v) < 1$  representa repulsión.

La densidad de covarianza de un proceso puntual espacial es:

$$\gamma(x) = \lambda_2(x, y) - \lambda(x)\lambda(y).$$

Si asumimos estacionariedad e isotropía:

$$\lambda(x) \equiv \lambda = \mathbb{E}[N(A)]/|A|$$

$$\lambda_2(x, y) \equiv \lambda_2(\|x - y\|).$$

### 3.3.5. Intensidad de primer y segundo orden para un proceso multivariante o multitypo

Se generalizan las funciones de intensidad de primer y segundo orden para el caso multivariado, es decir, cuando ocurren más de un tipo de puntos en un patrón dado y que además se pueden clasificar de alguna forma, los supuestos de que es un proceso homogéneo isotrópico así como ordenable, serán heredados del caso univariado. Se define así a la intensidad de primer orden de un proceso multivariado como se especifica a continuación:

En un proceso multivariante o multitypo los eventos son de dos o más tipos diferentes de puntos, si asumimos estacionariedad e isotropía, y con  $N_j(A)$  para el número  $j$  de tipos de eventos en una región plana  $A$ .

Se define a la intensidad de primer orden como  $\lambda_j = \mathbb{E}[N_j(A)]/|A|$  Y a la intensidad de segundo orden como:

$$\lambda_{ij}(r) = \lim_{|dx||dy| \rightarrow 0} \frac{\mathbb{E}[N_i(dx)]\mathbb{E}[N_j(dy)]}{|dx||dy|} (\|x - y\| = r),$$

cuya densidad de covarianzas será:

$$\gamma_{ij}(r) = \lambda_{ij}(r) - \lambda_i\lambda_j,$$

y su función de segundo momento se reduce a:

$$K_{ij}(h) = \lambda^{-1}\mathbb{E}[N_{oij}(h)].$$

La interpretación es similar al caso univariado. Es decir,  $N_{0ij}(h)$  denota el número de eventos de tipo  $j$  situados a una distancia menor que  $h$  de un evento arbitrario del tipo  $i$ . De igual forma se generaliza la relación  $K(h)$  y  $\lambda_2(h)$  para el caso multivariado:

$$K_{ij}(h) = 2\pi(\lambda_i\lambda_j)^{-1} \int_0^h \lambda_{ij}(x)x dx$$

### 3.3.6. Independencia y etiquetado aleatorio

La independencia y el etiquetado aleatorio serán las dos hipótesis de referencia para que se pueda determinar una relación o una asociación espacial entre dos tipos de eventos para un proceso bivalente.

La hipótesis de independencia es lo que queremos observar y sucede cuando dos tipos de eventos son generados por un par de procesos univariantes que cumplen la propiedad de ser independientes.

La hipótesis de etiquetado consiste en que dos tipos de eventos son generados por etiquetado de eventos de un proceso univariante en series de pruebas de Bernoulli mutuamente independientes. Es decir, que si generamos diferentes  $K$  funciones, cada una de estas hipótesis cumplirá que por cada dos procesos independientes tenemos que  $K_{ij}(h) = \pi h^2$ , es decir, que para cada proceso etiquetado aleatoriamente  $K_{ij}(h) = K(h)$ .

### 3.3.7. Tipos de patrones

Los procesos puntuales pueden tener variaciones con características diferentes y se localizan en formas muy diversas en la region de estudio. Se dice que un proceso puntual es completamente aleatorio cuando el comportamiento de los puntos es independiente en cualquier subregión. Sin embargo, se puede observar que la intensidad puede o no ser homogénea sobre toda el área de estudio.

En otros casos, los eventos pueden estar localizados o pueden ocurrir de forma constante, pero pueden mostrar algún tipo de interacción. En este caso se puede ver un comportamiento de agrupamiento o de atracción de los puntos, es decir, se forman grupos pequeños de puntos en conglomerados. Es aquí donde es importante realizar un análisis que pueda ayudar a entender los diferentes tipos de variaciones entre los patrones y si es que guardan algún tipo de relación entre sí.

La razón de la diferencia entre un tipo de patrón y otro se debe a la variación de la función de intensidad la cual puede ser homogénea o totalmente distinta para todo nuestro dominio  $D$ .

Los tres tipos de patrones puntuales son: regulares, aleatorios y conglomerados.

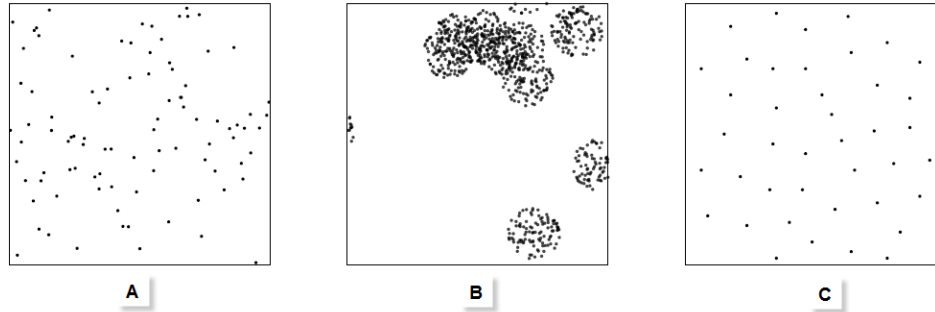


Figura 3.2: Simulación de los tres diferentes tipos de patrones puntuales: aleatorio panel (A); agregado panel (B) y regular panel (C).

La forma más sencilla de distinguir entre cada uno de ellos será por la distribución que presente con respecto de sus vecinos más cercanos estimando esta distancia por medio de las funciones de segundo orden.

En la figura 3.2 se observa un ejemplo con datos simulados sobre los tres tipos de patrones puntuales, en esta se pueden distinguir las diferencias entre cada patrón y si bien falta considerar pruebas estadísticas para afirmar la clasificación de cada uno de ellos, es un primer ejemplo para distinguirlos. Para un patrón agregado se observa que la distancia promedio entre cada uno de los eventos con respecto a sus vecinos más cercanos será más pequeña que la misma distancia para un patrón completamente aleatorio. De la misma forma, un patrón regular mostrara por el contrario que la distancia media entre puntos será mucho mayor a la de un proceso completamente aleatorio. Es de suma importancia identificar el tipo de patrón así como la intensidad para poder ajustar algún tipo de modelo.

Diggle (2003) y Moller and Waagepetersen (2003) dan las definiciones más apropiadas de los diferentes tipos de puntos y sus principales propiedades.

### 3.4. Intensidad condicional

La intensidad condicional para un proceso puntual se define en términos de la función de intensidad  $\lambda(x)$ . Esta intensidad también se conoce como la Intensidad condicional de Papangelou  $\lambda(x|X)$  donde  $x$  es una localización y  $X$  es un patrón puntual, dicha intensidad estima el número promedio de puntos dado una realización de  $X$ . Esta propiedad será fundamental para los procesos de Gibbs.

$$\mathbb{E}[\lambda(x|N)] = \lambda(x), x \in \mathbb{R}^2.$$

### 3.5. Teorema de Campbell

El teorema de Campbell aporta un resultado muy importante, ya que ofrece una alternativa para calcular el valor medio de una suma de densidades y establece que:

$$\mathbb{E}[\sum_{x \in N} f(x)] = \int f(x)\lambda(x), dx$$

En otras palabras, es suficiente conocer  $f(x)$  y la función de intensidad  $\lambda(x)$  para calcular la esperanza de la suma de densidades.

### 3.6. Análisis preliminar de un patrón puntual

El análisis de los patrones puntuales está enfocado en la distribución espacial de ciertos eventos observados con la finalidad de realizar inferencias sobre el proceso que generó dicho patrón. En particular se tienen dos aspectos importantes de interés: La distribución de eventos en el espacio y la posibilidad de que exista interacción entre los puntos. Para un análisis meramente descriptivo, se puede representar las localizaciones del patrón puntual en un área de estudio, esto muestra una idea grafica de la distribución de puntos, y puede ser de gran utilidad para poder concluir si éstos siguen un comportamiento determinado.

Para el estudio de los procesos puntuales es necesario estimar la intensidad, en párrafos anteriores se han mencionado que la intensidad la cual está representada cómo  $\lambda$ , será el número de eventos por unidad de área. La estimación de la función de intensidad se realiza a partir de las posiciones de los eventos se puede realizar de forma puntual, mediante una función llamada núcleo (estimación de Kernel), la cual se define más adelante, o por medio de los famosos métodos de distancias conocidos también como las propiedades de segundo orden.

Existen varias formas de estimar la función de intensidad de un proceso puntual: los métodos paramétricos y no paramétricos. Los métodos no paramétricos se basan únicamente en la posición de los puntos y no hay modelos involucrados, por lo general se utiliza el método de kernel o método de cuadrantes.

Los métodos paramétricos proponen un modelo paramétrico para el proceso puntual que genere el patrón observado, y por ende para la función de intensidad  $\lambda(B)$ , en dichos modelos se utiliza Máxima Verosimilitud o máxima pseudo verosimilitud, por lo que se necesita conocer la función de densidad del proceso.

## 3.7. Prueba de CSR (aleatoriedad espacial completa)

Se busca satisfacer el supuesto fundamental de la estadística espacial, que consiste en resolver el problema sin tener que utilizar estadística espacial. En el estudio de un proceso puntual la mínima característica es que cumpla la aleatoriedad espacial completa. Intuitivamente es el ruido blanco de un proceso puntual espacial, donde la principal característica es la ausencia de estructura dentro de los datos, lo que plantea la hipótesis CSR es que la media de eventos se distribuya de forma aleatoria e independiente sobre el área de estudio. Es decir, que no existan regiones con mayor o menor número de puntos y más aun no haya la interacción entre ellos, con lo cual podríamos concluir que terminamos el análisis.

Se parte del supuesto de que el proceso se distribuirá de forma aleatoria y sin interacción entre los puntos, para ello se asume que tendrá una distribución Poisson cuyas características son:

- Los puntos se distribuyen uniformemente en  $S$  con intensidad constante  $\lambda$ .
- El número de puntos en cualquier conjunto acotado de Borel  $B \subset S$  sigue una distribución Poisson con  $E[N(B)] = \lambda|B|$ .

Si un patrón puntual de estudio es completamente aleatorio, cumple con los requisitos anteriores, es decir, que la intensidad sea homogénea a lo largo de todo el conjunto acotado de Borel. De esta forma, el número de eventos por unidad cuadrada es el mismo a lo largo de toda el área, y que el número de eventos en cualquier subregión de estudio, tenga una distribución Poisson. Esto implica que implícitamente estaremos asegurando que no existe interacción entre ellos. Cuando se cumple con estas características se habla de un proceso de Poisson homogéneo.

Cuando se rechaza la hipótesis de aleatoriedad espacial completa, se tiene la alternativa de que se trata de un patrón es regular o agregado. En ellos recaerá el tipo de modelo que trataremos de ajustar a los datos. Primero se revisan los métodos para contrastar si el patrón de interés cumple o no la hipótesis de CSR.

### 3.7.1. Índice de dispersión

El índice de dispersión es el método para probar si el patrón es completamente aleatorio. Obtener dicho estadístico es bastante sencillo, bajo el supuesto de que el patrón se distribuye Poisson con media  $\lambda$  entonces,

$$\mathbb{E}[N(B_j)] = \lambda|B_j|, j = 1, \dots, k,$$

de esta manera

$$\mathbb{E}[N(B_j) = a] = \frac{(\lambda|B_j|)^a e^{-\lambda|B_j|}}{a!}, j = 1, \dots, k.$$

Sabemos que al ser una distribución Poisson, implica que tanto los valores para la esperanza y la varianza es igual al valor de dicho parámetro, por tanto el índice de dispersión se construye a partir de este supuesto. Es decir, similar a la construcción del coeficiente de variación, ya que se realiza el cociente de la varianza con respecto a la esperanza de la variable aleatoria, gracias a las características de dicha distribución, se puede concluir que cuando la media es similar a la varianza tendremos valores aproximados a 1.

$$\frac{Var[N(B_j)]}{\mathbb{E}[N(B_j)]} = 1.$$

### 3.7.2. Método de cuadrantes

Otro tipo de método para probar la aleatoriedad espacial completa es el método de cuadrantes. Dicho método consiste en dividir el dominio en subregiones independientes cuadradas de igual tamaño. Recordar el tipo de pruebas no paramétricas realizadas con las tablas de contingencia para poder probar independencia, el estadístico es más o menos similar, ya que utilizaremos una prueba de bondad de ajuste de la  $\chi^2$  cuadrada, donde la hipótesis nula será que los  $n$  puntos estarán distribuidos de forma independiente a lo largo de  $D$ . Es decir en términos prácticos que el número de puntos por subregión serán variables de Poisson independientes con media en común.

Este método se realizara de la siguiente forma:

- Se divide el área  $S$  en  $B_1, \dots, B_k$  cuadrantes cada uno de una cierta área  $B_j$ .
- Se cuenta el número de eventos  $N(B_j)$  en cada uno de los cuadrantes.
- Se estima la función de intensidad por cada uno de los cuadrantes  $\hat{\lambda} = N(B_j)/|B_j|$  mediante la razón del número de eventos situados en ese cuadrante entre el área del mismo.

El estadístico se define como:

$$I = \frac{(k-1) \sum_{j=1}^k [N(B_j) - \hat{N}(B)]^2}{\hat{N}(B)} \sim \chi^2_{(k-1)}.$$

La hipótesis nula es que los datos se distribuyan de forma aleatoria vs la hipótesis alternativa que se distribuyen de cualquier otra forma, dando indicios estadísticos de que se trata de un comportamiento de atracción o repulsión. En dicho contexto rechazar la hipótesis nula puede indicar una distribución no homogénea de los eventos,



que sugiere dependencia. Para valores del estadístico suficientemente grandes podemos decir que se trata de distribuciones agregadas, así como para valores pequeños tenemos evidencia de patrones regulares.

Entre las ventajas tenemos que es un método sencillo, rápido e intuitivo; sin embargo, existen desventajas entre las cuales tenemos que la calidad de la aproximación a  $\lambda$  depende del tamaño de cada uno de los cuadrantes, lo cual puede llevar a cometer algún tipo de error a la hora de contrastar las hipótesis.

Por ejemplo, veamos el caso de un proceso simulado de un patrón homogéneo. En la figura 3.3(izquierda) se muestra un proceso puntual de Poisson con 100 puntos. En la figura 3.3 (derecha) está el método de cuadrantes con una división de 3x3.

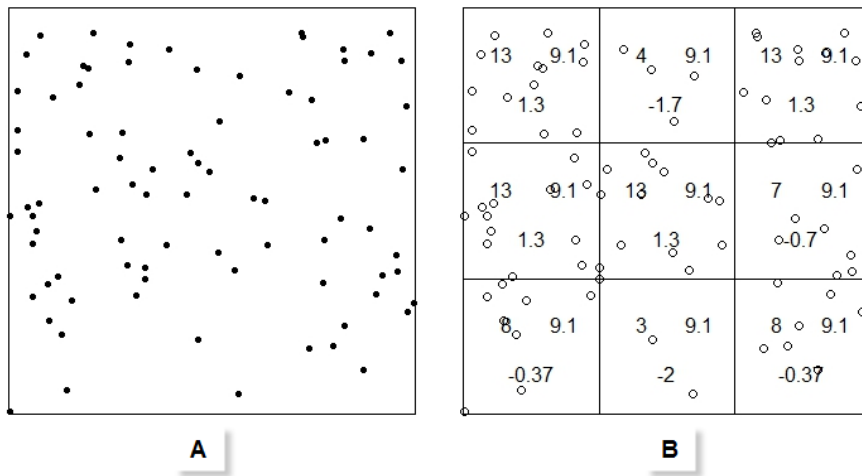


Figura 3.3: Ejemplo del método de cuadrantes para un patrón puntual homogéneo de poisson con 100 puntos: patrón puntual panel (A); método de cuadrantes con división de 3x3 panel (B).

Se observa el valor esperado y el observado para cada cuadrante así como el índice de dispersión. La siguiente salida en R Project refleja la prueba de hipótesis para el método de cuadrantes. Observamos que el valor de p es 0.1454 por lo cual no se rechaza la hipótesis de que se distribuye de forma aleatoria.

```
Chi-squared test of CSR using quadrat counts
Pearson X2 statistic
```

```
data: aleatorio
X2 = 14.366, df = 8, p-value = 0.1454
alternative hypothesis: two.sided
```

```
Quadrats: 3 by 3 grid of tiles
```

### 3.7.3. Estimación por Kernel

La estimación de Kernel es un método de estimación no paramétrico en el cual la función de densidad no imponen tantas restricciones como en otros tipos de métodos. La función de Kernel es una función ponderada la cual satisface ciertas condiciones de regularidad, es una función de densidad simétrica alrededor de cero que depende de un cierto parámetro  $r$ , el cual controla el grado de suavización en el cálculo de la intensidad, es decir cuando más grande sea el parámetro la suavidad es mayor. El Kernel es una función no negativa que satisface:

$$\int_{-\infty}^{\infty} k(u) du = 1$$

$$k(-u) = k(u)$$

donde:  $u = \|s_i - s_j\|$ .

Se tiene diferentes tipos de funciones de Kernel, algunas de ellas son: gaussiano, cuadrático, de mínima varianza, entre otros. Para nuestro análisis únicamente se utilizó el Kernel Gaussiano y dicho método se usa para localizar de manera gráfica y así tener un mejor comparativo visual para identificar las regiones de mayor concentración de incendios.

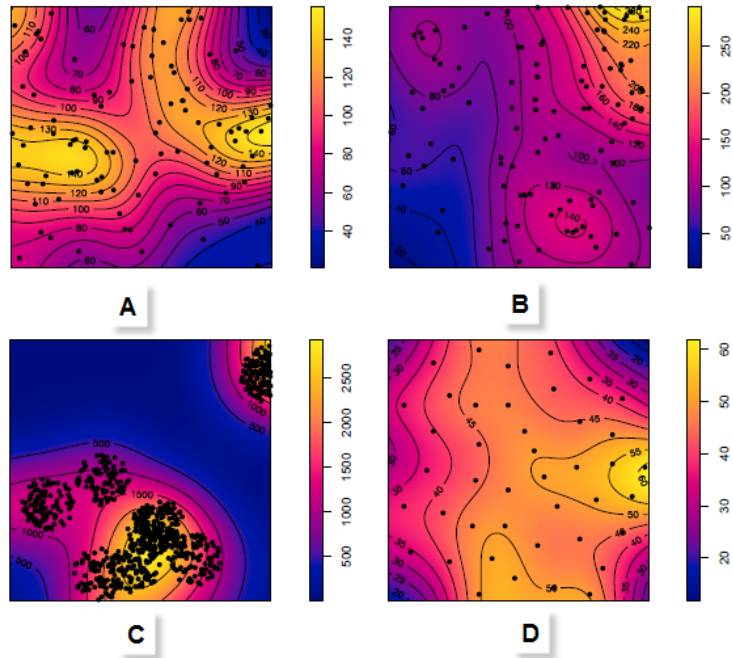


Figura 3.4: Estimación de Kernel gaussiano para los diferentes tipos de patrones puntuales: patrón aleatorio panel (A); patrón no homogéneo panel (B); patrón agregado panel (C) y patrón regular panel (D).

En la figura 3.4 se observan los diferentes tipos de patrones, así como su estimación de Kernel. En la parte superior se observan patrones completamente aleatorios, de lado izquierdo se observa un patrón homogéneo, mientras de lado derecho se observa un patrón no homogéneo. En cuanto a la estimación de Kernel se puede ver la diferencia al suponer homogeneidad, del lado izquierdo vemos una estimación de la intensidad más uniforme que la observada del lado derecho. En la parte inferior de la gráfica se observan los otros dos casos: un patrón agregado y un patrón regular. En el agregado se observa que la concentración de los puntos originan que la estimación del Kernel se concentre sobre la parte de la región donde se concentra la mayor parte de los puntos, mientras que del lado derecho se observa la repulsión entre los puntos. En cuanto a la estimación de Kernel, se refleja en el patrón la predominancia de la estimación de manera uniforme.

## 3.8. Método basado en distancias

Los métodos anteriores pueden ayudar a probar la hipótesis de aleatoriedad espacial completa (CSR). Sin embargo, muchas de ellas no son muy buenas o tienden a reflejar resultados no muy exactos los cuales dependen del criterio y la forma en la que son construidas. Existen otro tipo de métodos para hacer dicha prueba, estadísticos que se basan en las distancias entre los eventos o entre eventos muestreados seleccionados de un patrón. Este tipo de estadísticos son conocidos como métodos basados en distancias.

Entre las medidas más ampliamente utilizadas para construir resúmenes acerca de los patrones observados se tienen los métodos de distancias al vecino más cercano y las propiedades de momento de segundo orden, ya que para contrastar una hipótesis o estudiar la bondad de ajuste de un modelo serán de gran utilidad haber definido previamente dichas funciones.

La distancia al vecino más próximo puede medirse desde un origen arbitrario o desde un evento arbitrario a su vecino más cercano. Dado que existe una distancia entre eventos una forma de caracterizar a los procesos puntuales es mediante espacio vacío. A continuación definiremos dichas distribuciones las del espacio vacío así como la del vecino más cercano.

### 3.8.1. Función G (distribución del vecino más cercano)

Se define a la función G o también llamada distribución del vecino más cercano como:

$$G(h) = \mathbb{P}[\text{Distancia de un evento arbitrario al evento más cercano} < h].$$

Dicha función, como su nombre lo dice, determina la distancia mínima entre even-

tos donde  $d_i$  representa las distancias mínimas entre un evento y sus vecinos, una forma de calcular empíricamente la función  $\widehat{G}(h)$  será la siguiente:

$$\widehat{G}(h) = n^{-1} \#(d_i < h).$$

Bajo la hipótesis de CSR la distribución teórica de  $\widehat{G}(h)$  depende de  $n$ , y existe un problema para definir una expresión debido a los efectos de frontera, por lo cual ignorando este problema se puede definir a la distribución  $\widehat{G}(h) = \mathbb{P}(h < d) = 1 - \mathbb{P}(h > d)$  es decir utilizando los complementos definiremos a la distribución como:

$$\widehat{G}(h) = \mathbb{P}(h < d) = 1 - \mathbb{P}(h > d).$$

Lo que es equivalente a que sea igual a 1 menos la probabilidad de que no haya eventos en el área  $A$ . Cuando tenemos un modelo que es completamente aleatorio la función de distribución del vecino más cercano recae en un caso particular que es la distribución Poisson, por lo cual la función de distribución se puede escribir como:

$$\widehat{G}(d) = \mathbb{P}(h < d) = 1 - e^{(-\lambda\pi d^2)}.$$

Enseguida se describen cierto tipos de características para la función  $G$  empírica:

- Si  $\widehat{G}(h)$  crece rápidamente en distancias muy pequeñas, tenemos que los eventos están más concentrados que un proceso completamente aleatorio por lo cual existe un patrón agregado.
- Si por el contrario  $\widehat{G}(h)$  crece lentamente hasta una cierta distancia dada, se puede asociarlo con un patrón regular o de repulsión.

Para la comparación visual se le agregan bandas de confianza mediante simulaciones de Monte Carlo calculando  $n$  realizaciones de un proceso completamente aleatorio, y tomando como límite superior e inferior el valor máximo y mínimo respectivamente.

### 3.8.2. Función $F$ (distribución del espacio vacío)

Es aquí donde toma importancia la probabilidad de eventos vacíos, ya que la función  $F$  se define como la distribución del espacio vacío:

$$F(h) = \mathbb{P}[\text{distancia de un punto arbitrario al evento más cercano} < h].$$

Se define, al igual que en el caso anterior la función de distribución del espacio vacío, para ello partiremos de construir un vector en el cual se toma una serie de puntos aleatoriamente, y cuyos puntos sirven como referencia para poder encontrar la distancia de cada punto elegido de forma aleatoria al evento más cercano, a dicha

medida se llama  $d_i$ . Es decir, se construye el estimador para la función de distribución de la distancia de un punto a un evento como:

$$\widehat{F}(h) = n^{-1} \#(d_i < h).$$

Bajo la hipótesis de aleatoriedad espacial completa, la distribución  $F$  será la misma que la función  $G$ . Sin embargo, la diferencia radica en la forma en la cual se tomaron las distancias, puesto que en la función  $G$  fue a partir de eventos y la función  $F$  es a partir de eventos y puntos arbitrarios seleccionados al azar. Se definen ciertos tipos de características para la función  $F$  empírica:

- Si  $\widehat{F}(h)$  crece rápidamente en distancias muy pequeñas, a diferencia de la función  $G$  diremos que nuestro patrón es regular.
- Si por el contrario  $\widehat{F}(h)$  crece lentamente para distancias pequeñas y rápidamente cuando las distancias son largas, se asocia con un patrón agregado o de clúster.

Para la comparación visual y grafica se le agregan bandas de confianza similar a las de la función  $F$ : mediante simulaciones de Monte Carlo calculando  $n$  realizaciones de un proceso completamente aleatorio, y tomando como límite superior e inferior el valor máximo y mínimo respectivamente.

### 3.8.3. Función K de Ripley

La función K de Ripley es la función por excelencia que se utiliza para determinar el tipo de patrón de interés que se esta estudiando, dicha función está basada en la estructura de segundo orden de un proceso (Ripley, 1977), en otras palabras la función nos ayuda a determinar eficientemente el tipo, el rango y la intensidad del patrón espacial ya que describe las características del proceso puntual a diferentes escalas mediante el análisis de las distancias entre los puntos. La función K de Ripley se define como:

$$K(h) = \lambda^{(-1)} \mathbb{E}[N_o(h)],$$

donde  $N_o(h)$  denota el número de eventos que están situados a una distancia menor que  $h$  de un evento arbitrario. Lo cual también se interpreta como:

$$K(h) = \lambda^{(-1)} \mathbb{E}[\text{Eventos a una distancia} < h \text{ para un evento arbitrario}], \quad (3.5)$$

Donde  $\lambda$  determina la intensidad (Número de puntos por unidad). Para un proceso de Poisson homogéneo la K de Ripley es  $K(h) = \pi h^2$ , este será nuestro punto de comparación que contrastaremos con el patrón observado así tendremos que:

- $K(h) > \pi h^2 \Rightarrow$  Patrón agregado.

- $K(h) < \pi h^2 \Rightarrow$  Patrón regular.

Cuando el patrón no es completamente aleatorio nuestro análisis de segundo orden mediante la función  $K$  de Ripley puede llegar a sobrestimar la dependencia entre eventos, ya que la función  $K$  es sensible a la falta de homogeneidad y a la anisotropía, ya que proporciona funciones de distribución sesgadas cuando estos requisitos no se cumplen y que son perfectamente interpretables ya que está definida en términos de aleatoriedad (Dixon, 2002). Es por eso que para resolver esta dificultad se introduce la función  $K$  no homogénea donde la intensidad ya no es constante. Un estimador insesgado de la  $K$ -función inhomogénea propuesto por Baddeley (2000), es el siguiente:

$$K_{in}(t) = \frac{1}{|A|} \sum_{x_i \in X \cap A} \sum_{x_j \in X \cap A \setminus x_i} \frac{I(\|x_i - x_j\| \leq t)}{\lambda(x_i)\lambda(x_j)w_{ij}} \quad (3.6)$$

Ahora el análisis se enfoca en la interacción espacial entre dos tipos de patrones, ésta la veremos por medio de las distancias, cuando los eventos de ambos patrones se encuentran más próximos o lejanos de lo que parecen. Para este análisis se usa la extensión al caso bivariado de la función  $K$  de Ripley. La función  $K$  Cruzada ayuda a analizar la interacción entre patrones puntuales multitypo, la función  $K$  cruzada se define como:

$$K_{ij}(h) = \lambda_j^{(-1)} \mathbb{E}[\text{evento tipo } j \text{ a una distancia } < h \text{ de un evento tipo } i \text{ arbitrario}], \quad (3.7)$$

La interpretación será análoga al caso univariado, se observa que si la función sobrepasa las envolturas superiores del proceso tendremos indicios de atracción de los puntos o la formación de clúster. Por el contrario, si son inferiores a las envolturas inferiores se observa repulsión entre los procesos. Las envolturas serán generadas por simulaciones de Monte Carlo. Si los procesos son independientes se mantendrán dentro de las bandas de confianza. Existen otros métodos alternativos para detectar comportamientos de atracción y repulsión, pero solo la función  $K$  de Ripley permite detectar eficientemente el tipo de patrón tanto a pequeña como a gran escala espacial, el tamaño de los agregados y las distancias de inhibición, ya que se basa en el análisis de la distancia entre un punto con el resto del patrón (Dixon, 2002). Para una mejor interpretación de la función  $K$  de Ripley se puede utilizar una transformación de la misma:

$$L(h) = \sqrt{K(h)/\pi}.$$

La cual conocemos como la función  $L$ , dicha transformación tiene como objetivo el poder linealizar la función, además de que con esta transformación la varianza es constante ya que para un patrón de Poisson Homogéneo  $L(h) = h$  es decir representa a la identidad.

En las figuras 3.5, 3.6 y 3.7 se observa el comportamiento de cada una de los métodos de distancias (Función  $G$  y función  $F$  en la parte superior, función  $K$  y  $L$  en la parte inferior) aplicado a los diferentes tipos de patrones: aleatorio, de clúster y regular. Para el caso del patrón aleatorio se observa en la figura 3.5, en cada una de las

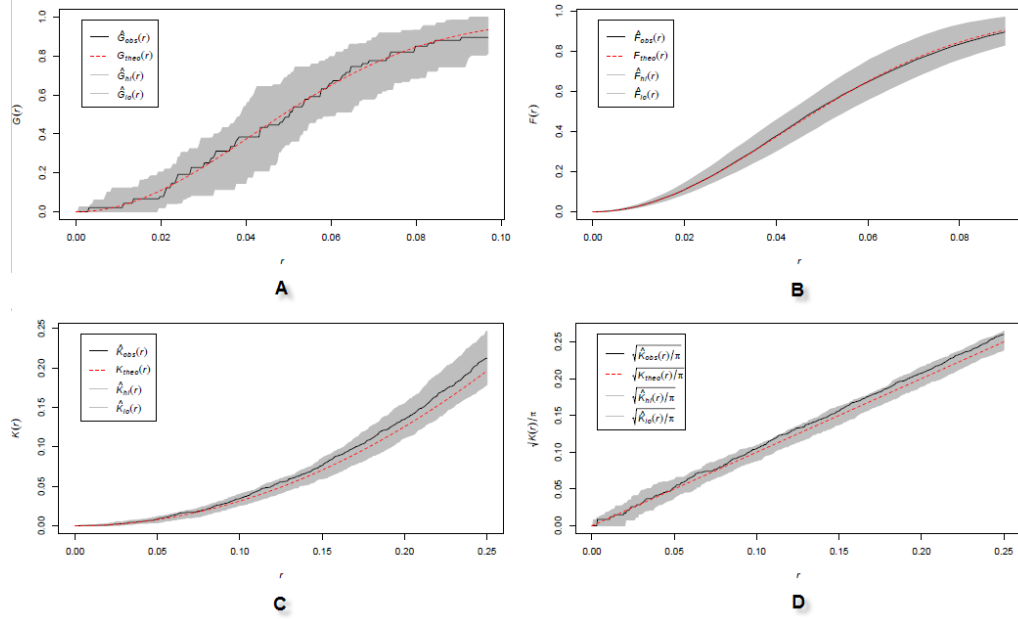


Figura 3.5: Funciones de distancias con intervalos de confianza construidos mediante simulaciones de Montecarlo correspondientes a un patrón aleatorio: función G panel (A); función F panel (B); función K panel (C) y función L panel (D).

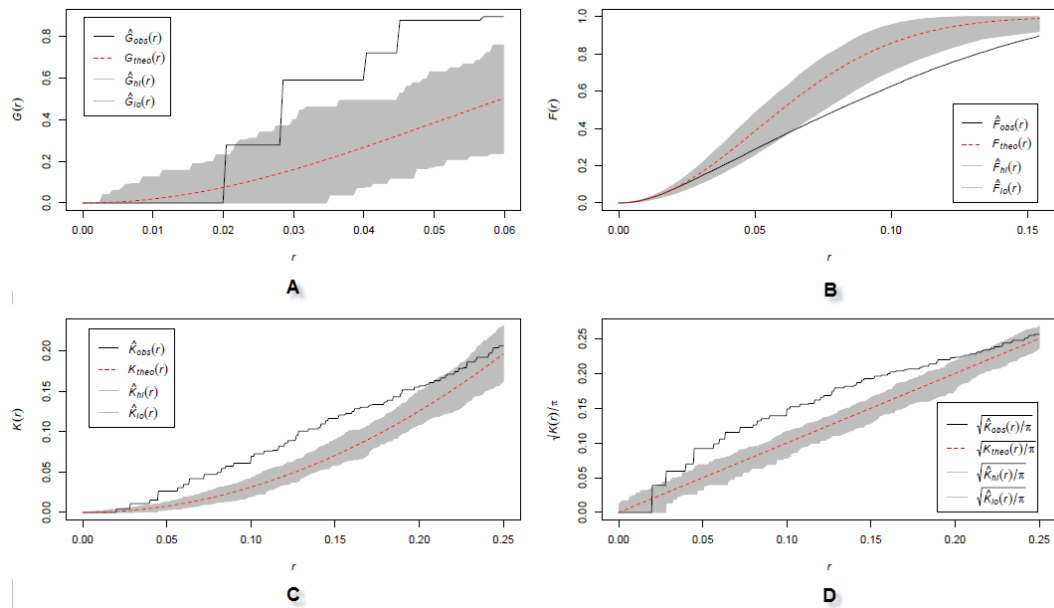


Figura 3.6: Funciones de distancias con intervalos de confianza construidos mediante simulaciones de Montecarlo correspondientes a un patrón cluster: función G panel (A); función F panel (B); función K panel (C) y función L panel (D).

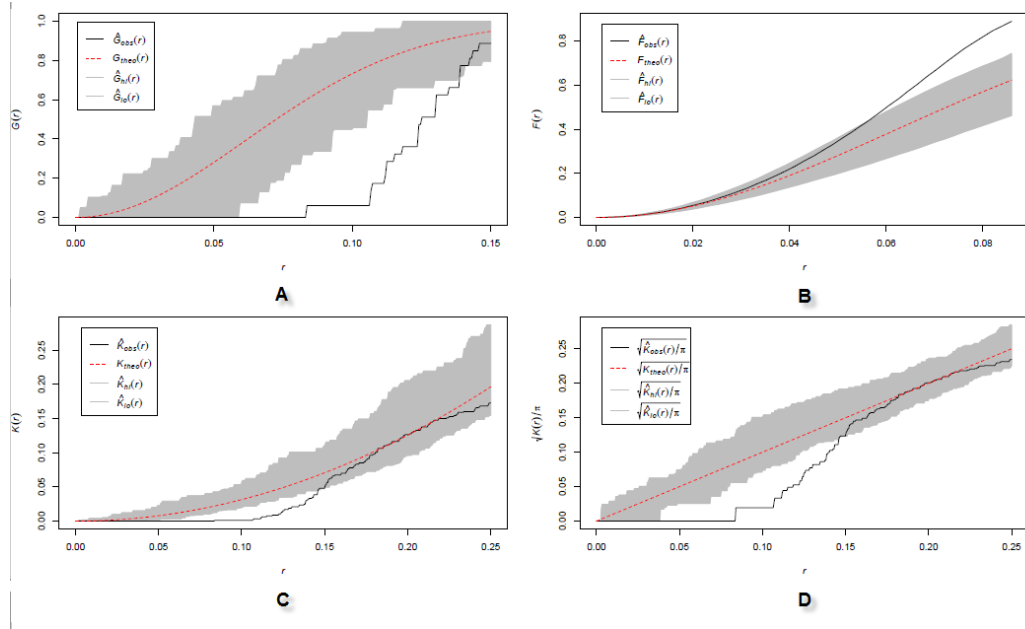


Figura 3.7: Funciones de distancias con intervalos de confianza construidos mediante simulaciones de Montecarlo correspondientes a un patrón regular: función G panel (A); función F panel (B); función K panel (C) y función L panel (D).

funciones se gráfica, la función observado (línea continua) vs la función teórica (línea punteada) para el caso de las cuatro funciones la función observada no sobrepasa ni se sale de las bandas de confianza, las cuales se generar mediante 99 simulaciones del proceso, por lo cual se concluye que el patrón es completamente aleatorio.

En la figura 3.6 se hace la comparación para un patrón de clúster, para la función G, la función observada crece más rápidamente que la función teórica a distancias pequeñas, y ésta sobrepasa las bandas de confianza con lo cual se argumenta que los eventos se ubican más cerca que en un patrón aleatorio. En la función F por el contrario la función observada crece más lentamente que la teórica a distancias pequeñas, para la función K de Ripley, se observa que la función observada es mayor que la teórica y que ésta además sobresale de las bandas de confianza, para el patrón regular (Figura 3.7) el comportamiento es totalmente contrario al del patrón clúster, la función G crece más lentamente que la teórica, mientras en la función F crece más rápido que la función teórica, la función K de Ripley presenta un comportamiento similar al de la función G. Estos ejemplos con patrones simulados son muy ilustrativos para ver el comportamiento de los diversos métodos de distancias y que serán aplicados al análisis de los datos de incendios.



## 3.9. Modelos de procesos puntuales

Los modelos de procesos puntuales los podremos clasificar en tres amplias clases: a) Modelos de procesos puntuales de Poisson, los cuales son caracterizados por tener la intensidad tanto homogénea como no homogénea  $\lambda$  y son la base de la cual se construyen los procesos puntuales espaciales. b) Procesos Puntuales de Cox los cuales son procesos de Poisson y que además están condicionados la principal utilidad es para modelar distribuciones espaciales de puntos regulares. c) Procesos puntuales de Gibbs son procesos que se definen con respecto a alguna especificación condicional, estos son útiles para modelar patrones espaciales más regulares que los procesos de Poisson donde cada punto conserva un espacio libre que los rodea.

Para el estudio de los diferentes tipos de modelos que existen dentro de los procesos puntuales comenzaremos planteando el más sencillo de ellos y del que parte el inicio natural el proceso de Poisson homogéneo, el supuesto de homogeneidad, el cual nos hace pensar que la distribución de puntos sobre la región se distribuye de forma aleatoria y más aun de forma independiente, con lo cual su mecanismo generador será un simple Modelo de Poisson.

### 3.9.1. Proceso puntual de Poisson

El modelo más simple de procesos puntuales es el Proceso de Poisson el cual tiene las siguientes características: Un proceso puntual  $X$  definido sobre  $S$  es un proceso puntual de Poisson con media  $\lambda$ , si cumple con las siguientes características:

- Para cualquier conjuntos de Borel  $B \subset S$  con media finita  $\mu(B) < \infty$ , y  $N(B) \sim Poisson(\mu(B))$
- Para cualquier  $n \subset N$ ,  $B \subset S$  con media finita y condicionada a  $N(B) = n$ ,  $X_B \sim Binom(B, n, f)$  donde  $f(\cdot) = \lambda(\cdot)/\mu(B)$ , dicho de otra forma que los  $n$  eventos que forman una muestra aleatoria de la distribución uniforme.

Otras características asociadas al proceso puntual de Poisson son:

$$\lambda_2 = \lambda^2$$

$$K(h) = \pi h^2$$

$$F(h) = G(h) = \mathbb{P}[N(\pi h^2) > 0] = 1 - e^{-\pi \lambda h^2}, h > 0.$$

Es decir, la función de intensidad de segundo orden es igual al cuadrado de la función de intensidad, la función de segundo orden reducida o la  $K$  de Ripley para un proceso de Poisson homogéneo, dependen del radio  $h$  que se esté tomando, las funciones del vecino más cercano y el espacio vacío son las mismas. Es por ello que este modelo

representa el mecanismo estocástico más sencillo de estudiar y de generar, es el modelo ideal a partir del cual se prueba la hipótesis de CSR, y del cual se puede partir un estudio.

Los procesos de Poisson son modelos muy sencillos, que rara vez aplican a un cierto conjunto de datos, lo cierto que en la vida real es muy común ver patrones donde el supuesto de estacionariedad no se cumple. Es decir, se sustituye la función de intensidad constante por una función de intensidad variable, gracias a estos modelos es posible introducir covariables que apoyen un diseño espacial a través de la función de intensidad variable entre sus diferentes ubicaciones. Es por ello que se definen los modelos de Poisson no homogéneos.

### 3.9.2. Proceso de Poisson no homogéneo

Los modelos de Poisson como se había planteado anteriormente se utilizan cuando el supuesto de estacionariedad no se cumple, es decir se sustituye el valor de la intensidad constante  $\lambda = \lambda(x)$  por una función de intensidad variable, este supuesto da origen a un proceso de Poisson no homogéneo el cual cumple con: Un proceso puntual  $X$  definido sobre  $S$  es un proceso puntual no homogéneo de Poisson con media  $\int_{\lambda}(x) dx$ , para alguna función  $\lambda(x)$  no negativa, si cumple con las siguientes características:

- Para cualquier conjuntos de Borel  $B \subset S$  con media finita  $\int_{\lambda}(x) dx < \infty$ , y  $\lambda > 0$ , y  $N(B) \sim Poisson\left(\int_{\lambda}(x) dx\right)$ ,
- Para cualquier  $n \in N$ ,  $B \subset S$  con media finita y condicionada a  $N(B) = n$   $X_B \sim Binom(B, n, f)$  donde  $f(.) = \lambda(.)/\mu(B)$ , dicho de otra forma que los  $n$  eventos que forman una muestra aleatoria de mi distribución uniforme.

Este tipo de modelos es de gran ayuda ya que permite incluir en el modelo información adicional por medio de covariables a través de la función de intensidad  $\lambda(x) = \lambda(z_1(x), z_2(x), \dots, z_n(x))$  con lo cual un posible ajuste para este tipo de modelos será  $\lambda(x) = exp\alpha + \beta z(x)$ , donde  $z(x)$  representa la covariable ajustada al modelo. Estos modelos serán llamados procesos de Poisson modulados (Cox, 1972), dentro del análisis que realizamos utilizamos covariables las cuales son descritas en el capítulo de análisis descriptivo.

### 3.9.3. Proceso de Cox

La utilidad de los procesos de Cox, surge de la necesidad de modelar una distribución de los eventos que se originan a partir de un proceso estocástico. Se define a

$\Lambda(x); x \in \mathbb{R}^d$  como un proceso estocástico localmente integrable con valores no negativos y que para cualquier conjunto de Borel  $B$ ,  $\int \Lambda(x) dx < \infty$ . Por lo general  $\Lambda(x)$  tendrá una distribución Gamma.

Este tipo de modelos los observaremos de forma natural en estimaciones bayesianas en donde nuestra intensidad depende de algún parámetro previo. El proceso de Cox hereda las propiedades de  $\Lambda(x)$  con lo cual suponemos aleatoriedad e isotropía, al igual que al proceso de Poisson homogéneo definiremos las propiedades de primer y segundo orden:

$$\begin{aligned}\lambda &= \mathbb{E}[\Lambda(x)] \\ \lambda_2 &= \mathbb{E}[\Lambda(x)\Lambda(y)] \\ K(h) &= \pi h^2 + 2\pi\lambda^{-1} \int_0^1 \gamma(r) dr,\end{aligned}$$

Donde la  $\gamma(h)$  representa la función de covarianza entre:

$$\gamma(h) = cov(\Lambda(x)\Lambda(y)), h = \|(x - y)\|$$

### 3.9.4. Modelo de Cluster

Este tipo de procesos son una familia muy importante dentro de la teoría de procesos puntuales, definidos como procesos que forman conglomerados o clúster los cuales son grupos de dos o más individuos que se localizan un de otro, más cerca de lo que se ubicarían con un proceso de Poisson homogéneo.

Dentro de estos modelos se encuentran el modelo de Hard Core y el de Neymann-Scott, este tipo de modelos son construidos suponiendo que se construyen a partir de un proceso puntual, el cual produce el centro de los puntos del clúster, donde el número y las localizaciones son distribuidas acorde a procesos reales. Cada uno de estos puntos es denominado padre, de ahí cada uno de estos puntos produce nuevos puntos descendientes cuya posición es acorde con un proceso real, un caso particular recae cuando los padres provienen de un proceso puntual de Poisson, si esto sucede el proceso es llamado proceso clúster Poisson, ejemplo de este tipo de procesos son los de Thomas y los procesos de Matern.

### 3.9.5. Modelo Hard Core

Asumimos que la configuración puntual en  $S$  que no permite vecinos cercanos. Los procesos puntuales donde permiten a las distancias  $d < r$  para  $r > 0$  son llamados procesos puntuales de Hard Core, este tipo de procesos se utilizan en aplicaciones de ecología forestal, ecología animal y econometría espacial entre otros, será uno de los modelos que se prueben.

### 3.9.6. Modelo de Neymann-Scott

El modelo de Neymann-Scott es un modelo doble estocástico con el objetivo de reproducir el patrón de cluster. Es la superposición de un proceso de Poisson Homogeneo  $U$  para puntos padres y condicionando a las realizaciones de  $U$  los puntos hijos son dibujados a partir de una distribución simétrica sobre  $\mathbb{R}^d$ , cuyos centros son los puntos padres.

### 3.9.7. Proceso de Gibbs

Los procesos de Gibbs se caracterizan por que el surgimiento del patrón es una consecuencia de las interacciones entre sus sucesos (Mateu & Montes 1998).

Los procesos de Gibbs se definen por intensidades condicionales: Para cada localización  $u \in \mathbb{R}^2$  y para cada realización  $x = x_1, x_2, \dots$  del proceso puntual  $X = X_1, X_2, \dots$ , se considera  $\lambda(u, x) > 0$ , donde  $\lambda(u, x)$  es la probabilidad condicional de que el proceso  $X$  tiene puntos en  $du$  dado que todos los puntos  $xn\{u\}$  del proceso  $X$  fuera de la región infinitesimal  $du$ .

# Capítulo 4

## Aplicación de la teoría de procesos puntuales al análisis de incidencia de incendios forestales

En este capítulo se aplica la teoría de procesos puntuales al análisis de incidencia de incendios forestales, ya que hoy en día representa una de las amenazas medioambientales más importantes a las que se enfrenta la sociedad, debido al daño ecológico y económico que generan (Asiain y Segovia, 2009).

La primera parte de este capítulo consiste en un análisis descriptivo de los incendios forestales con la finalidad de poder ajustar su comportamiento, mediante un modelo que describe su ubicación y que permite determinar la existencia de algún tipo de comportamiento, ya sea repulsión o atracción entre los eventos, o más aún si existe alguna tendencia o interacción con respecto de su ubicación geográfica. Con este análisis se puede obtener un mapa de riesgos para determinar las zonas más propensas y con mayor probabilidad donde se pueda originar un incendio forestal, ya sea de forma natural o por causas externas asociadas principalmente con los humanos.

### 4.1. Descripción de la base de datos

La base de datos utilizada para el desarrollo de este trabajo es la recopilación de los incendios forestales registrados en la provincia de Castellón España (ver Figuras 4.1 y 4.2), durante el periodo del año 2001-2006, adicionalmente se usan las siguientes variables:

- Año: Tiempo durante el cual se desarrolló el incendio.
- Tipo de causa: es una variable categórica la cual describe la causa que originó el incendio forestal entre las que tenemos: Intencionados, naturales, negligencia

y desconocidos.

- Coordenadas  $(X,Y)$ : Localización geográfica de cada incendio dentro de la región geográfica de estudio.
- Covariables espaciales: Dentro del análisis se incluyen 5 covariables cuatro de ellas continuas y una categórica.
  - Altura: Altitud con respecto al nivel del mar, a la cual se registra el incendio forestal dentro del área geográfica de estudio.
  - Pendiente: la pendiente se ha expresado en grados de inclinación con respecto de la horizontal.
  - Orientación: La orientación se ha medido en grados en sentido de las agujas del reloj, para un mejor categorización de esta variable se han clasificado en ocho diferentes puntos cardinales los cuales corresponden a Norte, Sur, Este Oeste, Noreste, Noroeste, Sureste y Suroeste.
  - Área: Tamaño del área quemada por el incendio registrado a partir de su punto de ignición en  $m^2$ .
  - Uso de suelo: Es nuestra variable categórica fue dividida en las siguientes categorías de acuerdo al uso de suelo: desconocido, actividades agrícolas, pradera, bosques y arbustos.



Figura 4.1: Localización de la provincia de Castellón en España.

## 4.2. Descripción general del análisis

Con la base de datos descrita anteriormente, el análisis se divide de la siguiente forma: se proponen modelos para cada uno de los patrones puntuales de estudio, los cuales son: el patrón general de incendios del 2001 a 2006, los patrones originados por los diferentes tipos de causas (naturales, negligentes o accidentales y los intencionados), así como por tamaño de incendio de acuerdo a la clasificación de la normativa europea (PLADIGA, 2008) en conatos e incendios(Figura 4.2).

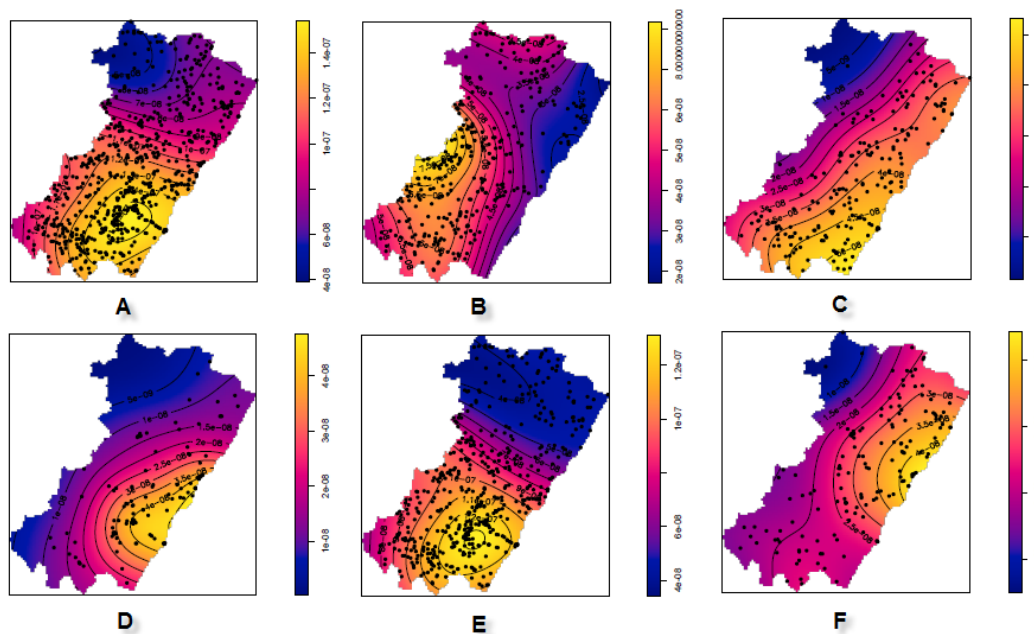


Figura 4.2: Patrones de interés para el desarrollo del estudio originados por diversas causas durante el periodo 2001-2006: patrón general de los incendios ocurridos en la provincia de Castellón panel (A); incendios originados por causas naturales panel (B); incendios originados por negligencia o accidente panel (C); incendios causados de forma intencional panel (D) incendios menores a una hectárea panel (E); incendios mayores a una hectárea panel (F).

También se estudian los patrones multitemporal por año, causa de incendios, tamaño y la interacción o relación que existe entre sus patrones puntuales, además adicionalmente se hace un análisis de la dependencia de los incendios de acuerdo al tamaño de área quemada. Al final se propone un mapa de riesgo adecuado para cada tipo de causa.

El análisis comienza con un análisis exploratorio de los datos con la finalidad de obtener una mejor descripción y un panorama más amplio de los mismos, así como observar las relaciones que existen entre las variables involucradas.

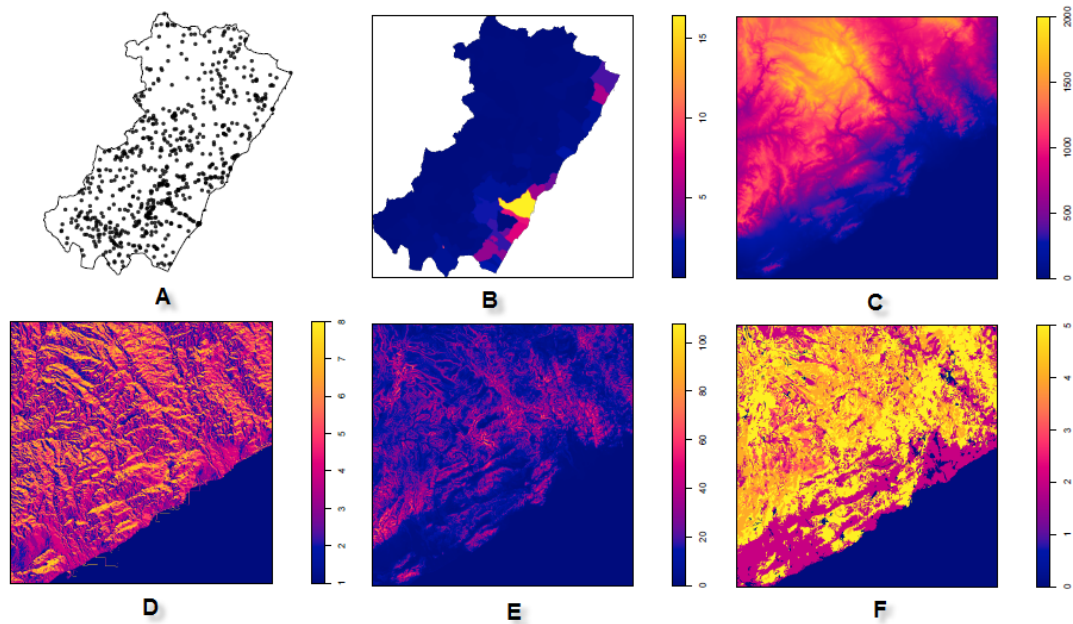


Figura 4.3: Covariables utilizadas como información adicional para el ajuste de los modelos: Ubicación de los incendios forestales en el periodo 2001-2006 panel (A); Densidad de población panel (B); Elevación panel (C); Orientación panel (D); Pendiente panel (E); Uso de suelo panel (F).

El análisis continúa con la prueba de CSR (aleatoriedad espacial completa), por medio de las funciones de distancias (función del evento más cercano, función de distribución del espacio vacío, la K de Ripley), así como por la estimación por kernel, de así se prosigue al ajuste de diferentes modelos: Modelos de Poisson homogéneos y no homogéneos, Modelos de Cluster y modelos de Gibbs apoyados de información adicional en las covariables (Figura 4.3).

Se escogen los tres modelos mejor ajustados para cada uno de los patrones y son validados por medio de los residuales y gráficos de diagnóstico así como por el criterio del AIC, a partir de los modelos escogidos para cada causa se generan los mapas de riesgos, en otra sección posterior se analizan las interacciones entre los diferentes tipos de patrones multitemporales por medio de la función K de Ripley para su caso bivalente, apoyado también de la función de correlación de marcas que es útil para observar la posible dependencia entre los tamaños de los incendios así como de su localización.

Con los resultados obtenidos en este capítulo se contrastan las hipótesis y los objetivos planteados al inicio del estudio.



## 4.3. Análisis exploratorio de datos

### 4.3.1. Análisis de la relación entre las variables

La figura 4.4 muestra la correlación que existe entre las variables, además de la distribución de las mismas y la relación que existe entre ellas. Se puede identificar que las variables pendiente y altura, concentra la correlación más alta de 0.5, Se observa también una baja correlación entre la orientación con respecto de la altura y la pendiente de 0.11 y 0.10 respectivamente. La variable con menos relación con las demás es el área quemada, en donde la correlación es casi nula con respecto de las otras variables. Por esta razón se utiliza para definir y trabajar con ella un patrón puntual marcado.

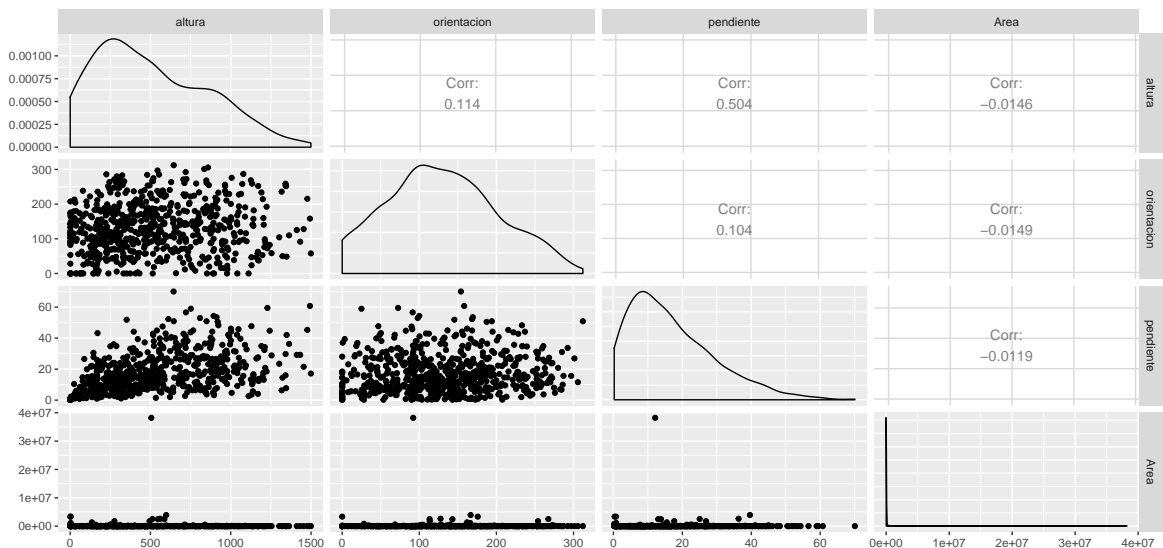


Figura 4.4: Grafica de correlación entre las variables asociadas a los datos de los incendios forestales.

En los gráficos de puntos se observa la relación positiva que existe entre altura y pendiente, manteniéndose casi constante la relación entre la orientación y la altura así como entre la orientación y la pendiente, el área quemada muestra un dato atípico el cual puede sesgar esta información motivo por el cual se ven relaciones muy pequeñas y un histograma totalmente sesgado a la derecha. El comportamiento de las otras variables se concentra de igual manera una acumulación importante de valores concentradas hacia la izquierda de la distribución donde se localizan los valores más pequeños.

### 4.3.2. Análisis del número de incendios forestales por causa y área quemada

Enseguida se analiza el número de incendios forestales por causa, área quemada y uso de suelo. La tabla 4.1 muestra el número total de los incendios por cada año de estudio, así como por tamaño del área quemada. Se observa que en el año 2005 se reporta el mayor número de incendios con un total de 160, seguidos de los años 2001, 2004 y 2006. El año donde se reportan menos incendios es 2002 con un total de 65.

De los 663 incendios registrados en los 6 años 508 son menores de 1 hectárea, y 155 son mayores de 1 hectárea, con respecto del área quemada se observa que el año 2001 reporta 4493.71  $m^2$ , seguido del año 2004 con 2338.50  $m^2$ .

Tabla 4.1: Número de incendios por el tamaño del mismo y área quemada por cada año.

Año	Conato	Incendio	Total	Área Quemada ( $m^2$ )
2001	89	31	120	4493.71
2002	52	13	65	278.41
2003	63	25	88	503.42
2004	93	27	120	2338.50
2005	122	38	160	1522.29
2006	89	21	110	232.83

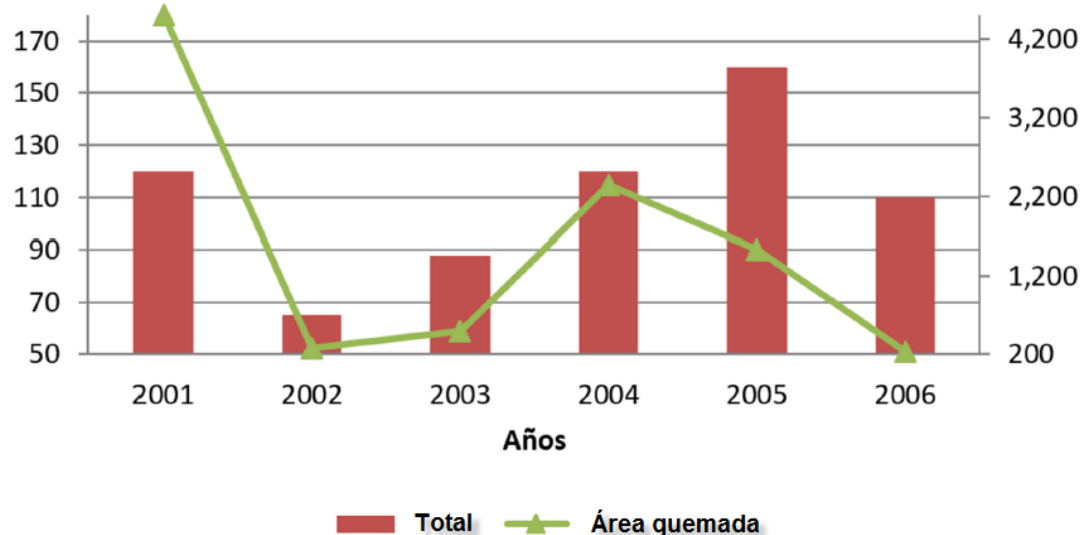


Figura 4.5: Número de incendios con respecto al área quemada para los años 2001 a 2006.

La información aportada por la tabla 4.1 así como la figura 4.2 muestra el comportamiento proporcional del área quemada con respecto al número de incendios para los años 2002, 2003 y 2004. Se observa un comportamiento inverso para los años 2001, 2005 y 2006 donde la proporción del área quemada no tiene relación directa con el número de incendios, tal es el caso del año 2005 cuando se reportó un mayor número de incendios. Sin embargo el total de área quemada es bajo, con respecto de otros años, como puede ser el 2001 o 2004. Lo mismo sucede para el año 2006 donde el área quemada es muy baja y es similar al del año 2002.

El caso de mayor relevancia ocurre en el año 2001, cuando el área quemada reportada es la mayor de todo el periodo de tiempo, en función del número de incendios ocasionados en ese año, esto se debe a que en 2001 se reportaron 31 incendios mayores a 1 hectárea lo cual supone que algún incendio de gran magnitud puede ser el causante de que el reporte de área quemada sea muy elevado.

Tabla 4.2: Número de incendios por causa de origen y el área quemada por cada año de estudio.

Año	Desconocido	Intencionado	Naturales	Negligencia	Total general
2001	1	16	56	47	120
2002	3	11	30	21	65
2003	8	15	31	34	88
2004	19	13	66	22	120
2005	18	41	68	33	160
2006	9	19	48	34	110
Total	58	115	299	191	663

La tabla 4.2 muestra el análisis del número de incendios por causa de origen, donde se observa que la mayoría de los incendios reportados son por causas naturales con un total de 299 incendios, seguido de los causados por negligencia los cuales suman 191 y los intencionados con un total de 115 para el periodo de 2001 a 2006.

Sin embargo hay que recordar que fuera de las causas naturales, las otras 3 clasificaciones son asociadas a causas humanas dando así un total de 364, lo que representa un 55 % del total de incendios, es decir más de la mitad de ellos se pudieron haber evitado o prevenido.

La figura 4.3 representa la relación entre área quemada y número total de incendios, en ella se observa una relación proporcional entre ambos, con lo cual se puede asegurar que el área quemada es acorde al número de incendios originado por cada una de las causas.

Con el análisis anterior no se pueden llegar a conclusiones objetivas acerca del comportamiento de los incendios el área quemada y otros aspectos de interés que se pudieran obtener al tratar estos datos.

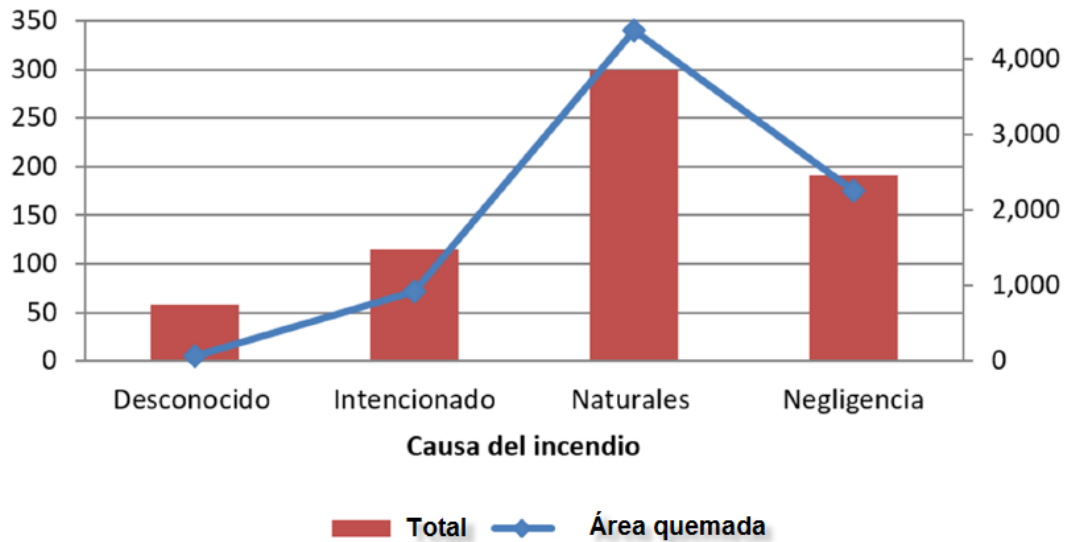


Figura 4.6: Número de incendios por causa de origen en relación con el área quemada.

Tabla 4.3: Área quemada por año y causa de incendio ( $m^2$ ).

Año	Desconocido	Intencionado	Naturales	Negligencia	Total general
2001	9.85	10.06	3944.79	529.01	4493.71
2002	5.24	19.32	18.58	235.27	278.40
2003	6.54	26.61	4.08	466.19	503.41
2004	10.71	42.37	67.33	78.13	583.53
2005	23.79	420.79	339.99	737.71	1522.29
2006	5.51	22.66	6.92	197.73	232.82
Total	61.64	926.82	4381.68	2244.05	7614.19

Es por ello que en la tabla 4.3 se separa el área quemada por cada causa que originó el incendio, para cada uno de los 6 años y se observa que la mayor área quemada la reportan los incendios de causas naturales y en especial el año 2001 con  $3944 m^2$  lo que apoya que el año 2001 se reportara como el año con mayor área quemada y el caso atípico en los gráficos pasados.

Seguido de las causas naturales, están los incendios causados por negligencia donde la mayor área quemada se reporta en el año 2005 con 737.71 metros cuadrados, con lo que se puede afirmar que la causa más devastadora son los incendios naturales.

Sin embargo, la tabla 4.4 muestra una razón del área quemada entre el número total de incendios por cada causa y los resultados son alarmantes y contrastan los supuestos anteriores.

Sin tomar en cuenta el año 2001, se observa en promedio que la causa más devastadora son los incendios por negligencia seguido de los incendios intencionados y en última instancia los naturales, el comportamiento longitudinal se observa en la figura 4.7 donde se ve como las líneas que representan a los incendios intencionados y por negligencia son en promedio más devastadores, tomando en cuenta el promedio de área quemada por cada incendio originado por esa causa.

La única excepción es la ocurrencia del incendio de mayor magnitud en el año 2001 que origina que en promedio el área quemada por incendio causado de forma natural sea de  $70.44m^2$ .

Tabla 4.4: Razón de área quemada entre número de incendios según tipo de causas para el año 2001 a 2006.

Año	Desconocido	Intencionado	Naturales	Negligencia	Total general
2001	9.85	0.63	70.44	11.26	37.45
2002	1.75	1.76	0.62	11.20	4.28
2003	0.82	1.77	0.13	13.70	5.72
2004	0.56	32.87	1.02	3.55	4.86
2005	1.32	10.26	5.0	22.35	9.51
2006	0.61	1.19	0.14	5.82	2.12
Total	1.06	8.06	14.65	11.75	11.48

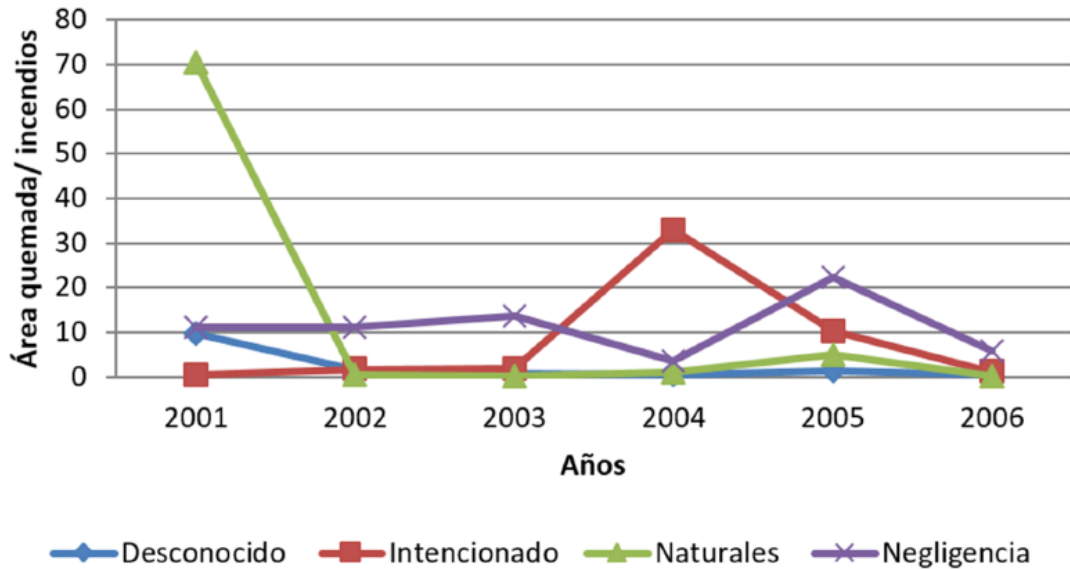


Figura 4.7: Trayectorias longitudinales del promedio de área quemada por número de incendios según tipo de causas para el año 2001 a 2006.

### 4.3.3. Dispersión y distribución de las principales variables de interés

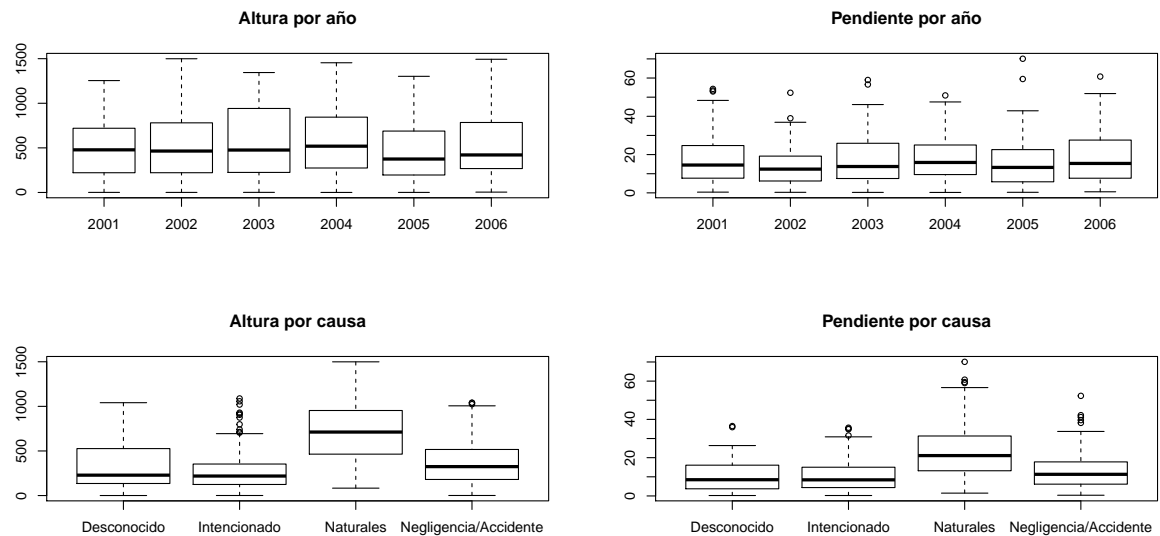


Figura 4.8: Graficas de caja para la variable pendiente y altura de cada registro de incendio dividido por año y tipo de causa.

En las gráficas de caja (Figura 4.8), se observa la distribución de los datos de incendios con respecto a la altura y pendiente donde fueron registrados, para el análisis por año se observa que el 50 % de los incendios se originaron cerca de los 500 metros de altura o menos, mientras que en el caso de la pendiente el 50 % se concentran hasta los 15 grados, no se observa variación en ninguno de los años, aunque en la distribución de la pendiente se muestran algunos casos atípicos que superan los 50 grados de inclinación los cuales se pueden asociar a causa naturales, ya que a esta pendiente es difícil que el ser humano tenga acceso fácilmente. En cuanto a las causas se observa una clara diferencia ya que el 50 % de los incendios provocados por causas naturales suelen presentarse a mayor altura así mismo como a mayor grado de inclinación, que los provocados por causas humanas (negligencia, intencionado y desconocido).

## 4.4. Prueba de aleatoriedad espacial completa (CSR)

El análisis inicia con el primer mandamiento de la estadística espacial: Resolver el problema sin utilizar estadística espacial. Se proponen las pruebas de primer y segundo orden para determinar si los patrones se distribuyen de forma aleatoria, es decir se puede contrastar la hipótesis de que la generación de dichos patrones provienen de una

distribución de Poisson con media constante la cual se llama función de intensidad. Cabe señalar que cuando se cumple CSR no existe la interacción entre los eventos.

Las medidas de primer orden son el método de cuadrantes y la estimación de Kernel. Los patrones de interés a analizar son los patrones que involucran todos los incendios con el fin de crear un modelo que ayude a observar las zonas más propensas y de mayor riesgo donde se pueda generar un incendio forestal. Se analiza también el patrón de incendios forestales causados de forma natural principalmente por rayos, con el fin de brindar información adicional que ayude en la en las políticas de prevención y control de incendios forestales, que muchas veces no se pueden evitar.

El otro tipo de patrón son los incendios causados por negligencia o accidente, y es que a veces olvidamos el peligro que generan al entorno factores externos como lo son, por ejemplo tendidos de cables eléctricos, carreteras, actividades agrícolas y la continúa destrucción de las zonas forestales para la construcción de unidades residenciales.

El último patrón, incendios intencionados, sin duda el más preocupante, el fuego es un factor importante dentro de la dinámica forestal ya que ayuda a mantener en equilibrio el medio ambiente. Sin embargo, el ser humano influye mucho en la destrucción de ecosistemas completos, el presente estudio espera generar conciencia para tomar responsabilidades y evitar seguir destruyendo ecosistemas que año con año consumen una gran parte de área quemada que tarda años a veces en volver a generarse.

#### 4.4.1. Método de Kernel

El método kernel es no paramétrico y constituye una forma de estimar la intensidad del patrón puntual. Para el presente análisis se realizaron los mapas de la densidad de kernel por cada uno de los patrones de estudio. Se usa el software R el cual utiliza el Kernel gaussiano. La librería *spatstat* utiliza un núcleo gaussiano definido como:

$$K_s = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\pi\sigma^2}\right)$$

El parámetro  $\sigma$  representa la desviación típica del núcleo, la cual actúa como el parámetro que suaviza la función; se utiliza adicionalmente la función *bwppl* implementada en la librería de *spatstat*, dicha función ayuda a calcular un valor apropiado para el ancho de banda es decir para el parámetro de  $\sigma$  con la cual se estima el núcleo de la intensidad del patrón, dicho proceso se hace mediante la validación cruzada por máxima verosimilitud, obteniéndose el valor para el cual se maximiza la función.

Bajo este criterio se considera el valor de la validación cruzada para el ancho de

banda para cada estimación del núcleo de kernel en cada uno de los patrones de estudio.

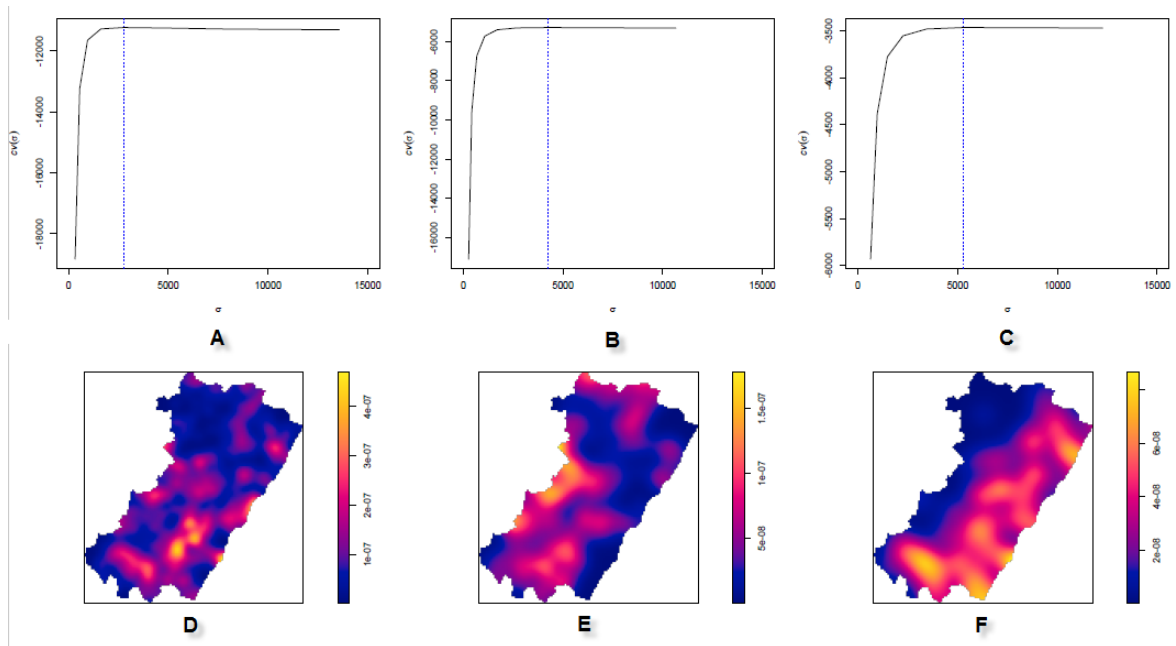


Figura 4.9: Graficas de verosimilitud de validación cruzada para el parámetro  $\sigma$  que maximiza la función de cada patrón de estudio: Total de los incendios de 2001 a 2006 panel (A); Naturales panel (B); Negligencia panel (C). Estimación del Kernel para la intensidad para cada patrón de estudio: Total de los incendios de 2001 a 2006 panel (D); Naturales panel (E); Negligencia panel (F).

En la figura 4.9 y 4.10 se muestra en la parte superior la gráfica de validación cruzada por el método de máxima verosimilitud con la cual se obtiene el parámetro adecuado para maximizar la función y que es utilizado para el cálculo de la estimación por Kernel del núcleo gaussiano.

Para el patrón puntual que modela todos los incendios registrados en la región de Castellón durante el periodo de 2001 a 2006 se observa que en la mayor parte de la región, la estimación es mayor a  $1 \times 10^{-7}$ , eso significa que al menos se han registrado 0.1 incendios por cada  $km^2$ , la región más conflictiva está en la parte sureste que es la región más densamente poblada, en donde se llegan a presentar hasta 0.4 incendios por cada  $km^2$ , casi el cuádruple de la mayor parte de la región.

Para el caso de los incendios naturales y por causas de negligencia se observa una división natural que está asociada a una cuestión de la geografía de la zona. Los incendios naturales se ven mayormente concentrados en la parte noroeste, es decir, la zona alta de la región alcanzando los 0.15 incendios por  $km^2$  mientras que los incendios por negligencia están muy marcados en la parte sureste de la región donde llegan a alcanzar los 0.06 por  $km^2$  principalmente en la región más poblada.



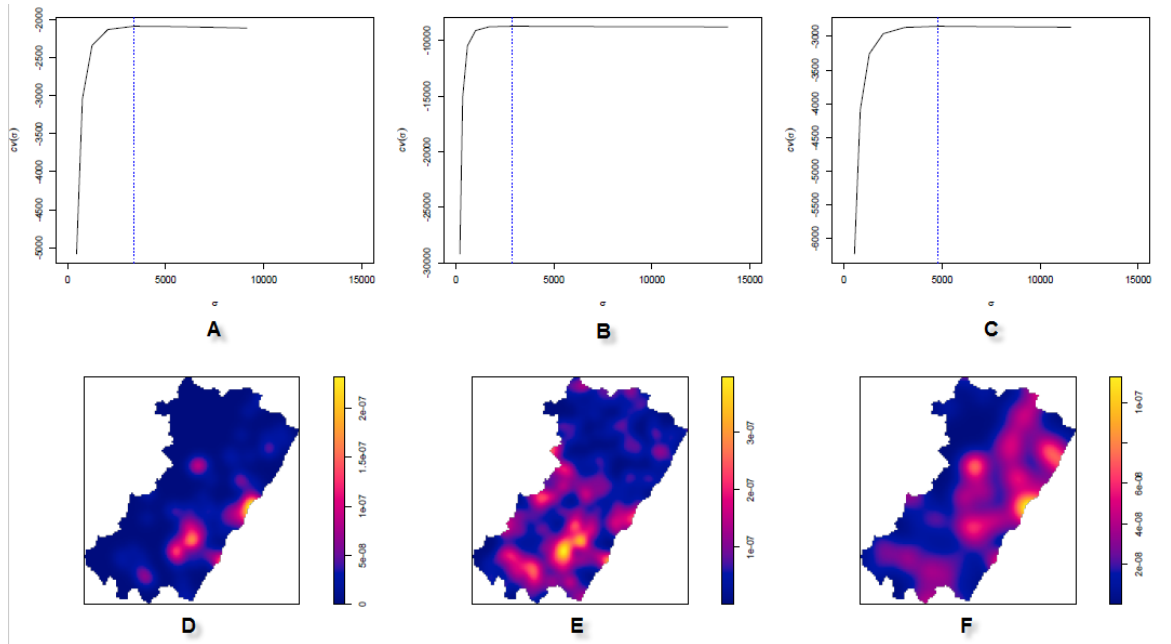


Figura 4.10: Graficas de verosimilitud de validación cruzada para el parámetro  $\sigma$  que maximiza la función de cada patron de estudio: Intencionados panel (A); Conatos (B); Grandes incendios panel (C). Estimación del Kernel para la intensidad para cada patrón de estudio: Intencionados panel (D); Conatos panel (E); Grandes incendios panel (F).

Los incendios intencionados (Figura 4.10) se observa una clara tendencia muy marcados cerca de las áreas de mayor población. En cuanto a los conatos y los incendios existe de igual forma una cierta división los conatos (menores a 1 ha.) se pueden ver más concentrados en dirección del sur y cercanos a las áreas pobladas mientras que los grandes incendios se localizan más al norte de la región.

En la figura 4.11 se realiza la estimación de kernel por año. Para el 2001 y el 2002 y se observa una distribución un poco más homogénea sobre toda la región, para el 2001 se observa una mayor concentración comparada con el año 2002; sin embargo, la estimación es de aproximadamente 0.01 a 0.02 incendios por  $km^2$  en cada año en el año 2005 y se observa una concentración mayor de incendios en la parte más poblada de la región.

La figura 4.12 muestra la estimación de Kernel para causas humanas y causas naturales por tamaño de incendio (conato y gran incendio) se observa que incendios pequeños causados por humanos son cercanos a las poblaciones mientras aquellos mayores a 1 ha. se localiza más en la región del norte, cercanos a la costa. Los incendios naturales tienen un comportamiento un poco más aleatorio. En la figura 4.13 se ejemplifica la división muy marcada por causas humana y causas naturales distribuido del centro hacia el poniente y oriente.

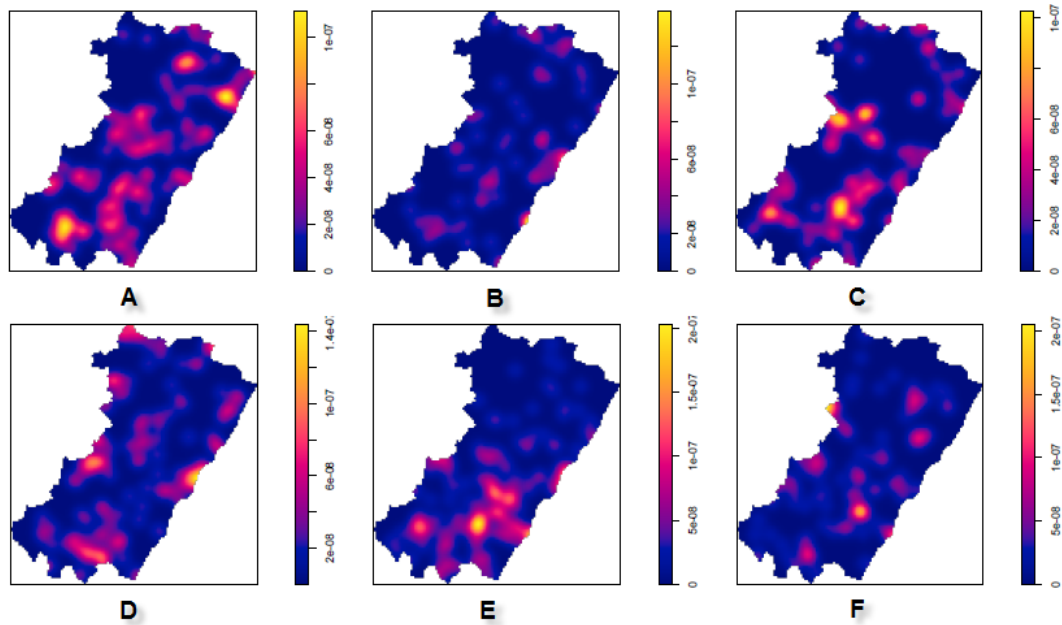


Figura 4.11: Estimación de Kernel para el total de incendios forestales correspondientes a cada uno de los años del estudio: 2001 panel (A); 2002 panel (B); 2003 panel (C); 2004 panel (D); 2005 panel (E); 2006 panel (F).

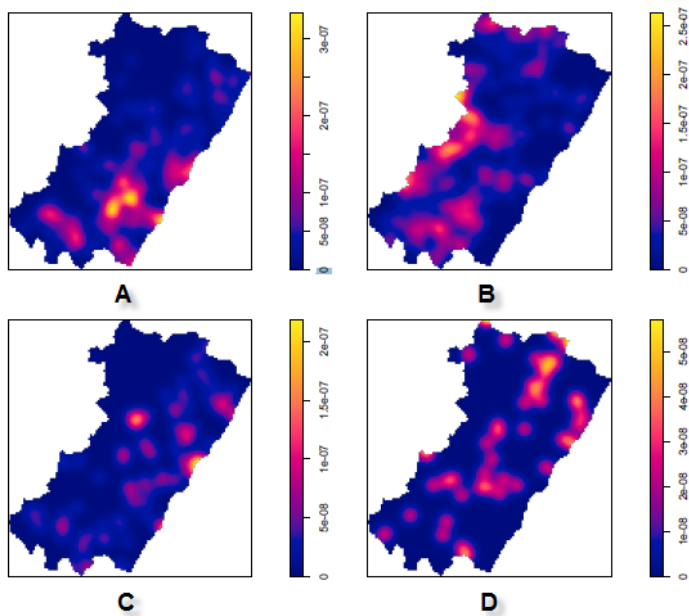


Figura 4.12: Estimación de Kernel por causas y tamaño de incendio: Conatos humanos panel (A); Conatos naturales panel (B); Incendios humanos panel (C); Incendios Naturales panel (D).

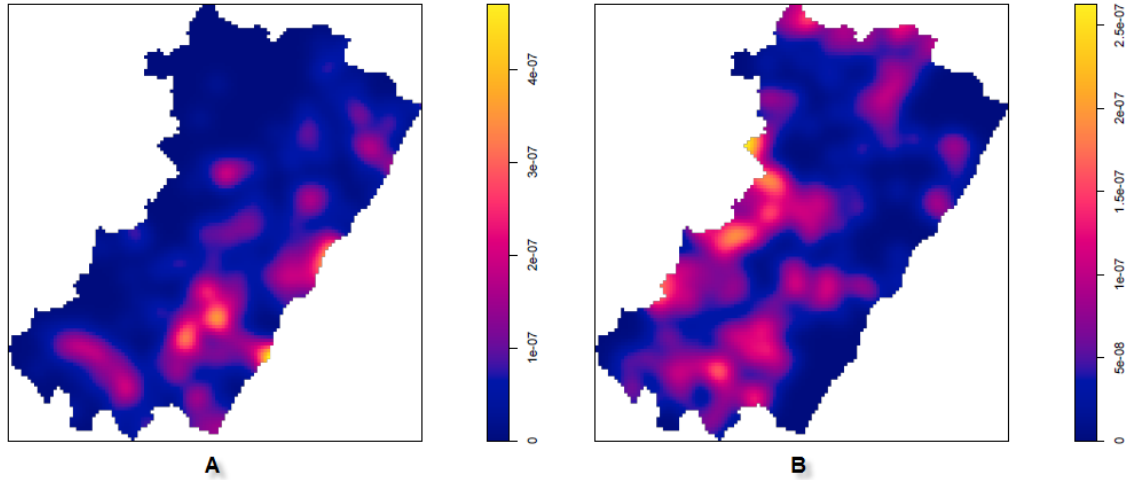


Figura 4.13: Estimación de Kernel por causas: Humanas panel (A); Naturales panel (B).

#### 4.4.2. Método de cuadrantes

Para el método de cuadrantes el área de estudio fue dividida a tres tamaños diferentes: a tamaño de  $km^2$  así como en una cuadrícula de 20 X 26 y de 10 X 13 para observar la sensibilidad y por que no es un buen método de estimación para contrastar la hipótesis de aleatoriedad espacial completa.

Tabla 4.5: Valor de los valores de  $p$  para contrastar la hipótesis de aleatoriedad espacial completa mediante el método de cuadrantes

Medida	Total	Naturales	Intencionado	Negligencia	Conato	Incendio
$Km^2$	$2.20 \times 10^{-16}$	0.4751	$2.20 \times 10^{-16}$	$1.74 \times 10^{-5}$	$2.20 \times 10^{-16}$	$9.05 \times 10^{-14}$
20x26	$2.20 \times 10^{-16}$	$7.67 \times 10^{-10}$	$2.20 \times 10^{-16}$	$6.41 \times 10^{-5}$	$2.20 \times 10^{-16}$	$2.20 \times 10^{-16}$
10x13	$2.20 \times 10^{-16}$	$1.04 \times 10^{-13}$	$2.20 \times 10^{-16}$	$1.11 \times 10^{-7}$	$2.20 \times 10^{-16}$	$4.75 \times 10^{-9}$

De acuerdo con la tabla 4.5 se concluye que para todos los casos la información proporcionada por el valor  $p$  indica que se debe rechazar la hipótesis nula. Es decir, afirmar que los patrones no son completamente aleatorios. La sensibilidad del método de cuadrantes la podemos observar para el caso del patrón de incendios causados de forma natural, donde la división por  $km^2$  nos da un valor  $p$  para afirmar que el patrón es completamente aleatorio. Sin embargo dicha hipótesis es rechazada cuando son observados los valores  $p$  de las otras dos cuadrículas.

En conclusión se puede asegurar que el análisis desarrollado a través de las funciones de primer orden tanto por el método de estimación Kernel, como el método de

cuadrantes en ambos casos no cumplen la hipótesis de aleatoriedad espacial completa. Enseguida se procede a analizar los patrones con las funciones de segundo orden, las cuales se basan en estimaciones no paramétricas de métodos de distancias, además de que tomar decisiones por medio de una gráfica o un estadístico muy sensible no es una buena opción para contrastar alguna hipótesis.

## 4.5. Método de distancias: funciones de segundo orden

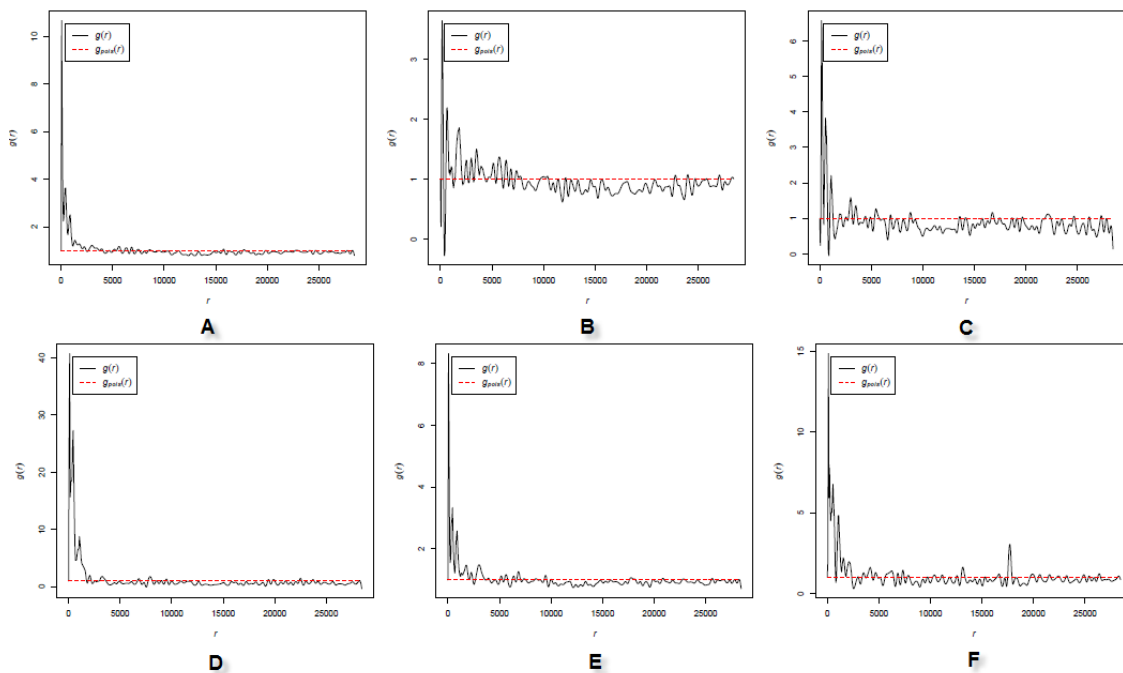


Figura 4.14: Función de correlación para los diferentes tipos de patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); Incendios por negligencia panel (C); Incendios intencionados panel (D); conatos (E); Grandes incendios (F).

En esta sección se utilizan los métodos de distancias para poder contrastar la hipótesis de CSR. Para la función de correlación (Figura 4.14) se observa un comportamiento muy similar entre las diferentes causas así como en los tamaños.

Para valores de la función mayores a la unidad indican que las distancias entre puntos son más frecuentes que bajo un proceso completamente aleatorio, con lo cual indica la existencia de agrupamiento. Para el caso del total de los incendios forestales, los provocados por causas intencionados al igual que los conatos e incendios, se observa una estabilización en la función de correlación a partir de cierta distancia  $r$ .

Para el total de los incendios esta estabilización se observa aproximadamente a los 2.5 km, lo mismo ocurre para los conatos y los grandes incendios. Por su parte los intencionados se estabilizan a los 2 km, por lo cual concluimos que con distancias menores a las mencionadas vemos la existencia de clúster o de agrupamiento para este tipo de incendios.

Estas mismas gráficas presentan un pico, este comportamiento permite estimar el tamaño medio de los agregados y el valor por el cual la función  $g(r)$  se estabiliza en 1. Para el caso de los grandes incendios se puede ver un ligero agrupamiento entre los 15 y 20 km.

Para los incendios provocados por causas naturales y negligencia se observa un comportamiento bastante curioso, una fluctuación sobre los valores de la función  $g(r)$  donde se aprecian cambios bruscos entre valores positivos y negativos de la función a cortas distancias con lo cual se concluye que por lo menos en ambos casos para valores menores a 5 km es difícil poder hablar de un comportamiento en agregados o de un comportamiento regular en estos patrones.

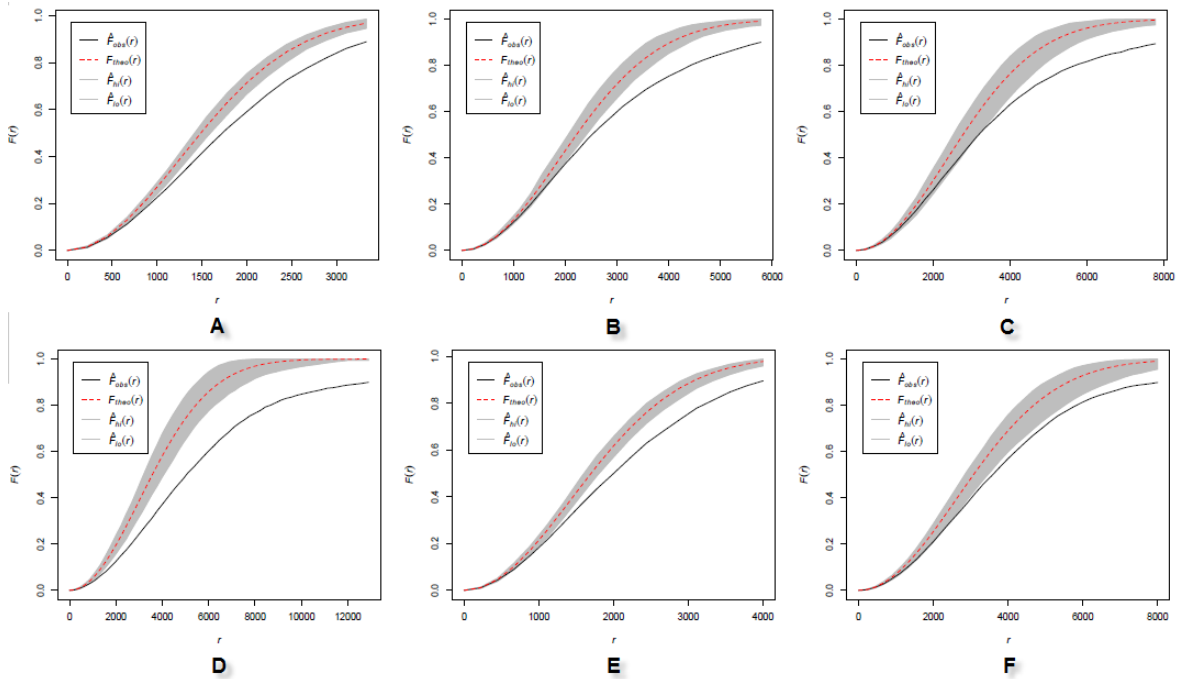


Figura 4.15: Función F o distribución del espacio vacío para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F).

En los métodos de segundo orden (también conocidos como métodos de distancias) se ajustaron las diferentes funciones (Función F, función G, función K de Ripley, y función L).

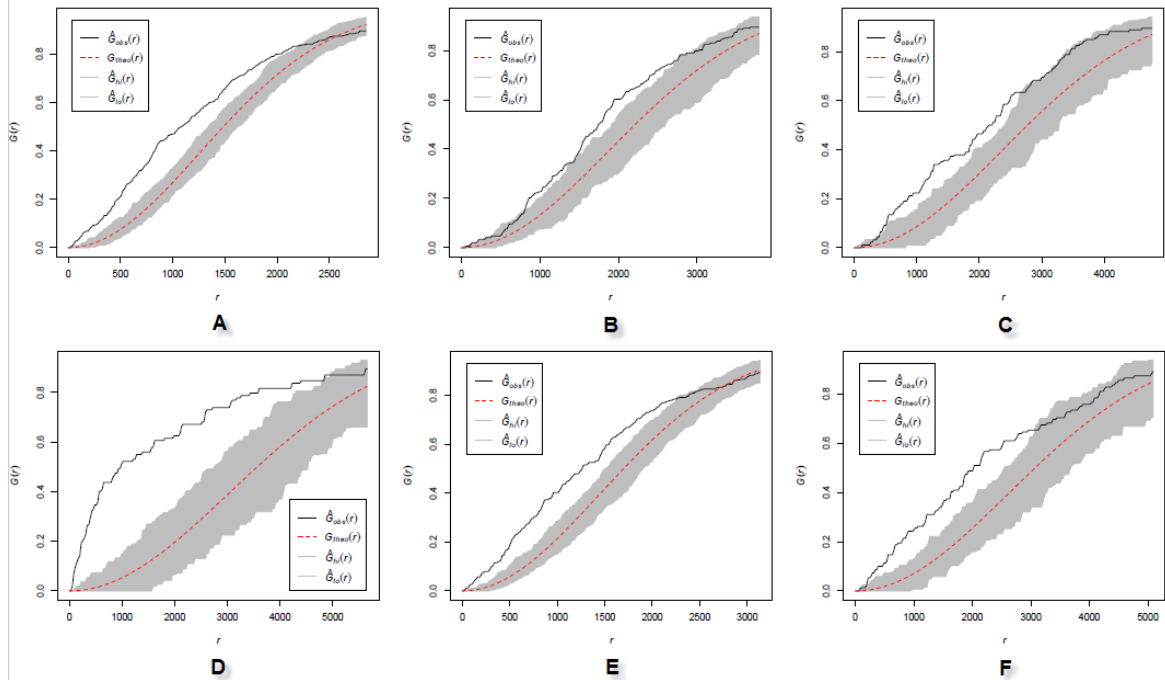


Figura 4.16: Función G o distribución de vecino más cercano para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F).

Para los seis patrones de estudio los gráficos se observa que se rechaza la hipótesis de CSR y que existe la evidencia de que se trata de patrones agregados o de Clúster. En el caso de la función F se concluye que de cada patrón la función observada es menor que las bandas de confianza a distancias que varían entre 500 metros hasta los 2 kilómetros dependiendo del patrón, con lo cual se concluye que es un patrón clúster. Un caso similar es el que presenta la función G donde se observa un rápido crecimiento de la función observada y que es mucho mayor que las bandas de confianza con lo cual se concluye de igual forma que es un patrón de clúster.

La función K no homogénea es un buen estimador de la estructura espacial del patrón ya que se puede observar más claramente el comportamiento de este a diferentes escalas, pareciera a simple vista que con la función K no homogénea los patrones de estudio cumplen la condición de CSR para valores menores a 10 km, y de ahí muestran un ligero comportamiento de repulsión para las distancias en promedio mayores a los 10 km, con lo cual tiende a que las conclusiones sean erróneas.

En la práctica se emplea con mayor frecuencia la función L, esta función es una transformación de la función K de Ripley la cual ofrece ciertas ventajas, la primera es que tiene una varianza constante y permite una interpretación más cómoda.

En la figura 4.18 se observa la función L no homogénea, que repite el mismo com-

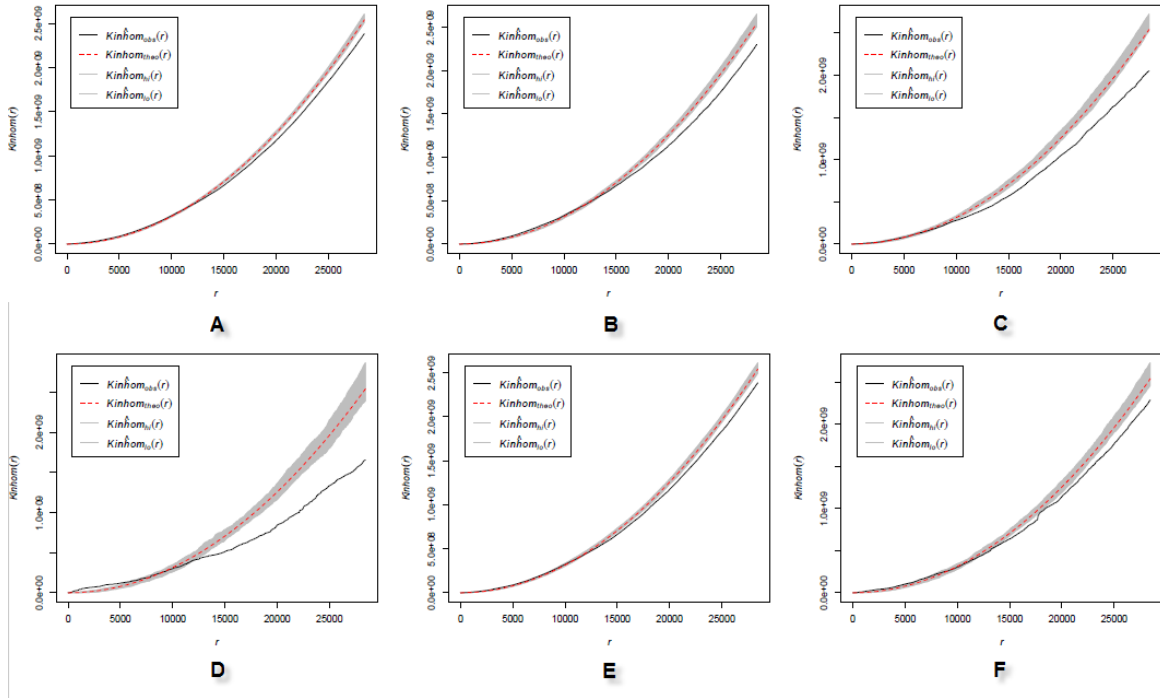


Figura 4.17: Función K de Ripley no homogénea para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F).

portamiento que la K no homogénea, la función tiende a crecer muy rápidamente a distancias pequeñas y esto lo observamos en los seis patrones de estudio, donde se observa un comportamiento de agregación para casi todos los patrones de estudio, a excepción de los naturales y de negligencia los cuales podríamos suponer un comportamiento aleatorio hasta los 15 km y 10 km respectivamente. Esta afirmación confirma lo observado en la función de correlación, donde fluctuaba mucho los valores a distancias muy pequeñas para estos patrones.

En los cuatro restantes se observa que a pequeñas distancias (10 km aprox.) éstos muestran agregación, en las distancias mayores se ve claramente la repulsión o el comportamiento regular en los patrones. Las graficas en general muestran que en todos los patrones de estudio se rechaza la hipótesis de CSR.

## 4.6. Ajuste de modelos para los datos de incendios forestales

Una vez descartada la hipótesis de CSR, y con la información adicional que aportan los métodos de distancias y las covariables se procede a realizar diversos tipos de

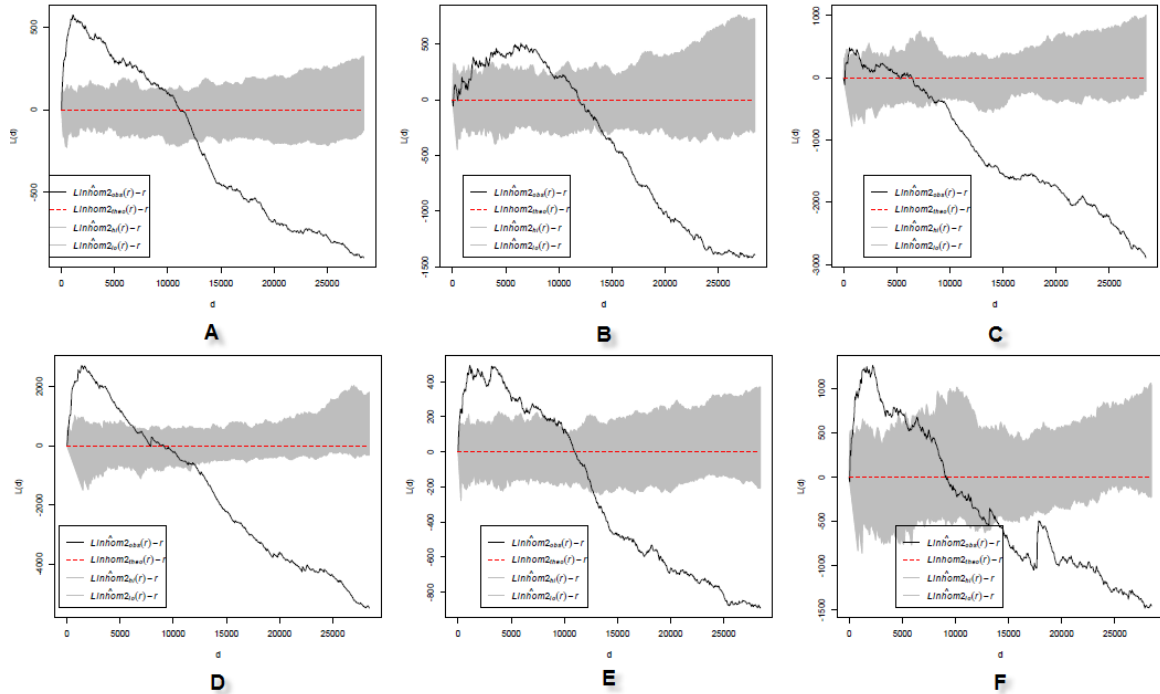


Figura 4.18: Función L no homogénea para los patrones de estudio: Total de incendios panel (A); Incendios naturales panel (B); incendios por negligencia panel (C); incendios intencionados panel (D); conatos (E); Grandes incendios (F).

ajustes para cada uno de los patrones de estudio. El análisis previo muestra indicios de que existen formación de conglomerados en todos los patrones a distancias cortas, con lo cual se propone un ajuste de la siguiente forma: modelos de Poisson no homogéneos, modelos de poisson Clúster, entre los que destacan los de Thomas y de Matern, modelos de Hard Core y de Strauss para los modelos de Gibbs.

Para los modelos de Poisson no homogéneos se utilizan las cinco covariables (densidad de población, altura, pendiente, orientación y uso de suelo). En el apéndice A se reportan todas las funciones L no homogéneas, esta función es el primer criterio para seleccionar el modelo más adecuado, para cada ajuste se generaron intervalos de confianza mediante 99 simulaciones de Montecarlo, aquellas funciones observadas que se encuentran dentro de la confianza son candidatas a ser un buen modelo ajustado.

Se ajustan a los modelos antes mencionados, polinomios de grado dos y cuatro ya que los análisis previos muestran una alta relación entre la intensidad de algunos patrones y su ubicación a lo largo de la provincia. De la misma manera se hace un ajuste con todas las covariables, con covariables seleccionadas por el método de backward, y la combinación de las mismas con las coordenadas geográficas.

Para los modelos de Gibbs el parámetro de interacción se propuso de acuerdo a la distancia promedio mínima entre los eventos de cada patrón. Una vez seleccionado el



modelo mediante el primer filtro, se procede a un segundo el cual es determinado por medio de criterio AIC que es un buen indicador de la bondad de ajuste del modelo y permite además comparabilidad con las demás selecciones.

Este método fue utilizado para 2 patrones los cuales se ajustan a modelos de Gibbs, donde se recurren a los residuales para el análisis de la bondad de ajuste. De igual manera solo un modelo utiliza la información de las covariables para el ajuste.

#### **4.6.1. Modelo ajustado para cada uno de los patrones de estudio**

De acuerdo a las funciones L no homogéneas los ajustes para cada patrón de estudio fueron los siguientes:

- Para todos los incendios del año 2001 – 2006 el mejor ajuste fue un modelo de Cauchy con las coordenadas geográficas.
- Para el patrón de los incendios por causas naturales el mejor ajuste fue un modelo de Thomas con las coordenadas geográficas.
- Para el patrón de los incendios causados por negligencia el ajuste se hizo con un polinomio de grado 2 ajustado a un modelo de Hard Core.
- Para el patrón de los intencionados, fue el único el cual utilizó la información adicional de las covariables (altura, pendiente y uso de suelo) dicha selección se hizo por el criterio de AIC y Backward.
- El patrón de los conatos fue ajustado por un modelo de Cauchy con un polinomio de grado 2.
- Los incendios mayores de 1 ha fueron ajustados por un modelo de Strauss con un polinomio de grado 4.

Las figuras 4.19 y 4.20 son un ejemplo de la función L no homogénea de los modelos mejor ajustados. La primera de ellas (Figura 4.19) corresponde a la función L no homogénea para el total de incendios con un modelo de Cauchy ajustados únicamente con las coordenadas geográficas.

#### **4.6.2. Análisis de Bondad de Ajuste**

Para los modelos de Gibbs se utiliza el análisis de bondad de ajuste para validar los modelos: el de Hard Core ajustado a los incendios causados por negligencia al cual se le ajustó un polinomio de grado 2, así como el modelo de Strauss ajustado patrón de puntos cuyo incendios reportan más de 1 ha quemada.

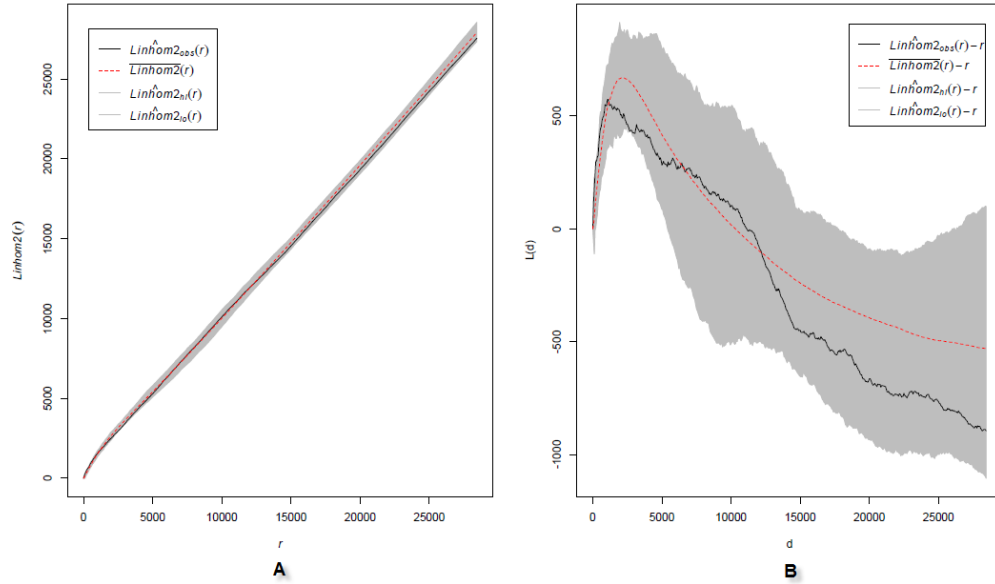


Figura 4.19: Función L no homogénea para el total de incendios con un modelo de Cauchy con las coordenadas geográficas: Funcion L no homogénea con bandas de confianza panel (A); Transformación de la función L no homogénea con bandas de confianza panel (B).

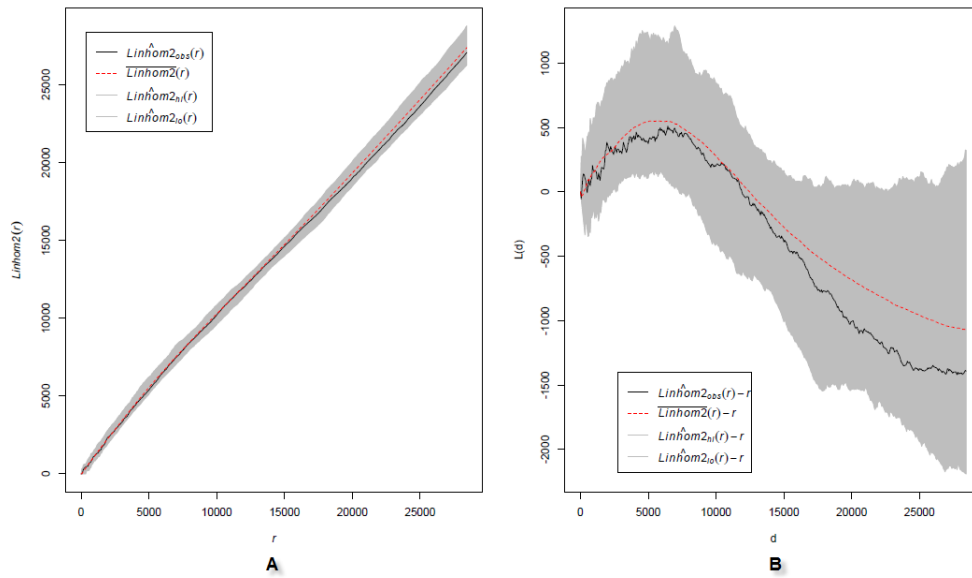


Figura 4.20: Función L no homogénea para los incendios naturales con un modelo de Thomas con las coordenadas geográficas: Funcion L no homogénea con bandas de confianza panel (A); Transformación de la función L no homogénea con bandas de confianza panel (B).

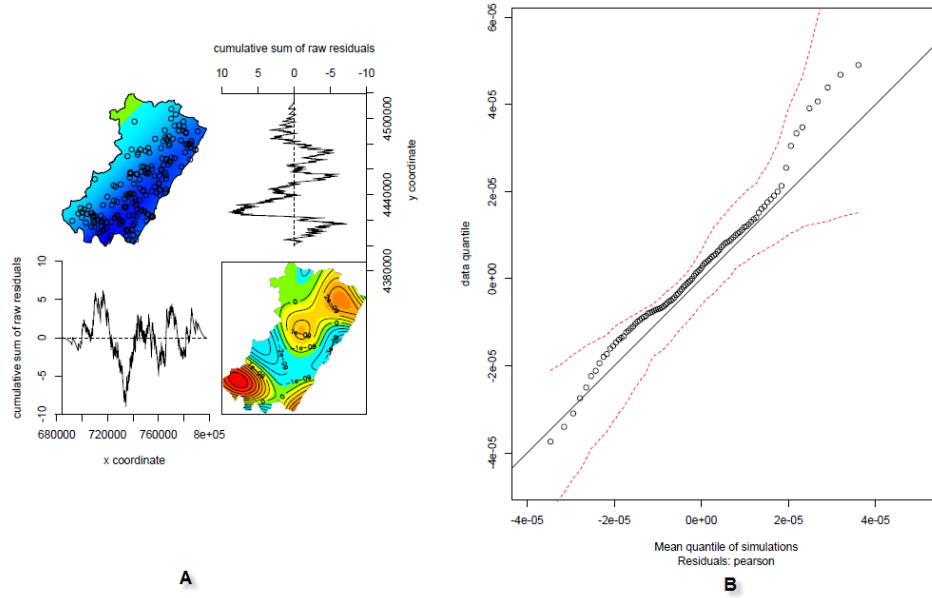


Figura 4.21: Gráficos de diagnóstico para medir la bondad de ajuste en el modelo de Hard Core con un polinomio de grado 2: Grafico de marcas y grafico de lurking panel (A); grafica de cuantil-cuantil panel(B).

En la figura 4.21 se observa el grafico de diagnóstico (lado izquierdo) y la gráfica cuantil-cuantil con bandas de confianza (lado derecho). Estos gráficos dan información acerca de los residuales del modelo, para el primer ajuste, la gráfica de marcas asocia cada evento el valor del residual que les corresponde, por lo que se ven residuales muy pequeños en casi todos los puntos.

Con respecto a las gráficas de lurking para cada coordenada, se muestra que la suma acumulada de los residuales se encuentra acotada entre valores de -10 y 10, lo cual indica un buen ajuste del modelo sobre toda la región de estudio y no se observa alguna tendencia hacia alguna coordenada geográfica.

La grafica cuantil-cuantil refleja un buen ajuste debido a la concentración de datos cerca de la identidad, sin embargo el único problema es que refleja colas pesadas en los extremos.

La figura 4.22 corresponde es el ajuste del modelo de Strauss, donde se observa un comportamiento es similar al descrito anteriormente, el ajuste de la gráfica de marcas muestra residuales muy pequeños, así como el grafico de lurking muestra un acumulado entre valores de -10 y 10 sin observar tendencia hacia alguna coordenada de la región.

De la misma forma la función cuantil-cuantil ajusta entre las bandas de confianza con el mismo problema que el anterior la sensibilidad y el registro de colas pesadas en valores extremos.

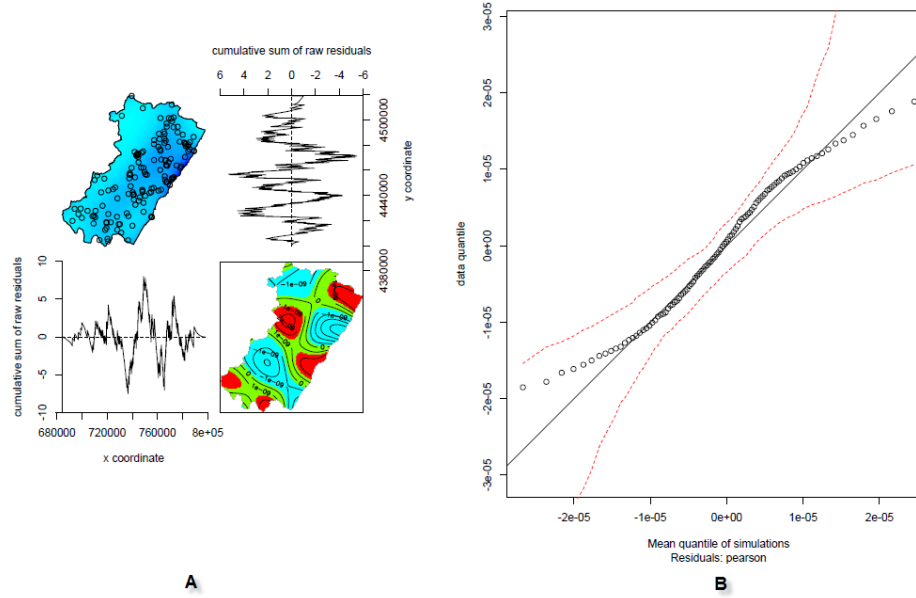


Figura 4.22: Gráficos de diagnóstico para medir la bondad de ajuste en el modelo de Strauss con un polinomio de grado 4: Gráfica de marcas y gráfica de lurking panel (A); Gráfica de cuantil-cuantil panel (B).

### 4.6.3. Mapas de predicción para los diferentes patrones ajustados

En la siguiente sección se presentan los mapas de predicciones con los modelos ajustados para los patrones de estudio: para el total de los incendios y los incendios por causas naturales, el resto de los mapas de predicción se localizan en la sección del apéndice A.

En cada gráfica se presentan de lado izquierdo el mapa de predicciones y del lado derecho el mapa de predicciones asociado con los puntos donde se registró un evento para el año 2001-2006.

## 4.7. Análisis de interacción mediante la k-cruzada

Para analizar la interacción entre los patrones de estudio se utilizó la función K cruzada, el análisis de dicha función muestra que los incendios ocurridos en el área de estudio para los años 2001- 2006 la mayor parte de las gráficas se observa que no existen interacción alguna por lo que se hace el supuesto de independencia entre cada año. Eso se ve reflejado por qué la función k cruzada no sobre pasa el intervalo de confianza generado por simulaciones de Montecarlo bajo un proceso de Poisson independiente.

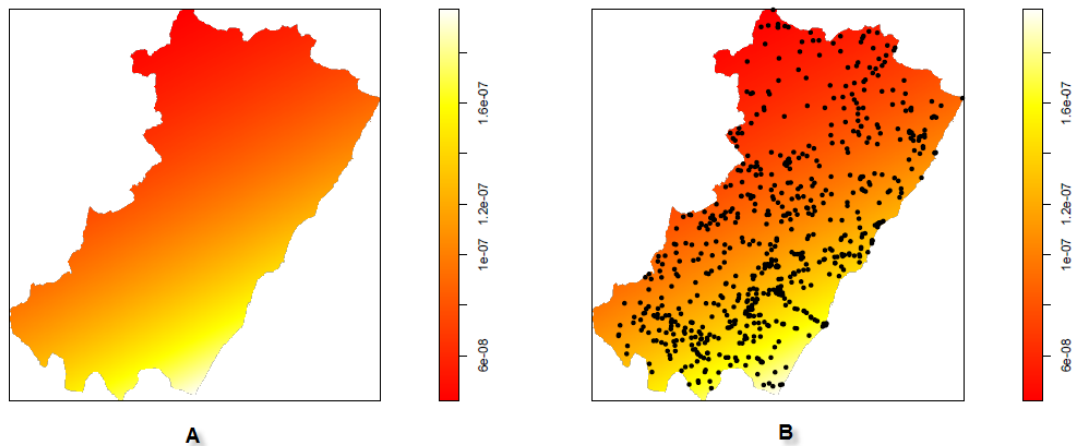


Figura 4.23: Mapa de predicciones ajustado con un modelo de Cauchy con coordenadas geográficas: Mapa de predicción para el total de incendios forestales panel (A); Mapa de predicción para el total de incendios forestales con la ubicación de los eventos ocurridos en periodo 2001-2006 panel (B).

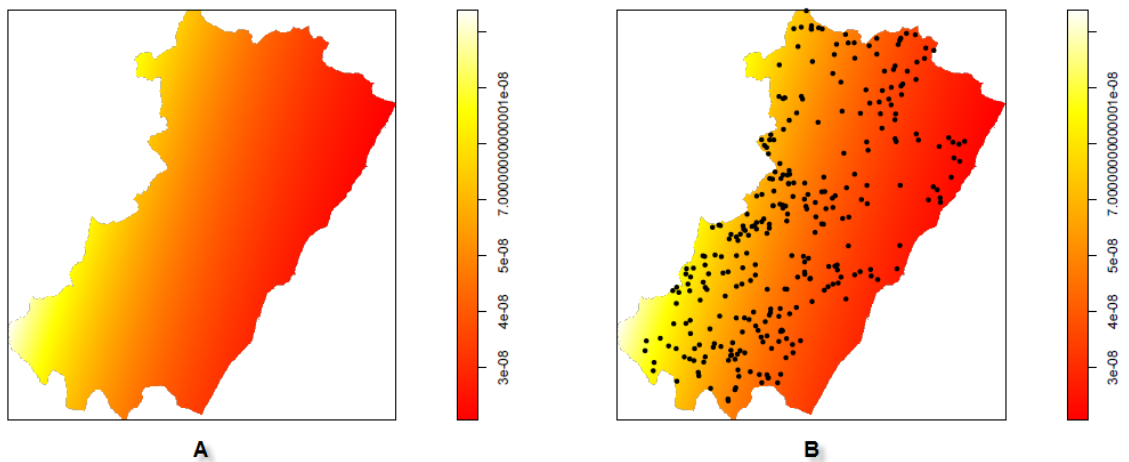


Figura 4.24: Mapa de predicciones ajustado con un modelo de Thomas con coordenadas geográficas: Mapa de predicción para los incendios naturales panel (A); Mapa de predicción para los incendios naturales con la ubicación de los eventos ocurridos en periodo 2001-2006 panel (B).

La figura 4.25 muestra los 4 casos más interesantes de la  $k$  cruzada en el panel (A) se observa la interacción entre los años 2002 y 2003 donde se muestra que existe una atracción a distancias muy pequeñas a menos de 1 km de radio, el mismo comportamiento de atracción se observa en el panel (B), correspondiente a la interacción entre los años 2002 y 2006 a una distancia de 1km muy similar al caso anterior.

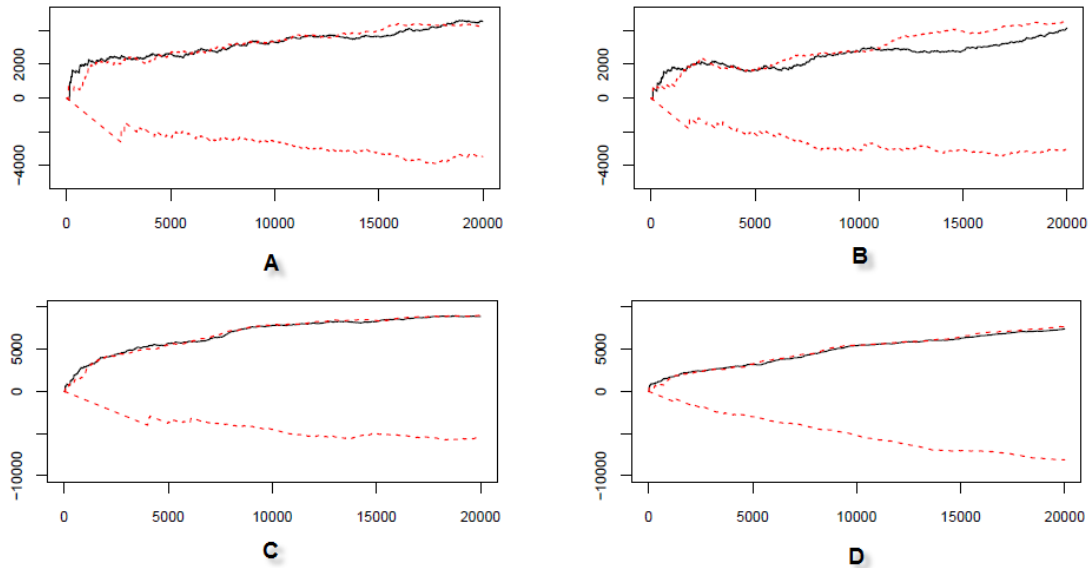


Figura 4.25: Función de  $k$  cruzada para los casos relevantes: Interacción entre los patrones de los años 2002 y 2003 panel (A); Interacción entre los patrones de los años 2002 y 2006 panel (B); Interacción entre los patrones de causas desconocidas y los incendios intencionados panel (C); Interacción entre los patrones de incendios por negligencia e intencionados panel (D).

En la Figura 4.25 panel (C) se encuentra la interacción entre los incendios de los cuales se desconoce su causa y los incendios intencionados al igual que los casos anteriores se observa una atracción a pequeña escala de aproximadamente 1 a 1.5 km lo cual podría suponer que los incendios desconocidos están ligados a aquellos ocurridos de forma intencionada, otro aspecto interesante de ver en este análisis es la repulsión muy definida de los patrones naturales con las otras causas (Ver gráfico en apéndice A).

La figura 4.26 refleja entender un poco el comportamiento de reincidencia, esta grafica muestra la atracción entre conatos e incendios (menores y mayores a una hectárea respectivamente) donde se observa una atracción a distancia muy pequeña entre los incendios lo que esto indica es que pueden ser asociados a comportamientos reiterativos de provocar un incendio en mayor magnitud hasta lograrlo.

## 4.8. Análisis de dependencia entre tamaño y localización

En esta sección se busca la evidencia de la presencia o falta de dependencia entre los diferentes tamaños del área quemada por marca asociada a cada uno de los

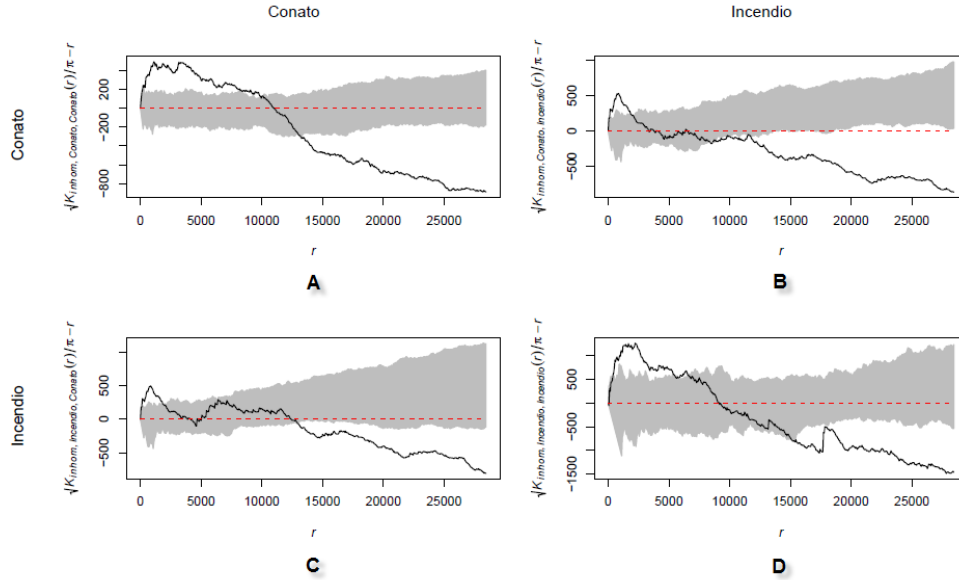


Figura 4.26: Función de k cruzada para la división de los incendios: Interacción entre conatos e incendios panel (C) y panel (B); Función k no homogénea para conatos panel(A); Función k no homogénea para incendios panel(D).

incendios y si estas dependen o no de su ubicación, el análisis contempla el periodo por cada año (2001-2006) así como causas naturales y humanas. Para medir la correlación entre el tamaño del área quemada y la localización se usan 2 funciones de la paquetería “spatstat” del software R “markcorr” y “E-mark”, las cuales calculan la función de correlación de marca y los reverse conditional moments respectivamente. En Baddeley (2010) se define a la función de correlación de marca como:

$$\rho_{f^*r} = \frac{\mathbb{E}[f(M_1, M_2)]}{\mathbb{E}[f(M, M_3)]} \quad (4.1)$$

Donde  $M_1$  y  $M_2$  representan las marcas, en nuestro caso el tamaño de los incendios asociados a 2 puntos que se encuentran a una distancia  $r$ .  $M$  y  $M_3$  son realizaciones independientes de la distribución marginal de las marcas.

La principal utilidad de la función es ayudar a medir la relación entre las marcas asociadas a los puntos y su ubicación, para este caso el tamaño de área quemada de los incendios y su ubicación geográfica.

No hay que confundir esta medida con la utilizada en estadística clásica para medir la correlación, ya que la medida previamente definida puede tomar cualquier valor positivo, en caso particular cuando la función toma el valor de 1 se sugiere la falta de correlación entre los valores de la marca de cada uno de los puntos.

Los momentos condicionales reversas, son los momentos asociados a cada una de las marcas, es decir la esperanza y la varianza respectivamente, condicionado a que

existe otro punto del proceso a una distancia  $r$ . Estas funciones se definen:

$$\mathbb{E}(r) = \mathbb{E}_{st}[M(s)] \quad (4.2)$$

$$Var(r) = \mathbb{E}_{st}[M(s) - \mathbb{E}(t)]^2 \quad (4.3)$$

Donde  $\mathbb{E}_{st}$  representa la esperanza condicional dado que existen 2 puntos  $s$  y  $t$  separados por una distancia  $r$ , además  $M(s)$  es la marca asociada al punto  $s$ , con estas funciones se puede diagnosticar la dependencia entre los puntos y las marcas, cuando esto sucede ambos valores son constantes y no depende la distancia  $r$ .

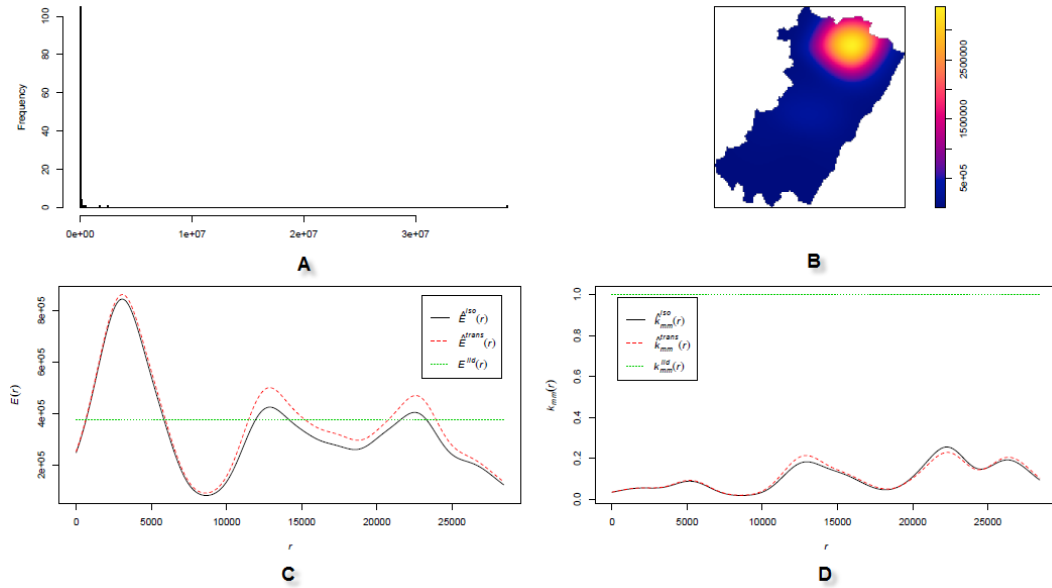


Figura 4.27: Gráficas correspondientes al análisis de marcas del año 2001: Histograma panel (A); Estimación de Kernel panel (B); función de correlación panel (C); Esperanza condicional inversa panel (D).

Para el año 2001 (Figura 4.27) se muestra por medio del histograma un sesgo muy marcado así valores pequeños así como la presencia de valores extremos que sesgan la distribución, en análisis previos se había detectado un caso extremo de un incendio en este año y por el análisis de estimación de Kernel se puede determinar que dicho incendio ocurrió en la parte noreste de la región de estudio ya que la intensidad es más marcada en esa parte de la región.

Para la función de correlación de marca en general en los 6 años vemos una disparidad entre valores positivos y negativos para todo tipo de distancias con lo cual se observa que no existe independencia entre las marcas, el año 2001 muestra los valores cercanos a cero, además de que toma valores positivos a cualquier distancia, lo cual se interpreta como marcas independientes sin embargo se considera que este resultado puede ser erróneo ya que el caso atípico de este año puede estar sesgando la estimación de esta función.



En cuanto a la media condicional reversa la función se localiza por arriba y por debajo del valor esperado donde se podría hablar de independencia, se ve una concauidad a una distancia de 5 km, y una ligera estabilidad entre 10 y 25 km.

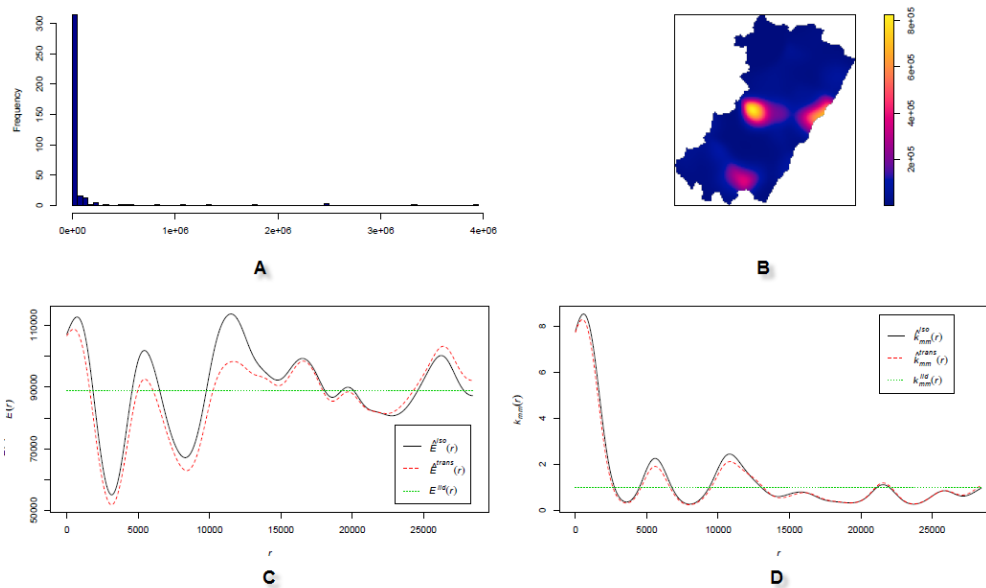


Figura 4.28: Graficas correspondientes al análisis de marcas para las causas humanas: Histograma panel (A); Estimación de Kernel panel (B); función de correlación panel (C); Esperanza condicional inversa panel (D).

Para la figura 4.28 y 4.29 los patrones de incendios por causas humanas y por causas naturales muestran algo interesante: los incendios causados por humanos presentan una alta correlación a distancias menores a 5 km (aprox. 4.5 km) con lo cual vemos esta alta asociación entre los incendios humanos y su localización asegurándonos son provocados cercanos unos de otros generando así una mayor área quemada, caso contrario de los incendios naturales los cuales la gráfica de kernel nos lo había anticipado, se comportan de forma independiente con su localización y el tamaño de área quemada.

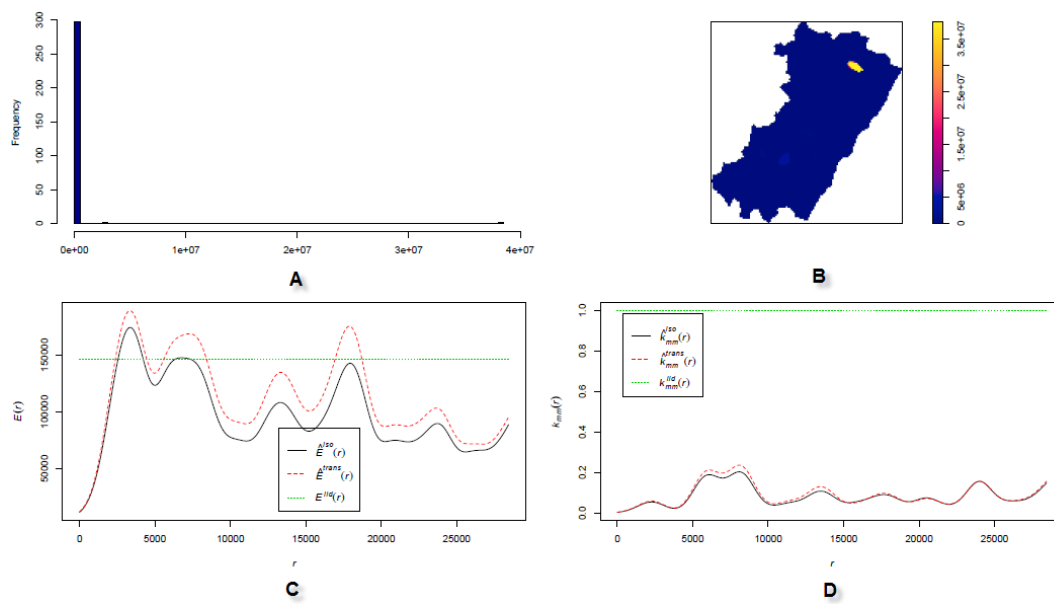


Figura 4.29: Graficas correspondientes al análisis de marcas para las causas naturales: Histograma panel (A); Estimación de Kernel panel (B); función de correlación panel (C); Esperanza condicional inversa panel (D).

# Capítulo 5

## Conclusiones

### 5.1. Objetivo general

- Aplicar los métodos desarrollados en procesos puntuales como una alternativa estadística para analizar la incidencia y el comportamiento del fenómeno de incendios forestales.

### 5.2. Objetivos específicos

- Construir modelos de la distribución geográfica de los incendios forestales y concluir si existe agrupamiento, repulsión o interacción entre ellos, esto se realiza en un contexto determinado por un espacio-tiempo, dimensión del área quemada y causas.
- Detectar las áreas con mayor probabilidad de ocurrencia de incendios forestales mediante un mapa de riesgo, tanto para causas naturales como humanas (negligencia o accidente e intencionado).
- Analizar el efecto de las covariables seleccionadas sobre la intensidad de los incendios forestales.
- Localizar tendencia de incendios forestales sobre la región de estudio.
- Identificar comportamientos reiterativos en la propagación de los incendios.

### 5.3. Hipótesis

- El total de incendios forestales para el periodo 2001-2006 en la región de estudio forman conglomerados ya sea por comportamientos de reincidencia, por factores externos como el tipo de suelo, altitud, clima o por la misma dinámica forestal.

- Los incendios por causas humanas (negligencia accidentes o intencionados) tienden a estar más concentrados cerca de áreas densamente pobladas.
- Los incendios por causas naturales tienden a ocurrir de forma aleatoria y no dependen de la información adicional que aportan las covariables.
- Existe una relación de dependencia por marcas entre el área quemada y por cada una de las causas mencionadas previamente.

## 5.4. Contraste entre las conclusiones los objetivos y las hipótesis

El desarrollo del presente trabajo permite contrastar las hipótesis así como ver el alcance de los objetivos tanto generales como específicos, obteniéndose así las siguientes conclusiones:

- Por medio de la estimación de Kernel y de los métodos de segundo orden se llega a la conclusión (apoyándose en los contrastes generados con los intervalos de confianza formados por las simulaciones de Monte Carlo), de que los incendios en general y cada patrón de interés en particular forman conglomerados de manera natural que tiene que ver más allá de la información adicional que se pueda agregar al modelo, ya que su distribución depende en mayor parte de sus coordenadas geográficas. Este hecho está fundamentado en que la mayoría de los modelos seleccionados se ajustan con funciones polinómicas dejando a un lado la información de las covariables.
- Por medio de la estimación de Kernel se puede contrastar que los incendios causados por humanos tienden a estar más cerca de las áreas con mayor población, las cuales se localizan en el sureste de la provincia, misma ubicación en que la estimación de Kernel reportan una mayor intensidad estimada para este tipo de incendios.
- Los incendios causados por la naturaleza tienen una distribución más aleatoria y no dependen de la información de las covariables, sino todo lo contrario, ya que en dicho patrón los mejores ajustes fueron obtenidos mediante los modelos que tomaban las coordenadas geográficas y funciones polinómicas de grado 2 y grado 4.
- Para los análisis de interacción mediante la función k cruzada se observó que los patrones por año presentan una independencia a distancias pequeñas y una repulsión a grandes distancias. El único año donde podemos ver una ligera dependencia o interacción a distancias muy pequeñas es entre el año 2002 y 2003. Con respecto a los incendios por tipo de causa, los que consideramos como desconocidos pueden asociarse como intencionados, esto debido a su relación

de dependencia a muy corta distancia, de igual forma se puede percibir una interacción a corta distancia entre negligencia e intencionados. Con respecto a los naturales era de esperarse que fueran independientes con respecto de los otros patrones y que mostraran una repulsión muy marcada a muy corta distancia. En la función  $k$  cruzada que compara los incendios por tamaño y causa podemos ver una atracción entre los conatos humanos y los incendios humanos, es decir, entre los incendios por causas humanas menores que 1 ha y mayores que la misma por lo cual da indicios de comportamientos reiterativos de provocar incendios que se salen de control.

- En cuanto a la función de correlación de marca se podría hablar de una independencia entre los tamaños de los incendios para el año 2001 y para aquellos causados de forma natural. Esto se muestra al tener una función de correlación de marcas estable y con muy pequeñas fluctuaciones a lo largo de todo el rango de distancias. El análisis es más certero para los incendios naturales que para los del año 2001 aunque el comportamiento en este año se puede asociar al dato atípico que puede estar modificando en cierta medida la correlación de marca.

Dentro de los objetivos generales y específicos se demostró que si es posible modelar estos fenómenos mediante un proceso estocástico y una estimación de la intensidad que es variante en cuanto a la ubicación geográfica que los eventos tengan. Este es un primer avance de esta metodología como una alternativa para plantear este y muchos otro tipos de problemas que se puedan presentar en fenómenos que son difíciles de comprender.

Llama la atención que el efecto de las covariables sobre los patrones no tuvo el peso que se esperaba, ya que la mayoría de los modelos ajustaron mejor con polinomios de grado 2 y 4 y las coordenadas geográficas, este efecto se puede asociar más a una tendencia geográfica donde ya existen áreas más propensas, ya sea por fácil acceso a las personas o porque delimitan áreas densamente pobladas, lo que explica la mayoría de los ajustes de poisson Clúster donde se origina el famoso punto padre y de ahí tiene descendencia.

Este último comentario sugiere algún tipo de estudio posterior en el que se tome en cuenta algún otro tipo de información adicional que se pueda utilizar para encontrar un verdadero efecto en incendios causados por humanos, de igual forma se deja abierto para un estudio posterior donde se pueda modelar e identificar comportamientos pirómanos o reiterativos, ya que el tipo de análisis realizado en primera instancia da lugar a concluir que si existen estos comportamiento. Sin embargo, no se puede acertar con seguridad debido a que la  $k$  cruzada solo analiza dependencia o interacción por medio de las distancias.

# Capítulo 6

## Apéndice A.

### 6.1. Funciones L no homogéneas.

#### 6.1.1. Modelos de Poisson no homogéneos.

En esta sección se presentan los modelos de poisson no homogéneos que fueron ajustados para los diferentes patrones de estudio. La función L no homogénea fue el primer criterio utilizado para la selección del grupo de modelos de cada patrón que se ajustaba al modelo propuesto.

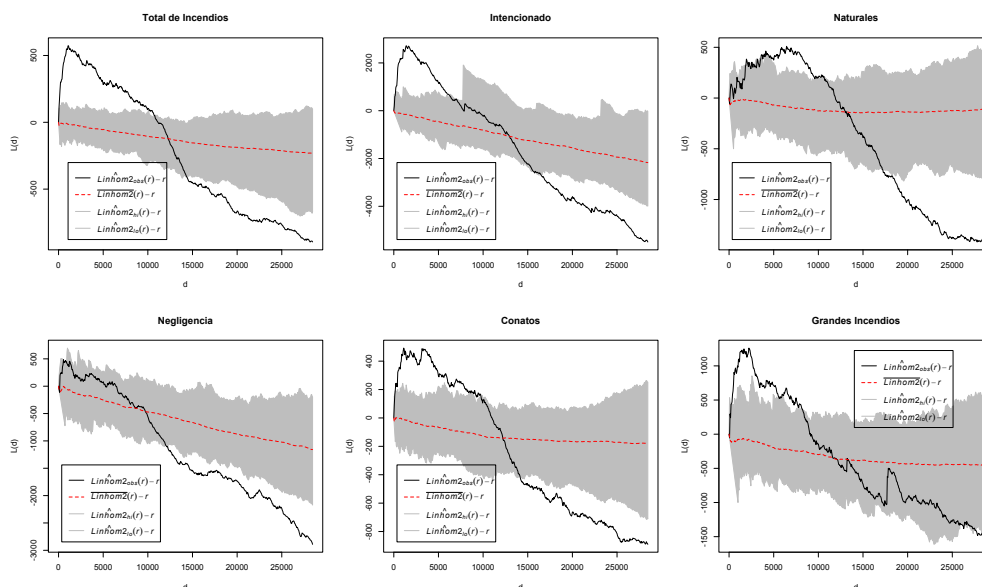


Figura 6.1: Funciones L no homogénea para un modelo de Poisson ajustado a los diferentes patrones de estudio.

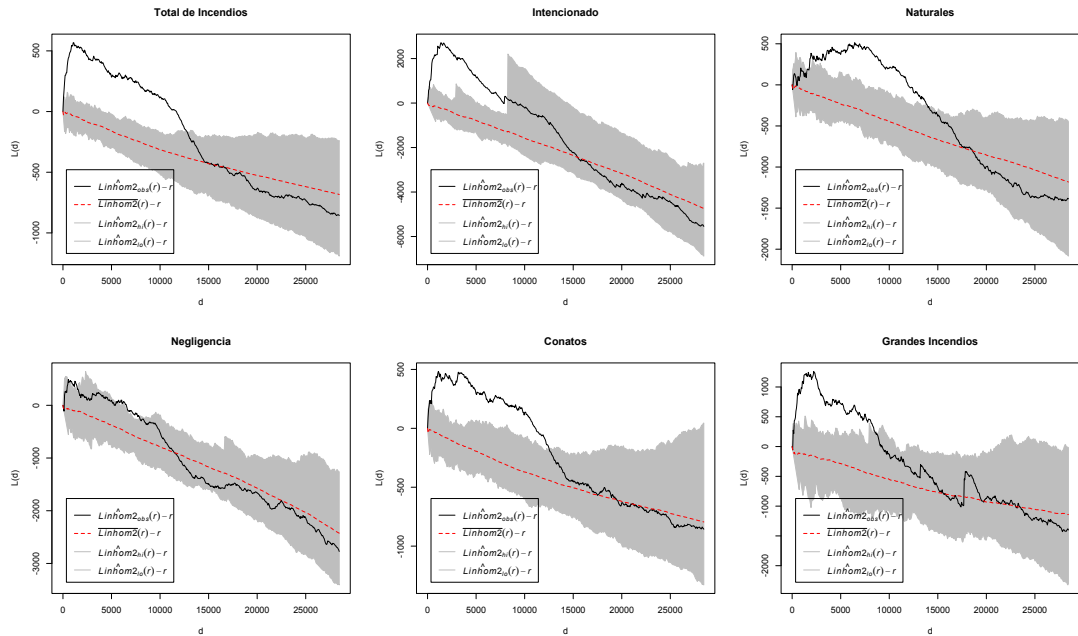


Figura 6.2: Funciones  $L$  no homogénea para un modelo de Poisson ajustado con las coordenadas geográficas para los diferentes patrones de estudio.

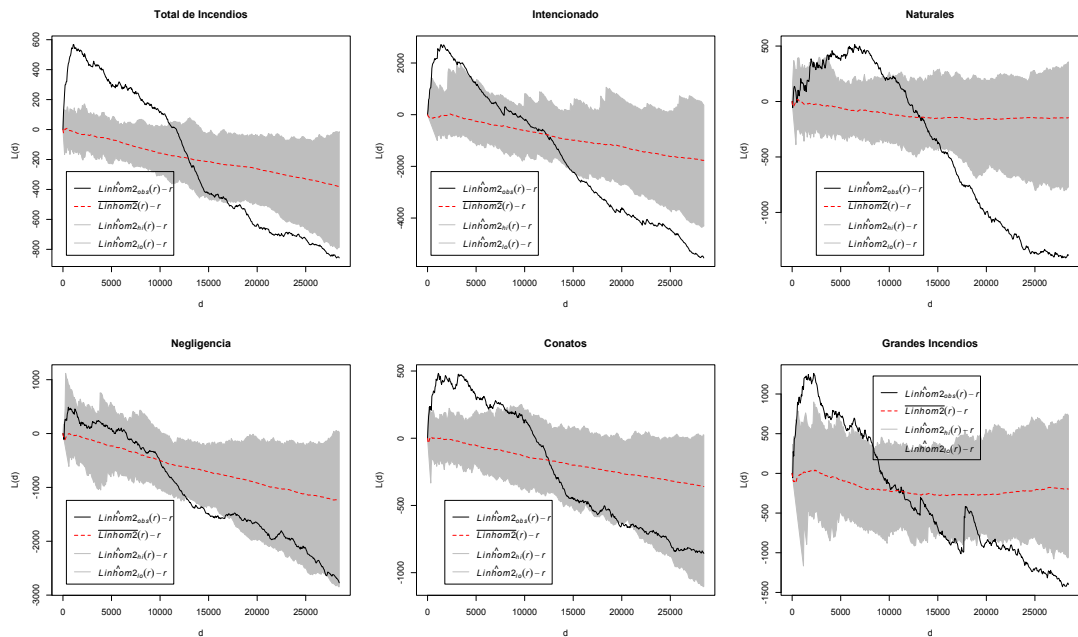


Figura 6.3: Funciones  $L$  para un modelo no homogéneo de Poisson ajustado con las covariables para los diferentes patrones de estudio.

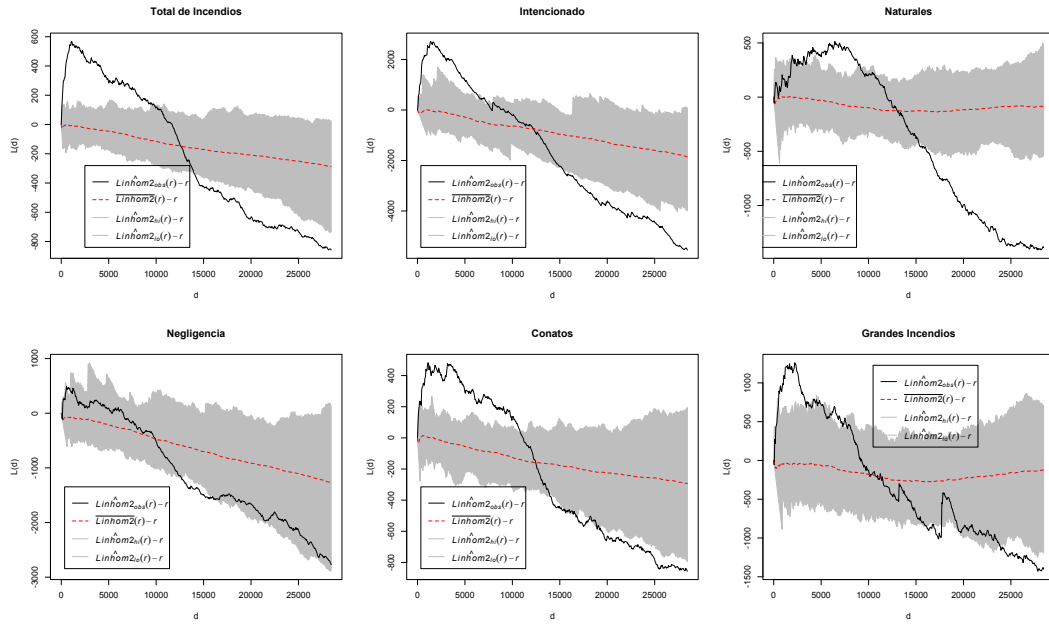


Figura 6.4: Funciones  $L$  para un modelo no homogéneo de Poisson ajustado con las covariables seleccionadas mediante el método de backward para los diferentes patrones de estudio.

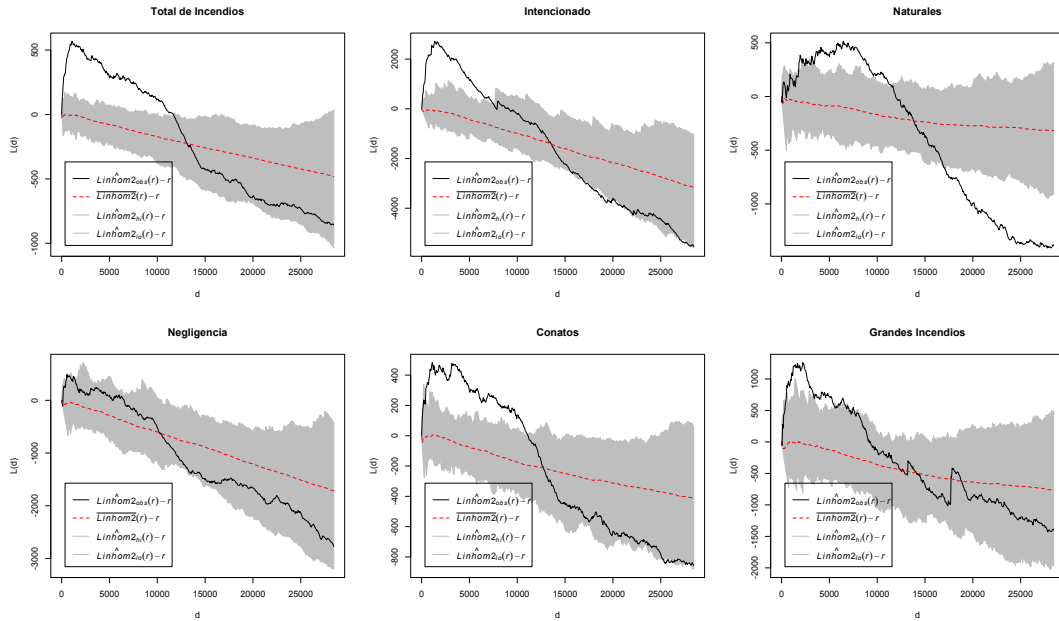


Figura 6.5: Funciones  $L$  para un modelo no homogéneo de Poisson ajustado con las covariables y coordenadas geográficas para los diferentes patrones de estudio.



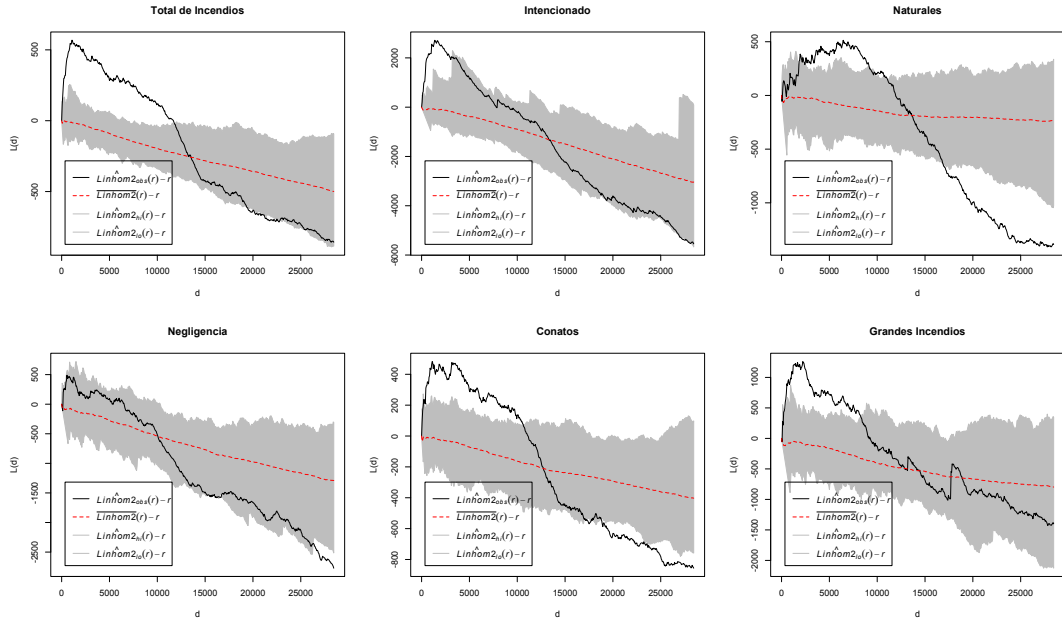


Figura 6.6: Funciones  $L$  para un modelo no homogéneo de Poisson ajustado con las covariables y coordenadas geográficas seleccionadas mediante el método backward para los diferentes patrones de estudio.

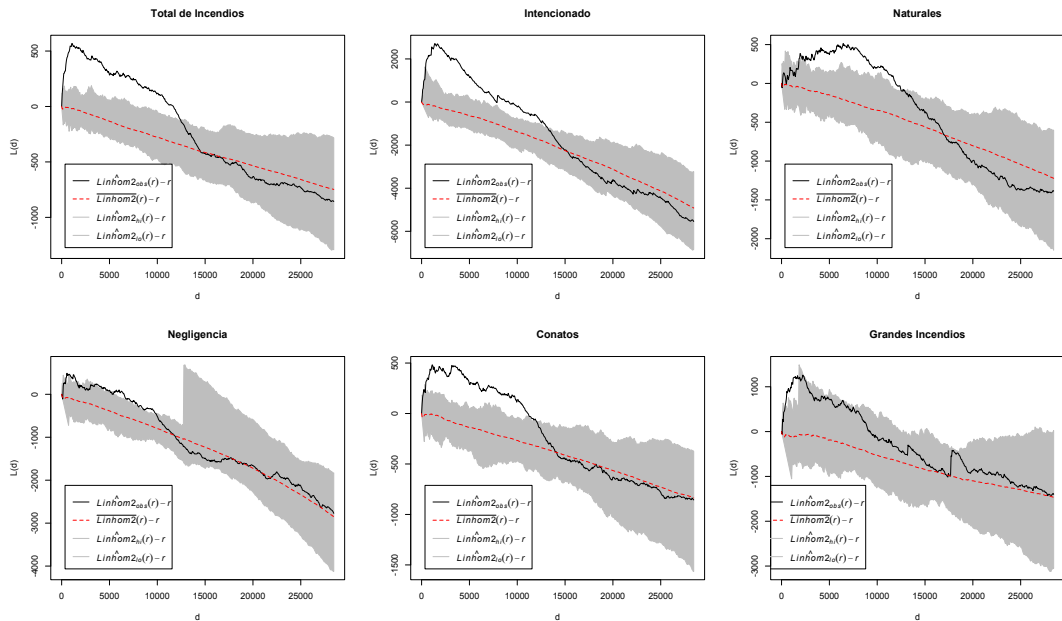


Figura 6.7: Funciones  $L$  para un modelo no homogéneo de Poisson ajustado con un polinomio de grado 4 para los diferentes patrones de estudio.

## 6.1.2. Modelos de cluster.

En esta sección se presentan los modelos de Poisson Clúster en los que tenemos los modelos de Thomas y los de Matern.

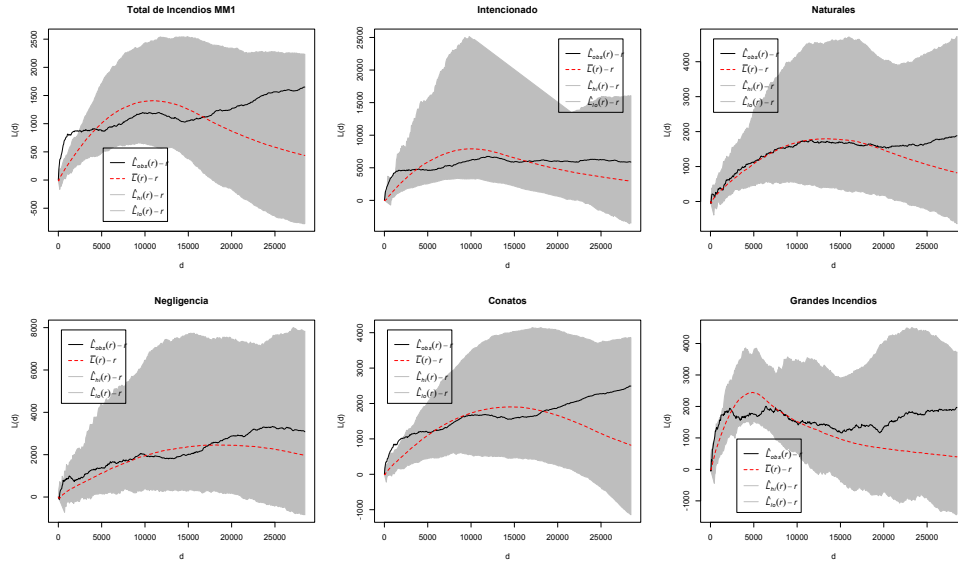


Figura 6.8: Funciones L para un modelo de Matern homogéneo para los diferentes patrones de estudio.

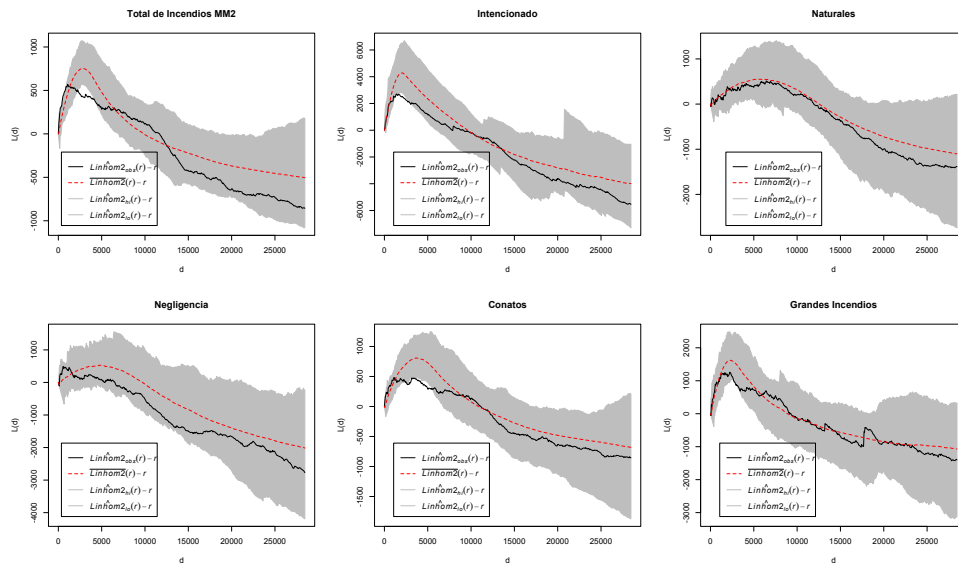


Figura 6.9: Funciones L para un modelo de Matern ajustado con las coordenadas geográficas para los diferentes patrones de estudio.

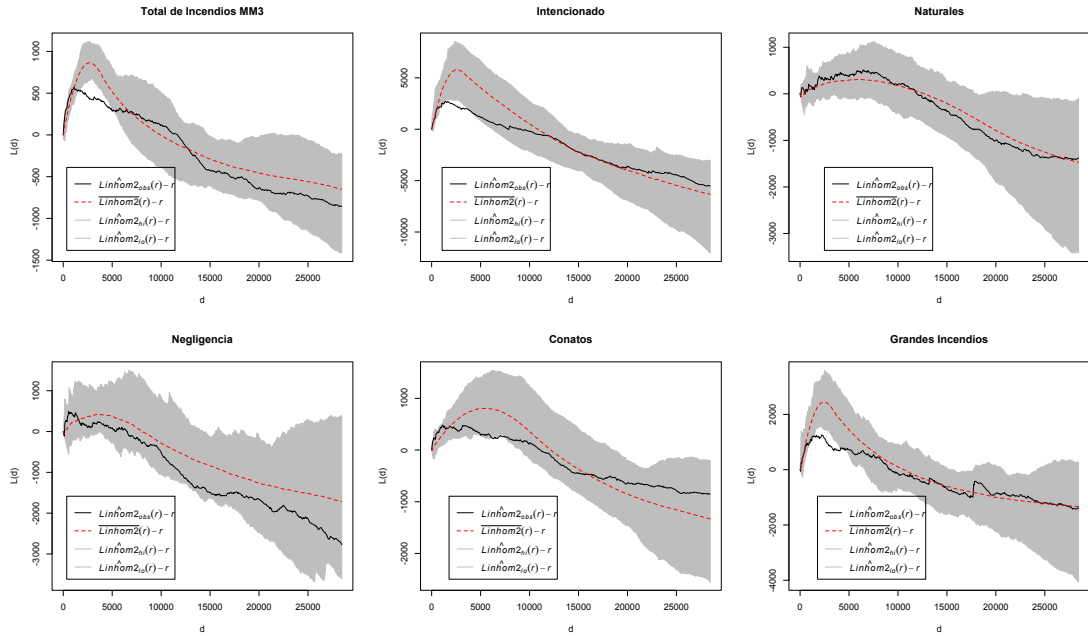


Figura 6.10: Funciones L para un modelo de Matern ajustado con las covariables para los diferentes patrones de estudio.

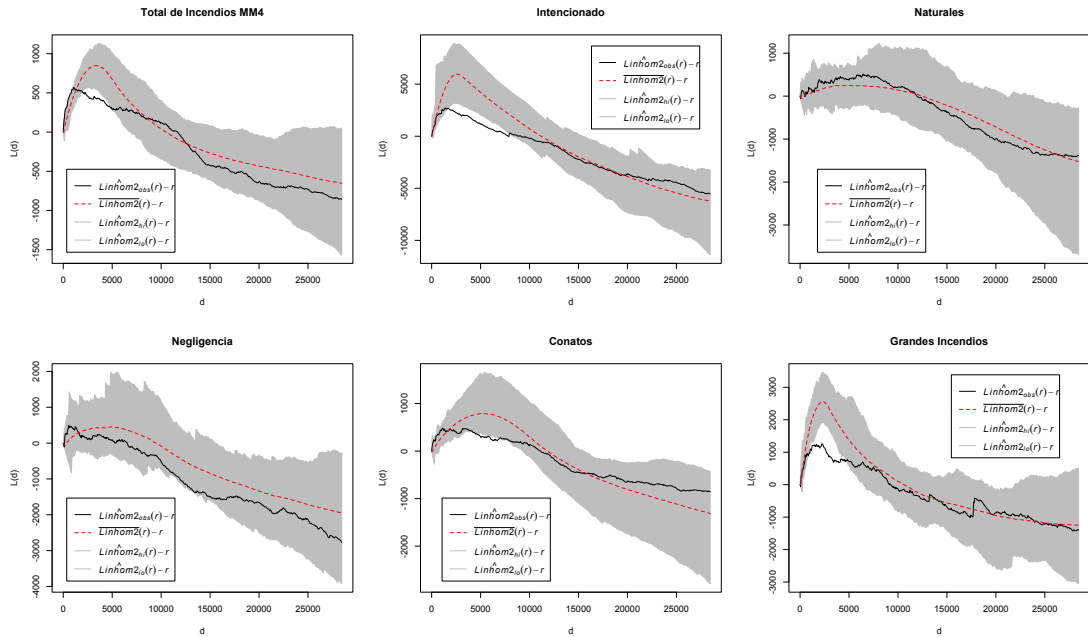


Figura 6.11: Funciones L para un modelo de Matern ajustado con las covariables seleccionadas por el método backward para los diferentes patrones de estudio.

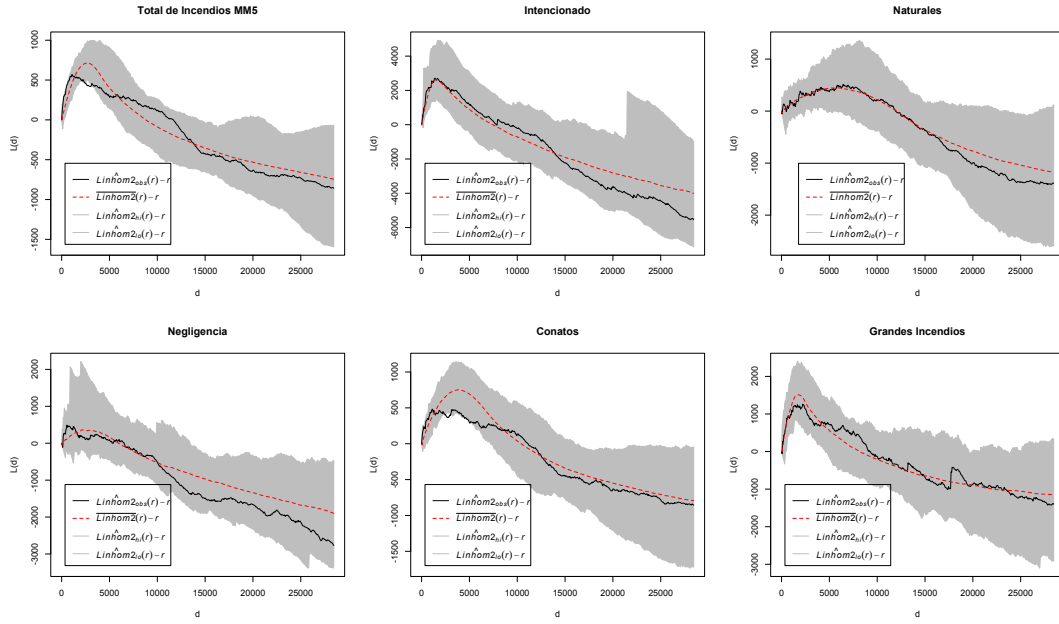


Figura 6.12: Funciones L para un modelo de Matern ajustado con las covariables y las coordenadas geográficas para los diferentes patrones de estudio.

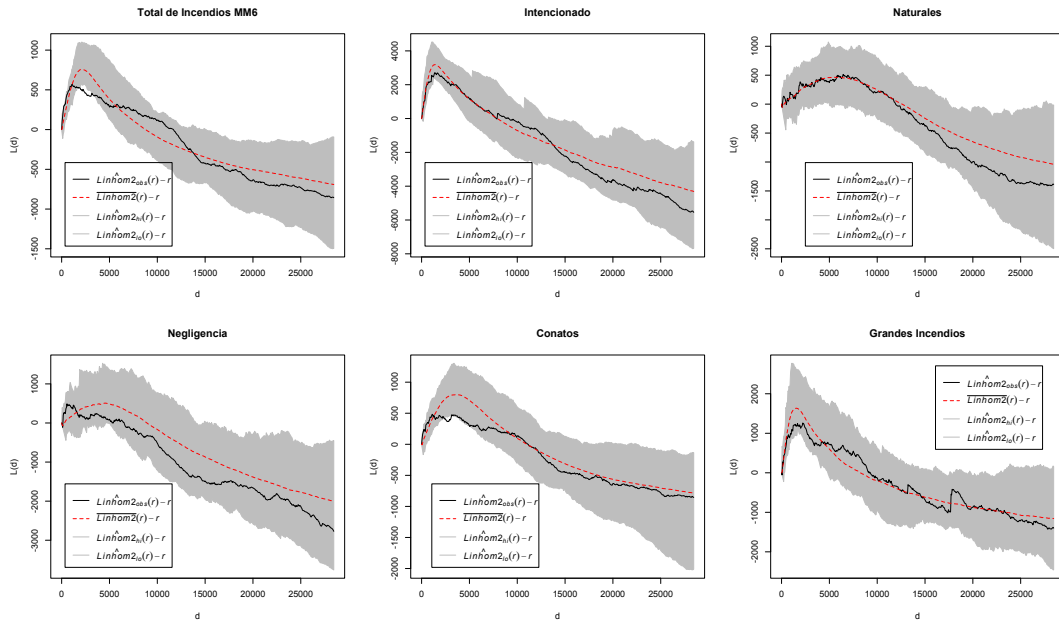


Figura 6.13: Funciones L para un modelo de Matern ajustado con las covariables y las coordenadas geográficas seleccionadas por el método de backward para los diferentes patrones de estudio.

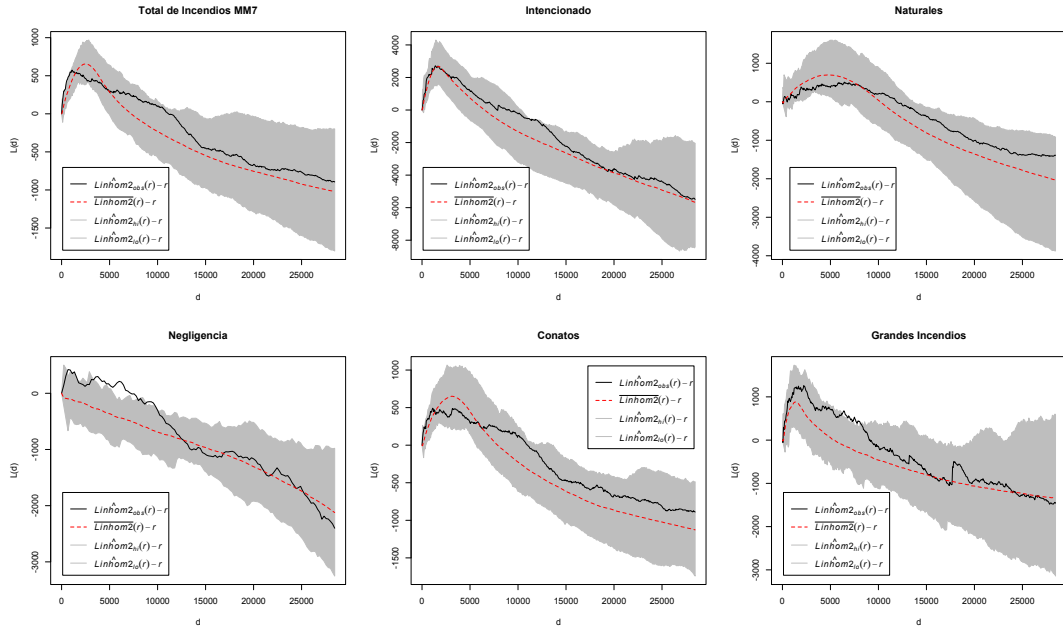


Figura 6.14: Funciones  $L$  para un modelo de Matern ajustado con un polinomio de grado 2 para los diferentes patrones de estudio.

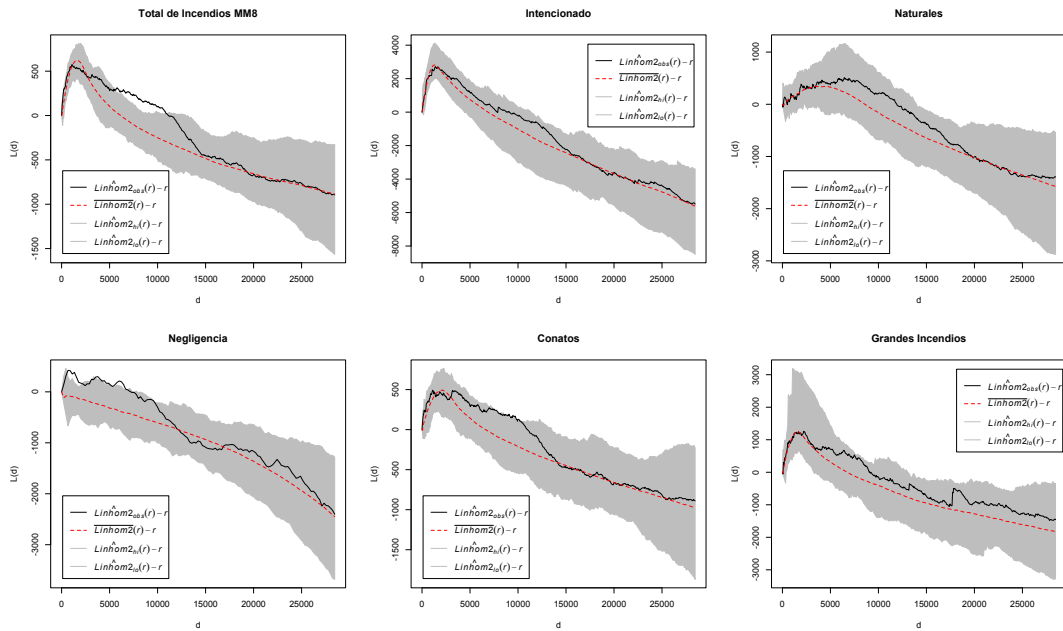


Figura 6.15: Funciones  $L$  para un modelo de Matern ajustado con un polinomio de grado 4 para los diferentes patrones de estudio.

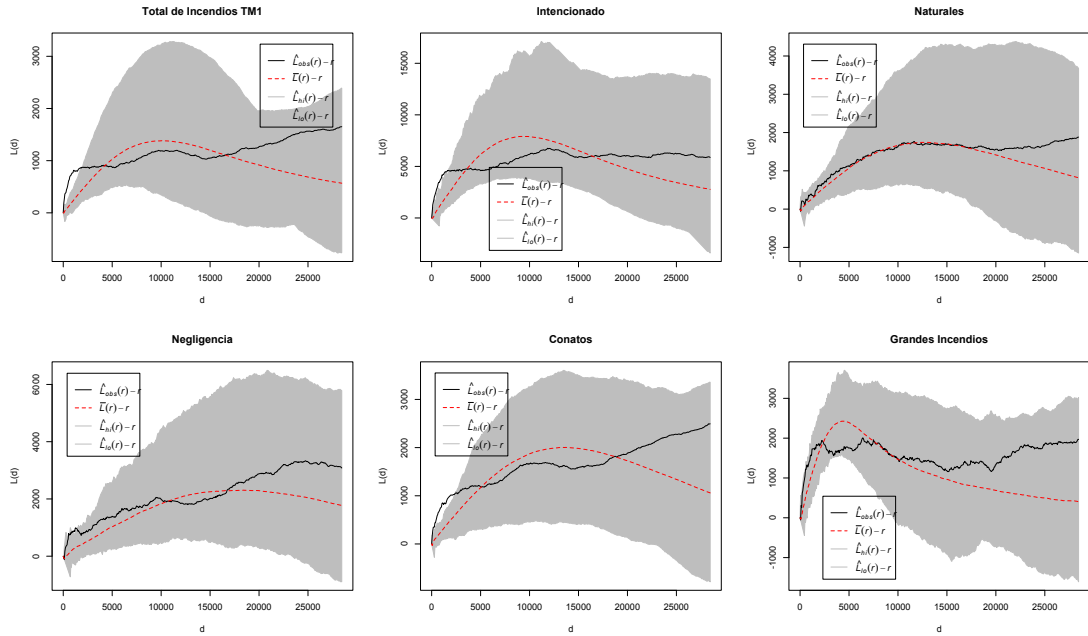


Figura 6.16: Funciones  $L$  para un modelo de Thomas homogéneo para los diferentes patrones de estudio.

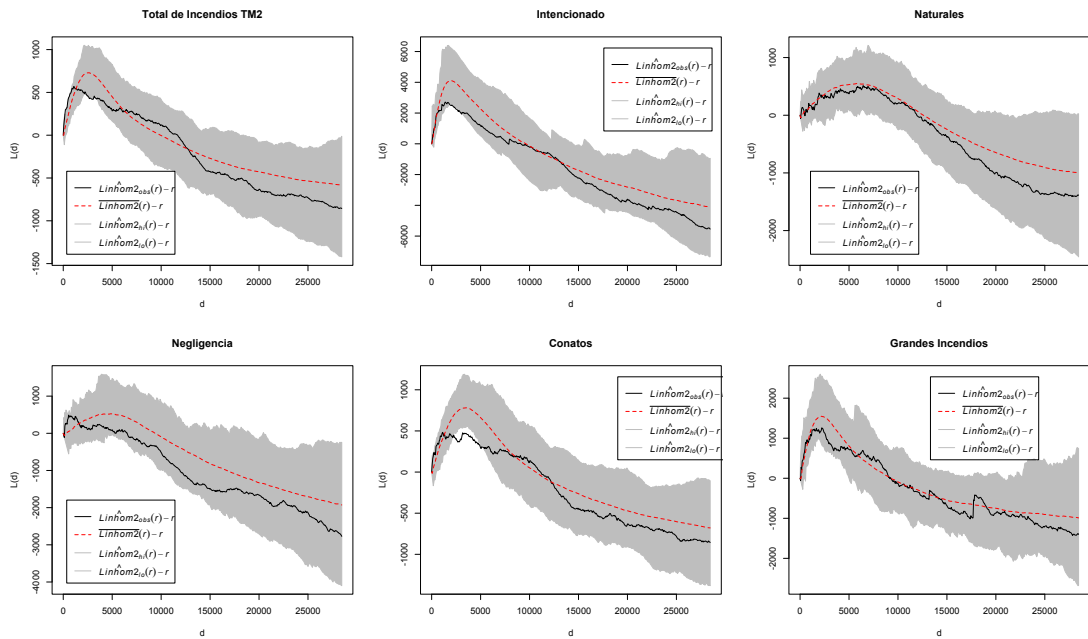


Figura 6.17: Funciones  $L$  para un modelo de Thomas ajustado con las coordenadas geográficas para los diferentes patrones de estudio.

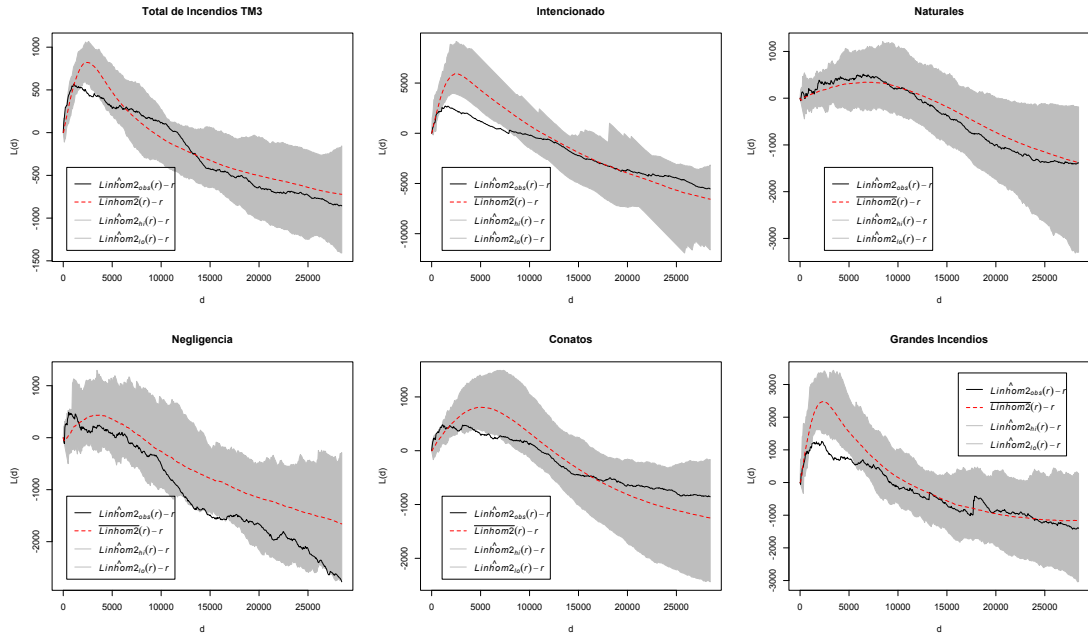


Figura 6.18: Funciones  $L$  para un modelo de Thomas ajustado con las covariables para los diferentes patrones de estudio.

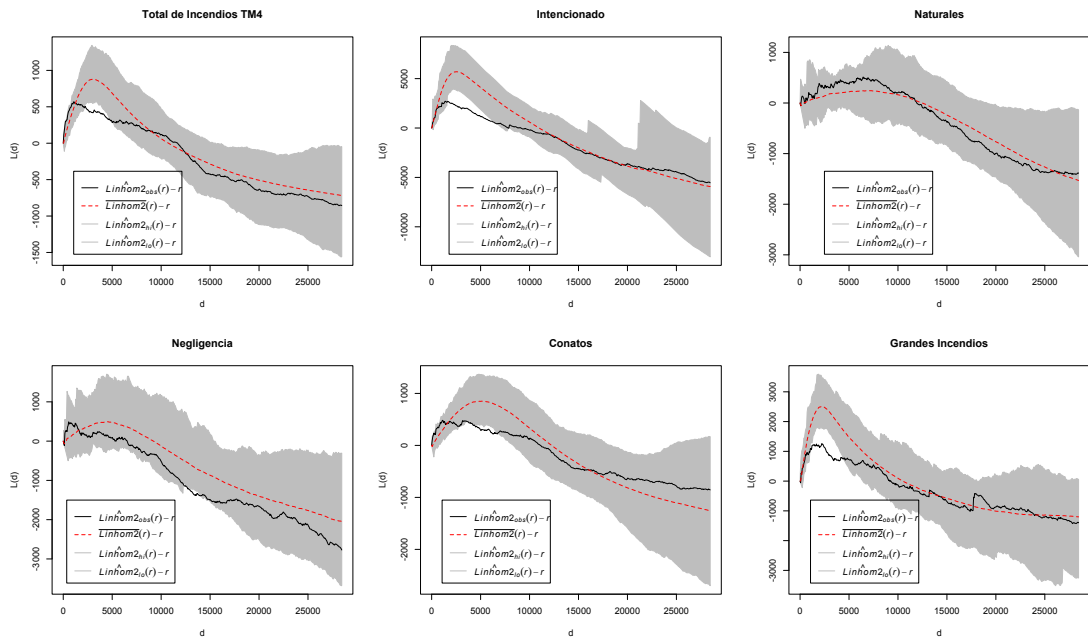


Figura 6.19: Funciones  $L$  para un modelo de Thomas ajustado con las covariables seleccionadas por el método backward para los diferentes patrones de estudio.

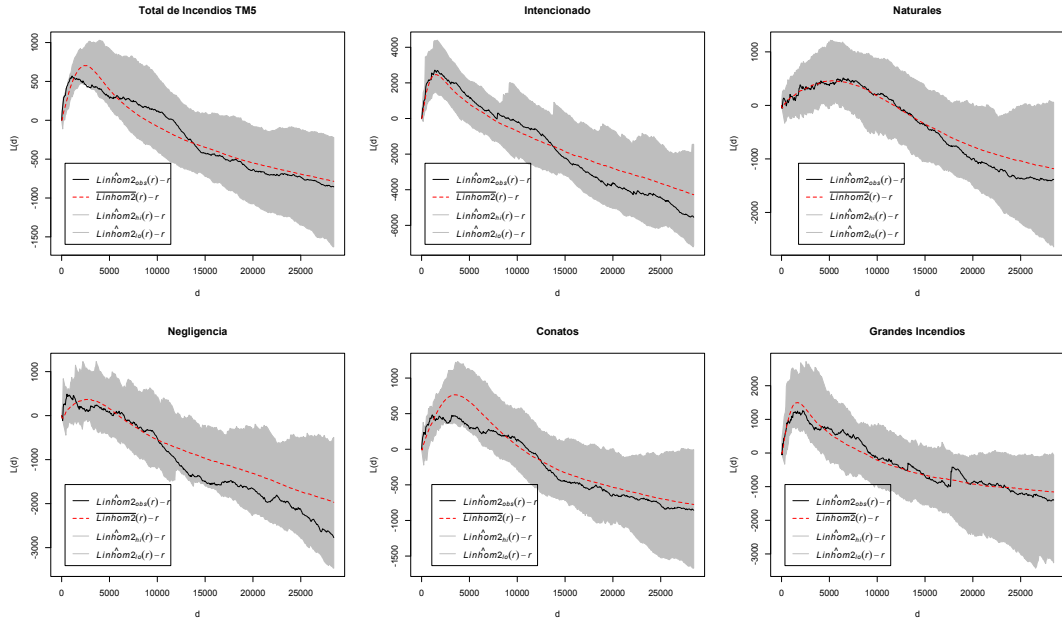


Figura 6.20: Funciones  $L$  para un modelo de Thomas ajustado con las covariables y las coordenadas geográficas para los diferentes patrones de estudio.

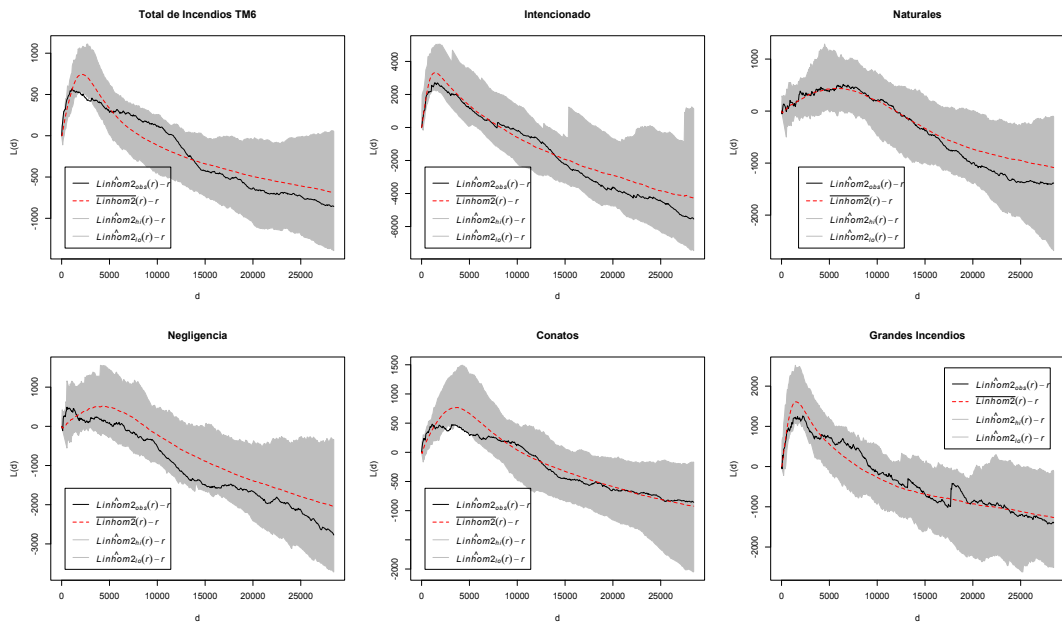


Figura 6.21: Funciones  $L$  para un modelo de Thomas ajustado con las covariables y las coordenadas geográficas seleccionadas por el método de backward para los diferentes patrones de estudio.



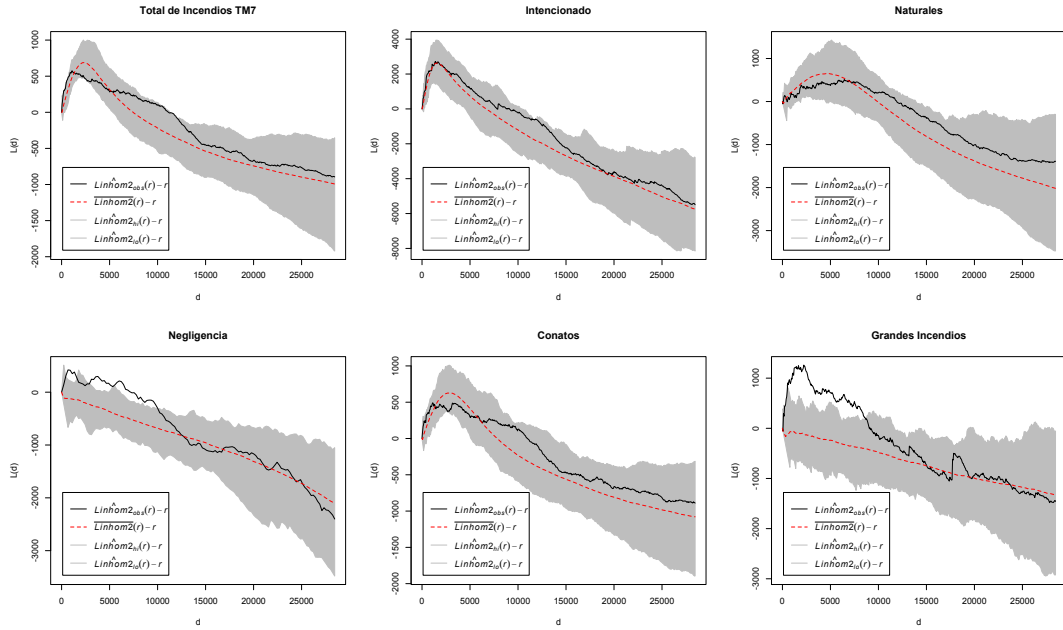


Figura 6.22: Funciones  $L$  para un modelo de Thomas ajustado con un polinomio de grado 2 para los diferentes patrones de estudio.

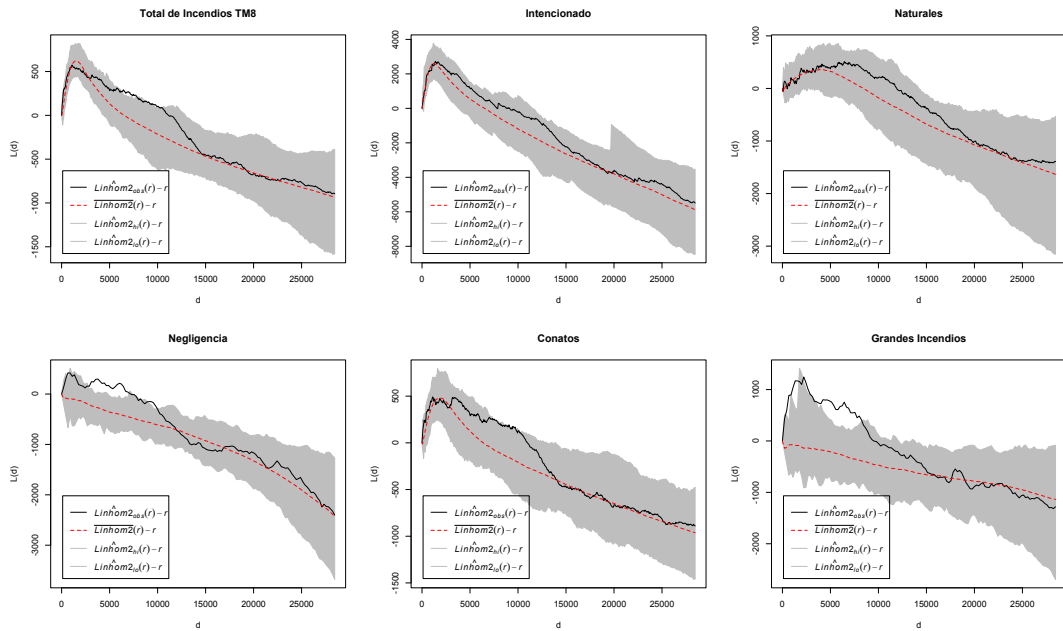


Figura 6.23: Funciones  $L$  para un modelo de Thomas ajustado con un polinomio de grado 4 para los diferentes patrones de estudio.

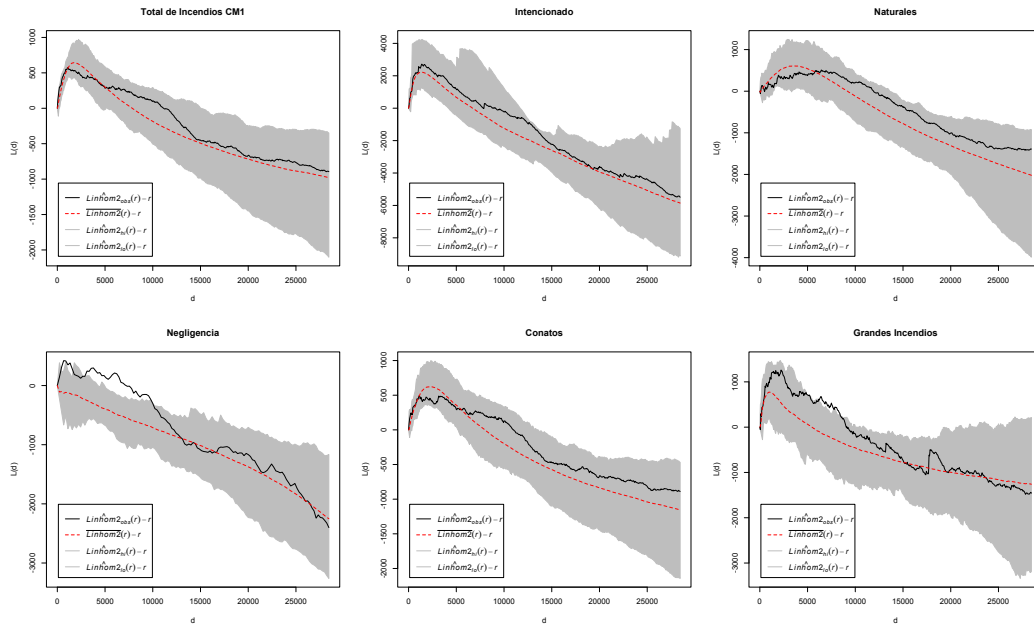


Figura 6.24: Funciones  $L$  para un modelo de Cauchy ajustado con las coordenadas geográficas para los diferentes patrones de estudio.

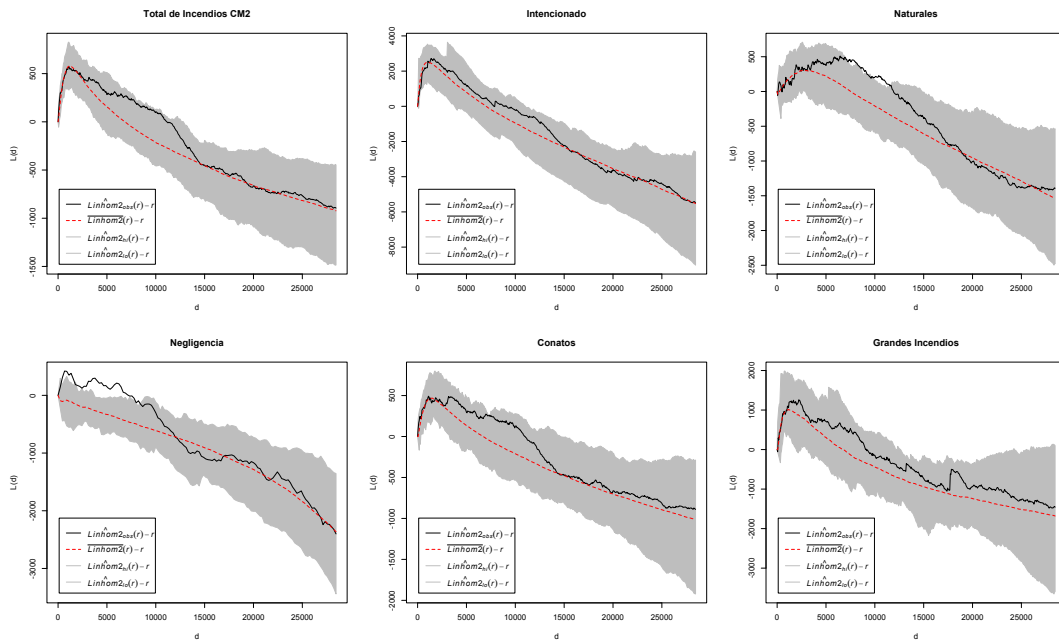


Figura 6.25: Funciones  $L$  para un modelo de Cauchy ajustado con un polinomio de grado 2 para los diferentes patrones de estudio.

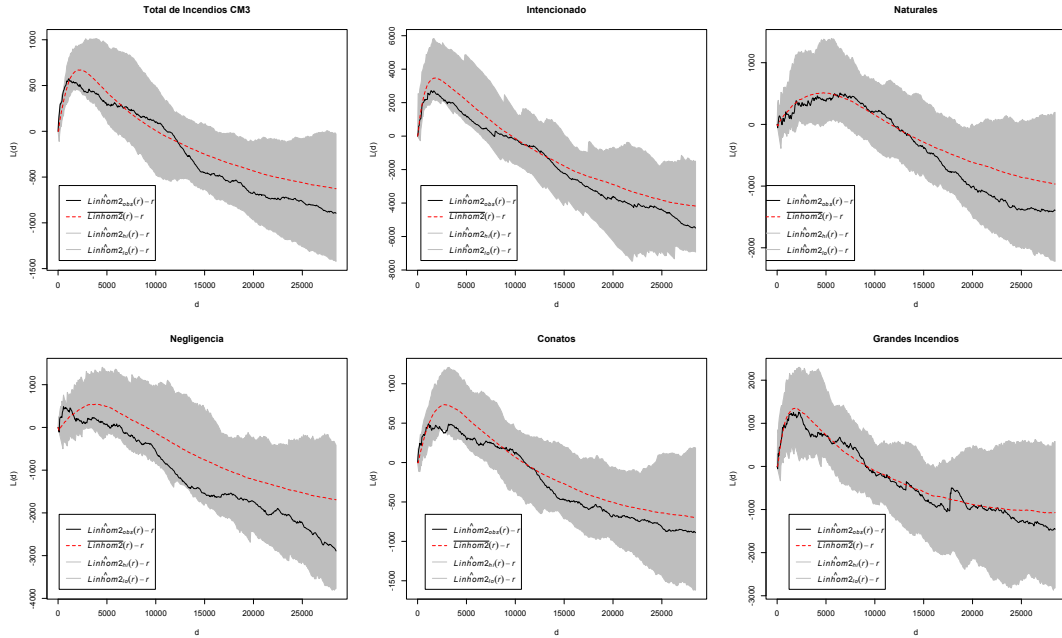


Figura 6.26: Funciones  $L$  para un modelo de Cauchy ajustado con un polinomio de grado 4 para los diferentes patrones de estudio.

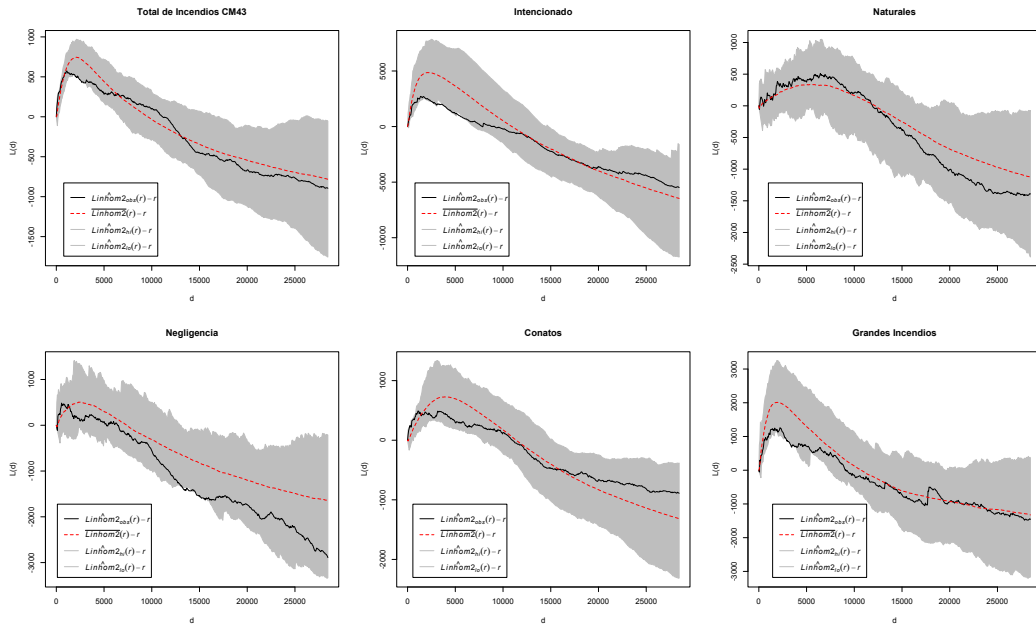


Figura 6.27: Funciones  $L$  para un modelo de Cauchy ajustado con las coordenadas geográficas y las covariables para los diferentes patrones de estudio.

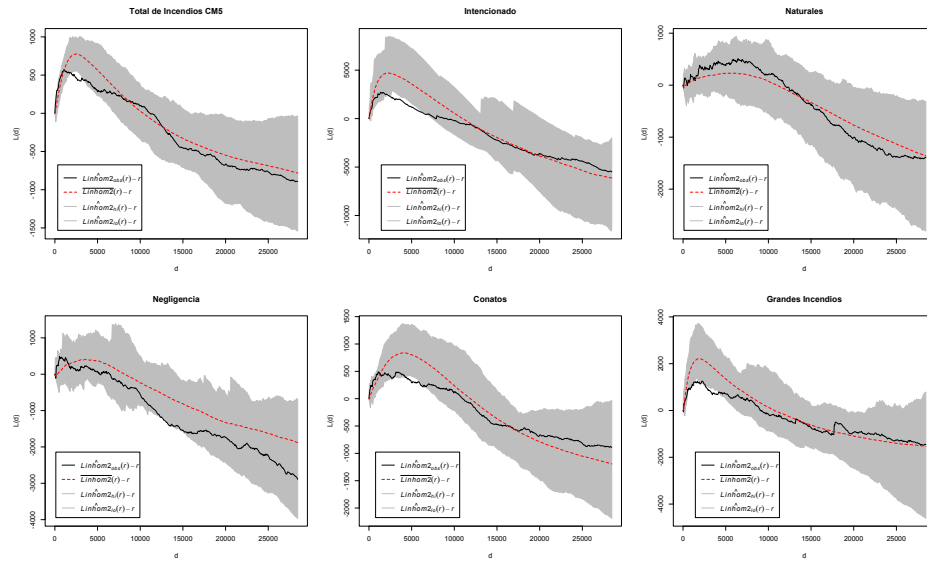


Figura 6.28: Funciones L para un modelo de Cauchy ajustado con las coordenadas geográficas y las covariables seleccionadas por el método de backward para los diferentes patrones de estudio.

### 6.1.3. Modelos de Gibbs.

En esta sección se presentan los modelos de Procesos de Gibbs entre los que tenemos el ajuste de los modelos de Hard Core y los de Strauss.

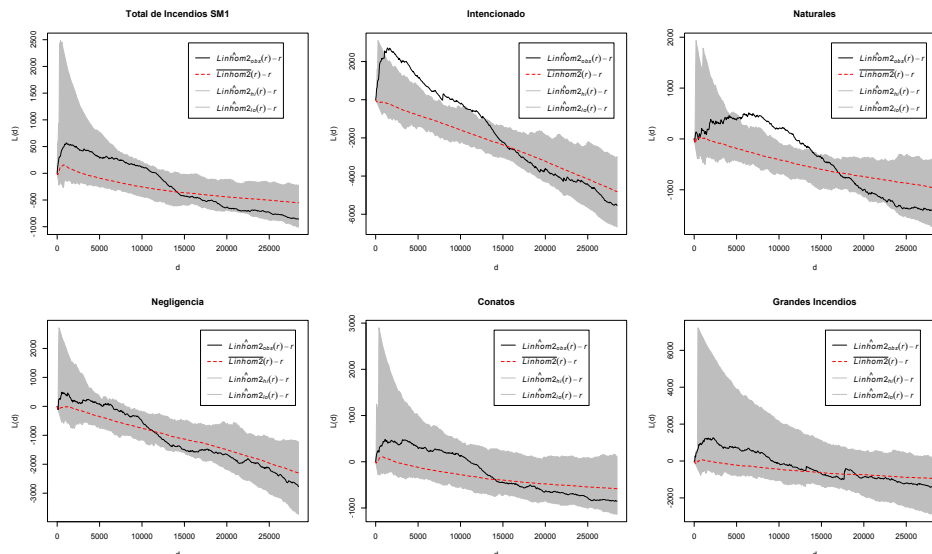


Figura 6.29: Funciones L para un modelo de Strauss ajustado con un polinomio de grado 2 para los diferentes patrones de estudio.

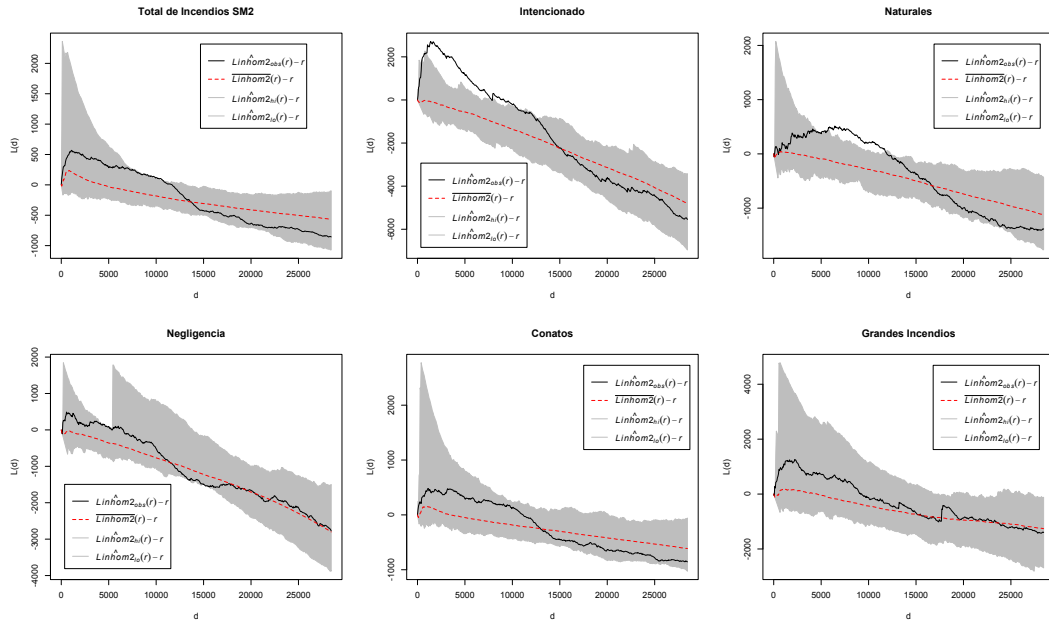


Figura 6.30: Funciones L para un modelo de Strauss ajustado con un polinomio de grado 4 para los diferentes patrones de estudio..

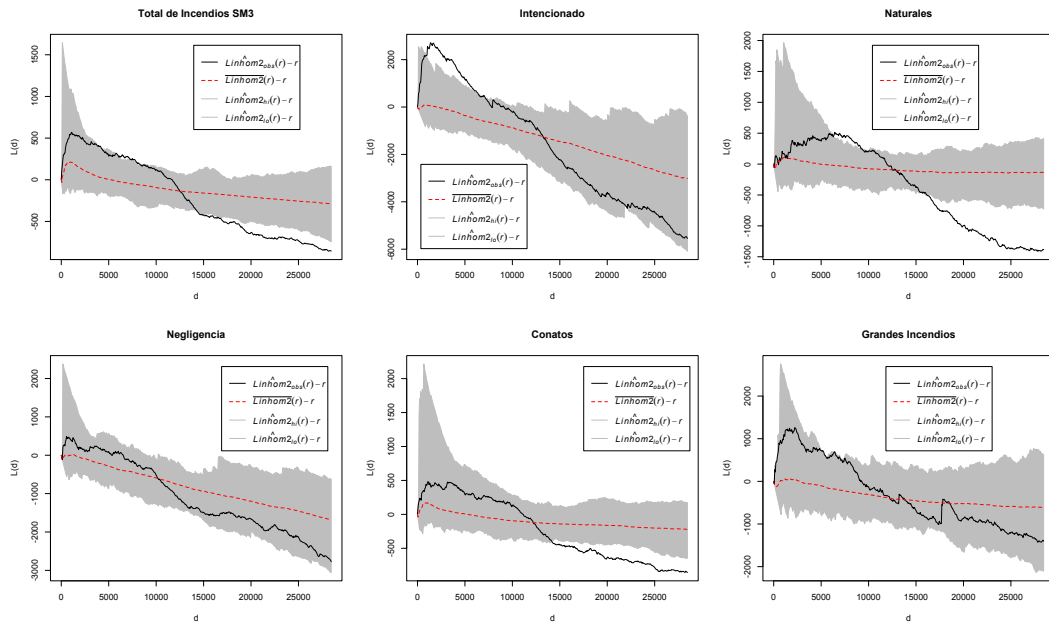


Figura 6.31: Funciones L para un modelo de Strauss ajustado con las coordenadas geográficas y las covariables para los diferentes patrones de estudio.

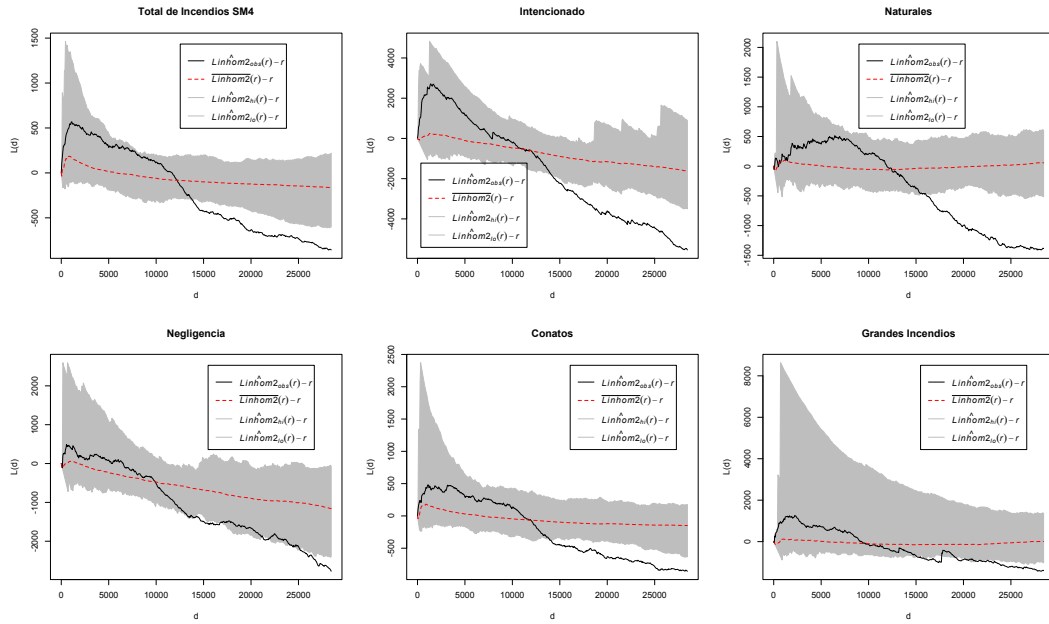


Figura 6.32: Funciones L para un modelo de Strauss ajustado con las covariables para los diferentes patrones de estudio.

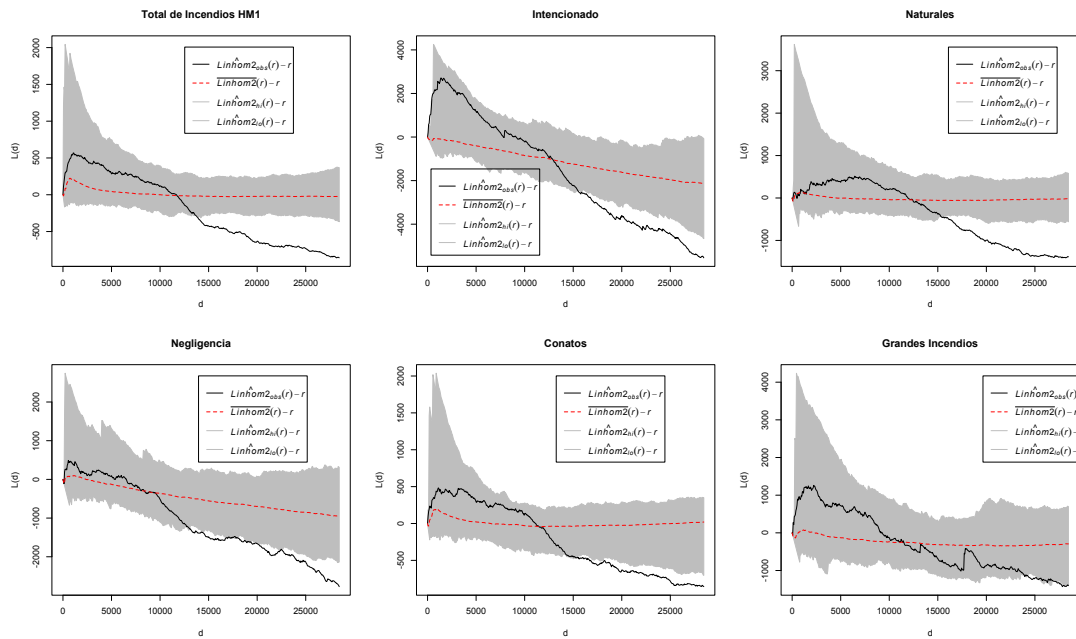


Figura 6.33: Funciones L para un modelo de Hard Core ajustado con las coordenadas geográficas para los diferentes patrones de estudio.

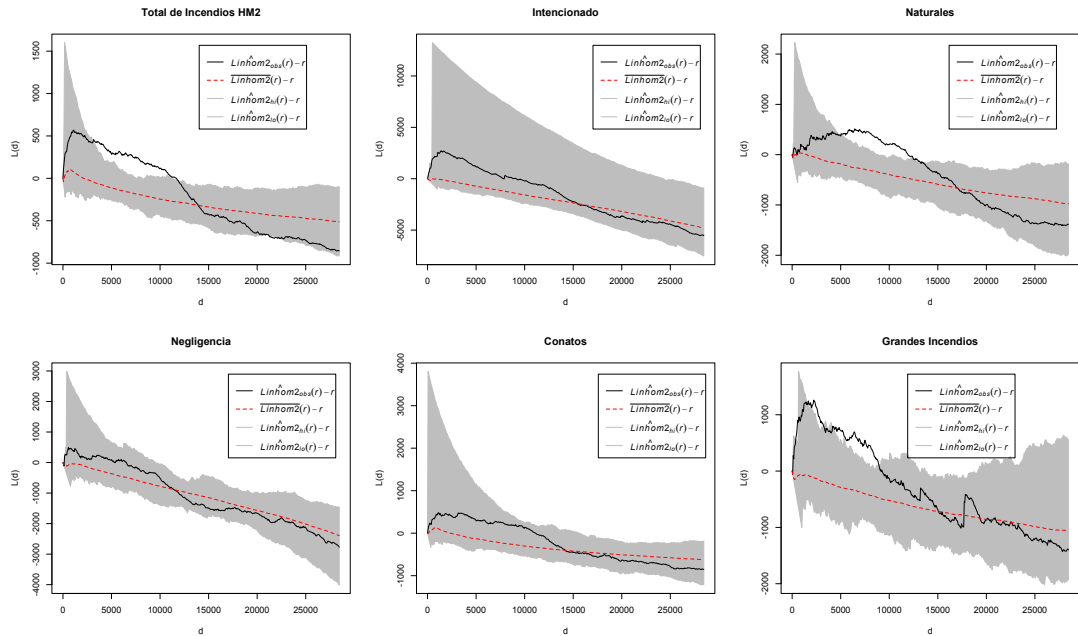


Figura 6.34: Funciones  $L$  para un modelo de Hard Core ajustado con un polinomio de grado 2 para los diferentes patrones de estudio.

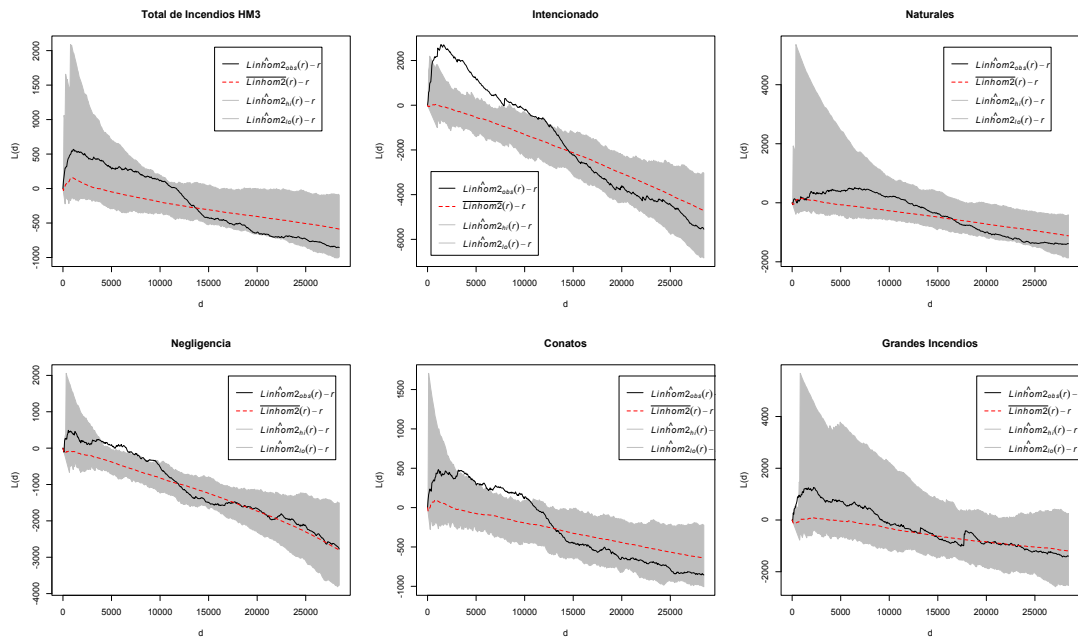


Figura 6.35: Funciones  $L$  para un modelo de Hard Core ajustado con un polinomio de grado 4 para los diferentes patrones de estudio.

### 6.1.4. Análisis de interacción mediante la k-cruzada.

En esta sección se presentan la interacción entre los diferentes patrones de estudio mediante la función k- cruzada.

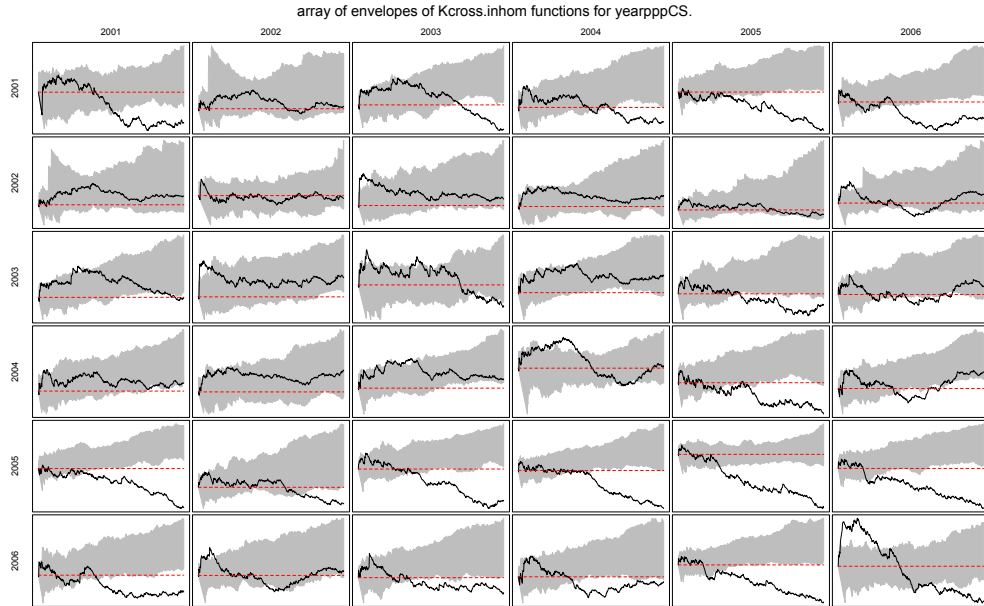


Figura 6.36: Función de k cruzada para los años 2001-2006.

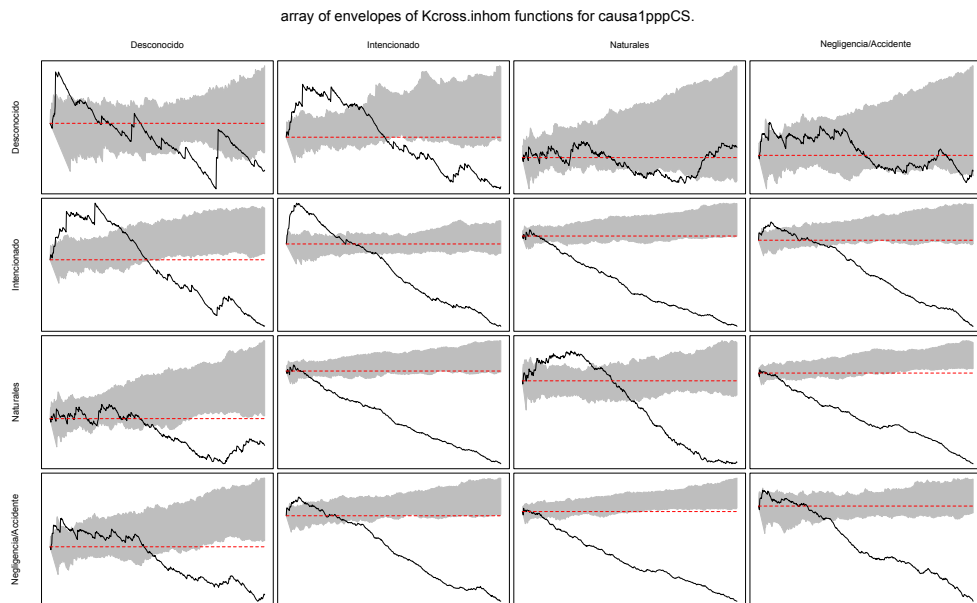


Figura 6.37: Función de k cruzada para las causas de origen de cada incendio.



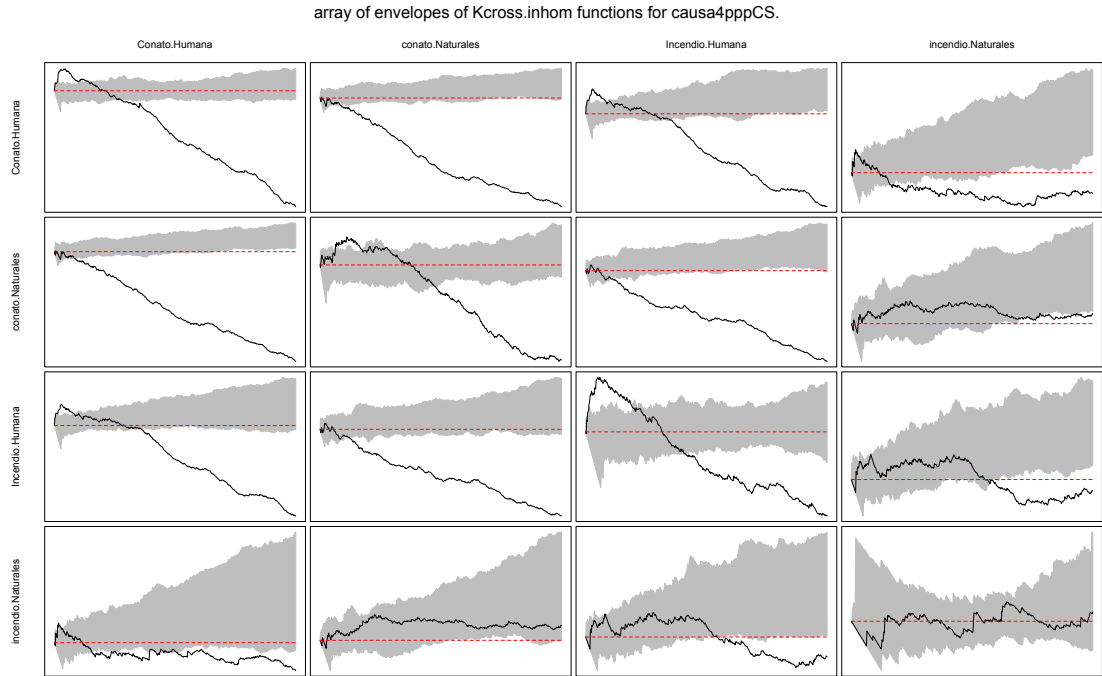


Figura 6.38: Función de k cruzada para los incendios mayores y menores a 1 ha para las causas humanas y naturales.

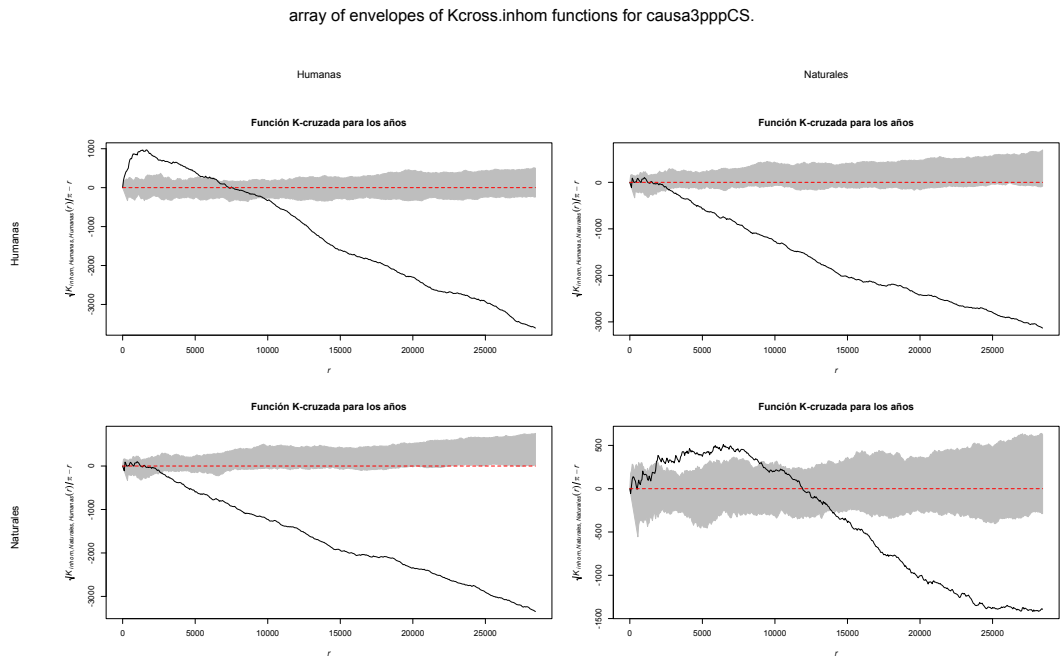


Figura 6.39: Función de k cruzada para las causas naturales y humanas.

array of envelopes of Kcross.inhom functions for tipppCS.

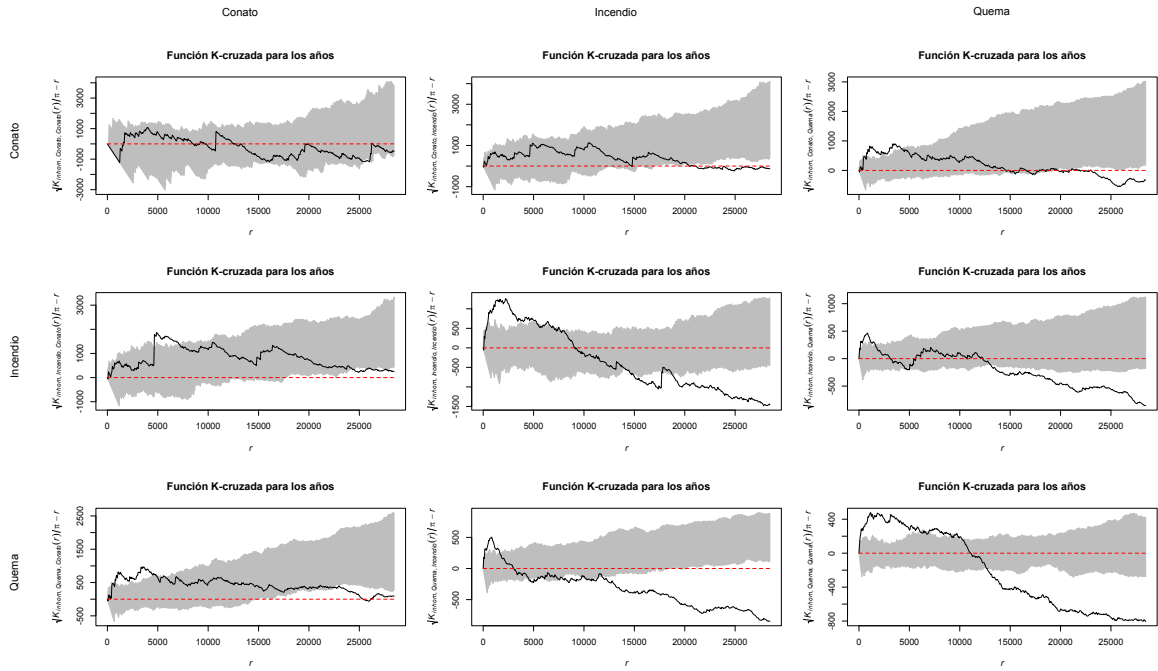


Figura 6.40: Función de k cruzada para los diferentes tamaños de incendios.

array of envelopes of Kcross.inhom functions for tipo2pppCS.

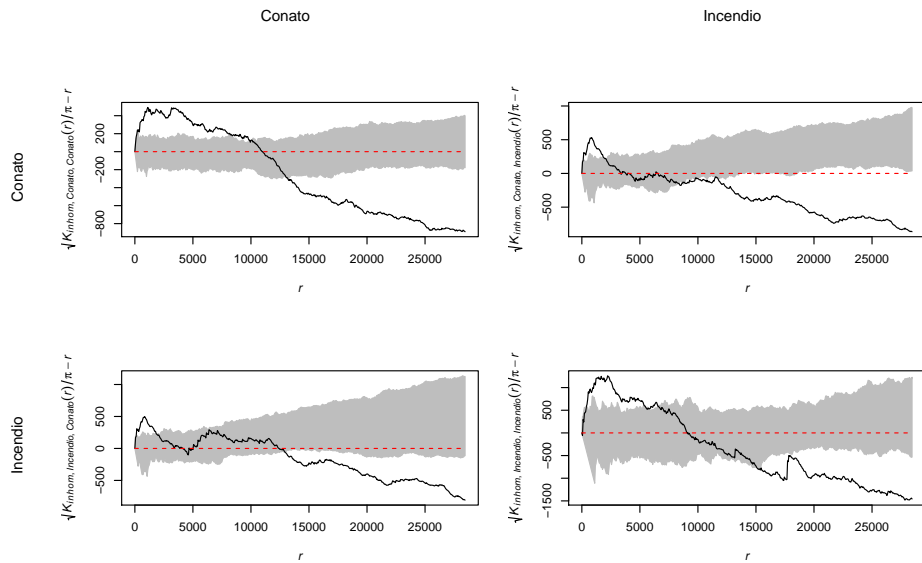


Figura 6.41: Función de k cruzada para los diferentes tamaños de incendios.

### 6.1.5. Análisis de dependencia entre tamaño y localización.

En esta sección se presentan las gráficas para analizar la presencia o falta de independencia entre el área quemada y la localización del incendio registrado.

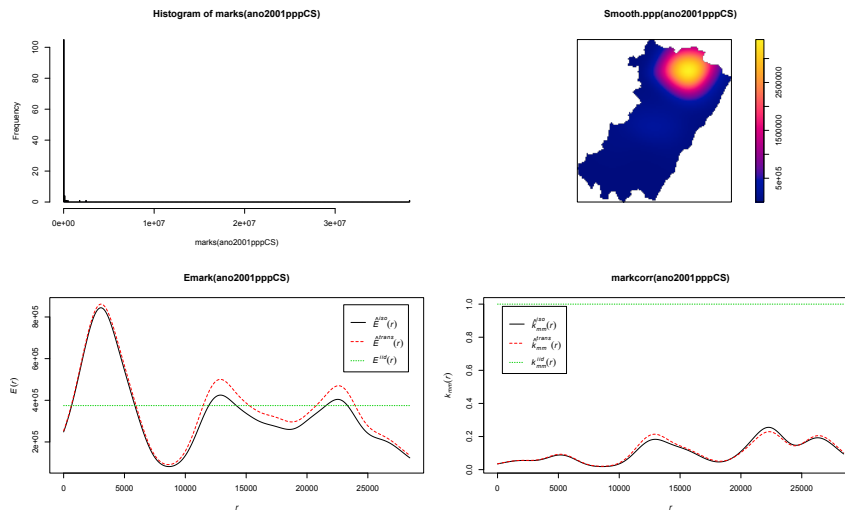


Figura 6.42: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2001.

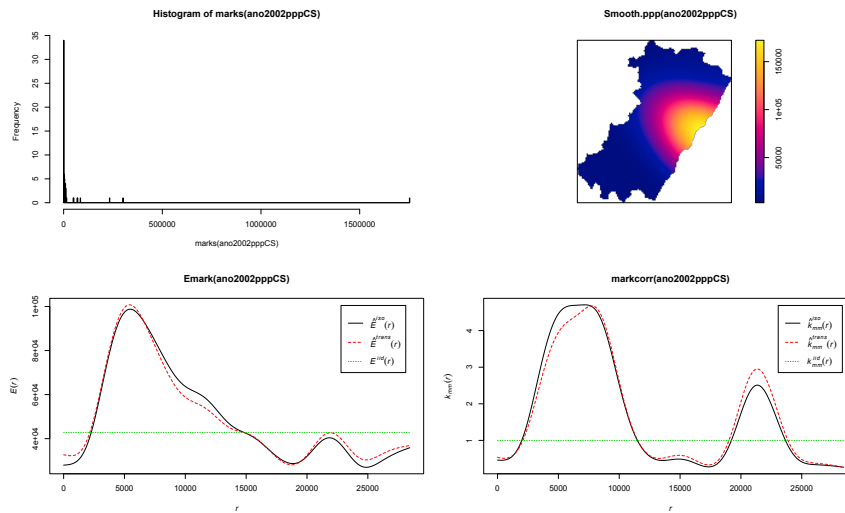


Figura 6.43: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2002.

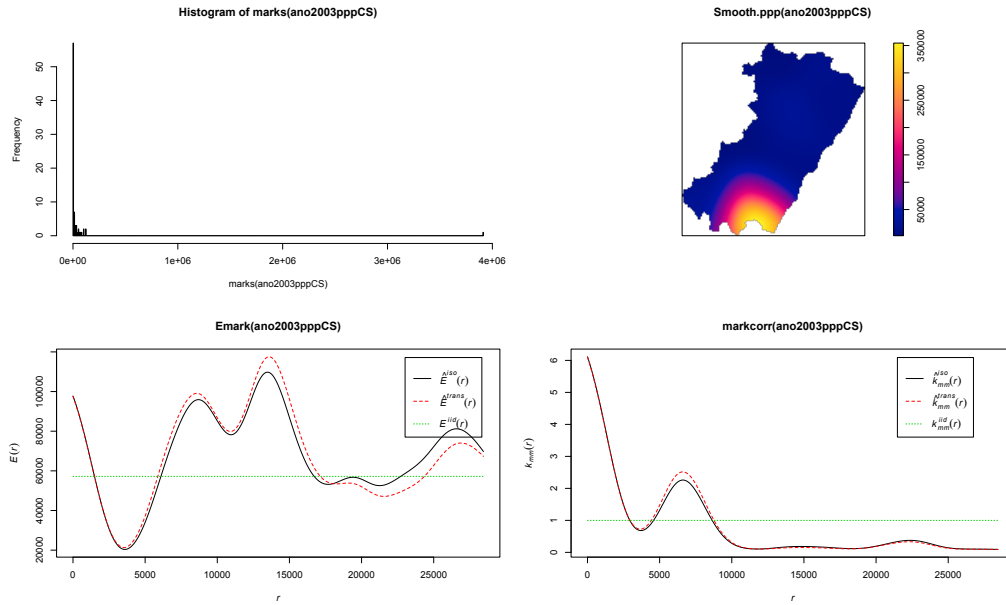


Figura 6.44: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2003.

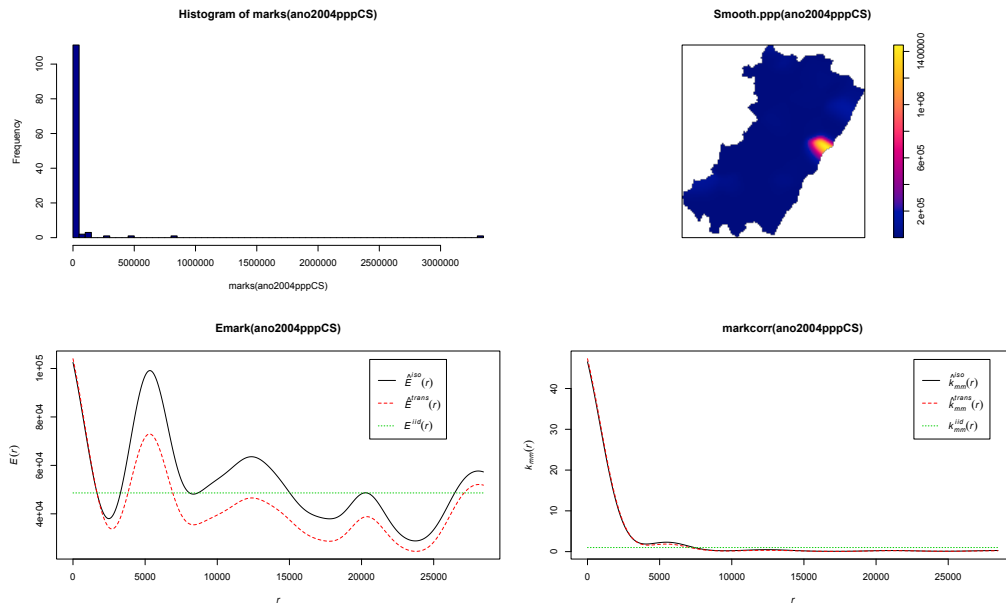


Figura 6.45: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2004.

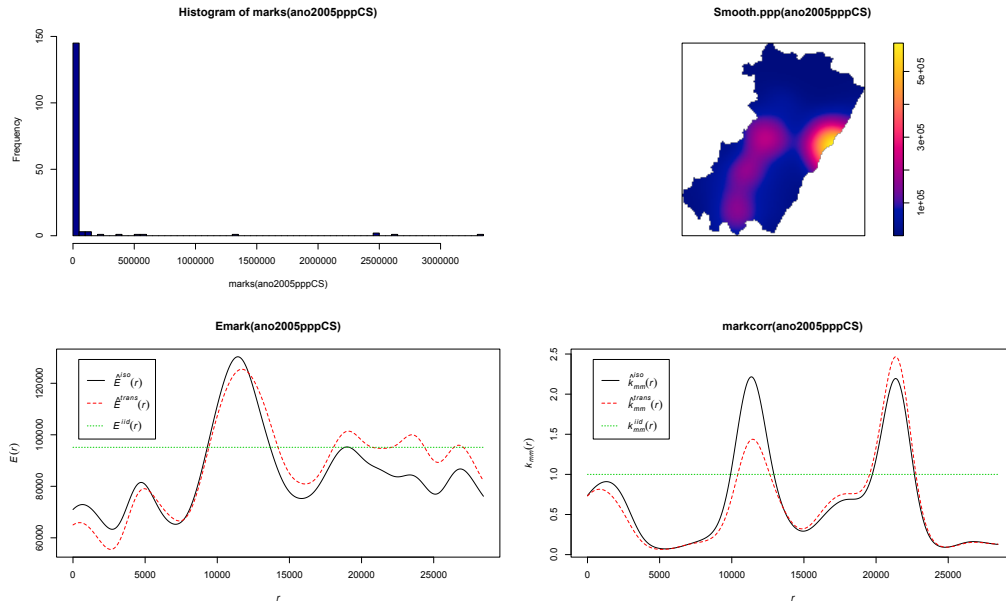


Figura 6.46: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2005.

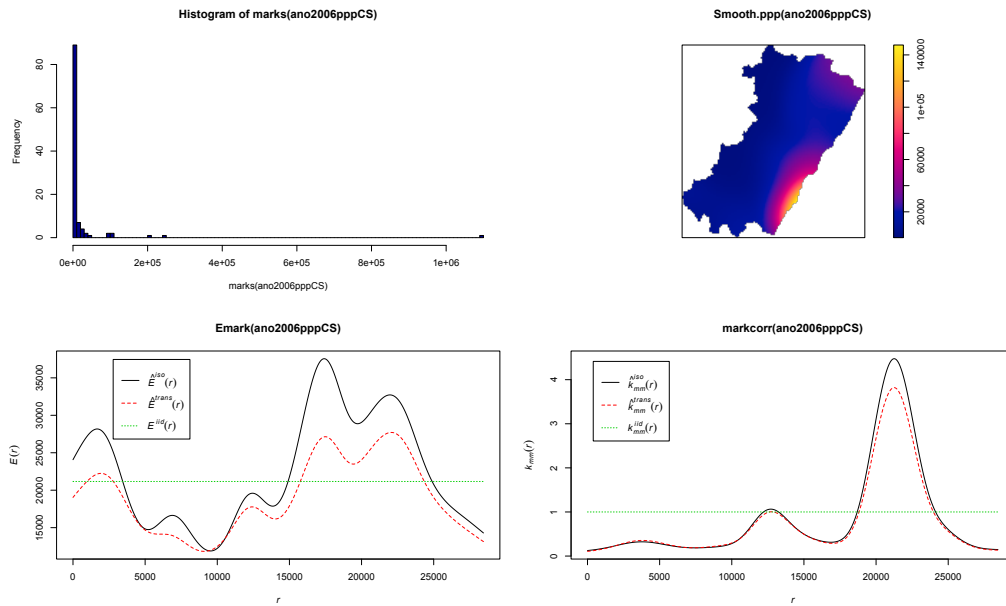


Figura 6.47: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas del año 2006.

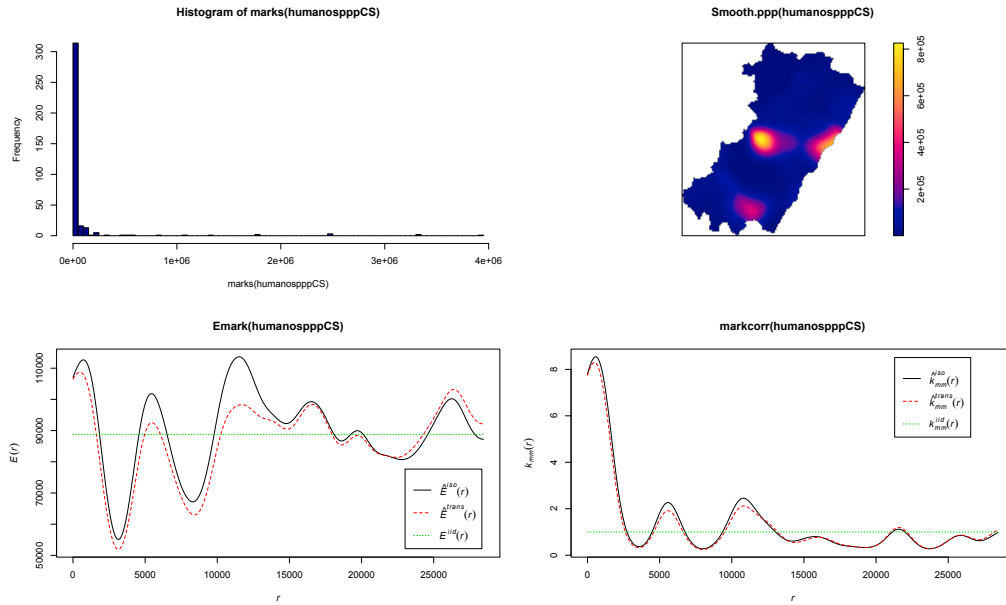


Figura 6.48: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas por causas humanas.

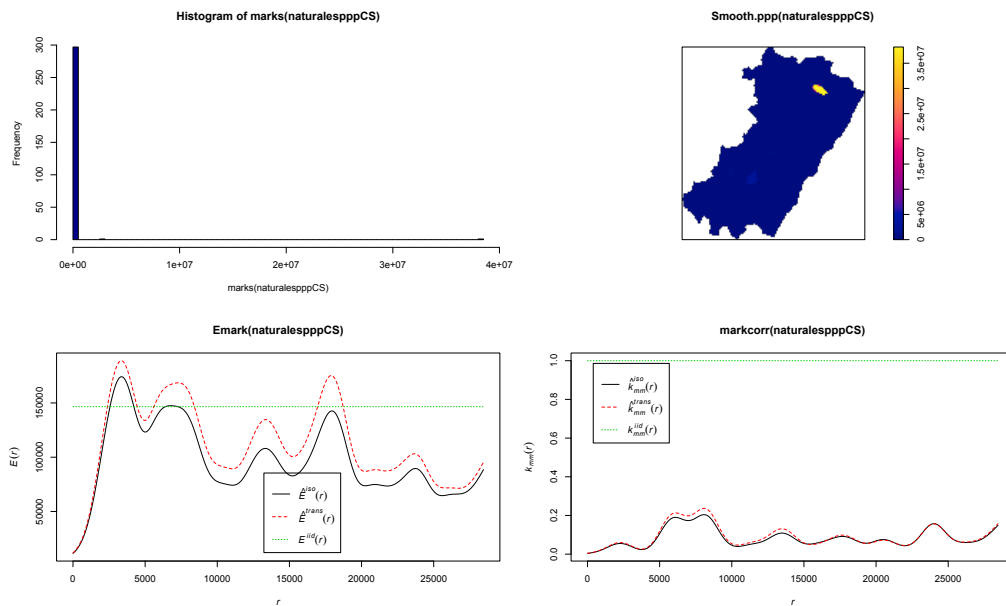


Figura 6.49: Histograma (superior izquierda), estimación de Kernel (superior derecha), función de correlación (inferior derecha) y la esperanza condicional inversa (inferior izquierda) para las marcas por causas naturales.

### 6.1.6. Mapas de predicción para los diferentes patrones ajustados.

En la siguiente sección se presentan los mapas de predicciones con los modelos ajustados para cada uno de los patrones: para el total de los incendios, los incendios por causas naturales, de negligencia intencionados y por tamaños.

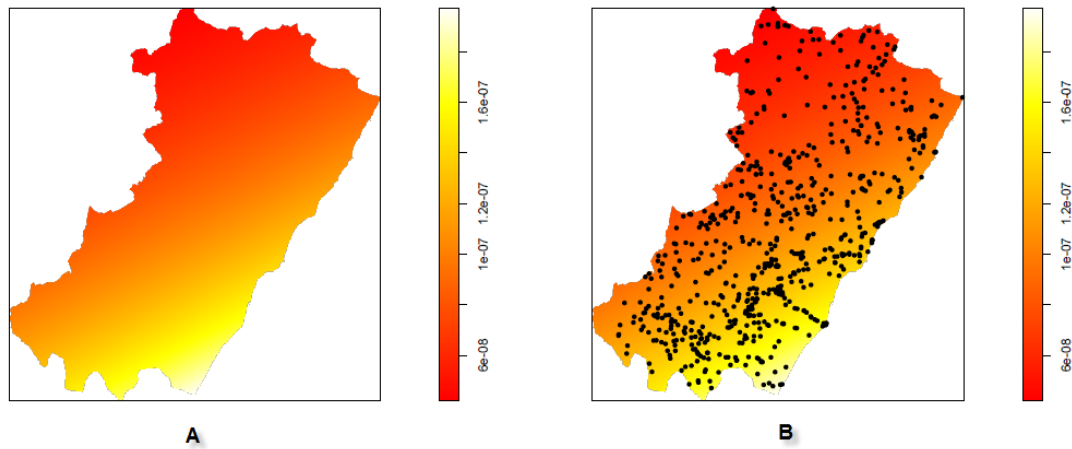


Figura 6.50: Mapa de predicciones ajustado con un modelo de Cauchy con coordenadas geográficas.

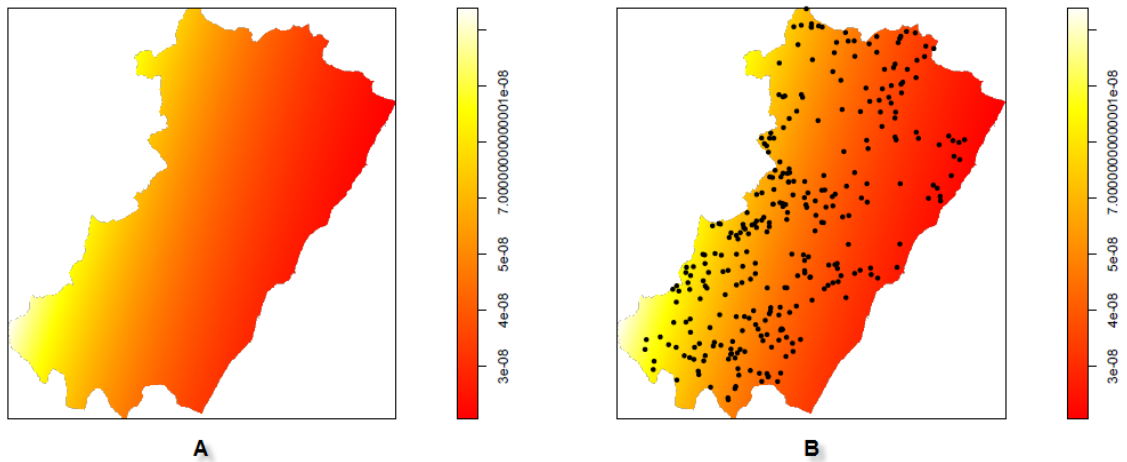


Figura 6.51: Mapa de predicciones ajustado con un modelo de Thomas con coordenadas geográficas.

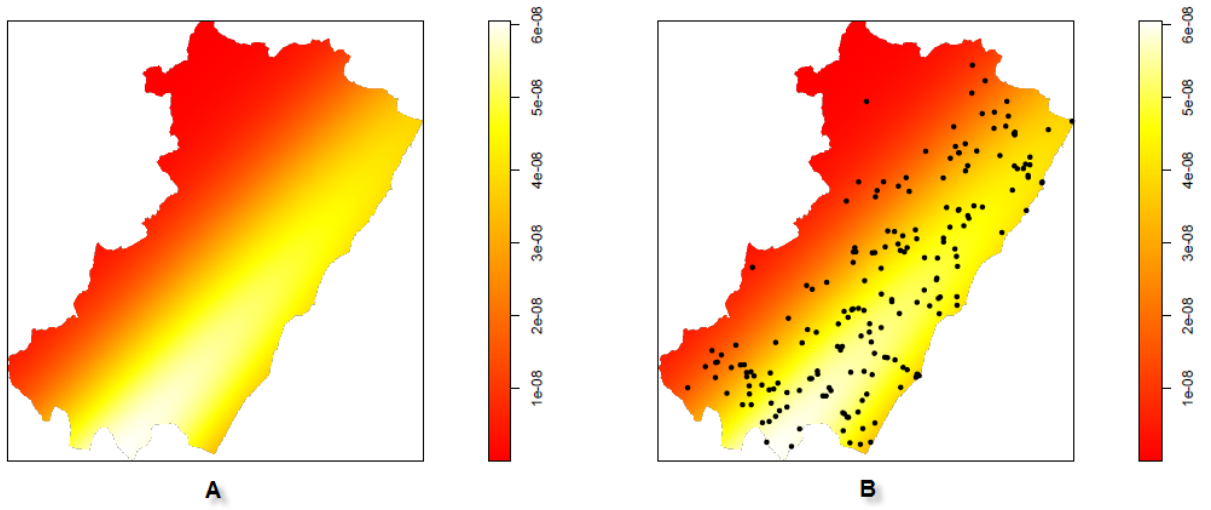


Figura 6.52: Mapa de predicciones ajustado con un modelo de Hard Core con un polinomio de grado 2.

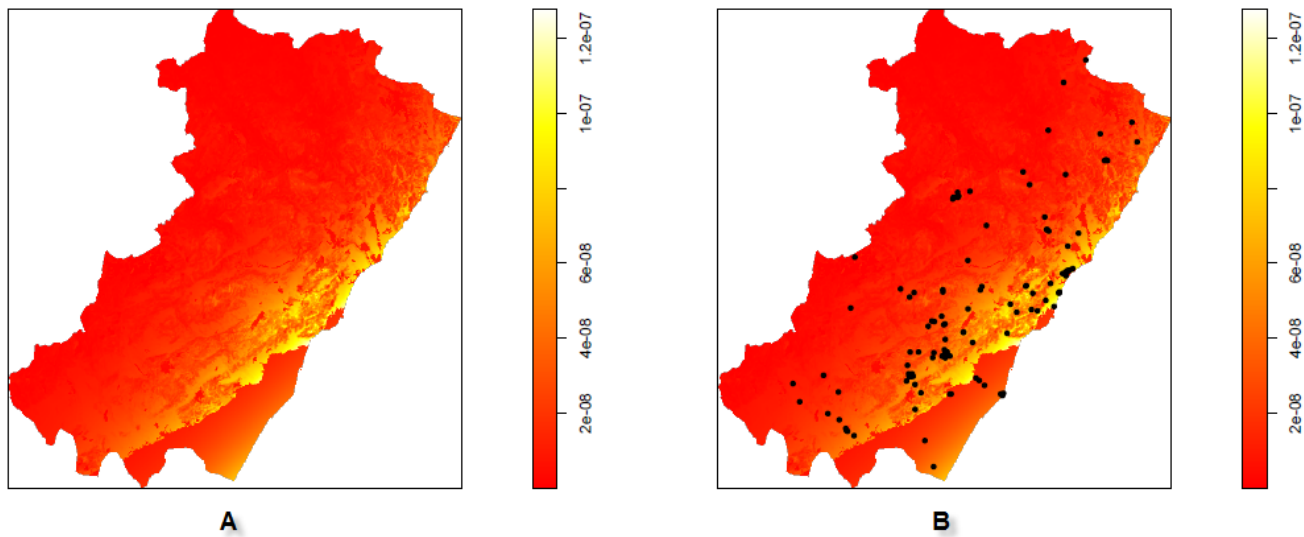


Figura 6.53: Mapa de predicciones ajustado con un modelo de Thomas con covariables seleccionadas.



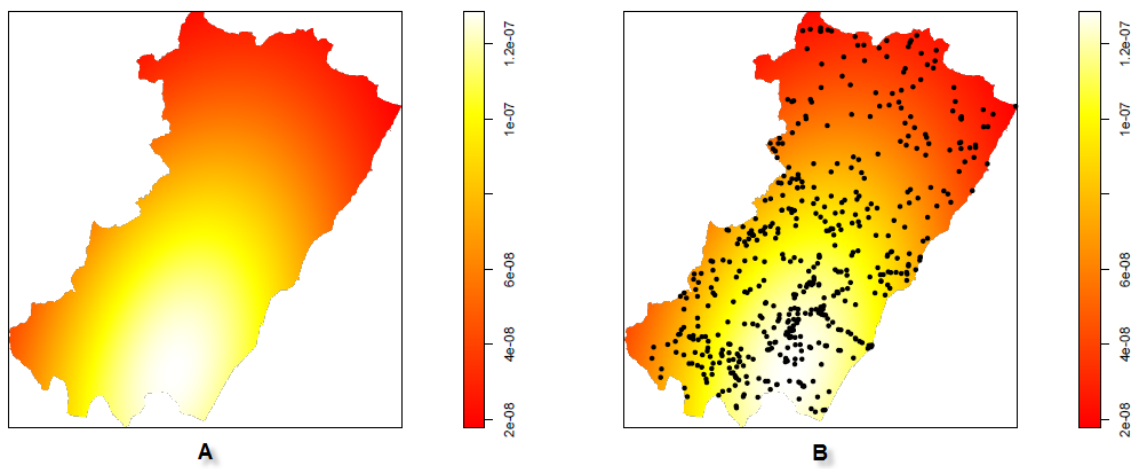


Figura 6.54: Mapa de predicciones ajustado con un modelo de Thomas con un polinomio de grado 2.

# Capítulo 7

## Apéndice B.

En el siguiente apéndice se muestra una parte de código en R donde se muestra como se cargan las bases de datos se construyen los patrones marcados y los patrones puntuales utilizados para el estudio así como la definición de las covariables, la estimación de kernel y de las propiedades de segundo orden, así como el ajuste de algunos modelos si se requiere mayor información del código servando.valdes.cruz@gmail.com

### 7.1. Código en R.

```
setwd("C:/Users/Tesis Servando")

library(TeachingDemos)
library(scatterplot3d)
library(rgl)
library(BiplotGUI)
library(spatstat)
library(splanCS)

#####
# Tipos de patrones (aleatorio, uniforme, agregado), Patrones marcados, #
# Covariables por tipo de causa y año #
#####

incendiosCS2<-read.table("incendiosCS0106.csv",sep=";",header=T)
contornoCS<-read.table("cast_cont.csv",sep=";",header=T)
regionCS_b<-read.table("region_cast_cov.csv",sep=";",header=T)
densidad<-read.table("covariables_densidad.csv",sep=";",header=T)
orientacionCS=read.table("orientacion_b.txt")
pendienteCS=read.table("pendiente.txt")
altitudCS=read.table("altura.txt")
```

```

landuseCS=read.table("landuse.txt")

polyowinCastellon=owin(poly=list(x=contornoCS$X[632:1],y=contornoCS$Y[632:1]))
yearpppCS=ppp(x=incendiosCS2$X,y=incendiosCS2$Y>window=polyowinCastellon,
marks=factor(incendiosCS2$year))
tipopppCS=ppp(x=incendiosCS2$X,y=incendiosCS2$Y>window=polyowinCastellon,
marks=factor(tipo))

#####
##Covariables##
#####

#####
##Densidad##
#####

vec.densidad<-densidad$hab.km2
vec.densidad[densidad$hab.km2<=-100]=NA
mapa.densidad=matrix(vec.densidad,600,600)
mapa.densidad<-im(mapa.densidad, xcol = seq(684100,803900,length = 600),
yrow = seq(4399300,4519100,length =600))
nueva.matriz=mapa.densidad$v
for(j in 1:600){
nueva.matriz[,j]=mapa.densidad$v[,601-j]
}
mapa.densidad$v=t(nueva.matriz)
plot(mapa.densidad, main="Densidad de Población")

#####
##Altura##
#####

vec.elevacion<-regionCS_b$altura
mapa.elevacion=matrix(vec.elevacion,600,600)
mapa.elevacion<-im(mapa.elevacion, xcol = seq(684100,803900,length = 600),
yrow = seq(4399300,4519100,length =600))
nueva.matriz=mapa.elevacion$v
for(j in 1:600){
nueva.matriz[,j]=mapa.elevacion$v[,601-j]
}
for(i in 1:600){
nueva.matriz[i,]=mapa.elevacion$v[601-i,]
}
mapa.elevacion$v=nueva.matriz
plot(mapa.elevacion)

```

```

#####
##tipo de patrones de estudio##
#####

#####
##Todos##
#####

par(mfrow=c(1,1))
incendioppCS=ppp(x=incendiosCS2$X,y=incendiosCS2$Y,
                 window=polyowinCastellon)
par(mfrow=c(1,1))
plot(density(incendioppCS),main="Incendios 2001-2006")
contour(density(incendioppCS),add= TRUE)
points(incendioppCS, pch=20)
plot(incendioppCS, pch=20,main="Incendios 2001-2006")

#####
##causas Naturales##
#####

selnaturales=incendiosCS2$codigo_causa==100|codigo_causa==600
naturalesCS=incendiosCS2[nelnaturales,]
naturalespppCS=ppp(x=naturalesCS$X,y=naturalesCS$Y,
                  window=polyowinCastellon)
plot(density(naturalespppCS),main="Causas Naturales")
contour(density(naturalespppCS),add= TRUE)
points(naturalespppCS, pch=20)

#####
##Propiedades de Primer Orden##
#####

#####
##Estimacion por Kernel##
#####

par(mfrow=c(2,3))
mse1<-bw.ppl(incendioppCS)
plot(mse1, main="Likelihood cross validation", xlim=c(0,15000))
plot(density(incendioppCS,mse1),main="Incendios 2001-2006")

mse2<-bw.ppl(naturalespppCS)
plot(mse2, main="Likelihood cross validation",xlim=c(0,15000))

```

```

plot(density(naturalespppCS,mse2),main="Naturales")

#####
##Propiedades de Segundo Orden##
#####

#####
##Función G##
#####

source("kinhom2.txt")

par(mfrow=c(2,3))
plot(envelope(incendioppCS, Gest), main="Total de Incendios")
plot(envelope(naturalespppCS, Gest), main="Naturales")

#####
##Función F##
#####

par(mfrow=c(2,3))
plot(envelope(incendioppCS, Fest), main="Total de Incendios")
plot(envelope(naturalespppCS, Fest), main="Naturales")

#####
##Función K##
#####

par(mfrow=c(2,3))
E <- envelope(incendioppCS, Kinhom2)
plot(E, main= "Total de Incendios")
E1 <- envelope(naturalespppCS, Kinhom2)
plot(E1, main="Naturales")

g <- pcf(incendioppCS)
g1 <- pcf(naturalespppCS)
g2 <- pcf(negligenciappCS)
g3 <- pcf(intencionadoppCS)
g4 <- pcf(conatoppCS)
g5 <- pcf(incgranpppCS)

plot(g, main= "Total de Incendios")
plot(g1, main="Naturales")
plot(g2, main= "Negligencia")
plot(g3, main= "Intencionado")

```

```

plot(g4, main= "Conatos")
plot(g5, main= "Grandes Incendios")

##Todos
fit02=kppm(Qincendiocov,~x + y, "Cauchy")
m02 <-envelope.kppm(fit02,Linhom2,nsim=99)
par(mfrow=c(1,2))
plot(m02, main="Modelo de Cauchy con coordenadas")
plot(m02, . - r ~ r, xlab="d", ylab="L(d)",
      main="Modelo de Cauchy con coordenadas")
predict.total2<-predict.kppm(fit02,ngrid=600,type="trend")
par(mfrow=c(1,2))
plot.im(predict.total2,main="Mapa de predicciones del
        total de incendios 2001-2006")
points(incendioppCS,pch=20)

##Naturales
fit25=kppm(Qnaturalescov, ~x + y, "Thomas")
m25 <-envelope.kppm(fit25,Linhom2,nsim=99)
par(mfrow=c(1,2))
plot(m25, main="Modelo de Thomas con coordenadas")
plot(m25, . - r ~ r, xlab="d", ylab="L(d)",
      main="Modelo de Thomas con coordenadas")
predict.natural<-predict.kppm(fit25,ngrid=600,type="trend")
par(mfrow=c(1,2))
plot.im(predict.natural,main="Mapa de predicciones
        de incendios Naturales de 2001-2006")
points(naturalesppCS,pch=20)

```

# Bibliografía

- [1] Abellán, J. J., Martínez-Beneito, M. A., Zurriaga, O., Jorques, G., Ferrándiz, J., & López-Quílez, A. (2002). Procesos puntuales como herramienta para el análisis de posibles fuentes de contaminación. *Gac Sanit*, 16(5), 445-449.
- [2] Arunachalam, V. (2005). Procesos puntuales, densidades producto y biología celular. *Revista Colombiana de Estadística*, 28(1), 1-16.
- [3] Asiain, A., & Segovia, E. (2009 ). *incendimetro 2009: Bosques en peligro frente al cambio climático*. España WWF.
- [4] Baddeley, A., Bárány, I., Schneider, R., & Weil, W. (2007). *Stochastic Geometry* Springer.
- [5] Baddeley, A., Moller, J., & Waagepetersen, R. P. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3). doi: 10.1111/1467-9574.00144
- [6] Baddeley, A., & Turner, R. (2000). Practical Maximum Pseudolikelihood for Spatial Point Patterns. *Australian & New Zealand Journal of Statistics*, 42(3), 283-322. doi: 10.1111/1467-842X.00128
- [7] Baddeley, A., & Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6), 1-42.
- [8] Baddeley, A., Turner, R., Moller, J., & Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Statistical Methodology*, 67(5), 617-666. doi: 10.1111/j.1467-9868.2005.00519.x
- [9] Butry, D. T., Gumpertz, M., & Genton, M. G. (2008). The production of large and small wildfires. *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, 79, 79.
- [10] Casella, G., & Berger, R. (2002). *Statistical Inference*. United States of America: Duxbury Thomson Learning.
- [11] Cox, D. R., & Isham, V. (1980). *Point Processes*. New York Chapman & Hall.
- [12] Cressie, N. A. (1993). *Statistics for Spatial*. Data New York Wiley.

- [13] Daley, D. J., & Vere-Jones, D. (2003). *An introduction to Theory of Point Processes (Vol. I: Elementary Theory and Methods.)* New York Springer.
- [14] Diggle, P. J. (1983). *Statistical analysis of spatial point patterns.* London: Academic Press.
- [15] Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns.* England UK: Chapman & Hall Book.
- [16] Dixon, P. M. (2002). *Ripley's K function (Vol. 3).* Chichester: John Wiley & Sons, Ltd.
- [17] Gaetan, C., & Guyon, X. (2010). *Spatial Statistics and Modeling.* New York: Springer-Verlag New York.
- [18] Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Transactions of the Institute of British Geographers*, 21(1), 246-274.
- [19] Gonzalez, J. R., Palahi, M., Trasobares, A., & Pukkala, T. (2006). A fire probability model for forest stands in Catalonia (north-east Spain). *Annals of Forest Science*, 63, 169-176. doi: dx.doi.org/10.1051/forest:2005109
- [20] Gregori, P., & Mateu, J. *Spatial Point Process: An Overview.* Mathematics. University Jaume I. Castellon Spain.
- [21] Hering, A. S., Bell, C. L., & Genton, M. G. (2009). Modeling spatio-temporal wildfire ignition point patterns. *Environmental and Ecological Statistics*, 16(2), 225-250. doi: 10.1007/s10651-007-0080-6.
- [22] Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models.* McGraw-Hill/Irwin: New York.
- [23] Larson, A. J., & Churchill, D. (2012). Tree spatial patterns in fire-frequent forest of western North America, including mechanisms of pattern formation and implications for designing fuel reduction and restoration treatments. *Forest Ecology and Management*, 267, 74-92. doi: 10.1016/j.foreco.2011.11.038.
- [24] Lopez-Abente, G., & Ibañez, C. (2001). *Aplicación de técnicas de análisis espacial a la mortalidad por cáncer en Madrid.* Madrid.
- [25] Martell, D. L., Otukol, S., & Stocks, B. J. (1987). A logistic model for predicting daily people-caused fire occurrence in Ontario, Canadian. *J. Forest Res.*, 17, 394-401.
- [26] Mateu, J., & Montes, F. (1998). Modelización de la distribución espacial de quistes en el estómago de la Marsopa mediante un proceso de Gibbs. *Qüestió*, 22(1), 175-194.



- [27] Moller, J., & C., D.-A. (2010). Structured Spatio-Temporal Shot-Noise Cox Point Process Models. *Scandinavian Journal of Statistics*, 37, 2–25. doi: 10.1111/j.1467-9469.2009.00670.x
- [28] Moller, J., & Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. United States of America Chapman & Hall/CRC.
- [29] Moritz, M. A. (2003). Spatiotemporal analysis of controls on shrubland and fire regimes: age dependency and fire hazard. *Ecology*, 84, 351-361.
- [30] Neter, J., Wasserman, W., & Kutner, M. (1983). *Applied Linear Regression Models*. Illinois: Richard D. Irwin, INC.
- [31] Oliver, C. D., & Larson, A. J. (1996). *Forest stand dynamics*. New York: McGraw Hill.
- [32] Pastor, E., Zárata, L., Planas, E., & Arnaldos, J. (2003). Mathematical models and calculation systems for the study of wildland fire behaviour. *Progress in Energy and Combustion Science*, 29(2), 139-153.
- [33] Pérez-Verdín, G., Márquez-Linares, M. A., Cortés-Ortiz, A., & Salmerón-Macías, M. (2013). Análisis espacio-temporal de la ocurrencia de incendios forestales en Durango, México. *Madera y Bosques*, 19(2), 37-58.
- [34] Rincón, L. (2007). *Curso Intermedio de Probabilidad*. Facultad de Ciencias UNAM: Las prensas de las Ciencias.
- [35] Rincón, L. (2012). *Introducción a los procesos estocásticos*. Facultad de Ciencias UNAM Las prensas de las ciencias.
- [36] Ripley, B. D. (1981). *Spatial Statistics*. New York Wiley.
- [37] Rodrigues, M., & De la Riva, J. (2012). *Modelado de la variación espacial de los factores explicativos de la causalidad humana en incendios forestales mediante Regresión Logística Ponderada Geográficamente*. Paper presented at the XV Congreso Nacional de Tecnologías de la Información Geográfica, Madrid.
- [38] Rodríguez-Trejo, D. A., & Fulé, P. Z. (2003). Fire ecology of mexican pines and fire management proposal. *International Journal of Wildlife Fire*, 12, 23-37.
- [39] Sanmartín, P. (2011). Estadística Espacial *ETISIT-UPCT*, 2, 133-141.
- [40] Todd, B., & Kourtz, P. H. (1991). Predicting the Daily Occurrence of People-Caused Forest Fire. Canada Petawa National Forestry Institute.
- [41] Turner, R. (2009). *Point patterns of forest fire locations*. Environmental and Ecological Statistics, 16(2), 197–223. doi 10.1007/s10651-007-0085-1
- [42] Valdés-Cruz, S., & Díaz-Avalos, C. (2016). *Spatial interactions of natural and human caused forest fires*. Paper presented at the METMA VIII Proceedings of the 8th International Workshop on Spatio-Temporal Modelling, Valencia Spain.