



UNIVERSIDAD
DE GRANADA

ESCUELA DE DOCTORADO DE CIENCIAS,
TECNOLOGÍAS E INGENIERÍAS

PROGRAMA DE DOCTORADO EN ESTADÍSTICA MATEMÁTICA Y APLICADA

Línea de Investigación en Teoremas límites funcionales
para campos aleatorios y procesos Hilbert-valuados

TESIS DOCTORAL

ESTIMACIÓN ESPACIO-TEMPORAL DE PROCESOS HILBERT-VALUADOS

Aplicación a la estimación y predicción funcional de mapas de riesgo de enfermedades

Antoni Torres Signes

Directoras de tesis:

María del Pilar Frías Bustamante

María Dolores Ruiz Medina

Tutora:

María Dolores Ruiz Medina

Coordinador:

Francisco De Asís Torres Ruiz

Granada, 2021

Editor: Universidad de Granada. Tesis Doctorales
Autor: Antoni Torres Signes
ISBN: 978-84-1306-891-6
URI: <http://hdl.handle.net/10481/69093>

Índice general

Portada	I
Índice general	VIII
Índice de figuras	IX
Índice de tablas	XVI
Abstract	XX
Resumen	XXV
Introducción	2
Objetivos	15
Objetivos generales	17
Estructura de la tesis	18
1. PRELIMINARES	27
1.1. Preliminares para el Capítulo 2	27
1.1.1. Estimadores de mínimo contraste y datos <i>tapered</i>	27
1.1.2. Definición general de estimadores de mínimo contraste	30
1.1.3. Datos y periodograma <i>tapered</i>	33
Esperanza y covarianza del periodograma	35
1.1.4. Ejemplo de estimación a partir de datos <i>tapered</i>	37
1.1.5. Sobre los estimadores Whittle con datos <i>tapered</i>	39
Caso de campos aleatorios débilmente dependientes	39
Caso de campos aleatorios fuertemente dependientes	41
1.2. Preliminares para los capítulos 3, 6 y 7	43
1.2.1. Procesos de Poisson doblemente estocásticos	43

1.2.2. Procesos de Poisson homogéneos en el espacio ℓ^2	46
Predicción clásica	46
1.2.3. El proceso de Wiener fuertemente cilíndrico inducido por un proceso de Wiener Hilbert-valuado	47
1.2.4. El proceso Ornstein-Uhlenbeck en espacios de Hilbert . . .	49
1.2.5. Procesos ARH(1)	50
1.3. Preliminares para el Capítulo 4 y 5	52
1.3.1. Estimación de momentos del proceso autorregresivo es- pacial de orden uno evaluado en un espacio de Hilbert . .	54
METODOLOGÍA Y RESULTADOS	59
2. Análisis de las propiedades asintóticas de estimadores espaciales de mínimo contraste basados en el periodograma tapered	60
2.1. Condiciones para el análisis de propiedades asintóticas	61
2.1.1. Teoremas sobre la consistencia y la normalidad asintótica de los estimadores	66
2.2. Demostraciones	68
2.3. Ejemplos numéricos	74
2.3.1. Procesos autorregresivos fraccionarios espaciales	75
2.3.2. Campos aleatorios de Gegenbauer espaciales	79
2.3.3. Campos aleatorios espaciales de Gegenbauer: singulari- dades en el origen	83
2.4. Teoremas centrales del límite	88
3. Procesos de Cox log-gaussianos temporales dirigidos mediante un proceso de Ornstein-Uhlenbeck Hilbert-valuado	98
3.1. Proyección del proceso O-U Hilbert-valuado	99
3.2. Aproximación ARH(1) en tiempo discreto del proceso O-U Hilbert- valuado	101
3.3. Estimación paramétrica y predicción del proceso O-U Hilbert- valuado	104
3.4. Proceso de Cox log-gaussiano en espacios de dimensión infinita .	105
3.5. Proceso de Cox log-gaussiano espacio temporal en el mapeo de enfermedades	110

3.5.1. Observación de la tasa aleatoria funcional de incidencia de la enfermedad	110
3.6. Estudio de simulación condicionada	111
3.6.1. Algoritmo de simulación condicionada para el enfoque de intensidad log-gaussiana con aproximación ARH(1)	115
3.6.2. Intensidad espacio-temporal con dependencia de largo rango (DLR)	117
Algoritmo de simulación condicionada DLR	122
4. Procesos de Cox espaciales en un marco infinito-dimensional	131
4.1. Procesos de Cox dirigidos por una log- intensidad aleatoria \mathcal{H} -valuada espacial	132
4.1.1. Predicción por mínimos cuadrados y momentos condicionados marginales de segundo orden	135
4.2. Estimación de la estructura de correlación espacial de la log-intensidad infinito-dimensional	137
4.3. Estudio de simulación	147
4.4. Ejemplo con datos reales	156
5. Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo	165
5.1. Procesos de Cox log-gaussianos espaciales de dimensión infinita	166
5.2. Análisis de segundo orden espacial a diferentes escalas temporales	168
5.3. Estimación funcional espacial multiresolución en el dominio espectral	171
5.3.1. Una ecuación de espacio de estados curva espacial	171
5.3.2. El enfoque de estimación	174
5.4. Estudio de simulación	179
5.5. Ejemplo de datos reales	188
5.5.1. Estimación multiescalar	190
5.5.2. Resultados de la validación	193

APLICACIONES A DATOS REALES	197
6. Una metodología funcional bayesiana para el mapeo del riesgo de dengue en América Latina y el Caribe	198
6.1. Métodos	200
6.1.1. El modelo	200
6.1.2. Modelo ARH(1)	202
6.1.3. Datos	203
6.2. Resultados	205
6.2.1. Una comparativa con enfoques previos	207
6.3. Discusión	210
6.4. Comentarios finales al capítulo	213
7. Análisis de la mortalidad por COVID-19 mediante curvas multivariantes de regresión a partir de datos blandos y mediante aprendizaje automático	216
7.1. Modelización de los datos	219
7.2. Aplicación del modelo de curvas de regresión y del análisis residual espacial	222
7.3. Análisis estadístico de la mortalidad por COVID-19	226
7.4. Un estudio empírico comparativo	232
7.4.1. Resultados de la validación cruzada aleatoria <i>k</i> -fold	233
7.4.2. Resultados de clasificación basados en <i>bootstrap</i>	237
7.5. Implementación de datos duros y blandos en modelos ML	243
7.6. Comentarios finales al capítulo	251
CONCLUSIONES Y LÍNEAS ABIERTAS	259
CONCLUSIONS AND OPEN LINES	266
BIBLIOGRAFÍA	272

Índice de figuras

1.1. Función <i>taper</i> de Tuckey-Hanning	38
2.1. Diagrama de cajas de valores muestrales de \hat{d}_1 y \hat{d}_2	78
2.2. Gráfico Q-Q de normalidad de \hat{d}_1 y \hat{d}_2	79
2.3. Probabilidades muestrales $P_0(\hat{\theta}_T - \theta_0 < \varepsilon)$	80
2.4. Diagrama de cajas de valores muestrales de $\hat{\sigma}_T^2$	80
2.5. Probabilidades muestrales $P_0(\hat{\sigma}_T^2 - \sigma^2(\theta) < \varepsilon)$	81
2.6. Diagrama de cajas de valores muestrales de \hat{d}_1 y \hat{d}_2	82
2.7. Gráficos Q-Q de normalidad de \hat{d}_1 y \hat{d}_2	83
2.8. Probabilidades muestrales $P_0(\hat{\theta}_T - \theta_0 < \varepsilon)$	84
2.9. Diagramas de valores muestrales de $\hat{\sigma}_T^2$	84
2.10. Probabilidades muestrales $P_0(\hat{\sigma}_T^2 - \sigma^2(\theta) < \varepsilon)$	85
2.11. Diagrama de cajas de los valores muestrales de \hat{d}_1 y \hat{d}_2 con $u_i = 1, i = \{1, 2\}$	86
2.12. Gráfico Q-Q de normalidad de \hat{d}_1 y \hat{d}_2 con $u_i = 1, i = \{1, 2\}$	86
2.13. Probabilidades muestrales $P_0(\hat{\theta}_T - \theta_0 < \varepsilon)$	87
2.14. Diagrama de cajas de valores muestrales de $\hat{\sigma}_T^2$	87
2.15. Probabilidades muestrales $P_0(\hat{\sigma}_T^2 - \sigma^2(\theta) < \varepsilon)$	88
3.1. Mapas de riesgo relativo de mortalidad por cáncer de prósta- ta: para una generación y su estimación, basada en la aproxi- mación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana	118
3.2. Mapas de riesgo relativo de mortalidad por cáncer de mama: pa- ra una generación y su estimación, basada en la aproximación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana	119

3.3. Mapas de riesgo relativo de mortalidad por cáncer de encéfalo: para una generación y su estimación, basada en la aproximación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana	120
3.4. Mapas de riesgo relativo de mortalidad por cáncer de próstata: para una generación y su estimación, basada en el modelo de intensidad log-gaussiana con DLR	125
3.5. Mapas de riesgo relativo de mortalidad por cáncer de mama: para una generación y su estimación, basada en el modelo de intensidad log-gaussiana con DLR	126
3.6. Mapas de riesgo relativo de mortalidad por cáncer de encéfalo: para una generación y su estimación, basada en el modelo de intensidad log-gaussiana con DLR	127
3.7. Enfoque O-U log-gaussiano. Mapas de España con el log-riesgo de mortalidad por tipo de cáncer: de próstata (año 1975), de mama (año 1975), y de encéfalo (año 1986); en base a una generación de un proceso O-U y sus respectivas estimaciones con el filtrado de Kalman	128
3.8. Enfoque log-gaussiano con DLR. Mapas de España con el log-riesgo de mortalidad por tipo de cáncer: de próstata (año 1975), de mama (año 1975), y de encéfalo (año 1986); en base a una generación del proceso de Cox con intensidad log-gaussiana con DLR y sus respectivas estimaciones de mínimos cuadrados de DLR	129
4.1. Diagramas de cajas de los valores muestrales de $\hat{\theta}_N$, para $\theta_0 = 1$ basados en 100 generaciones, con el error cuadrático medio empírico y desviación la típica	151
4.2. Diagramas de cajas de los valores muestrales de $\hat{\theta}_{1,1,N}$, para $\theta_{1,1,0} = 1$, basados en 100 generaciones, con el error cuadrático medio empírico y la desviación típica	153
4.3. Diagramas de cajas de los valores muestrales de $\hat{\theta}_{2,1,N}$, para $\theta_{2,1,0} = 1.5$, basados en 100 generaciones con el error cuadrático medio empírico y desviación típica	153

4.4.	Diagramas de cajas de los valores muestrales de $\widehat{\theta}_{1,2,\mathcal{N}}$, para $\theta_{1,2,0} = 1.6$, basados en 100 generaciones con el error cuadrático medio empírico y desviación típica	154
4.5.	Diagramas de cajas de los valores muestrales de $\widehat{\theta}_{2,2,\mathcal{N}}$, para $\theta_{2,2,0} = 1.2$, basados en 100 generaciones con el error cuadrático medio empírico y desviación típica	154
4.6.	Diez proyecciones normalizadas de la log-intensidad residual funcional espacial	159
4.7.	Estimaciones de las proyecciones normalizadas de la log-intensidad residual funcional espacial	159
4.8.	Gráfico de espectros puntuales estimados de $L_{i,\widehat{\theta}_{i,\mathcal{N}}}$, $i = 1, 2, 3$, para $M = 10$, y $\mathcal{N} = 400$	160
4.9.	Curvas de log-intensidad de mortalidad por enfermedades respiratorias originales, y sus respectivas estimaciones, para las 24 primeras provincias españolas por orden alfabético	161
4.10.	Curvas de log-intensidad de mortalidad por enfermedades respiratorias originales, y sus respectivas estimaciones, para las 24 últimas provincias españolas por orden alfabético	161
4.11.	Zoom. Provincias españolas con mayor y menor norma $L^1([0, 1725])$ del error relativo absoluto medio funcional asociado, con las curvas originales y estimadas	162
4.12.	Mapas de riesgo de mortalidad por enfermedad respiratoria observada y estimada, promediados anualmente, durante el período 1980–2015	163
5.1.	Escala 10, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30	181
5.2.	Escala 9, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30	181
5.3.	Escala 8, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30	182
5.4.	Escala 7, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30	182
5.5.	Errores cuadráticos medios empíricos (EMSE), asociados con $\{\widehat{\lambda}_{N,p,1}, \widehat{\lambda}_{N,p,2}, p = 1, \dots, k_N\}$, $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$	184

5.6. Diagrama de cajas con los valores muestrales de $\widehat{\lambda}_{N,p,i}$, $p = 1, \dots, 10, i = 1, 2$ de izquierda a derecha, y de arriba a abajo), basados en 100 generaciones de muestras funcionales de tamaño $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$	185
5.7. Verdadero L_1, L_2 y sus respectivas estimaciones multiescalares, para las escalas $j = 7, 8, 9, 10$, sobre una rejilla regular espacial 30×30	186
5.8. Campo X de log-intensidad espacial original y estimado, para el tiempo $t = 1/2$, a través de las escalas $j = 7, 8, 9, 10$, sobre una rejilla regular espacial 10×10 , a partir de los datos curva suavizados	187
5.9. Campo X de log-intensidad espacial original y estimado, para el tiempo $t = 1/2$, a través de las escalas $j = 7, 8, 9, 10$, sobre una rejilla regular espacial 30×30 , a partir de los datos curva suavizados	187
5.10. Datos temporales y espaciales interpolados en una rejilla regular espacial 20×20	188
5.11. Datos temporales y espaciales interpolados en una rejilla regular espacial 20×20 a escala (nivel de resolución) 7	189
5.12. Datos curva suavizados mediante B-splines en una rejilla regular espacial 20×20	189
5.13. Estimación multiescalar del operador L_1 y L_2 a través de escalas (niveles de multiresolución) $j = 7, 8, 9, 10$	190
5.14. Gráficos de contorno del campo de log-intensidad observado y estimado para los valores mensuales $t = 108$ y $t = 216$. Tanto los valores observados como los estimados para los tiempos $t = 108$ y $t = 216$ se muestran a través de las escalas $j = 7, 8, 9, 10$, en el sistema wavelet de Haar, a partir de los datos interpolados temporales y espaciales sobre una rejilla regular 20×20	191
5.15. Gráficos de contorno del campo de log-intensidad observado y estimado para los valores mensuales $t = 108$ y $t = 216$. Tanto los valores observados como los estimados para los tiempos $t = 108$ y $t = 216$ se muestran a través de las escalas $j = 7, 8, 9, 10$, en el sistema wavelet de Haar, a partir de los datos temporales interpolados y suavizados sobre una rejilla regular 20×20	192

5.16. Promedio anual del número de muertes por enfermedades respiratorias observadas y estimadas en cada una de las 48 provincias españolas desde enero de 1980 hasta diciembre de 2015	194
6.1. Evolución del número de DF observado desde 1998 hasta 2018 en los países bajo estudio	204
6.2. Log-riesgo estimado a partir de la estimación bayesiana en el modelo propuesto desde 1998 hasta 2018 en los países bajo estudio	208
6.3. Log-riesgo estimado a partir de la estimación clásica, bayesiana y el log-riesgo observado desde 1998 hasta 2018 para los países: Argentina, México, Brasil, Guyana Francesa, Costa Rica y Venezuela	208
6.4. Log-riesgo estimado a partir de los modelos espacio-temporales tradicionales, la metodología de estimación y el log-riesgo observado, desde 1998 hasta 2018, para los países: Barbados y Paraguay	209
7.1. Curvas acumulada, riesgo y log-riesgo de mortalidad por COVID-19 en España, desde el 8 de marzo de 2020 hasta el 13 de mayo de 2020, e intervalos de confianza <i>bootstrap</i>	228
7.2. Valores muestrales de la media espacial del riesgo mínimo empírico en la regresión trigonométrica, histograma <i>bootstrap</i> y densidad de probabilidad <i>bootstrap</i> de la media espacial del riesgo mínimo empírico de regresión, basado en 1000 muestras <i>bootstrap</i>	230
7.3. Proyecciones empíricas del operador de autocorrelación ρ , que reflejan la autocorrelación temporal y la correlación cruzada entre las 17 comunidades españolas analizadas., y comparación de la densidad de probabilidad a-priori considerada de una variable aleatoria con distribución Beta escalada por el factor $1/3$, con parámetros de forma con valores 14 y 13, con el ajuste <i>bootstrap</i> a priori de una variable empírica	231

7.4. Mapas de riesgo de mortalidad por COVID–19, desde el 8 de marzo hasta el 13 de mayo de 2020, observados y estimados, calculados a partir de la regresión trigonométrica, combinados con predictores residuales clásicos y bayesianos	232
7.5. Categoría de datos duros. A partir de 1000 muestras <i>bootstrap</i> , se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para GRNN, MLP y SVR lineal	239
7.6. Categoría de datos duros. A partir de 1000 muestras <i>bootstrap</i> , se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para BNN, RBF y GP .	240
7.7. Categoría de datos blandos. A partir de 1000 muestras <i>bootstrap</i> , se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para GRNN, MLP y SVR no lineal	241
7.8. Categoría de datos blandos. A partir de 1000 muestras <i>bootstrap</i> , se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para BNN, RBF y GP	242
7.9. Categoría de datos blandos. A partir de 1000 muestras <i>bootstrap</i> , se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para la regresión trigonométrica combinada con la predicción residual, basada en los momento empíricos, clásica y bayesiana	243
7.10. Categoría de datos duros. Curvas de log-riesgo de mortalidad por COVID–19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), y Procesos Gaussianos (GP)	252

7.11. **Categoría de datos blandos.** Curvas de log-riesgo de mortalidad por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), Procesos Gaussianos (GP), y regresión trigonométrica combinada con predicción residual clásica y bayesiana 252

7.12. **Categoría de datos duros.** Curvas de mortalidad acumulada por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), y Procesos Gaussianos (GP) 253

7.13. **Categoría de datos blandos.** Curvas de mortalidad acumulada por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), Procesos Gaussianos (GP), y regresión trigonométrica combinada con predicción residual clásica y bayesiana 253

Índice de tablas

2.1. Media, desviación estándar y ECM de \hat{d}_1 y \hat{d}_2	78
2.2. Media, desviación estándar y ECM de \hat{d}_1 y \hat{d}_2	82
2.3. Media, desviación estándar y ECM de \hat{d}_1 y \hat{d}_2	85
3.1. Promedio espacial de los errores cuadráticos medios empíricos por año, obtenido a partir de 100 generaciones de la muestra funcional de tamaño T del proceso de intensidad ARH(1), para los datos generados de tres tipos de cáncer	116
3.2. Estimaciones paramétricas con DLR	121
3.3. Promedio espacial de los errores cuadráticos medios empíricos por año, obtenidos a partir de 100 generaciones de la muestra funcional de tamaño T del modelo de intensidad aleatoria con DLR, para tres tipos de datos de cáncer generados	123
4.1. Estimaciones de los parámetros de localización. Tamaños de muestra, media, desviación típica, y errores cuadrático medio empíricos, a partir de 100 generaciones	151
4.2. Estimaciones de los parámetros de escala y localización. Tamaños de muestra, media, desviación típica, y errores cuadrático medio empíricos, a partir de 100 generaciones	155
4.3. Estimaciones de espectros puntuales de $L_{i,\hat{\theta}_{i,\mathcal{N}}}$, $i = 1, 2, 3$ para $M = 10$, $\mathcal{N} = 400$	160
4.4. Promedio puntual anual del CVFARE obtenido a partir de la validación	162
5.1. Autovalores λ_{p1} , λ_{p2} , $p = 1, \dots, 10$	179

5.2.	Promedio por escala de los errores cuadráticos medios empíricos asociados a los estimadores paramétricos implicados en la aproximación de L_1 , a partir de 100 generaciones de las muestras funcionales de tamaño $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$. Se consideran las escalas $j = 6, 7, 8, 9$, y el parámetro de truncamiento $k_N = 10$	183
5.3.	Promedio por escala de los errores cuadráticos medios empíricos asociados a los estimadores paramétricos implicados en la aproximación de L_2 , a partir de 100 generaciones de las muestras funcionales de tamaño $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$. Se consideran las escalas $j = 6, 7, 8, 9$, y el parámetro de truncamiento $k_N = 10$	183
5.4.	Promedio puntual anual de los errores de validación cruzada	193
6.1.	Errores absolutos empíricos Promediados para el riesgo relativo de DF según el promedio de los países bajo estudio, para la estimación bayesiana y clásica, (BEAEA) y (CEAEA), respectivamente	206
6.2.	Errores Cuadráticos Empíricos Promediados para el riesgo relativo de DF según el promedio de los años bajo estudio, para la estimación bayesiana y clásica	206
6.3.	Comparación del Error Absoluto Medio Empírico (EMAE) y el Error Cuadrático Medio Empírico (EMSE) desde 1998 hasta 2018 para todos los países, según las cinco metodologías aplicadas	209
7.1.	Estimaciones de los parámetros de regresión $\hat{A}_k(\cdot)$, $k = 1, \dots, 6$, en las 17 comunidades españolas	227
7.2.	Estimaciones de los parámetros de regresión $\hat{B}_k(\cdot)$, $k = 1, \dots, 6$, en las 17 comunidades españolas	227
7.3.	Valores observados $L_{265}(\hat{\theta}_{265}(p))$, $p = 1, \dots, 17$	229
7.4.	Intervalos de confianza <i>bootstrap</i> para $\overline{L_{265}}$ (nivel de confianza $1 - \alpha = 0.95$)	230
7.5.	Intervalos de confianza <i>bootstrap</i> para el error estándar de entrenamiento esperado de los predictores de log-riesgo de mortalidad por COVID-19 clásico y bayesiano ($1 - \alpha = 0.95$)	231

7.6. Categoría de datos duros. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 10-fold	235
7.7. Categoría de datos blandos. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 10-fold	235
7.8. Nuestro enfoque. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 10-fold de los análisis residuales clásico y bayesiano probados	236
7.9. Categoría de datos duros. Intervalos de confianza <i>bootstrap</i> ($1 - \alpha = 0.95$) para los SMAPE promediados espacialmente a partir de 1000 muestras <i>bootstrap</i> ($T = 265, P = 17$)	237
7.10. Categoría de datos blandos. Intervalos de confianza <i>bootstrap</i> ($1 - \alpha = 0.95$) para los SMAPE promediados espacialmente a partir de 1000 muestras <i>bootstrap</i> ($T = 265, P = 17$)	238
7.11. Categoría de datos duros. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold	250
7.12. Categoría de datos blandos. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold	251
7.13. Categoría de datos duros. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold	254
7.14. Categoría de datos blandos. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold	255

Abstract

The application of techniques for the analysis of functional data correlated in time and/or space is a relatively recent area of research, where a number of problems have arisen and remain open. In particular, the derivation of probabilistic (point processes in function spaces) and statistical (functional spatial and time series) models is required for the analysis of high-dimensional data that often exhibit complex correlation structures in time and/or space. Point processes are used to explain the distribution of points generated by random mechanisms in time and/or space. Such processes allow to model and analyze the incidence or mortality associated with a disease.

In this thesis, we have considered the context of doubly stochastic counting processes or Cox processes. In particular, an infinite-dimensional statistical approach, based on functional linear models, has been adopted for the statistical description of the random log-intensity.

The spatio-temporal dynamics of these models are analyzed through temporal or spatial processes with values in an appropriate function space. The complexity of these models, given the high dimension of the parameter space (on many occasions we work with infinite-dimensional spaces), makes it essential to implement appropriate dimension reduction techniques, as well as the implementation of model selection procedures.

From the theoretical point of view, in the following chapters, new scenarios are introduced in order to apply different estimation methodologies. On the one hand, log-Gaussian Cox processes in Hilbert spaces with random intensity given by an Ornstein-Uhlenbeck process approximated by an autoregressive Hilbertian process (ARH) are developed. These temporal patterns are analyzed from a time-correlated functional data perspective. On the other hand, Cox processes driven by linear infinite-dimensional spatial log-intensities are developed. In this case, these spatial patterns are analyzed from a spatially correlated fun-

ctional data perspective.

Regarding the methodological approaches adopted for the estimation, in the case of Cox processes driven by an O-U Hilbert-valued log-intensity, approximated by an ARH(1) process, the method of empirical moments has been used. In the case of spatial Cox processes driven by an infinite-dimensional spatial linear random log-intensity, to estimate the parameters modeling the parametric structure of the spectral density operator, under the condition of spatial stationarity, functional spectral techniques based on the periodogram operator extending the Whittle functional have been applied. As a preliminary analysis, we contribute, in the case of spatially stationary real-valued spatial processes, to obtaining sufficient conditions that guarantee the consistency and asymptotic normality of minimum-contrast estimators based on the *tapered* periodogram.

Specifically, in this thesis, from the perspective of infinite-dimensional Cox processes, or Cox processes driven by infinite-dimensional linear log-intensities, not necessarily Gaussian, within the field of functional statistical analysis of point patterns in time and/or space, the following contributions have been established:

- Study of consistency and asymptotic normality of minimum-contrast estimators in spatial processes.
- Introduction of the class of temporal log-Gaussian Cox processes with random log-intensity defined by an Ornstein-Uhlenbeck Hilbert-valued process.
- Approximation of the Ornstein-Uhlenbeck Hilbert-valued processes by ARH(1) processes, using the estimation from the method of empirical moments and calculation of the associated *plug-in* predictor.
- Introduction of a new class of Cox processes driven by a linear Hilbert-valued log-intensity. Here, the log-Gaussian process condition, or Gaussian log-intensity, is not required. Neither is it required in the introduction, nor for the consistency result, that the log-intensity is SARH(1). It is only considered in that way in the simulation and application.
- Introduction of new estimation techniques by minimum componentwise contrast for the previously introduced family of processes (in particular, with SARH intensity).

- Development of conditions guaranteeing the strong consistency of the proposed estimators.
- Fitting linear and non-linear trend models in an infinite-dimensional statistical framework for spatio-temporal log-risk processes of disease incidence and mortality. Residual linear correlation in an autoregressive Hilbertian process framework, under a Bayesian approach.
- Comparison, via cross-validation and bootstrapping techniques, of the presented approaches with regression or prediction models based on *machine learning*.

Epidemiology and the study in general of the evolution, both spatial and temporal, of several diseases has been the fundamental framework considered for the contributions indicated. Specifically, real data have been used for the estimation and functional prediction in time and space of prostate, breast and brain cancer, as well as respiratory diseases, in Spanish provinces, from annual or monthly observations, over periods ranging around thirty years. Furthermore, by implementing the techniques presented throughout the thesis, an application to real data has been carried out to analyze the incidence of a disease in a foreign territory. In particular, the evolution of dengue fever in American countries in recent years has been modeled. On the other hand, given the social emergency situation caused by the COVID-19 pandemic in the last stage of development of the thesis, it has been considered pertinent to include a statistical study on the estimation of the spatio-temporal evolution of the mortality risk, as well as of the daily mortality cases caused by this disease in the Autonomous Communities, which allows reflecting, among other aspects, the effect of the first state of alarm on the behavior of this evolution. In this way, the daily mortality due to COVID-19 in the Spanish Communities during the first wave, specifically from 8 March 2020 to 13 May 2020, has been modeled.

The latter practical applications have been developed, based on the infinite-dimensional statistical techniques proposed in the development of the thesis, under a classical and Bayesian approach, with modifications in the estimation methodology. Subsequently, in both cases, an empirical comparison has been made with other approaches. In the case of the risk of annual incidence of dengue fever in American countries, a comparison has been made with traditional spatio-temporal models, including a Leroux model, an Intrinsic Condi-

tional Autoregressive model and a Besag, York and Mollie model. In the case of daily mortality risk by COVID-19 in the Spanish Autonomous Communities, the proposed approaches have been compared with another methodology based on the estimation by confidence intervals and probability densities using *bootstrap* techniques, as well as with a battery of models in the context of *Machine Learning*, including Generalized Regression Neural Networks, Multilayer Perceptron, Support Vector Regression, Bayesian Neural Networks, Neural Networks from Radial Function Bases, and Gaussian Processes. In addition, model selection in the context of parametric non-linear regression is addressed.

Resumen

La aplicación de técnicas de análisis de datos funcionales correlados en el tiempo y/o espacio constituye un área de investigación relativamente reciente, donde surgen diversos problemas que aún permanecen abiertos. En particular, se requiere la deducción de modelos probabilísticos (procesos puntuales en espacios de funciones) y estadísticos (series espaciales y temporales funcionales), para el análisis de datos de dimensión elevada que suelen presentar estructuras complejas de correlación en el tiempo y/o espacio. Los procesos puntuales se utilizan para explicar la distribución de los puntos generados por mecanismos aleatorios en el tiempo y/o espacio. Dichos procesos permiten modelizar y analizar la incidencia o mortalidad asociados a una enfermedad.

En esta tesis, se ha considerado el contexto de los procesos de recuento doblemente estocásticos o procesos de Cox. En particular, se ha adoptado un enfoque estadístico infinito-dimensional, basado en modelos lineales funcionales, para la descripción estadística de la log-intensidad aleatoria.

La dinámica espacio-temporal de estos modelos se analiza mediante procesos temporales o espaciales, con valores en un espacio de funciones apropiado. La complejidad de dichos modelos, dada la elevada dimensión del espacio de parámetros (en bastantes ocasiones se trabaja con espacios infinito-dimensionales), hace imprescindible la implementación de técnicas apropiadas de reducción de la dimensión, así como la implementación de procedimientos de selección de modelos.

Desde el punto de vista teórico, en los siguientes capítulos, se introducen nuevos escenarios para poder aplicar diferentes metodologías de estimación. Por un lado, se desarrollan los procesos de Cox log-gaussianos en espacios de Hilbert con intensidad aleatoria dada por un proceso Ornstein-Uhlenbeck que se aproxima mediante un proceso autorregresivo hilbertiano (ARH). Estos patrones temporales se analizan desde una perspectiva de datos funcionales co-

relacionados en el tiempo. Por otro lado, se desarrollan los procesos de Cox dirigidos por log-intensidades espaciales infinito-dimensionales lineales, en este caso, estos patrones espaciales se analizan desde una perspectiva de datos funcionales correlados en el espacio.

En cuanto a los enfoques metodológicos adoptados para la estimación, en el ámbito de procesos de Cox dirigidos por una log-intensidad O-U Hilbert-valuada, aproximada mediante un proceso ARH(1), se ha utilizado el método de los momentos empíricos. En el caso de procesos de Cox espaciales dirigidos mediante una log-intensidad aleatoria lineal espacial infinito-dimensional, para la estimación de los parámetros que modelizan la estructura paramétrica del operador de densidad espectral, bajo la condición de estacionariedad espacial, se han aplicado técnicas espectrales funcionales basadas en el operador periodograma que extienden el funcional de Whittle. Como análisis preliminar, se contribuye, en el caso de procesos espaciales real-valorados estacionarios en el espacio, con la obtención de condiciones suficientes que garantizan la consistencia y normalidad asintótica de estimadores de mínimo contraste, basados en el periodograma *tapered*.

De forma concreta, en esta tesis, a partir de una perspectiva de procesos de Cox infinito-dimensionales, o bien, procesos de Cox dirigidos por log-intensidades lineales infinito-dimensionales, no necesariamente gaussianas, dentro del ámbito del análisis estadístico funcional de patrones puntuales en el tiempo y/o espacio, se han establecido las siguientes contribuciones:

- Estudio de consistencia y normalidad asintótica de los estimadores de mínimo contraste para procesos espaciales.
- Introducción de la clase de procesos temporales de Cox log-gaussianos con log-intensidad aleatoria definida por un proceso Ornstein-Uhlenbeck Hilbert-valuado.
- Aproximación de los procesos Ornstein-Uhlenbeck Hilbert-valorados mediante procesos ARH(1), utilizando la estimación a partir del método de momentos empíricos y cálculo del predictor *plug-in* asociado.
- Introducción de una nueva clase de procesos de Cox dirigidos mediante una log-intensidad lineal Hilbert-valuada. Aquí, la condición de proceso log-gaussiano, o de log-intensidad gaussiana, no es necesaria. Tampoco

se requiere en la introducción, ni para el resultado de consistencia, que la log-intensidad sea SARH(1). Sólo se considera de esta forma en la simulación y aplicación.

- Introducción de nuevas técnicas de estimación por mínimo contraste componente a componente para la familia de procesos anteriormente introducida (en particular, con intensidad SARH).
- Desarrollo de las condiciones que garantizan la consistencia fuerte de los estimadores propuestos.
- Ajuste de modelos de tendencia lineal y no lineal en un marco estadístico infinito-dimensional para procesos espacio-temporales de log-riesgo de incidencia y mortalidad en enfermedades. Análisis de los residuos de regresión mediante un enfoque autorregresivo hilbertiano en el contexto bayesiano.
- Comparación, mediante validación cruzada y técnicas *bootstrap*, de los enfoques presentados con modelos de regresión o predicción basados en aprendizaje automático.

La epidemiología y el estudio en general de la evolución, tanto espacial como temporal, de diversas enfermedades ha sido el marco fundamental considerado para plasmar las contribuciones indicadas. En concreto, se han utilizado datos reales para la estimación y predicción funcional en el tiempo y en el espacio del cáncer de próstata, mama y encéfalo, así como enfermedades respiratorias, en las provincias españolas, a partir de observaciones anuales o mensuales, en periodos que oscilan en torno a treinta años. Además, mediante la implementación de técnicas vistas a lo largo de la tesis, se ha llevado a cabo una aplicación a datos reales, para el análisis de la incidencia de una enfermedad en territorio extranjero. En particular, se ha modelizado la evolución de fiebre de dengue en países americanos durante los últimos años. Por otro lado, dada la situación de emergencia social provocada por la pandemia de COVID-19 en la última etapa de desarrollo de la tesis, se ha considerado pertinente incluir un estudio estadístico sobre la estimación de la evolución espacio-temporal del riesgo de mortalidad, así como de los casos de mortalidad diaria ocasionados por dicha enfermedad en las comunidades autónomas, que

permite reflejar, entre otros aspectos, el efecto del primer estado de alarma sobre el comportamiento de dicha evolución. De esta manera, se ha modelizado la mortalidad diaria por COVID-19, en las comunidades españolas, durante la primera ola, en concreto, desde el 8 de marzo de 2020 hasta el 13 de mayo de 2020.

Estas últimas aplicaciones prácticas se han desarrollado, a partir de las técnicas estadísticas infinito-dimensionales propuestas en el desarrollo de la tesis, bajo un enfoque clásico y bayesiano, con modificaciones en la metodología de estimación. Posteriormente, en ambos casos se ha realizado una comparativa empírica con otros enfoques. En el caso del riesgo de incidencia anual por fiebre del dengue en países americanos, se ha realizado una comparativa con modelos espacio-temporales tradicionales, incluyendo un modelo de Leroux, un modelo Autorregresivo Condicionado Intrínseco y otro modelo de Besag, York y Mollie. En el caso del riesgo de mortalidad diario por COVID-19 en las comunidades autónomas españolas, los enfoques propuestos se han comparado con otra metodología basada en la estimación por intervalos de confianza y densidades de probabilidad mediante técnicas *bootstrap*, así como con una batería de modelos en el contexto de *Machine Learning*, incluyendo Redes Neuronales de Regresión Generalizada, Perceptrón Multicapa, Regresión de Soporte Vectorial, Redes Neuronales Bayesianas, Redes Neuronales a partir de Bases de Funciones Radiales, y Procesos Gaussianos. Además se aborda la selección de modelos en el contexto de la regresión no lineal paramétrica.

Introducción

La estimación paramétrica de procesos estocásticos estacionarios y campos aleatorios a partir de datos correlacionados temporalmente o espacialmente, constituye un tema central en la inferencia estadística. Dentro de este campo de estudio, el método de estimación por máxima verosimilitud (EMV) es una de las herramientas de estimación más populares. El problema de la estimación paramétrica de procesos integrados fraccionarios con componentes estacionales se trató en [Reisen et al. \(2006\)](#). Para estimar los parámetros fraccionarios, ellos proponen varios estimadores de regresión mediante el log-periodograma con diferentes anchos de banda seleccionados alrededor o entre las frecuencias estacionales.

La misma metodología se utilizó en [Li y McLeod \(1986\)](#) para procesos autorregresivos de medias móviles diferenciados fraccionalmente en el contexto de series temporales estacionarias. También se han hecho varias contribuciones para la EMV de los procesos espaciales de memoria larga; véase, por ejemplo, [Anh y Lunney \(1995\)](#). Para los datos espaciales bidimensionales, el artículo de [Basu y Reinsel \(1993\)](#) introdujo un modelo espacial unilateral autorregresivo de medias móviles de primer orden, ARMA(1,1). Para implementar la EMV proporcionaron un tratamiento adecuado a los valores de las celdas fronterizas, con un efecto sustancial en la estimación de los parámetros. La estimación de máxima verosimilitud gaussiana, en el contexto de los modelos ARMA, se aplicó en [Yao y Brockwell \(2006\)](#).

Los funcionales lineales y no lineales del periodograma desempeñan un papel clave en el diseño de técnicas de Estimación de Mínimo Contraste (EMC); véase, por ejemplo, [Taniguchi \(1987\)](#). El procedimiento de estimación de Whittle se ha tratado ampliamente en la literatura; véase [Guyon \(1982, 1995\)](#); [Ludeña y Lavielle \(1999\)](#), para el caso de campos aleatorios. Se ha prestado especial atención a la estimación paramétrica de mínimo contraste de modelos fraccio-

narios, fractales y con dependencia de largo rango; véase, por ejemplo, Chung (1996). En particular, WeiLin et al. (2012) aplicaron la estimación paramétrica de mínimo contraste para aproximar el parámetro de desviación del proceso de O-U cuando la correspondiente ecuación diferencial estocástica viene dada por el movimiento browniano fraccionario.

La consistencia y normalidad asintótica del estimador de máxima verosimilitud de Whittle para los procesos autorregresivos de medias móviles integrados fraccionalmente estacionales y estacionarios (SARFIMA) se ha demostrado en Guo et al. (2009). La maximización de la verosimilitud de Whittle también se ha considerado en diversos autores (Boissy et al., 2005; Chan y Tsai, 2012). En Robinson y Sanz (2006), se consideran modificaciones a la estimación de Whittle de modelos multilaterales sobre rejilla. En Leonenko y Taufer (2013), se ha estudiado la estimación de la densidad espectral para campos aleatorios gaussianos agregados, con posible dependencia fuerte, a partir de una expansión en términos de polinomios ortogonales de Gegenbauer. Estos campos se construyen a partir de la acumulación de campos aleatorios de memoria corta independientes y con idéntica distribución, a través de una densidad mixta desconocida (a estimar).

Una alternativa a la familia de funcionales lineales de Whittle se propuso en Ibragimov (1967), véase también Anh et al. (1999, 2004a,b, 2007a,c,d); Espejo et al. (2015), entre otros. Un desarrollo de estos métodos de estimación en el espectro se presenta en el Capítulo 2, a partir del periodograma *tapered*. Los resultados de dicho capítulo sirven de base para la extensión al caso funcional que se estudia en el Capítulo 4. En el contexto de los modelos lineales en espacios funcionales, motivado por el análisis estadístico de datos de grandes dimensiones, se puede consultar Bosq (2000); Bosq y Blanke (2007); Ferraty y Vieu (2006); Ramsay y Silverman (2005), entre otros. Un tratamiento completo de la estimación basada en el método de los momentos y la proyección en bases ortogonales de funciones se puede ver en Antoniadis y Spatinas (2003); Bosq (2000); Bosq y Blanke (2007); Damon y Guillas (2005); Soltani y Hashemi (2011). En el contexto de series autorregresivas espaciales hilbertianas destacan los trabajos de Ruiz-Medina y Espejo (2012, 2013).

En las dos últimas décadas, ha habido un desarrollo creciente de técnicas que permiten el análisis estadístico de datos funcionales (curvas, superficies, etc.) que están correlados en el tiempo o espacio, es decir, con una perspectiva

de series temporales o espaciales funcionales. En este sentido, en Bosq (2000), se desarrolla la estimación paramétrica de series funcionales que obedecen una ecuación de estados mientras que Ferraty y Vieu (2006) extienden el análisis estadístico no paramétrico al contexto infinito-dimensional. Paralelamente, ha habido un desarrollo creciente de los métodos o técnicas de la estadística espacial, formulados en el ámbito de la estadística infinito-dimensional, para el análisis de curvas correladas en el espacio, ver, por ejemplo, Delicado et al. (2010); Giraldo et al. (2010); Ignaccolo et al. (2014). Cabe mencionar Ruiz-Medina (2011) y Ruiz-Medina (2012) que abordan el problema de modelización y estimación de series espaciales de curvas, bajo un enfoque de ecuaciones de estados autorregresivos. En Hörmann y Kokoszka (2010, 2012, 2013); Hörmann et al. (2018), se desarrolla el análisis de datos funcionales correlados en el tiempo o espacio.

Una introducción, que incluye aspectos básicos sobre el análisis de datos funcionales, así como resultados relativamente recientes, se recogen en Ferraty y Romain (2010); Hsing y Eubank (2015); Ramsay y Silverman (2005), véase también Goia y Vieu (2016); Horváth et al. (2013); Kidziński et al. (2018); Kokoszka et al. (2010), con contribuciones en análisis de datos funcionales no correlados y correlados. Hasta el momento, según nos consta (desde el análisis bibliográfico realizado), no se han introducido procesos de recuento que modelicen la distribución de puntos o patrones puntuales aleatorios en el tiempo y espacio desde una perspectiva funcional. Según nuestro conocimiento actual, solamente se tiene constancia de la definición de proceso de Poisson homogéneo con valores en ℓ^2 , en Bosq y Ruiz-Medina (2014). Este hecho ha motivado el desarrollo de las contribuciones que aparecen en los capítulos 3, 6 y 7, para poder realizar el análisis estadístico de dichos patrones puntuales desde la perspectiva de las técnicas de análisis de datos funcionales.

La distribución de los campos ambientales ocultos que dirigen procesos de recuento suele mostrar una variabilidad e incertidumbre significativas a través del espacio y el tiempo. La caracterización de estos campos depende de la escala espacial en la que se considera el fenómeno, que podría ser diferente de la escala de medición. El efecto de las heterogeneidades a diferentes escalas geográficas en la distribución espacial de los recuentos ya ha sido examinado en Christakos et al. (2001); Congdon (2017); Li et al. (2008).

Otra cuestión que debe abordarse, cuando la inferencia entra en juego, es el

tamaño y la resolución de la ventana temporal, cuantificando las fluctuaciones de la tasa temporal en las regiones espaciales, véase, por ejemplo, Banks et al. (2007); Kennedy y Eberhart (1995); Salap-Ayca y Jankowski (2018). El Capítulo 5 trata este problema en un marco de Análisis de Datos Funcionales (FDA), incorporando correlaciones espaciales entre los parámetros curva de las tasas, en las regiones consideradas. Las predicciones funcionales resultantes reflejan la evolución del patrón de puntos espaciales en cualquier momento. Obsérvese que las técnicas de FDA son muy adecuadas para estimar estadísticos de resumen, que son de naturaleza funcional. En particular, la clasificación de datos de procesos puntuales, basada en estadísticos de segundo orden, puede realizarse aplicando las metodologías de FDA.

Las técnicas de análisis de datos funcionales aplicadas a la estimación de procesos puntuales son una rama de investigación, relativamente novedosa. En el marco de procesos de Poisson doblemente estocásticos, Bouzas et al. (2006a) y Bouzas et al. (2006b) abordan el problema de obtener un modelo óptimo y estimar la función de intensidad desde una perspectiva del análisis de datos funcionales. Primero se aplica un paso de alisado previo, basado en splines. Las técnicas de análisis de datos funcionales también se aplican a la resolución de problemas de clasificación en patrones puntuales, véase, por ejemplo, Illian et al. (2008), cuando se aplican estadísticos de segundo orden. En Wu et al. (2013), cuando no se conoce la forma de las funciones de intensidad que generan los sucesos observados en los tiempos, se propone un enfoque funcional, para obtener la estructura de covarianza de las densidades aleatorias asociadas. En concreto, en el contexto de sucesos observados sobre un intervalo de tiempo, se extrae una fórmula de reconstrucción para las funciones de densidad aleatorias asociadas, que reflejan la distribución de tales sucesos en el tiempo.

La aplicación de la estadística espacial infinito-dimensional al análisis de patrones puntuales espaciales dentro del marco de FDA constituye un campo abierto, con diversos problemas pendientes de abordar. Entre los desafíos planteados en esta área, la definición adecuada del proceso que genera los patrones puntuales juega un papel crucial. En este sentido, como ya se ha comentado anteriormente, en Bosq y Ruiz-Medina (2014) se introduce un proceso de Poisson homogéneo evaluado en ℓ^2 , donde la estimación y predicción de sus parámetros funcionales se aborda desde un marco bayesiano y clásico componente a componente. También se muestran la eficiencia asintótica y la equivalencia de

ambos enfoques de estimación.

La literatura existente sobre el análisis de patrones puntuales tiene una historia extensa, constituyendo una rama importante en la modelización estocástica y estimación de conjuntos numerables de puntos en un espacio plano, generados por un mecanismo aleatorio.

Estos procesos se aplican en muchos campos diferentes como la geología, la sismología, la economía, el procesamiento de imágenes, la ecología o la biología. En particular, la estrecha relación entre los procesos puntuales y los datos geoestadísticos se ha explotado en gran medida en el campo de la estadística espacio-temporal. De hecho, existe una extensa literatura sobre la modelización estadística y el análisis de procesos puntuales, puede consultarse, por ejemplo, Diggle (1983); Diggle y Milne (1983); Møller y Waagepetersen (2004); Ripley (1977); se puede encontrar un tratamiento exhaustivo en el libro Daley y Vere-Jones (1988), véase también Daley y Vere-Jones (2003, 2008). Respecto a la inferencia en procesos espaciales y espacio-temporales, se puede consultar, por ejemplo, Baddeley et al. (2000); Baddeley y Turner (2000); Chakraborty y Gelfand (2001); Guan et al. (2006); Møller et al. (2016); Rodrigues y Diggle (2012).

En términos generales, la literatura en el campo de los procesos puntuales espaciales y espacio-temporales ha aumentado considerablemente en contextos paramétricos (verosimilitud, pseudo-verosimilitud, verosimilitud compuesta), semiparamétricos y no paramétricos, desde una perspectiva clásica y bayesiana, ver Baddeley et al. (2006); Diggle et al. (2010b); Gonçalves y Gamerman (2018); Guan (2006); Jalilian et al. (2019), por mencionar algunos.

Los datos espacio-temporales se han hecho más accesibles en muchos campos científicos, lo que ha impulsado una aceleración de los avances metodológicos. Esto incluye patrones puntuales en el espacio y el tiempo, véase una revisión en González et al. (2016), patrones puntuales espaciales para el caso real o patrones puntuales espaciales multivariantes (Jalilian et al., 2015; Waagepetersen et al., 2016).

Hasta hoy en día, la literatura toma como soporte el plano euclidiano, y desarrolla las herramientas teóricas bajo este paradigma. En particular, Waagepetersen et al. (2016) proponen una versión multivariante de los procesos de Cox log-gaussianos para realizar análisis estadísticos de conjuntos de datos de patrones puntuales multivariantes, y desarrollar un marco inferencial pa-

ra, en principio, un número arbitrario de especies sin una jerarquía conocida. Proporcionan funciones de correlación par cruzadas o K cruzadas en el espacio euclidiano. La función de correlación par, o función de correlación de pares de puntos (*pair correlation function*, véase el Apartado 1.2.1), constituye uno de los estadísticos de uso más común en el campo de los procesos puntuales espaciales (Møller y Waagepetersen, 2004).

La estimación no paramétrica tipo núcleo (o *kernel*) de la función de correlación par, se ha tratado ampliamente en la literatura (González et al., 2016; Møller et al., 1998), también se pueden encontrar alternativas a esta metodología de estimación, como la que utiliza un enfoque bayesiano (Yue y Loh, 2013), y la basada en series ortogonales (Jalilian et al., 2019). En particular, Jalilian et al. (2019) adaptan los estimadores de densidad de series ortogonales para la aproximación de la función de correlación par. Además, deducen estimadores no sesgados de los coeficientes en una expansión de series ortogonales de la función de correlación par, y proponen un criterio para elegir un determinado esquema de suavizado óptimo. Finalmente, establecen la consistencia y la normalidad asintótica del nuevo estimador de series ortogonales.

Los estimadores *kernel* son rápidos en los cálculos pero sufren un fuerte sesgo en los retardos espaciales cercanos a cero. Esto es un gran inconveniente si se intenta inferir un modelo paramétrico a partir de una estimación no paramétrica previa, dado que el comportamiento cercano a cero es importante para determinar el modelo paramétrico correcto (Jalilian et al., 2013). En particular, se debe abordar la estimación paramétrica para el análisis basado en el momento de primer y segundo orden. No hay muchos trabajos en esta línea.

- Un ejemplo lo podemos encontrar en Guan et al. (2008) donde se consideran pares de procesos puntuales espaciales con funciones de intensidad que comparten un término multiplicativo común, e introducen un enfoque de estimación por máxima verosimilitud condicionada en el ajuste de un modelo paramétrico que aproxime la función de correlación par. Se establece asimismo la consistencia, así como se aborda diferentes aspectos relacionados con el diagnóstico estadístico del modelo y la estimación de la función de intensidad.
- Otro ejemplo lo podemos encontrar en Waagepetersen y Guan (2009) donde proponen la estimación paramétrica para procesos puntuales espaciales no homogéneos con un modelo de regresión para la función de

intensidad y las propiedades de segundo orden (función K). En este caso, los parámetros de regresión se estiman a partir de una función de verosimilitud de Poisson. En un segundo paso, se aplica la estimación de mínimo contraste para los parámetros de clusters residuales. Adicionalmente, se establece la normalidad asintótica de los estimadores paramétricos bajo ciertas condiciones.

El proceso de Poisson es un modelo básico y simple en la modelización teórica de procesos puntuales. Este proceso puede ser usado para construir una clase más flexible y fundamental de modelos espaciales, llamados procesos de Cox (véase el Apartado 1.2.1). Un proceso de Cox (o proceso de Poisson doblemente estocástico) se obtiene como una extensión de un proceso de Poisson considerando la función de intensidad de un proceso de Poisson no homogéneo como una realización de una función aleatoria (es decir, un proceso estocástico en el caso uniparamétrico, o un campo aleatorio en el caso multiparamétrico). Los procesos de Cox son modelos naturales para fenómenos de procesos puntuales originados en el medio ambiente, pero son menos naturales para fenómenos dirigidos principalmente por interacciones entre puntos, véase, por ejemplo, [Møller y Waagepetersen \(2004\)](#).

Los procesos de Cox fueron introducidos y estudiados en [Cox \(1955\)](#), considerando una variable aleatoria positiva en la definición de la intensidad, en lugar de una función determinística, ver también [Grandell \(1976\)](#); [Stoyan et al. \(1995\)](#). Esta generalización permite que el proceso exógeno al modelo influya en las transiciones del proceso puntual. Los primeros enfoques utilizados para estos procesos tenían como inconveniente el reducido número de aplicaciones que permitían, ya que se centraban en las propiedades generales de dichos procesos, sin explorar la forma funcional de la intensidad del proceso.

Cuando la forma funcional de la intensidad aleatoria se define explícitamente se obtienen las expresiones analíticas de las funciones de densidad de probabilidad para diferentes tipos de procesos de Cox, véase, por ejemplo, [De Genero y Simonis \(2011\)](#). En este contexto, las restricciones y limitaciones, resultantes de la condición de no negatividad de la intensidad, sugieren un modelo log-normal para el proceso de intensidad, aunque también es posible considerar el proceso tipo ruido de disparo para garantizar la no negatividad de la intensidad, o bien suponer que la intensidad se rige por un proceso de Feller unidimensional. Los procesos de Feller fueron introducidos y estudiados en la

literatura financiera, después de la obra pionera de Cox et al. (1985). Una de las propiedades del proceso de Feller es que se da en \mathbb{R}_+ , cumpliendo la condición de no negatividad.

La simplicidad de un proceso de Cox radica en que las propiedades de los momentos de un proceso de Cox son equivalentes a las propiedades de los momentos de su correspondiente campo aleatorio de intensidad. De hecho, la función de intensidad de un proceso de Cox es igual a la esperanza de la intensidad en un campo aleatorio, y la densidad de covarianza de un proceso de Cox es igual a la función de covarianza de su intensidad aleatoria.

Se pueden destacar los siguientes trabajos, que inicialmente promovieron el desarrollo de la literatura en el ámbito de procesos puntuales y procesos de Cox. En particular, en el contexto de la bioestadística y fiabilidad, mencionamos los trabajos iniciales de Gail et al. (1980), donde se aplica la teoría de procesos puntuales para estudiar el tamaño de los tumores en ratas durante un período de tiempo; Lando (1998), donde la descripción de ocurrencias de obtención de créditos se logra a partir de los procesos puntuales; y Dalal y McIntosh (1994), donde la prueba y la validación del software se realiza en la teoría de la fiabilidad. En Rathbun y Cressie (1994), se considera el contexto de procesos de Cox log-gaussianos, dirigidos por una intensidad aleatoria con valores constantes sobre los cuadrantes de una rejilla regular; véase también Besag (1974). El muestreo de Gibbs se puede aplicar para explorar estos modelos discretizados, véase Rathbun y Cressie (1994).

La literatura ofrece varias subclases de procesos de Cox de particular interés. En la presente tesis se presta atención sobre los procesos de Cox log-gaussianos (LGCP), véase, por ejemplo, Møller et al. (1998), definidos como procesos de Cox con una intensidad aleatoria que es la exponencial de un campo aleatorio gaussiano, o Rostami et al. (2012), donde se considera un proceso puntual de Cox log-gaussiano para modelizar la distribución espacio-temporal de la mortalidad por abuso de sustancias en Irán. Este modelo de log-intensidad normal proporciona un marco flexible en el análisis de patrones puntuales espaciales y espacio-temporales, véase Diggle et al. (2013); González et al. (2016).

La caracterización completa de esta clase de procesos por la intensidad y la densidad producto de segundo orden hace posible su aplicación en diferentes campos, ver, por ejemplo, Rathbun y Cressie (1994) en bosques de pinos; Serra et al. (2014) en incendios forestales. También se puede señalar algún ejem-

plo sobre áreas de aplicación adicionales como ecología, con trabajos como Serra et al. (2014); Waagepetersen et al. (2016); sismología (Møller y Toftaker, 2014); o reconocimiento de texturas de escenas visuales (Simpson et al., 2016). En la presente tesis se va más allá del caso real-valuado, introduciendo procesos de Cox dirigidos por un proceso (o campo) aleatorio de log-intensidad Hilbert-valuada.

En un contexto matemático y probabilístico, modelos lineales y no lineales de ecuaciones de estados en el espacio y tiempo para procesos funcionales se han considerado en los trabajos de Kelbert et al. (2005); Leonenko y Ruiz-Medina (2006); Leonenko et al. (2011), entre otros. La estimación paramétrica de modelos de este tipo mediante mínimo contraste se puede ver, por ejemplo, en Anh et al. (2004a), en el caso espacial. Este tipo de modelos permiten representar procesos espacio-temporales con difusión anómala espacial. En particular, permiten representar la distribución espacial anómala de riesgos de mortalidad (respectivamente incidencia) de enfermedades cuando se observan patrones espaciales cuyas log-intensidades aleatorias son localmente singulares, o bien, presentan discontinuidades, así como un comportamiento local errático que corresponde a un modelo diferencial fraccionario, en los correspondientes mapas de evolución de la enfermedad.

En el ámbito de la salud, la epidemiología investiga la evolución de los patrones espaciales de distribución de enfermedades en la población humana, entre otros aspectos. Es interesante ver las posibilidades de aplicar la Estadística a la modelización de tales sucesos. Se consideran los registros de ocurrencias para determinar los riesgos de mortalidad e incidencia asociados a una enfermedad cuando el objetivo es un análisis de la evolución. Así, la determinación de patrones espaciales es de particular interés para identificar los territorios de riesgo y, en base a los mapas de riesgo, se puede lograr un mejor desempeño de los planes nacionales de emergencia de salud pública (Sánchez-Gómez et al., 2017). La inversión en la vigilancia y respuesta a epidemias y pandemias, junto con el mapeo de riesgos y su comunicación, son esenciales para la detección temprana de los brotes. Lawson (2008) sugiere que los usuarios y lectores se sienten más cómodos si los datos sanitarios agregados se analizan en mapas.

Tradicionalmente, los riesgos de mortalidad (o incidencia de una enfermedad) suelen estar representados en mapas que resumen la distribución espacial o espacio-temporal de los patrones de mortalidad en diferentes regiones. La me-

Metodología estadística utilizada en epidemiología para la estimación de riesgos de mortalidad se basa principalmente en las extensiones del modelo autorregresivo real-valuado de Besag et al. (1991). Modelos bayesianos espacio-temporales alternativos que extienden el modelo desarrollado por Besag et al. (1991), son evaluados en Ugarte et al. (2012, 2009a, 2010), donde se utilizan bases de P-spline para suavizar los riesgos de la enfermedad en el espacio y tiempo. Un estudio comparativo entre el modelo autorregresivo espacial condicionado y los modelos basados en la proyección en bases de P-spline se puede ver en Goicoa et al. (2012); ver también Nguyen et al. (2012), sobre aplicación de algoritmos de suavizamiento local y algoritmo EM, para resolver el problema de “datos faltantes”.

A lo largo de la tesis, como aplicación de las técnicas y metodologías presentadas en cada capítulo, se estudia la evolución en el espacio y tiempo de diversas enfermedades, como son el cáncer de próstata, mama y encéfalo, así como enfermedades respiratorias. En la última parte de la tesis, se llevan a cabo de forma más profunda, aplicaciones a enfermedades epidémicas importantes como son la fiebre del dengue (DF) y la covid.

Para la fiebre del dengue, trabajos recientes se han centrado en la dinámica de la DF y la incidencia geográfica. Las propuestas más utilizadas son modelos mixtos lineales generalizados con efectos aleatorios espaciales propios de modelos CAR (autoregresivos condicionados) o con efectos aleatorios espaciales, temporales y espacio-temporales (Yu et al., 2014). Otros, como Yu y Christakos (2011), estiman el riesgo relativo de transmisión del dengue basándose en un tiempo y espacio discretos mediante un modelo de recuperación infecciosa susceptible (SIR) para poblaciones humanas con un diseño bayesiano de máxima entropía.

Existen trabajos con enfoques bayesianos para modelizar la fiebre del dengue en Asia (Jaya et al., 2016; Yu et al., 2014; Zhu et al., 2016) y consideran aspectos socioeconómicos, geográficos, entomológicos y covariables climáticos para explicar los casos de DF.

En América, la DF en Brasil es estudiada por Pastrana et al. (2014), que se ocupa de las metodologías espaciales y estadísticas para analizar la distribución geográfica del dengue y relacionar su incidencia con el Índice de Vulnerabilidad Sanitaria (HVI). Lowe et al. (2014) formulan otro modelo mixto lineal generalizado, utilizando una distribución binomial negativa para el recuento de casos

de dengue para predecir la epidemia de dengue en Brasil durante 2014. Otros, como [Pepin et al. \(2015\)](#) utilizan un modelo de regresión bayesiana jerárquica para examinar el papel de los datos de vigilancia de vectores en toda la ciudad para predecir los casos humanos de dengue en el espacio y el tiempo en Vitoria, Brasil. [Lorenz et al. \(2020\)](#) estudian el mosquito *Aedes aegypti* y su diseminación del dengue, Zika, chikungunya y la fiebre amarilla urbana. Utilizando un modelo bayesiano y el método de la Ecuación Diferencial Parcial Estocástica, determinan las características de la tierra asociadas con las infestaciones de mosquitos en São José do Rio Preto/SP, Brasil.

En Colombia, [Martínez-Bello et al. \(2018\)](#) genera estimaciones suavizadas del riesgo relativo aplicando modelos espacio-temporales bayesianos jerárquicos, añadiendo covariables obtenidas a partir de imágenes de satélite que contienen la temperatura de la superficie terrestre y un índice de vegetación diferencial normalizado, para el período comprendido entre enero de 2009 y diciembre de 2015 en Bucaramanga. [Restrepo et al. \(2014\)](#) examinan la variación de la distribución espacial de los casos de dengue notificados en Colombia de enero de 2007 a diciembre de 2010 con un modelo autorregresivo condicionado espacio-temporal bayesiano. Venezuela fue analizada por [Cabrera y Taylor \(2019\)](#) con una modelización espacio-temporal de dengue a través de un modelo mixto aditivo generalizado. En Argentina, [Estallo et al. \(2014\)](#) estudian la dinámica espacio-temporal del brote, en la ciudad de Córdoba, durante el 2009, pero carecen de la visualización de un modelo bayesiano. Los modelos espacio-temporales descritos anteriormente han sido ajustados, en su mayoría, por los métodos de Monte Carlo en base a Cadenas de Markov (MCMC) pero, recientemente, se ha utilizado una técnica novedosa llamada Aproximaciones Integradas Anidadas de Laplace (INLA) para aligerar el coste de cálculo.

La enfermedad de coronavirus 2019 (COVID-19) surge en diciembre de 2019, y rápidamente se extiende por los países ([Sivakumar, 2020](#); [Wang et al., 2008](#); [Zhou et al., 2020](#)). La asignación eficaz de los recursos médicos requiere la derivación de técnicas de predicción que describan la dinámica espacio-temporal de la COVID-19 (véase, por ejemplo, [Du et al. \(2020\)](#); [Khan y Atangana \(2020\)](#); [Nishiura et al. \(2020\)](#); [Remuzzi y Remuzzi \(2020\)](#), por mencionar algunos). Los modelos epidemiológicos pueden contribuir al análisis de las causas, la dinámica y la propagación de esta pandemia (véanse, por ejemplo, [Huppert y Katriel \(2013\)](#); [Keeling y Rohani \(2008\)](#); [Laaroussi et al. \(2018\)](#), y

las referencias en ellos). Las predicciones a corto plazo pueden obtenerse adoptando el marco de los modelos compartimentados SIR (susceptible-infeccioso-recuperado), basados en ecuaciones diferenciales ordinarias, véanse, por ejemplo, Angulo et al. (2013); Elhia et al. (2014); Ji et al. (2012); Kermack y McKendrick (1927); Kuznetsov y Piccardi (1994); Milner y Zhao (2008); Pathak et al. (2010); Tornatore et al. (2005); Yu et al. (2009); Zhang et al. (2008).

Existe una extensa literatura, que incluye diferentes versiones de modelos compartimentados, como el SIR-susceptible, SIRS (Dushoff et al., 2004), y formulaciones basadas en ecuaciones diferenciales de retardo, ver Beretta et al. (2001); McCluskey (2010); Sekiguchi y Ishiwata (2010). Las extensiones espaciales, basadas en modelos de reacción-difusión, que reflejan la propagación de la enfermedad infecciosa en una región espacial pueden encontrarse, por ejemplo, en Guin y Mandal (2014); Webb (1981). También se analizan los modelos SEIRD (susceptible, expuesto, infectado, recuperado, fallecido), que incorporan la propagación espacial de la enfermedad con términos de difusión no homogéneos (Roques y Bonnefon, 2016; Roques et al., 2011).

La versión estocástica de los modelos de tipo SIR pretende cubrir varias limitaciones detectadas en cuanto a la incertidumbre, en las observaciones, y el proceso epidémico dinámico oculto. La modelización basada en cadenas de Markov SIR (Anderson et al., 2017; Xu et al., 2007), y algunas formulaciones estocásticas que implican redes complejas (Volz, 2008; Zhou et al., 2006) o la gripe resistente a los medicamentos (Chao et al., 1996) constituyen algunas alternativas. En Aalen et al. (2008); Abboud et al. (2019); Anderson et al. (2017); Fleming y Harrington (1991), se adopta un marco de modelo estadístico bayesiano SIRS teniendo en cuenta el error de observación en los recuentos, y la incertidumbre en el espacio de parámetros. Más allá de la modelización SIR, mencionamos aquí los enfoques basados en el análisis multivariante y de supervivencia para modelizar, por ejemplo, la infección, la incubación y los períodos aleatorios de recuperación, que afectan a la contención del COVID-19, véase, por ejemplo, Bolker y Grenfell (1996); Keeling et al. (1997); Pak et al. (2020); Wasiur et al. (2019).

En una primera etapa, la mayoría de los modelos referidos han sido adaptados y aplicados para aproximar la evolución espacio-temporal de la incidencia y la mortalidad del COVID-19. Ese es el caso, por ejemplo, de los tres modelos presentados en Roosa et al. (2020), que fueron validados con brotes de otras

enfermedades diferentes a la COVID-19. En Kucharski et al. (2020) se formulan modelos alternativos de tipo SEIR, con componentes estocásticos. También se ha propuesto un modelo SEIR revisado en Zhang et al. (2020), véase también He et al. (2020). En Ivorra et al. (2020) se introduce un modelo θ -SEIHRD, capaz de estimar el número de casos, muertes y necesidades de camas en los hospitales, adaptado a COVID-19, basado en el modelo Be-CoDiS (Ivorra et al., 2015).

Debido a la baja calidad de los registros disponibles, y a la información muestral oculta, lo más destacable en esta área de investigación es el equilibrio entre la complejidad y la indentificabilidad de los parámetros del modelo. Recientemente, en Ramos et al. (2020) se presenta un intento de simplificar las estrategias de modelización, aplicadas al análisis de los datos de COVID-19, en términos del modelo θ -SEIHQRD. La mitigación del submuestreo se propone en Langousis y Carsteanu (2002), basándose en el reescalado de los estadísticos que intervienen en el análisis de las propiedades de la epidemia, como el análisis de la mortalidad, útil entre países con niveles similares de asistencia sanitaria.

Los métodos de regresión lineal, perceptrón multicapa y autoregresión vectorial se han aplicado en Sujath et al. (2020) para predecir la propagación de COVID-19, y anticipar los patrones potenciales de los efectos de COVID-19 (véase también la Sección 2 de Sujath et al. (2020), sobre el trabajo relacionado). La localización de COVID-19 en las primeras fases se aborda en Barstugan et al. (2020), aplicando estrategias de aprendizaje automático actualizadas en imágenes de Tomografía Computarizada de estómago. Chien y Chen (2002) evalúa la asociación entre los factores meteorológicos y la propagación del COVID-19. Estos autores concluyeron que la temperatura media, la humedad relativa mínima y las precipitaciones eran los mejores predictores, mostrando posibles correlaciones no lineales con las variables de COVID-19. Estas conclusiones son cruciales en el posterior análisis basado en la regresión *Machine Learning*.

En la actualidad, los modelos ML se han establecido como serios contendientes de los modelos estadísticos clásicos en el ámbito de la predicción. La investigación comenzó en los años ochenta con el desarrollo del modelo de red neuronal. Posteriormente, la investigación extendió este concepto a modelos alternativos, como las máquinas de vectores soporte, los árboles de decisión y otros, véase, por ejemplo, Alpaydin (2004); Blanquero et al. (2020); Hastie

et al. (2001); Mohammady et al. (2021).

En general, las técnicas de curvas de regresión basadas en una función, normalmente en el espacio de funciones de cuadrado integrable con respecto a una medida de probabilidad adecuada, permiten realizar predicciones a corto y largo plazo. Así, dependiendo de nuestra elección de la base de funciones y de la medida de probabilidad seleccionada, podrían combinarse las visiones de *partícula* y *campo*. Obsérvese que los modelos clásicos de difusión estocástica ofrecen una visión de *partícula* más que de *campo*, véase, por ejemplo, Malesios et al. (2016).

Objetivos

La motivación fundamental para el desarrollo de esta tesis reside en el análisis estadístico de la evolución de patrones aleatorios espaciales de enfermedades, esencialmente en el marco epidemiológico. Se intenta proporcionar herramientas estadísticas apropiadas para establecer la evolución en el tiempo así como en el espacio, de mapas de riesgo o, en su defecto, mapas de mortalidad o incidencia de una enfermedad.

La temática teórica se desarrolla en el contexto de modelos que utilizan los procesos de Cox log-gaussianos en espacios infinito-dimensionales, en particular, se consideran espacios de Hilbert separables. Se analizan las propiedades asintóticas de las estimaciones paramétricas para dichos modelos y también, de una forma más general, para campos aleatorios lineales estacionarios a partir de funcionales espectrales, basados en el periodograma y periodograma *tapered*, así como su versión infinito-dimensional, es decir, el operador periodograma.

La teoría espectral de operadores compactos y autoadjuntos juega un papel fundamental en todos estos desarrollos, cuando se considera una diagonalización de la estructura de dependencia del proceso latente de interés, que usualmente se trata de un proceso temporal funcional o espacial funcional. En el contexto log-gaussiano, centramos la atención en las medidas gaussianas Hilbert-valuadas, definidas mediante el producto infinito de medidas gaussianas univariantes.

En el ámbito computacional y aplicado, se desarrollan e implementan nuevos algoritmos para el análisis estadístico y la construcción de mapas de riesgo de enfermedades y mapas de mortalidad e incidencia causada por enfermedad,

abordando el problema de predicción funcional asociado en un ámbito infinito-dimensional.

En la literatura existente, se han propuesto modelos estadísticos espacio-temporales para el análisis de la evolución temporal de la distribución geográfica de los riesgos de mortalidad (o incidencia). Sin embargo, los enfoques estadísticos funcionales, basados en la teoría de procesos Hilbert-valuados, mediante datos funcionales que evolucionan en el tiempo o datos funcionales que evolucionan en el espacio, hasta donde conocemos, constituyen un terreno prácticamente inexplorado.

Se necesita generar modelos flexibles estadísticos en el espacio y tiempo para solventar las limitaciones y dificultades presentes en los modelos estadísticos tradicionales clásicos que se han desarrollado en el contexto de modelos markovianos discretos en el espacio y, más recientemente, en el espacio y tiempo. Por otra parte, las herramientas que aportan los modelos lineales y no lineales de ecuaciones de estados para procesos evaluados en espacios de funciones, definidos en el tiempo y/o espacio, son bastante más generales y potentes para solventar los problemas mencionados. Estas dificultades suelen aparecer en el suavizamiento, estimación y predicción de riesgos de mortalidad por enfermedad mediante las técnicas y modelos estadísticos tradicionales en epidemiología.

Existe una importante laguna en las técnicas estadísticas usuales que se aplican en el suavizamiento y estimación de riesgos de mortalidad o incidencia de una enfermedad en el espacio y tiempo. A los comentados modelos markovianos discretos en el tiempo y espacio, actualmente, se ha implementado el suavizamiento mediante splines. Los intervalos de confianza que se extraen para la detección de zonas de alto riesgo se basan fundamentalmente en la teoría de modelos mixtos lineales gaussianos, y modelos bayesianos y empíricos bayesianos (regiones de credibilidad). La falta de precisión de estos métodos, dado el excesivo suavizamiento de los mismos ocasiona la ocurrencia de falsos negativos y dificulta la detección de zonas de alto riesgo.

Esta es una motivación importante para el desarrollo de los siguientes capítulos, en los que se busca aportar nuevos modelos espacio-temporales funcionales, más flexibles, para la representación de la distribución espacial y temporal del riesgo, mortalidad o incidencia de enfermedades y su predicción funcional mediante el mapeo de enfermedades en diferentes tiempos. Las herramientas

estadísticas funcionales que se extraen para el desarrollo de la inferencia a partir de estos modelos supone un avance importante en el ámbito teórico (estadística matemática) y aplicado (análisis estadístico de riesgo, mortalidad o incidencia de enfermedades que evolucionan en el tiempo y/o en el espacio).

Objetivos generales

El desarrollo de la presente tesis doctoral tiene como objetivo fundamental el análisis estadístico y estimación paramétrica de procesos temporales y espaciales infinito-dimensionales, en el marco de procesos puntuales, representados mediante la familia de procesos de Cox dirigidos por una intensidad temporal o espacial infinito-dimensional. La teoría desarrollada, en las últimas décadas, sobre modelos lineales en espacios de funciones, se aplicará en la implementación de herramientas estadísticas apropiadas para la predicción funcional de dichos procesos.

Los objetivos generales que se persiguen son:

- (i) Desarrollar técnicas de estimación paramétrica de mínimo contraste para campos aleatorios con parámetro discreto: caso real-valuado y Hilbert-valuado.
- (ii) Obtener los resultados teóricos que definen las condiciones que garantizan un buen comportamiento asintótico de los estimadores y predictores formulados (es decir, consistencia débil y fuerte y normalidad asintótica).
- (iii) Proporcionar técnicas de estimación paramétrica de mínimo contraste para la aproximación funcional de mapas de riesgo, mortalidad o incidencia de enfermedades.
- (iv) Desarrollar diferentes escenarios para la estimación paramétrica de procesos de Cox. Por un lado, a partir de datos funcionales correlados en el tiempo y por otro, a partir de datos funcionales correlados en el espacio.
- (v) Estudiar la clase de procesos temporales de Cox log-gaussianos en espacios de Hilbert con intensidad aleatoria dada por un proceso Ornstein-Uhlenbeck Hilbert-valuado, aproximado por un proceso ARH(1).

- (vi) Introducir una nueva clase de procesos espaciales puntuales definidos mediante la familia de procesos de Cox estudiados, dirigidos por una intensidad lineal infinito-dimensional espacial. Se considerará, como ilustración de los resultados de estimación derivados, la sub-clase de procesos de Cox dirigidos mediante una intensidad espacial SARH(1).
- (vii) Extraer técnicas de predicción de mapas de riesgo, mortalidad o incidencia de enfermedad, en el contexto de la Estadística Funcional, para procesos y campos Hilbert-valuados.
- (viii) Abordar el estudio de simulación para validar las técnicas de inferencia propuestas en el ámbito de la teoría de campos aleatorios discretos y continuos real y Hilbert-valuados.
- (ix) Llevar a cabo aplicaciones prácticas reales con datos reales espacio-temporales donde ilustrar el estudio, metodología y resultados indicados en los objetivos anteriores.

Estructura de la tesis

Para el desarrollo de la tesis, se comienza con una introducción en la que se realiza una motivación de la misma que incluye la literatura relacionada destacada, los principales objetivos señalados y su estructura. Posteriormente, en las secciones 1.1, 1.2 y 1.3, se establecen los preliminares necesarios para el desarrollo en la obtención de los resultados de los capítulos posteriores.

De forma más concreta, en la Sección 1.1 se establecen los preliminares sobre los estimadores de mínimo contraste a partir de datos *tapered*¹. Los apartados de esta sección evidencian los problemas que surgen, en la estimación, cuando se trabaja con campos aleatorios respecto a las metodologías de estimación, basadas en el espectro unidimensional, en el caso de procesos estocásticos temporales.

En la Sección 1.2 se presentan los preliminares y el fundamento teórico de los procesos de Cox log-gaussianos. Primeramente se definen los procesos de Poisson doblemente estocásticos, o procesos de Cox, y posteriormente se

¹*Tapered* se podría traducir como una disminución gradual del diámetro, anchura y grosor de un objeto. En nuestro caso se trata de la disminución gradual en el borde de los campos aleatorios a partir de una función *taper*, véase el Apartado 1.1.3.

introducen estos procesos cuando el logaritmo de su intensidad sigue una distribución gaussiana. Para completar los fundamentos teóricos de los procesos introducidos en la tesis, se hace un análisis de los procesos de Poisson homogéneos en el espacio ℓ^2 , el proceso de Wiener fuertemente cilíndrico inducido por un proceso de Wiener Hilbert-valuado, el proceso Ornstein-Uhlenbeck en espacios de Hilbert y, finalmente, los procesos ARH(1).

Este primer capítulo concluye con la Sección 1.3, donde se establecen los preliminares sobre las series espaciales autorregresivas hilbertianas. Esta sección viene a completar los preliminares indicados en la Sección 1.2, extendiendo al caso de datos espaciales e incluyendo las posibilidades que ofrece la Sección 1.1 mediante la metodología de estimación de mínimo contraste.

Los capítulos 2, 3, 4 y 5 conforman el bloque principal de la tesis puesto que contienen los principales resultados de la misma. En el Capítulo 2 se estudia la estimación paramétrica de procesos estocásticos estacionarios y campos aleatorios a partir de datos correlacionados temporalmente o espacialmente. Se estudian los estimadores de mínimo contraste de Ibragimov basados en datos *tapered*. Estos primeros resultados sobre estimación de mínimo contraste en campos aleatorios, se desarrolla en el contexto de campos aleatorios real-valorados, con parámetro discreto en \mathbb{Z}^d .

Los beneficios de los datos *tapered* han sido ampliamente ilustrados en la literatura. Es bien sabido que el *tapering*² reduce los efectos de posibles pérdidas de información muestral, por el truncamiento del comportamiento asintótico del proceso subyacente, debido a su observación en un dominio acotado, cuando el análisis de las propiedades asintóticas de los estimadores formulados se realiza, bajo un enfoque “*increasing domain asymptotics*”. Y lo que es aún más importante, el uso de datos *tapered* nos lleva a la reducción del sesgo, asociado a la estimación del espectro, a partir del periodograma, ver, por ejemplo, Dahlhaus y Künsch (1987) o Guyon (1995).

Las principales contribuciones del Capítulo 2 son:

1. Se enuncian los teoremas sobre consistencia y normalidad asintótica de los estimadores de mínimo contraste de Ibragimov para campos aleatorios con dependencia de largo rango en un contexto general, es decir, campos aleatorios discretos real-valorados bajo condiciones más generales que las asumidas en Anh et al. (2004a,b, 2007c). Más específicamente:

²*Tapering* se puede entender como la utilización de datos *tapered*.

- a) Se dan las condiciones sobre el comportamiento de la densidad espectral y sus derivadas en el punto de singularidad.
- b) Se dan las condiciones exactas que garantizan buenas propiedades en la aplicación de funciones *taper*.

De esta manera los resultados obtenidos se pueden aplicar a modelos particulares, según convenga. A partir de la definición de las condiciones comentadas se obtiene una función peso, la cual se incorpora en el funcional de Ibragimov para compensar las singularidades de la densidad espectral.

2. Para el caso de campos aleatorios gaussianos con dependencia de largo rango, se establece el teorema central del límite para funcionales espectrales (o formas cuadráticas) basado en datos *tapered*.

En dicho Capítulo 2, las declaraciones de los principales resultados van acompañadas de discusiones para aclarar las condiciones utilizadas y también de algunos incisos destinados a comentar los resultados relacionados existentes en la literatura. Finalmente, la parte numérica del capítulo contribuye a la investigación sobre tasas de convergencia de los estimadores de mínimo contraste. También se prueba que los resultados teóricos obtenidos pueden extenderse a otras clases de modelos fraccionarios.

En el Capítulo 3 se desarrollan los procesos de Cox log-gaussianos en un espacio de Hilbert real separable H , desarrollando condiciones suficientes para la definición apropiada de una medida de recuento en el espacio ℓ^2 isométrico a H . El ingrediente principal requerido es la representación espectral diagonal, según un sistema común de autovectores, de los operadores implicados en la definición del operador de autocovarianza de un proceso de O-U Hilbert-valuado. También se obtiene la aproximación en tiempo discreto del proceso O-U H -valuado en términos de un proceso ARH(1), a partir de la proyección en dicho sistema de autovectores.

La estimación componente a componente del operador de autocorrelación y del operador de autocovarianza del proceso de innovación dado por un proceso ARH(1) lleva a la correspondiente estimación paramétrica del proceso O-U aproximado. Los resultados de la predicción *plug-in* se obtienen entonces, a partir de la forma integral del proceso O-U H -valuado, y de la formulación correspondiente del predictor ℓ^2 -valuado para el proceso de Cox log-gaussiano.

Finalmente, en dicho Capítulo 3, se muestra una aplicación al mapeo de enfermedades. En concreto, se realiza un estudio de simulación condicionada a partir de datos de cáncer de mama, próstata y encéfalo, en las provincias de España. También se desarrolla un estudio comparativo en el caso de intensidad log-gaussiana de dependencia de largo rango (DLR), aplicando la metodología de estimación paramétrica basada en wavelets, presentada en Frías y Ruiz-Medina (2011), para los procesos gaussianos DLR.

En los capítulos 4 y 5 se estudian nuevamente los procesos de Cox con intensidad log-gaussiana. Como ya se ha comentado, estos modelos ofrecen un marco adecuado para la modelización y estimación de sucesos que se producen de forma aleatoria en el espacio y/o tiempo, tal es el caso del número de muertes ocasionadas por una enfermedad, en una determinada región, o bien, del número de casos que se han producido en las regiones espaciales analizadas, en relación con una determinada enfermedad infecciosa. Motivado por el problema que surge en el análisis estadístico de datos cuando se trabaja con una gran dimensión, en el contexto del análisis de procesos puntuales espaciales, dichos capítulos amplían los resultados desarrollados en el Capítulo 3 sobre los procesos de Cox log-gaussianos infinito-dimensionales al caso espacial.

El Capítulo 4 introduce una nueva familia de procesos de Cox, dirigidos por una intensidad aleatoria infinito-dimensional, cuyo logaritmo obedece un modelo lineal funcional dinámico y estacionario, en el espacio, evaluado en un espacio de Hilbert separable. La contribución fundamental del Capítulo 4 a la estadística funcional consiste en la obtención de una versión infinito-dimensional del funcional de Whittle, basada en el operador periodograma, que permite la estimación paramétrica, en el dominio espectral, de una familia de campos aleatorios espaciales estacionarios Hilbert-valuados. Asimismo, se deduce la consistencia en sentido fuerte de dicho estimador. Como aplicación, se obtiene un estimador *plug-in* del proceso de Cox dirigido por la exponencial de los elementos de la familia considerada. Los estudios de simulación realizados y las aplicaciones con datos reales abordadas muestran el interés de los resultados derivados, en el contexto del análisis espacio-temporal de la evolución de enfermedades, en particular, enfermedades respiratorias.

Es bien sabido que los procesos de Cox log-gaussianos proporcionan modelos útiles para representar patrones agregados, Diggle (2013); Diggle et al. (2013). Como se comenta en Diggle et al. (2013), los procesos de Cox son

modelos naturales para fenómenos de procesos puntuales que aparecen en el medio ambiente, y no tan naturales para fenómenos que se dan principalmente a partir de interacciones puntuales. Dicho Capítulo 4 ilustra este hecho en un marco de variables aleatorias infinito-dimensionales. Todos estos aspectos se reflejan y describen en las últimas secciones de simulación y aplicación a datos reales del capítulo. En particular, se ilustran las propiedades asintóticas de los diferentes estimadores y predictores analizados, en el contexto de procesos de Cox dirigidos por la exponencial de un proceso SARH(1).

En el Capítulo 5, se adopta un marco de modelo dinámico de curva espacial para la predicción funcional de los recuentos en un modelo de proceso de Cox log-gaussiano espacio-temporal. Este enfoque de estimación funcional espacial maneja tanto el análisis de heterogeneidad basado en wavelets en el tiempo, como el análisis espectral en el espacio. Específicamente, el ajuste del modelo se logra minimizando la divergencia de información o la entropía relativa entre el modelo multiescalar subyacente a los datos, y los candidatos correspondientes en el dominio espectral espacial. Se lleva a cabo un estudio de simulación dentro de la familia de procesos ℓ^2 -valuados autorregresivos espaciales log-gaussianos (procesos SAR ℓ^2) para ilustrar las propiedades asintóticas de los estimadores funcionales espaciales propuestos. Esta estrategia de modelización se aplica a la predicción espacio-temporal de la mortalidad por enfermedades respiratorias.

El dicho capítulo, se establecen condiciones suficientes para introducir una nueva clase de proceso de Cox log-gaussiano ℓ^2 -valuado espacial. Estas condiciones conllevan que el correspondiente proceso de intensidad aleatoria se dé en un espacio real separable de Hilbert. Las condiciones obtenidas permiten realizar un análisis multiescalar de la varianza funcional del proceso de intensidad aleatoria. Posteriormente, se analiza el rango de fluctuaciones temporales a través de diferentes escalas. En nuestro caso, elegimos una base wavelet de soporte compacto. Con este análisis multiescalar, se obtiene un ajuste más preciso de la variabilidad local mostrada por los datos curva. Nótese que las bases B-splines han sido ampliamente utilizadas en el preprocesamiento del Análisis de Datos Funcionales (FDA), lo que ha llevado, en algunos casos, a un sobreajuste de los datos curva analizados.

Siguiendo en el Capítulo 5, se propone una metodología alternativa de estimación funcional espacial multiescalar basada en el espectro, en contraste con

la paramétrica basada en el estimador de Whittle, adoptada en el Capítulo 4. En efecto, esta metodología implica el criterio de minimización relativa de la entropía, para obtener el modelo multiescalar óptimo, subyacente a los datos, en el dominio espectral espacial, a partir del cálculo del operador periodograma a diferentes niveles de resolución temporal. Las propiedades de los estimadores multiescalares desarrollados se analizan en el estudio de simulación. Los resultados de validación obtenidos en la aplicación de datos reales ilustran las buenas propiedades del enfoque de estimación presentado en la reconstrucción del campo de log-intensidad a diferentes escalas temporales.

La siguiente parte de la tesis se dedica a la aplicación a datos reales de las técnicas vistas en los capítulos previos. En este sentido, en un contexto más amplio en cuanto a regiones geográficas, en el siguiente capítulo se considera la fiebre del dengue, que se ha convertido en una de las enfermedades infecciosas más extendidas por el mundo. Además, su incidencia está aumentando en las zonas endémicas de las regiones tropicales y subtropicales.

Los modelos de mapeo de enfermedades en el espacio y el tiempo son instrumentos comunes para explicar los patrones de recuentos de enfermedades, donde los modelos bayesianos constituyen un marco adecuado para su formulación. Estos sucesos aleatorios reflejan interacciones entre lugares geográficos cercanos, así como correlaciones entre instantes temporales cercanos. Las técnicas de análisis de datos funcionales pueden describir mejor la evolución del mapeo de enfermedades. En ese sentido, en el Capítulo 6, se estudia el riesgo de dengue en México, América Central y América del Sur, desde un enfoque funcional a través de un modelo de estimación bayesiano centrado en los procesos autorregresivos Hilbert valuados. Así pues, se aproxima la evolución funcional temporal de los patrones geográficos espaciales de riesgo de incidencia en el mapeo de enfermedades durante 1998-2018.

Siendo la tesis sobre el mapeo de enfermedades, con la coyuntura actual provocada por el virus SARS-CoV-2, en el Capítulo 7 se han desarrollado e implementado técnicas de análisis de datos funcionales (FDA) multivariantes bajo un enfoque bayesiano y clásico para la predicción espacio-temporal de mortalidad por COVID-19. También se aplican modelos de aprendizaje automático dentro del marco de FDA utilizando técnicas de *bootstrap* y validación cruzada para la selección de parámetros que permiten una comparación y clasificación de los modelos empíricos utilizados.

En concreto, se presenta un enfoque de predicción espacio-temporal a partir de información muestral blanda que incluye el análisis de log-curvas de regresión cíclicas y el análisis de la correlación residual espacial de series temporales multivariantes bajo un enfoque bayesiano. Esto nos lleva a un problema de optimización en dos etapas. Primeramente, la función de pérdida cuadrática media se minimiza en el marco de regresión trigonométrica. En nuestro posterior análisis de correlación residual espacial, la maximización de la verosimilitud nos permite calcular la moda de la distribución a posteriori, en un marco bayesiano de series temporales multivariantes.

El enfoque presentado se aplica al análisis de la mortalidad de COVID-19 en la primera ola que afecta a las comunidades españolas, desde el 8 de marzo de 2020 hasta el 13 de mayo de 2020. Se realiza un estudio empírico comparativo con la regresión *Machine Learning* (ML), basado en la validación cruzada aleatoria k -fold, y técnicas de estimación *bootstrap*. Este análisis empírico también investiga el rendimiento de los modelos de regresión ML en un marco de datos duros y blandos, referidos a las dos categorías correspondientes (definidas cuando se considera información muestral dura y blanda). Como posibles aplicaciones, las técnicas utilizadas podrían extrapolarse a otros recuentos, países y olas posteriores de COVID-19. Como resultado de este trabajo, se comprueba que el enfoque funcional multivariante bayesiano, adoptado en la predicción del riesgo de mortalidad por COVID-19, proporciona una aproximación similar a las técnicas de regresión, basadas en redes neuronales bayesianas, cuando estas últimas se implementan a partir de información muestral blanda.

Finalmente, en la última parte de la tesis, de forma resumida se destacan las conclusiones y líneas futuras de investigación que se desprenden del desarrollo de los capítulos 2, 3, 4 y 5. Así, se pone de manifiesto la importancia de la metodología de estimación seguida en los capítulos, así como los resultados obtenidos, destacando los de consistencia y normalidad asintótica. Durante los mencionados capítulos, se llevan a cabo estudios de simulación para analizar las propiedades de las metodologías de estimación utilizadas. También, se abren nuevas líneas de investigación, en el contexto de los procesos de Cox log-gaussianos en el tiempo y en el espacio, dentro de los patrones puntuales y procesos de recuento. Las técnicas utilizadas de análisis de datos funcionales hacen posible la reducción de dimensiones que ayuda a evitar problemas de cálculo que surgen al trabajar con matrices de elevado orden. Esto permite el

análisis de propiedades locales para datos de elevada dimensión o de naturaleza continua.

Capítulo 1

PRELIMINARES

En este capítulo de la tesis se proporcionan los elementos preliminares necesarios, para el desarrollo de sus principales contribuciones, presentadas en los capítulos 2–7.

1.1. Preliminares para el Capítulo 2

1.1.1. Estimadores de mínimo contraste y datos *tapered*

Seguidamente se consideran los elementos básicos que intervienen en la introducción de estimadores de mínimo contraste basados en el periodograma espacial (ver Capítulo 2, sobre las contribuciones derivadas en la presente tesis, en este ámbito).

Definición 1.1. Sea $Z(\mathbf{s})$, $\mathbf{s} \in S \subseteq \mathbb{R}^d$, un proceso de segundo orden. Se dice que Z es un proceso estacionario en sentido débil si satisface:

$$\forall \mathbf{s}, \mathbf{t} \in S : E(Z(\mathbf{s})) = m, c(\mathbf{s}, \mathbf{t}) = \text{Cov}(Z(\mathbf{s}), Z(\mathbf{t})) = C(\mathbf{t} - \mathbf{s}).$$

Por tanto, se tiene que la función de covarianza C satisface

$$\forall \mathbf{s}, \mathbf{t}, \mathbf{h} \in S : c(\mathbf{s} + \mathbf{h}, \mathbf{t} + \mathbf{h}) = \text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{t} + \mathbf{h})) = C(\mathbf{t} - \mathbf{s}).$$

A $\rho : S \mapsto \mathbb{R}$ se le llama “función de correlación” de Z ,

$$\mathbf{h} \rightarrow \rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})},$$

donde $C(\mathbf{h}) = \text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s}))$, $C(\mathbf{0}) = \text{Var}(Z(\mathbf{s}))$, $\forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in S$.

Para el caso $S = \mathbb{R}^d$, asociamos a C una medida simétrica $F \geq 0$, acotada en conjuntos de Borel, $\mathcal{B}(\mathbb{R}^d)$, tales que

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} e^{i\mathbf{h}'\mathbf{u}} F(d\mathbf{u}).$$

C es integrable con respecto a una medida Lebesgue en \mathbb{R}^d , F es absolutamente continua con función de densidad f , llamada *densidad espectral* de X . La transformada inversa de Fourier expresa f en función de C :

$$f(\mathbf{u}) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\mathbf{h}'\mathbf{u}} C(\mathbf{h}) d\mathbf{h},$$

el par de Fourier se completa con

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} e^{i\mathbf{h}'\mathbf{u}} f(\mathbf{u}) d\mathbf{u}$$

En el caso de procesos estacionarios discretos, es decir, cuando $Z(\mathbf{s})$ toma valores en un conjunto de puntos discretos, $\mathbf{s} \in \mathbb{Z}^d$, la función de covarianza se define, para $\mathbf{h} \in \mathbb{Z}^d$, asociada a una medida $F \geq 0$, acotada en conjuntos de Borel $\mathcal{B}(\mathbb{T})$, donde \mathbb{T} denota un toro d -dimensional, $\mathbb{T} = [0, 2\pi)^d$. Si los valores de la covarianza forman una secuencia de cuadrados sumables, es decir, $\sum_{\mathbf{h} \in \mathbb{Z}^d} C(\mathbf{h})^2 < \infty$, entonces, se define la transformada Fourier como:

$$f(\mathbf{u}) = \frac{1}{(2\pi)^d} \sum_{\mathbf{h} \in \mathbb{Z}^d} C(\mathbf{h}) e^{-i\mathbf{h}'\mathbf{u}}, \quad \mathbf{u} \in \mathbb{T}$$

y la función de covarianza en un desfase \mathbf{h} puede recuperarse mediante la transformada inversa de Fourier:

$$C(\mathbf{h}) = \int_{\mathbb{T}} f(\mathbf{u}) e^{i\mathbf{h}'\mathbf{u}} d\mathbf{u}.$$

Ver, por ejemplo, Crujeiras (2006); Gaetan y Guyon (2010).

El periodograma espacial se introduce a continuación, como estimador no paramétrico básico, en la aproximación de la estructura de dependencia espacial de procesos estacionarios, en el dominio espectral. Sus limitaciones, respecto al sesgo y velocidad de convergencia a cero del mismo se analizarán con más

detalle posteriormente, en la introducción del “periodograma tapered”.

Para el caso bidimensional, un proceso espacial Z observado en una rejilla regular $D = \{\mathbf{s} = (s_1, s_2) : s_1 = 1, \dots, n_1, s_2 = 1, \dots, n_2\}$, $D \subset \mathbb{R}^2$, con $N = n_1 n_2$ puntos, el periodograma para una frecuencia $\boldsymbol{\lambda}$ se define como:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2 N} \left| \sum_{\mathbf{s} \in D} Z(\mathbf{s}) e^{i\mathbf{s}'\boldsymbol{\lambda}} \right|^2, \quad \boldsymbol{\lambda} \in \pi^2.$$

Se sabe que el periodograma espacial es un estimador asintóticamente insesgado de la densidad espectral, en el que la varianza es proporcional al cuadrado de la densidad espectral en cada frecuencia. Sin embargo, la característica del periodograma que lo hace útil es que los valores del periodograma a diferentes frecuencias son asintóticamente incorrelados. Este hecho permite tratar los valores del periodograma, en el caso gaussiano, como si fueran datos independientes para una muestra lo suficientemente grande.

Se ha definido el periodograma en términos de los datos observados. Como la densidad espectral es la transformada de Fourier de la función de covarianza, podemos escribir el periodograma en términos de un estimador de covarianza. Se define la covarianza muestral como:

$$\hat{C}(\mathbf{h}) = \frac{1}{N} \sum_{\mathbf{s} \in D(\mathbf{h})} Z(\mathbf{h}) Z(\mathbf{s} + \mathbf{h}),$$

donde $D(\mathbf{h}) = \{\mathbf{s} \in D : \mathbf{s} + \mathbf{h} \in D\}$. Entonces, el periodograma en términos de la covarianza se define como:

$$I(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^2} \sum_{\mathbf{h} \in \mathcal{H}} \hat{C}(\mathbf{h}) e^{i\mathbf{h}'\boldsymbol{\lambda}},$$

donde $\mathbf{h} \in \mathcal{H} = \{(h_1, h_2) : -n_1 \leq h_1 \leq n_1, -n_2 \leq h_2 \leq n_2\}$; n_1, n_2 , son los puntos a partir de los cuales se define una rejilla regular $D \subset \mathbb{R}^2$.

Si definimos la transformada de Fourier discreta de los datos como

$$J(\boldsymbol{\lambda}) = \frac{1}{2\pi\sqrt{n_1 n_2}} \sum_{\mathbf{s} \in D} Z(\mathbf{s}) e^{i\mathbf{s}'\boldsymbol{\lambda}},$$

entonces, el periodograma puede ser obtenido como

$$I(\boldsymbol{\lambda}) = J(\boldsymbol{\lambda})\overline{J(\boldsymbol{\lambda})} = |J(\boldsymbol{\lambda})|^2,$$

donde $\overline{(\cdot)}$ denota al conjugado.

1.1.2. Definición general de estimadores de mínimo contraste

El método de estimación de *mínimo contraste* para modelos paramétricos (o semiparamétricos) también se conoce como estimación *pseudo-verosímil* o *cuasi-verosímil*. Este método sigue el principio subyacente de cualquier método de estimación: el valor del parámetro estimado maximiza algún funcional “pseudo-verosímil” o de contraste, véase Gaetan y Guyon (2010). Este funcional reemplaza la verosimilitud cuando no es posible utilizarla, ya sea porque el modelo es semiparamétrico especificado de forma no completa, o porque la verosimilitud es imposible de calcular.

Este procedimiento de estimación, bajo condiciones apropiadas de regularidad presenta buenas propiedades tales como la consistencia y normalidad asintótica.

De forma general, siguiendo lo comentado en Gaetan y Guyon (2010), un funcional de contraste debería tener las siguientes características:

1. Codificar de manera simple la información de interés del modelo.
2. Resultar simple en su cálculo numérico.
3. Habilitar los parámetros del modelo para ser identificados.
4. Permitir el control de las propiedades estadísticas de los estimadores.

Sea $Y = \{Y(\mathbf{t}), \mathbf{t} \in \mathbb{Z}^d\}$, un campo aleatorio observado sobre una secuencia L_T de dominios finitos crecientes. Suponemos que L_T es un hipercubo: $L_T = [-T, T]^d = \{\mathbf{t} \in \mathbb{Z}^d : -T \leq t_i \leq T, i = 1, \dots, d\}$.

Consideramos un modelo estadístico paramétrico, con una familia de distribuciones $P_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$, donde Θ es un subconjunto compacto de \mathbb{R}^q , $q \in \mathbb{N}$, y el verdadero valor del parámetro $\boldsymbol{\theta}_0 \in \text{int } \Theta$ (el interior de Θ). Denotamos $P_0 = P_{\boldsymbol{\theta}_0}$. Nuestro objetivo es estimar $\boldsymbol{\theta}_0$ a partir de las observaciones

$Y(\mathbf{t}), \mathbf{t} \in L_T$. El estudio asintótico está asociado con la secuencia estrictamente creciente (L_T) de dominios de observación. En general, $\boldsymbol{\theta} \in \Theta$ denota un punto en Θ .

1. Se define **Función de contraste** para $\boldsymbol{\theta}_0$ como una función no aleatoria real-valuada $K(\cdot; \boldsymbol{\theta}_0) : \Theta \mapsto \mathbb{R}$, tal que para cualquier punto del espacio paramétrico $\boldsymbol{\theta} \mapsto K(\boldsymbol{\theta}; \boldsymbol{\theta}_0) \geq 0$, y tiene un único mínimo en $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. El valor de $K(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$ puede ser interpretado como una pseudo-distancia entre el modelo que está bajo $\boldsymbol{\theta}_0$ y el modelo bajo $\boldsymbol{\theta}$.
2. Se define **Proceso o campo aleatorio de contraste**, para una función de contraste $K(\cdot; \boldsymbol{\theta}_0)$ y un conjunto de observaciones $Y(\mathbf{t}), t = 1, \dots, T$, como una secuencia de variables aleatorias o campos aleatorios $U_T(\boldsymbol{\theta}), T \geq 1, \boldsymbol{\theta} \in \Theta$, relacionado con las observaciones $Y(\mathbf{t}), U_T(\boldsymbol{\theta}) = U_T(\boldsymbol{\theta}, Y(\mathbf{t})), \mathbf{t} \in L_T$, definido para todo $\boldsymbol{\theta} \in \Theta$ tal que cumple la siguiente relación:

$$\liminf_{T \rightarrow \infty} [U_T(\boldsymbol{\theta}) - U_T(\boldsymbol{\theta}_0)] = U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}_0) \geq K(\boldsymbol{\theta}; \boldsymbol{\theta}_0), \quad \forall \boldsymbol{\theta} \in \Theta, \quad (1.1)$$

en P_0 -probabilidad.

Esta Condición (1.1) de *subergodicidad* transfiere el hecho de que el valor $U_T(\boldsymbol{\theta}) - U_T(\boldsymbol{\theta}_0)$ que se obtiene al estimar el contraste de $\boldsymbol{\theta}$ con $\boldsymbol{\theta}_0$ en base a $Y(\mathbf{t})$, asintóticamente separa los parámetros. Esta Condición (1.1) se puede fortalecer mediante la condición *ergódica*¹, obteniendo:

$$\lim_{T \rightarrow \infty} [U_T(\boldsymbol{\theta}) - U_T(\boldsymbol{\theta}_0)] = K(\boldsymbol{\theta}; \boldsymbol{\theta}_0), \quad \text{en } P_0\text{-probabilidad.} \quad (1.2)$$

Entonces, se define el **estimador de mínimo contraste** como el valor $\hat{\boldsymbol{\theta}}_T \in \Theta$ que minimiza el funcional U_T , véase Anh et al. (2004a), esto es,

$$\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} U_T(\boldsymbol{\theta}).$$

Obsérvese que la forma habitual de realizar la definición anterior en la práctica es construir, basándose en las observaciones, una función $U_T(\boldsymbol{\theta})$ que conver-

¹La ergodicidad es una propiedad que fortalece la idea de estacionariedad y nos permite obtener casi seguramente la convergencia empírica cuando el dominio de observación tiende a infinito. No siempre resulta fácil comprobar que se cumple, con lo que para probar la consistencia de estimadores se recurre a la propiedad de subergodicidad o incluso a condiciones sobre L^2 . Para más detalle, véase Apéndice B.1 de Gaetan y Guyon (2010).

ge en P_0 -probabilidad a alguna función $U(\boldsymbol{\theta})$ tal que $U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}_0) = K(\boldsymbol{\theta}; \boldsymbol{\theta}_0)$.

Seguidamente, se consideran algunos ejemplos sobre la estimación de mínimo contraste, ver, por ejemplo, Gaetan y Guyon (2010).

Ejemplo 1.1. *Primeramente, se plantea el contraste de mínimos cuadrados para la regresión. Consideramos el modelo de regresión ya sea lineal o no:*

$$X_i = m(Z_i, \theta_0) + \varepsilon_i, \quad i = 1, \dots, n,$$

donde $X_i \in \mathbb{R}$ se expresa a partir de las variables independientes o predictoras Z_i y un error ε_i que forma un ruido blanco de varianza $\sigma^2 < \infty$. Este modelo es semi-paramétrico ya que no se tiene ninguna hipótesis sobre la distribución de los errores excepto que son un ruido blanco. El contraste de mínimos cuadrados estándar se define como:

$$U_n(\theta) = \sum_{i=1}^n (X_i - m(Z_i, \theta))^2.$$

A continuación se define $K(\theta, \theta_0) = \liminf_{n \rightarrow \infty} \sum_{i=1}^n (m(x_i, \theta) - m(x_i, \theta_0))^2 / n$. Si el diseño experimental $\mathcal{Z} = \{Z_i, i = 1, 2, \dots\}$ es tal que $K(\theta, \theta_0) > 0$ para cualquier $\theta \neq \theta_0$, entonces (U_n) es un proceso de contraste asociado a la función de contraste $K(\cdot, \theta_0)$. Por ejemplo, esta condición se satisface cuando:

1. Z_i son i.i.d. cuyas distribuciones son tales que \mathcal{Z} es ergódica.
2. El modelo $\theta \mapsto m(\cdot, \theta)$ es identificable, es decir, que para $\theta \neq \theta'$, $P_{\mathcal{Z}}\{z : m(z, \theta) \neq m(z, \theta')\} > 0$.

Para ver más características de este ejemplo puede consultarse Gaetan y Guyon (2010).

Ejemplo 1.2. *Otro ejemplo que puede resultar interesante es el contraste gaussiano en procesos de segundo orden.*

Consideremos que $X = (X(t), t \in \mathbb{Z})$ es una serie temporal centrada estacionaria de segundo orden con densidad espectral f_θ . El periodograma asociado a las autocovarianzas empíricas, $\hat{r}_n(k)$, a partir de las observaciones $X_n = (X(1), X(2), \dots, X(n))$, es la estimación de la densidad espectral

$$I_n(\lambda) = \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \hat{r}_n(k) e^{i\lambda k},$$

donde $\hat{r}_n(k) = \hat{r}_n(-k) = n^{-1} \sum_{i=1}^{n-|k|} X_i X_{i+k}$.

La estimación de $f(\lambda)$ a partir del periodograma $I_n(\lambda)$ resulta bastante pobre. Sin embargo, se puede utilizar el contraste de Whittle, definido a partir de la siguiente regularización de $I_n(\lambda)$:

$$U_n(\theta) = \frac{1}{2\pi} \int_0^{2\pi} \log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)} d\lambda. \quad (1.3)$$

Este es un buen funcional para la estimación de θ_0 . Si se puede aplicar la hipótesis gaussiana, $-2U_n(\theta)$ se aproxima a la log-verosimilitud. En caso contrario, U_n sigue teniendo buenas propiedades de estimación, bajo condiciones bastante generales, véase Dahlhaus y Künsch (1987) o Guyon (1995). La función de contraste asociada a U_n viene dada por:

$$K(\theta, \theta_0) = \frac{1}{2\pi} \int_0^{2\pi} \log \frac{f_\theta(\lambda)}{f_{\theta_0}(\lambda)} - 1 + \frac{f_\theta(\lambda)}{f_{\theta_0}(\lambda)} d\lambda, \quad K(\theta, \theta_0) > 0 \text{ si } \theta \neq \theta_0.$$

La condición $K(\theta, \theta_0) \neq 0$ si $\theta \neq \theta_0$ se satisface si la parametrización de f_{θ_0} por θ_0 es identificable.

1.1.3. Datos y periodograma tapered

Supongamos que Y es un campo aleatorio centrado y estacionario de segundo orden real-valuado definido sobre una rejilla discreta en \mathbb{Z}^d . Por simplicidad, pensemos que las observaciones $Y(\mathbf{t})$ de Y se dan sobre un hipercubo $D_T = \{1, 2, \dots, T\}^d$, con $T^d = \text{card}(D_T)$ observaciones. Los resultados siguientes son la generalización de aquellos obtenidos para el caso unidimensional en series temporales estacionarias.

Para una distancia $\mathbf{k} \in \mathbb{Z}^d$, la covarianza empírica es:

$$\hat{C}_T(\mathbf{k}) = \frac{1}{T^d} \sum_{\mathbf{t}, \mathbf{t}+\mathbf{k} \in D_T} Y(\mathbf{t})Y(\mathbf{t}+\mathbf{k}), \quad \mathbf{k} \in \mathbb{Z}^d,$$

cuyo soporte para la función $\hat{C}_T(\cdot)$ se da en $\Delta_T = \{\mathbf{i} - \mathbf{j} : \mathbf{i}, \mathbf{j} \in D_T\}$.

En la estadística espacial, los efectos de borde aumentan conforme la dimensión aumenta. Siguiendo Gaetan y Guyon (2010), para la covarianza empírica,

podemos ver que:

$$\lim_{T \rightarrow \infty} \sqrt{T^d} E[\hat{C}_T(\mathbf{k}) - C(\mathbf{k})] = \begin{cases} 0 & \text{si } d = 1 \\ -\{|k_1| + |k_2|\} & \text{si } d = 2 \\ \infty & \text{si } d > 2 \text{ y } \mathbf{k} \neq \mathbf{0} \end{cases}$$

Aunque hay otras maneras de reducir el sesgo, distintas a la utilización de datos *tapered*, estas técnicas tienen efectos contrarios para la estimación de la covarianza, ya que no se garantiza que sea semidefinida positiva, para más detalles véase Gaetan y Guyon (2010). En el Capítulo 2, se consideran los datos *tapered*:

$$\{h_T(\mathbf{t}) Y(\mathbf{t}), \mathbf{t} \in L_T\},$$

donde $h_T(\mathbf{t}) = h(\mathbf{t}/T)$, $\mathbf{t} = (t_1, \dots, t_d)$, $L_T = [-T, T]^d = \{\mathbf{t} \in \mathbb{R}^d : -T \leq t_i \leq T, i = 1, \dots, d\}$, y la función *taper* $h(\mathbf{t})$ se factoriza como $h(\mathbf{t}) = \prod_{i=1}^d \tilde{h}(t_i)$, $t_i \in \mathbb{R}$, con $\tilde{h}(\cdot)$ satisfaciendo la siguiente suposición.

H1. $\tilde{h}(t)$, $t \in \mathbb{R}$, es una función positiva de variación acotada con soporte acotado: $\tilde{h}(t) = 0$ para $|t| > 1$.

Se denota

$$\begin{aligned} \tilde{H}_{k,T}(\lambda) &= \sum_{t=-T}^T [\tilde{h}_T(t)]^k e^{-i\lambda t}, \\ H_{k,T}(\boldsymbol{\lambda}) &= \sum_{\mathbf{t} \in L_T} [h_T(\mathbf{t})]^k e^{-i(\boldsymbol{\lambda}, \mathbf{t})} = \prod_{i=1}^d \tilde{H}_{k,T}(\lambda^{(i)}), \end{aligned} \quad (1.4)$$

donde $\tilde{h}_T(t) = \tilde{h}(t/T)$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$, y k es un número entero positivo.

Obsérvese que la evaluación del comportamiento asintótico de las estimaciones espectrales se basa en las propiedades de las funciones $\tilde{H}_{k,T}(\lambda)$, que, a su vez, se basa en las propiedades de las funciones $\tilde{h}(t)$. Por ejemplo, la suposición de que $\tilde{h}(t)$ es de variación acotada permite escribir los límites superiores útiles para $\tilde{H}_{k,T}(\lambda)$. Ejemplos de estas funciones se pueden ver en la Sección 3 de Brillinger (1970), donde, además, se razona sobre su uso según interese en distintos casos. En lo que sigue, se introducen algunos supuestos adicionales sobre el *tapering*.

Se define la transformada de Fourier finita de los datos *tapered*

$\{h_T(\mathbf{t})Y(\mathbf{t}), \mathbf{t} \in L_T\}$:

$$d_T^h(\boldsymbol{\lambda}) = \sum_{\mathbf{t} \in L_T} h_T(\mathbf{t}) Y(\mathbf{t}) e^{-i(\boldsymbol{\lambda}, \mathbf{t})}, \boldsymbol{\lambda} \in \mathbb{T},$$

y el periodograma *tapered* de segundo orden (siempre que $H_{2,T}(\mathbf{0}) \neq 0$):

$$I_T^h(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^d H_{2,T}(\mathbf{0})} d_T^h(\boldsymbol{\lambda}) d_T^h(-\boldsymbol{\lambda}). \quad (1.5)$$

En la práctica, para los resultados numéricos, el periodograma se calcula para las frecuencias de Fourier. Para que las condiciones de independencia e insesgadez se cumplan será necesario que T^d sea suficientemente grande. Por lo tanto, estas características se cumplen cuando se trabaja con una rejilla densa, cosa que en la práctica no suele ocurrir. Esto hace que el error al utilizar el periodograma como estimador de la densidad espectral no pueda tratarse como un ruido aleatorio. Será interesante poder medir de algún modo, cuánta información se pierde bajo el supuesto de independencia. Para ello, se puede considerar el valor esperado y la covarianza del periodograma *tapered* para una rejilla finita D .

Esperanza y covarianza del periodograma a partir de cumulantes

De la misma manera que la función generatriz de momentos se utiliza para generar los momentos de una variable aleatoria, el logaritmo de la función generatriz de momentos se utiliza para generar los cumulantes. Los cumulantes tienen propiedades que resultan interesantes, son simétricos y multilineales en sus argumentos, véase Brillinger (1970). Si cualquier subconjunto de $\{Y_1, \dots, Y_r\}$ es estadísticamente independiente del conjunto restante, entonces, su cumulante conjunto, $cum(Y_1, \dots, Y_r) = 0$. Así pues, los cumulantes proporcionan medidas útiles para el estudio de la dependencia estadística conjunta de variables aleatorias. Todo conjunto de cumulantes de una colección de variables independientes serán nulos.

Sea (Y_1, \dots, Y_r) una variable aleatoria multivariante de dimensión r , con $E|Y_j|^r < \infty$, $j = 1, \dots, r$, el cumulante conjunto de orden r , $cum(Y_1, \dots, Y_r)$ viene dado por el coeficiente de $t_1 \cdots t_r$, en la expansión de la función generatriz de cumulantes

$$\log E \left(e^{\sum_{j=1}^r Y_j t_j} \right).$$

Una definición alternativa viene dada por

$$\text{cum}(Y_1, \dots, Y_r) = \sum (-1)^{p-1} (p-1)! E \left(\prod_{j \in \nu_1} Y_j \right) \dots E \left(\prod_{j \in \nu_p} Y_j \right),$$

donde la sumatoria se extiende a todas las particiones (ν_1, \dots, ν_p) , $p = 1, \dots, r$.

El valor esperado del periodograma *tapered* está condicionado por el tamaño de la rejilla de observación y el nivel de disminución que se haga al aplicar la función *taper* en los datos. Bajo ciertas condiciones de regularidad sobre la función *taper* y el orden polinomial de su caída a cero, $E(I_T^h(\boldsymbol{\lambda})) = f(\boldsymbol{\lambda}) + o(T^{-1})$, véase, por ejemplo, [Dahlhaus y Polonik \(2002\)](#). Así pues, el periodograma *tapered* puede funcionar como estimador de la densidad espectral cuando T es suficientemente grande, ya que asintóticamente es un estimador insesgado, véase, por ejemplo, [Brillinger \(1970\)](#), entre otros.

Podemos utilizar la teoría de cumulantes para obtener una expresión más concreta de la esperanza del periodograma para el caso bidimensional. Consideremos Z , un campo aleatorio estacionario de segundo orden, con media μ y función de covarianzas C . Tomemos una rejilla de tamaño T^2 como dominio de observación del campo aleatorio Z . Sea I_T^h el periodograma *tapered* definido en la ecuación (1.5). Se asume también que $\sum_{\mathbf{h} \in \mathcal{H}} |C(\mathbf{h})| < \infty$. Entonces, para $\boldsymbol{\lambda} \in \pi^2$,

$$E[I_T^h(\boldsymbol{\lambda})] = \left((2\pi)^2 \int_{\pi^2} |H_{1,T}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \right)^{-1} \times \left[\int_{\pi^2} |H_{1,T}(\boldsymbol{\lambda} - \boldsymbol{\omega})|^2 f(\boldsymbol{\omega}) d\boldsymbol{\omega} + |H_{1,T}(\boldsymbol{\omega})|^2 \mu^2 \right],$$

donde H viene definida en (1.4).

Siguiendo el mismo desarrollo anterior, podemos ver la expresión para la covarianza asociada con el periodograma *tapered* donde se refleja que dicha covarianza entre valores del periodograma no es estacionaria. Consideremos Z un campo aleatorio estacionario de segundo orden, con media μ y función de covarianzas C . Tomamos una rejilla de tamaño T^2 como dominio de observación del campo aleatorio Z . Sea I_T^h el periodograma *tapered* definido en la ecuación (1.5). Entonces, siguiendo [Crujeiras \(2006\)](#), la covarianza asociada al

periodograma *tapered* viene dada por

$$\begin{aligned} Cov(I_T^h(\boldsymbol{\lambda}), I_T^h(\boldsymbol{\omega})) &= |H_{2,T}(\mathbf{0})|^{-2} f(\boldsymbol{\lambda})^2 \\ &\times [|H_{2,T}(\boldsymbol{\lambda} - \boldsymbol{\omega})|^2 + |H_{2,T}(\boldsymbol{\lambda} + \boldsymbol{\omega})|^2] + O(N^{-1}), \end{aligned}$$

donde H viene definida en (1.4).

En la Sección 2.7.2 de Crujeiras (2006) se puede ver la obtención de los resultados anteriores, tanto para la esperanza del periodograma *tapered* como para la covarianza.

1.1.4. Ejemplo de estimación a partir de datos *tapered*

Finalmente, a modo de ejemplo, se puede reflejar la metodología seguida para la obtención de propiedades asintóticas importantes en la estimación de mínimo contraste, siguiendo lo comentado en la Sección 5.3.1 de Gaetan y Guyon (2010). Para solventar las dificultades en la obtención de propiedades buenas asintóticas, debidas al sesgo y la falta de regularidad en la estimación de la covarianza, vamos a utilizar una función *taper* adecuada que, para una dimensión menor o igual a tres, haga desaparecer los problemas descritos y así lograr mantener su eficiencia. Además, este método ayuda a mejorar los análisis estadísticos al disminuir los pesos de los puntos límite, que a menudo no son representativos del modelo bajo estudio. En este caso, vamos a seguir la función *taper* propuesta por Tukey-Hanning:

$$h(u) = \begin{cases} \frac{1}{2}(1 - \cos \frac{2\pi u}{\rho}) & \text{si } 0 \leq u \leq \frac{\rho}{2} \\ 1 & \text{si } \frac{\rho}{2} \leq u \leq \frac{\rho}{2} \\ h(1 - u) & \text{si } \frac{\rho}{2} \leq u \leq \rho \end{cases}$$

En la Figura 1.1, véase Guyon (1995), se puede observar de forma gráfica cómo la utilización de una función *taper* disminuye los valores obtenidos en el borde. Esta función se aplica a los datos para así obtener los datos *tapered* $\{h_T(\mathbf{t})Y(\mathbf{t}), \mathbf{t} \in L_T\}$ de la siguiente forma:

$$h_T(\mathbf{t})Y(\mathbf{t}) = a_T(\mathbf{t})Y(\mathbf{t}), \text{ donde } a_T(\mathbf{t}) = \prod_{k=1}^d h\left(\frac{i_k - 1/2}{T}\right), \quad \mathbf{t} \in L_T.$$

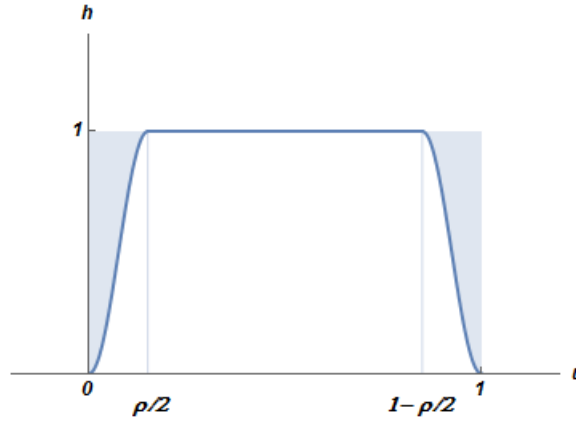


Figura 1.1: Función *taper* de Tuckey-Hanning, véase Guyon (1995)

La covarianza empírica *tapered* \hat{C}_T^h se puede obtener utilizando los datos *tapered* $\{h_T(\mathbf{t})Y(\mathbf{t}), \mathbf{t} \in L_T\}$. Para $d \leq 3$ y una elección de $\rho_T = o(T^{1/4})$, el sesgo de un estimador *tapered* es despreciable, véase Dahlhaus y Künsch (1987) o el Capítulo 4 de Guyon (1995).

Supongamos que Y se parametriza por su densidad espectral f_{θ_0} , donde θ_0 está en el interior de un espacio paramétrico compacto $\Theta \in \mathbb{R}^p$. El periodograma *tapered* viene expresado por

$$I_T^h(\boldsymbol{\lambda}) = \frac{1}{(2\pi)^d} \sum_{\mathbf{k} \in L_T} \hat{C}_T^h(\mathbf{k}) e^{i\boldsymbol{\lambda}'\mathbf{k}}.$$

Esta expresión es la estimación de la densidad espectral, $f_{\theta_0}(\boldsymbol{\lambda})$, asociada a las covarianzas empíricas *tapered*, para más información se puede consultar la Sección 4.1.1 de Guyon (1995). Al igual que en series temporales, para una frecuencia fijada $\boldsymbol{\lambda}$, $I_T^h(\boldsymbol{\lambda})$ es un estimador pobre de $f_{\theta_0}(\boldsymbol{\lambda})$. Sin embargo, en su forma integral, el periodograma *tapered* nos lleva a buenas estimaciones. En particular, para un campo aleatorio gaussiano Y , una buena aproximación a la log-verosimilitud de $Y(\mathbf{t})$ es, salvo constantes aditivas, igual a $-2U_T(\boldsymbol{\theta}_0)$, donde

$$U_T(\boldsymbol{\theta}) = \frac{1}{(2\pi)^d} \int_{\mathbb{T}} \left\{ \log f_{\boldsymbol{\theta}}(\boldsymbol{\lambda}) + \frac{I_T^h(\boldsymbol{\lambda})}{f_{\boldsymbol{\theta}}(\boldsymbol{\lambda})} \right\} d\boldsymbol{\lambda}, \quad \mathbb{T} = [0, 2\pi)^d,$$

(Gaetan y Guyon, 2010). Además, ya sea Y un campo aleatorio gaussiano o no, minimizar U_T lleva a una buena estimación de $\boldsymbol{\theta}_0$, bajo hipótesis razonables. A la expresión $-2U_T(\boldsymbol{\theta}_0)$ se le conoce como el *proceso de contraste de Whittle*

o proceso de contraste gaussiano de Y . Consideremos el estimador de mínimo contraste $\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} U_T(\boldsymbol{\theta})$. Entonces, bajo ciertas condiciones de regularidad, si Y es un campo aleatorio gaussiano estacionario, $\hat{\boldsymbol{\theta}}_T \xrightarrow{P} \boldsymbol{\theta}_0$. Además, para dimensión menor o igual a tres, $T^{d/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$ tiende en distribución a una normal multivariante centrada. Para más detalles, véase Gaetan y Guyon (2010), donde, además, se contempla el caso de campos aleatorios no gaussianos.

1.1.5. Sobre los estimadores Whittle con datos tapered

En este apartado se hace un breve repaso de algunos resultados de la literatura sobre los estimadores de Whittle basados en datos tapered, con un enfoque particular en las condiciones de las funciones taper que ayuden a controlar el sesgo de los estimadores.

Caso de campos aleatorios débilmente dependientes

Como ya se ha comentado, para realizar este apartado se han seguido los resultados a partir de Guyon (1995). Uno de los primeros trabajos donde se aborda el problema de los efectos de borde se da en Guyon (1982). Considerando las estimaciones paramétricas habituales de Whittle para los datos tapered, encontró que tales estimaciones tenían sesgo de orden $N^{-1/d}$, para campos observados en un rectángulo $P_N = \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_d\}$ en Z^d , siendo $N = \prod_{i=1}^d n_i$, o del orden n^{-1} si P_N es un cubo de borde n . Es decir, para $d \geq 2$, el sesgo es del mismo orden o de un orden superior que la desviación típica, que usualmente es $O(N^{-1/2})$. Una posible solución para el problema descrito de los efectos de borde es la disminución progresiva de los datos en los bordes del dominio de observación (tapering).

Revisamos los resultados que aparecen en Guyon (1982, 1995). Allí se considera que el dominio L_T es de la forma $L_T = \prod_{i=1}^d [1, T_i]$. Sin embargo, en nuestro caso, simplificamos lo anterior y consideramos el cubo, es decir, consideramos $T_i = T$ y $L_T = [1, T]^d$.

En lugar de los datos observados $\{Y(\mathbf{t}), \mathbf{t} \in L_T\}$, se consideran los datos tapered $\{h_T(\mathbf{t})Y(\mathbf{t}), \mathbf{t} \in L_T\}$, donde las funciones taper son de la siguiente

forma:

$$h_T(\mathbf{t}) = h\left(\frac{\mathbf{t} - 1/2}{T}\right) = \prod_{i=1}^d \tilde{h}\left(\frac{t_i - 1/2}{T}\right),$$

y

$$\tilde{h}(u) = \begin{cases} g\left(\frac{2u}{\rho}\right) & \text{si } 0 \leq u \leq \frac{1}{2}\rho \\ 1 & \text{si } \frac{1}{2}\rho \leq u \leq \frac{1}{2} \\ \tilde{h}(1-u) & \text{si } \frac{1}{2} \leq u \leq 1 \end{cases}$$

donde la función $g : [0, 1] \rightarrow [0, 1]$ es tal que $g(0) = 0$ y $g(1) = 1$. Esta función es creciente y pertenece a \mathcal{C}^2 , con $0 \leq \rho \leq 1$. De esta forma, se obtiene el denominado *g-taper*, el cual disminuye un $100(1 - \rho)\%$ los valores del borde.

A continuación se formulan los resultados sobre el control del sesgo para los estimadores *tapered* indicados por Guyon; véase, [Guyon \(1995\)](#).

Consideremos $J(\phi) = \int_{\mathbb{T}} f(\boldsymbol{\lambda})\phi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$, y el correspondiente estimador basado en los datos *tapered* $J_T(\phi) = \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda})\phi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$.

Proposición 1.1. *Sea $\phi \in \mathcal{C}(\mathbb{T})$, supongamos que h y la densidad espectral, f , pertenecen a \mathcal{C}^2 . Entonces, cuando $T \rightarrow \infty$,*

$$E [J_T(\phi) - J(\phi)] = CT^{-2} (1 + o(1)),$$

donde C es una constante.

Por lo tanto, bajo las condiciones anteriores, si ρ está fijo, el sesgo es de orden T^{-2} , en consecuencia menor que $T^{-d/2}$, para $d = 1, 2, 3$. Si $\rho = \rho_T \rightarrow 0$, C se comporta como ρ_T^{-1} de manera que el sesgo es de orden $o\left(T^{-2+\frac{1}{4}}\right)$, el cual todavía es más pequeño que $T^{-d/2}$. Esto es, para $d = 1, 2, 3$,

$$\lim_{T \rightarrow \infty} E [T^{d/2}(J_T(\phi) - J(\phi))] = 0.$$

Además, si $\phi_1, \phi_2 \in \mathcal{C}(\mathbb{T})$, la función *taper*, h , y la densidad espectral, f , pertenecen a \mathcal{C}^2 , entonces, para un campo aleatorio gaussiano tenemos el siguiente comportamiento asintótico de la covarianza:

$$\lim_{T \rightarrow \infty} T^d \text{Cov} (J_T(\phi_1), J_T(\phi_2)) = 2(2\pi)^d e(h) \int_{\mathbb{T}} \phi_1(\boldsymbol{\lambda})\phi_2(\boldsymbol{\lambda})f^2(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$$

donde el factor *taper* $e(h)$ viene dado por (2.9), $e(h) \geq 1$, y $e(h) = 1$, si no se produce *tapering*. Se puede elegir un ρ_T tal que el factor *taper* tienda a 1.

Nota 1.1. En vista de la anterior Proposición 1.1, con el fin de controlar el sesgo para el estimador de Whittle, se pueden usar las siguientes condiciones:

la función taper $\tilde{h}(t)$ está en $C^2[-1, 1]$; la densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ está en $C^2(\mathbb{T})$ y la función $\frac{\partial}{\partial \theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ es continua con respecto a $(\boldsymbol{\lambda}, \boldsymbol{\theta})$.

Cabe señalar que la normalidad asintótica de los estimadores de Whittle se consideró en Guyon (1995) bajo varias condiciones.

Caso de campos aleatorios fuertemente dependientes

Para realizar este apartado se ha seguido los resultados desarrollados en Ludeña y Lavielle (1999). En dicho artículo, se consideran los siguientes dos modelos de densidades espectrales con singularidades:

A1 Existen funciones $\alpha_i : \Theta \rightarrow (0, 1)$, $1 \leq i \leq d$, tales que tenemos $f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = f_0(\boldsymbol{\lambda}, \boldsymbol{\theta}) \prod_{i=1}^d f_i(\lambda_i, \boldsymbol{\theta})$ donde asumimos, para todo $\delta > 0$, que $f_i(\lambda_i, \boldsymbol{\theta}) = O(|\lambda_i|^{-\alpha_i(\boldsymbol{\theta})-\delta})$, $|\lambda_i| \rightarrow 0$ para cada $i = 1, \dots, d$, y que $f_0(\boldsymbol{\lambda}, \boldsymbol{\theta})$ es una función de $(\boldsymbol{\lambda}, \boldsymbol{\theta})$ positiva con segunda derivada continua.

A2 En este caso, suponemos que existe una función $\alpha : \Theta \rightarrow (0, d)$, tal que la densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ satisface, para todo $\delta > 0$: $f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \|\boldsymbol{\lambda}\|^{-\alpha(\boldsymbol{\theta})+\delta}(1 + o(1))$, donde el $o(1)$ es uniforme con respecto a $\boldsymbol{\lambda}$. Para más detalles, véase Ludeña y Lavielle (1999).

Los autores del citado artículo estudian el estimador de Whittle $\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} U_T(\boldsymbol{\theta})$ para ambos modelos **A1** y **A2** usando los datos tapered y el funcional empírico de contraste $U_T(\boldsymbol{\theta})$, que viene dado por (1.3), donde la función $w(\boldsymbol{\lambda})$, llamada por los autores función de suavizamiento o regularización, depende de la forma de singularidad. Bajo **A1**, $w(\boldsymbol{\lambda}) = \prod_{i=1}^d w_i(\lambda_i)$, $w_i(\lambda_i) = |\lambda_i|^\nu$, o bajo **A2**, $w(\boldsymbol{\lambda}) = \mathbb{1}_{\mathbb{T}}(\boldsymbol{\lambda})$, esto es, la función indicadora en el d -toro.

La función taper $h_{T,\varepsilon}(t) = \prod_{i=1}^d \tilde{h}_\varepsilon\left(\frac{t_i}{T}\right)$, donde $\tilde{h}_\varepsilon : [0, 1] \rightarrow [0, 1]$ es de la siguiente forma:

$$h_\varepsilon(u) = \begin{cases} u/\varepsilon & \text{si } u \leq \varepsilon \\ 1 & \text{si } \varepsilon \leq u \leq 1 - \varepsilon \\ h_\varepsilon(1 - u) & \text{si } u > 1 - \varepsilon \end{cases}$$

con $\varepsilon = T^{-\gamma}$.

En el curso de la deducción de la normalidad asintótica para el estimador de Whittle, se obtuvo el siguiente resultado sobre el sesgo.

Proposición 1.2. *Supongamos que se cumplen **A1** o **A2**, y que $w(\boldsymbol{\lambda}) = \prod_{i=1}^d w_i(\lambda_i)$ bajo **A1**, o $w(\boldsymbol{\lambda}) = \mathbb{1}_{\mathbb{T}}(\boldsymbol{\lambda})$ bajo **A2**. Además, se considera que $\gamma < (1 + \nu - d/2)/\nu$ bajo **A1**, o $\gamma < 2 - d/2$ bajo **A2**. Entonces, cuando $T \rightarrow \infty$*

$$E[J_T(\varphi) - J(\varphi)] = O(T^{-(1+\kappa(1-\gamma))}),$$

donde $\kappa = \nu$ bajo **A1**, o $\kappa = 1$ bajo **A2**, y por lo tanto, el sesgo es de orden $o(T^{-d/2})$, si $d \leq 3$.

Obsérvese que en el desarrollo del resultado anterior, se utilizó esencialmente la forma de φ (según se necesitaba para un caso particular de la función de Whittle). Con esta función, y bajo las condiciones de la Proposición 1.2, se obtuvieron algunas propiedades de regularidad de convolución, $\int_{\mathbb{T}} \varphi(\boldsymbol{\lambda} - \boldsymbol{\mu}) f(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$, y se usaron para la evaluación del sesgo, así como para una interacción entre una función *taper* y la función $w(\boldsymbol{\lambda})$ bajo **A1**. La estructura multiplicativa de las funciones que definen $J(\varphi)$ también fue esencial en **A1**.

Cabe señalar que la normalidad asintótica de los estimadores de Whittle fue establecida en Ludeña y Lavielle (1999), bajo la Suposición **A1**, utilizando el teorema central del límite, de forma análoga a Fox y Taqqu (1987). También, bajo la Suposición **A2**, y utilizando el teorema central del límite, se estableció la normalidad asintótica en Doukhan et al. (1996).

1.2. Preliminares para los capítulos 3, 6 y 7

En las secciones restantes, consideraremos que todas las variables aleatorias introducidas a continuación se definen en el espacio de probabilidad básico (Ω, \mathcal{A}, P) .

1.2.1. Procesos de Poisson doblemente estocásticos

Dentro de los procesos de Poisson hay una gran variedad de tipos. El caso homogéneo es el más simple de ellos, sin embargo, resulta insuficiente para modelizar la mayoría de casos reales. Para ampliar más las posibilidades del proceso de Poisson se utilizan los procesos de Poisson doblemente estocásticos, donde la intensidad no solo es no homogénea sino que a su vez es un proceso estocástico. A estos procesos también se les conoce como procesos de Cox, quien los introdujo en su artículo, Cox (1955). Un proceso de Poisson doblemente estocástico puede ser visto como un procedimiento de aleatorización en dos pasos. Un proceso λ_t , $t > 0$, se utiliza para generar otro proceso, N_t , actuando como su intensidad. En particular, se supone que N_t es un proceso de Poisson condicionado a λ_t . Se pueden dar muchas posibles definiciones de un proceso de Poisson doblemente estocástico. De todas las posibles, se ha seguido la utilizada en la Sección 2.1, pp. 2–3, en Møller (2005), por ser muy general, bien fundamentada y adaptada al marco espacial.

Primeramente, se define un proceso puntual espacial en \mathbb{R}^d como un conjunto aleatorio localmente finito $\mathbf{X} \subset \mathbb{R}^d$, lo que significa que para cualquier conjunto de Borel acotado $B \subset \mathbb{R}^d$, el número de puntos en $\mathbf{X}_B = \mathbf{X} \cap B$ es una variable aleatoria finita denotada por $N(B)$.

Se dice que \mathbf{X} es estacionario, o respectivamente isotrópico, si su distribución es invariante bajo traslaciones en \mathbb{R}^d , o respectivamente rotaciones sobre el origen en \mathbb{R}^d .

Para cualquier entero $n \geq 1$, la medida del momento factorial de orden n ,

$\mu^{(n)}$, de \mathbf{X} se define como

$$\mu^{(n)}(A) = E \sum_{u_1, \dots, u_n \in \mathbf{X}}^{\neq} \mathbf{1}[(u_1, \dots, u_n) \in A]$$

para conjuntos de Borel $A \in \mathbb{R}^{dn}$, donde \neq sobre la suma significa que la suma recorre todos los puntos distintos, dos a dos, u_1, \dots, u_n en \mathbf{X} , y $\mathbf{1}[\cdot]$ es la función indicadora. Se asume que $\mu^{(n)}$ es localmente finita y tiene función de densidad $\rho^{(n)}$ con respecto a la medida de Lebesgue en \mathbb{R}^{dn} , donde llamamos a $\rho^{(n)}$ la densidad producto de orden n . La medida de momento es $\mu(A) = \mu^{(1)}(A) = E[N(A)]$, y $\rho = \rho^{(1)}$ es la *función de intensidad*. Intuitivamente, si $u_1, \dots, u_n \in \mathbb{R}^d$ son distintos dos a dos, $\rho^{(n)}(u_1, \dots, u_n) du_1 \cdots du_n$ es la probabilidad de observar la ocurrencia conjunta de n puntos de \mathbf{X} en cada n regiones infinitesimales con áreas du_1, \dots, du_n , conteniendo u_1, \dots, u_n .

Una extensión natural del proceso de Poisson es considerar un proceso estocástico no negativo $\Lambda = (\Lambda(u))_{u \in \mathbb{R}^d}$ tal que \mathbf{X} condicionado a Λ es un proceso de Poisson con intensidad Λ , Cox (1955). Entonces, decimos que \mathbf{X} es un proceso de Cox (o doblemente estocástico) con intensidad Λ .

Como extensión inmediata de las propiedades de un proceso de Poisson $\mathbf{X} \mid \Lambda$, se tiene que \mathbf{X} es estacionario o isotrópico si Λ es estacionaria o isotrópica respectivamente. Además, la densidad producto

$$\rho^{(n)}(u_1, \dots, u_n) = E[\Lambda(u_1) \cdots \Lambda(u_n)],$$

de manera que

$$\rho(u) = E[\Lambda(u)], \quad g(u, v) = \frac{E[\Lambda(u)\Lambda(v)]}{\rho(u)\rho(v)}.$$

La función de correlación par se define como $g(u, v) = \rho^{(2)}(u, v)/(\rho(u)\rho(v))$, donde $\rho(u) > 0$ y $\rho(v) > 0$, por ser medias de variables positivas. Este tipo de normalización es útil, ya que cuando $g \equiv 1$ estamos en el caso de no interacción, cuando $g > 1$ se interpreta como atracción entre los puntos del procesos ubicados en u y v , y si $g < 1$ se interpreta como repulsión entre las ubicaciones. La estacionariedad de \mathbf{X} implica que podemos asumir $\rho(u)$ como constante y $g(u, v) = g(u - v)$ como invariante ante traslaciones. Si se cumple la isotropía, entonces, $g(u, v) = g(\|u - v\|)$ depende solo de la distancia entre u y v .

Consideremos \mathbf{X} como un proceso de Cox con intensidad Λ . Supongamos que

$$\log \Lambda(u) = \Psi(u), \quad u \in \mathbb{R}^d$$

donde $\Psi = \Psi(u), u \in \mathbb{R}^d$, es un proceso gaussiano con media $\xi(u) = E[\Psi(u)]$ y función de covarianza $C(u, v) = Cov(\Psi(u), \Psi(v))$. Entonces llamamos a \mathbf{X} un *proceso de Cox log-gaussiano*. Es decir, un proceso de Cox donde el logaritmo de la intensidad aleatoria es un proceso gaussiano.

Para asegurar la integrabilidad local de $\Lambda(u)$ se tienen que cumplir ciertas condiciones suaves, como que $\Lambda(u)$ posea realizaciones o trayectorias no negativas y localmente integrables casi seguramente (Møller, 2005).

La clase de procesos estacionarios de Cox log-gaussianos posee varias propiedades atractivas:

- La distribución se caracteriza completamente por la intensidad y la función de correlación par del proceso de Cox. Esto hace que los modelos paramétricos sean fáciles de interpretar y se disponga de métodos sencillos para la estimación paramétrica y la verificación de modelos.
- Las propiedades teóricas son fácilmente extraídas. Por ejemplo, los momentos de orden superior se expresan simplemente por la intensidad y la función de correlación par del proceso de Cox log-gaussiano.
- El proceso gaussiano subyacente y la intensidad aleatoria pueden predecirse a partir de la realización de un proceso de Cox gaussiano observado dentro de una ventana limitada usando métodos bayesianos.
- No hay ningún problema con los efectos de borde ya que la distribución de un proceso de Cox log-gaussiano restringido a un subconjunto acotado es conocida.

Ver pp. 451–452, Møller et al. (1998), para más detalles.

Estas propiedades definen suficientemente los procesos de Cox log-gaussianos. Además, se puede observar que dichos procesos son modelos flexibles, se les puede aplicar transformaciones, trabajar con técnicas *cluster*, utilizar fácilmente la simulación. Así mismo, la definición univariante de dichos procesos puede ser extendida de forma natural al caso multivariante, véase Møller et al. (1998).

Consideremos el subconjunto aleatorio localmente finito $\mathbf{X} \subset \mathbb{R}^d$, proceso de Cox con proceso de intensidad aleatoria $\Lambda = \Lambda(s) : s \in \mathbb{R}^d$, es decir, la distribución condicionada de \mathbf{X} dada Λ es un proceso de Poisson con función de intensidad $\Lambda(\cdot)$. Por lo tanto, según se ha comentado anteriormente, para un conjunto de Borel acotado $B \subset \mathbb{R}^d$ se tiene que $\mathbf{X}_B = \mathbf{X} \cap B$ condicionado a Λ se distribuye como una Poisson de media $\int_B \Lambda(s) ds$,

1.2.2. Procesos de Poisson homogéneos en el espacio ℓ^2

Se introducen a continuación, los procesos de Poisson en el espacio ℓ^2 , su predicción y estimación paramétrica clásica; véase [Bosq y Ruiz-Medina \(2014\)](#).

Sea $\{N_{t,j}, t \in \mathbb{R}_+, j \geq 1\}$ una secuencia de procesos de Poisson homogéneos independientes con intensidades respectivas $\lambda_j > 0, j \geq 1$, tales que $\sum_j \lambda_j < \infty$. Ya que

$$E \left(\sum_j N_{t,j}^2 \right) = \sum_j E(N_{t,j}^2) = \sum_j [\lambda_j t + (\lambda_j t)^2] < \infty,$$

se sigue que $\sum_j N_{t,j}^2 < \infty$ casi seguramente (c. s.). Entonces, $M_t = \{N_{t,j}, j \geq 0\}$ define una variable aleatoria con valores en ℓ^2 , y satisface $E\|M_t\|_{\ell^2}^2 < \infty$. Así, $\{M_t, t \geq 0\}$ es un proceso en tiempo continuo ℓ^2 -valuado.

Predicción clásica

Asumamos que se observa M_T y se quiere predecir M_{T+h} ($h > 0$). Para este propósito, primero se define la estimación por máxima verosimilitud (EMV) para $\{\lambda_j, j \geq 1\}$. Sea $\mathcal{N} \subset \ell^2$ la familia de secuencias $\{x_j, j \geq 1\} = (x_j)$ tales que x_j es un entero para cada j , y $x_j = 0$, para un j suficientemente grande. Se puede escribir

$$\mathcal{N} = \bigcup_k \mathcal{N}_k,$$

donde $\mathcal{N}_k = \{(x_j) : (x_1, \dots, x_k) \in \mathbb{N}^k; x_j = 0, j > k\}$. Claramente, \mathcal{N}_k es numerable para cada k . Entonces se sigue que \mathcal{N} es numerable ya que se trata de una unión numerable de conjuntos numerables. Así, se puede definir la medida de conteo como μ sobre \mathcal{N} , y extenderlo a ℓ^2 mediante $\mu(\ell^2 - \mathcal{N}) = 0$. Entonces, la medida obtenida es σ -finita.

Se considera M_t como variable aleatoria \mathcal{N} -valuada. En realidad, puesto que $N_{t,j}^2$ es un entero, $\sum_j N_{t,j}^2 < \infty$ (c. s.) implica que $N_{t,j} = 0$ c. s., para un j suficientemente grande. De forma más precisa, existe Ω_0 tal que $P(\Omega_0) = 1$, y para todo $\omega \in \Omega_0$, existe un $j_0(\omega, \lambda, T)$ tal que $N_{T,j}(\omega) = 0$, para $j > j_0$. Por lo tanto, se puede definir la verosimilitud de M_T con respecto a μ estableciendo

$$L(M_T(\omega), \lambda) = \prod_{j=1}^{j_0(\omega, \lambda, T)} \exp(-\lambda_j T) \frac{(\lambda_j T)^{N_{T,j}(\omega)}}{N_{T,j}(\omega)!}.$$

Así pues, la EMV de (λ) viene dado por:

$$(\hat{\lambda})_T = \left(\frac{N_{T,j}}{T}, j \geq 1 \right) = (\hat{\lambda}_{T,j}, j \geq 1). \quad (1.6)$$

Claramente, $\sum_{j=1}^{\infty} \hat{\lambda}_{T,j} < \infty$ (c. s.) y $(\hat{\lambda})_T$ es insesgado (considerando (λ) como un parámetro con valores en ℓ^2). Se sigue que

$$f(M_T) = \frac{T+h}{T} M_T \quad (1.7)$$

es un predictor insesgado eficiente de $E_{\lambda}(M_{T+h}|M_T) = h(\lambda) + M_T$, $(\lambda) \in \ell^2$; véase Bosq y Ruiz-Medina (2014).

1.2.3. El proceso de Wiener fuertemente cilíndrico inducido por un proceso de Wiener Hilbert-valuado

Sea H y H^* un espacio de Hilbert real separable y su dual, respectivamente. Se denota por $\mathcal{B}(H)$ la σ -álgebra de Borel en H . Para $f_1^*, \dots, f_n^* \in \Gamma \subseteq H^*$, y $B \in \mathcal{B}(\mathbb{R}^n)$, se define un conjunto cilíndrico o cilindro con respecto a (H, Γ) como sigue:

$$\mathcal{Z}(f_1^*, \dots, f_n^*, B) := \{g \in H : (\langle g, f_1^* \rangle, \dots, \langle g, f_n^* \rangle) \in B\}.$$

El conjunto de todos los conjuntos cilíndricos se denota como $Z(H, \Gamma)$, el cual resulta ser un álgebra. La generada σ -álgebra se denota por $C(H, \Gamma)$, y se denomina σ -álgebra cilíndrica con respecto a (H, Γ) . Si $\Gamma = H^*$, escribimos $C(H) := C(H, H^*)$. Ya que H es separable, se tiene que la σ -álgebra de Borel sobre H y la σ -álgebra cilíndrica $C(H)$ coinciden.

Una función $\mu : C(H) \rightarrow [0, \infty]$ se dice medible cilíndrica en $C(H)$ si, para cada subconjunto finito $\Gamma \subseteq H^*$, la restricción de μ sobre la σ -álgebra $C(H, \Gamma)$ es una medida. Una medida cilíndrica se dice finita si $\mu(H) < \infty$.

Un proceso cilíndrico X en H es una familia $\{X(t), t > 0\}$ de variables aleatorias cilíndricas en H . Nótese, que la función característica $\varphi_X : H^* \rightarrow \mathbb{C}$, con $\varphi_X(f^*) = E[\exp(iX(f^*))]$, $f^* \in H^*$, de una variable aleatoria cilíndrica es definida positiva y continua en subespacios finitos. Entonces, existe una medida cilíndrica μ_X con la misma función característica. Llamamos μ_X a la distribución cilíndrica de X . El recíproco también es cierto. Una medida cilíndrica fuertemente gaussiana μ , tiene como función característica φ dada por

$$\varphi(f^*) = \exp\left(-\frac{1}{2} \langle Qf^*, f^* \rangle\right), \quad \forall f^* \in H^*,$$

para cierto operador positivo simétrico $Q : H^* \rightarrow H$. Un proceso de Wiener débilmente cilíndrico, $\{W(t) : t > 0\}$, es fuertemente cilíndrico si la distribución cilíndrica $\mu_{W(1)}$ de $W(1)$ es fuertemente gaussiana; véase, por ejemplo, la Definición 6.7 en Riedle (2011).

Un proceso estocástico adaptado H -valuado $\{W(t) : t > 0\}$ se dice proceso de Wiener si:

- (a) $W(0) = 0$ casi seguramente con respecto a una medida de probabilidad P ;
- (b) W tiene incrementos estacionarios independientes;
- (c) existe Q , un operador positivo simétrico tal que $Q : H^* \rightarrow H$, y $W(t) - W(s) \sim \mathcal{N}(0, (t - s)Q)$, para todo $0 \leq s \leq t$.

Un proceso de Wiener fuertemente cilíndrico $\{W(t) : t > 0\}$, con operador de covarianza $Q = i_Q i_Q^*$, es inducido por un proceso de Wiener H -valuado si y solo si i_Q es Hilbert-Schmidt, donde i_Q denota la inclusión continua del rango de Q , con respecto a la norma $\|f\|_Q = Q(f)(f) = \langle Q(f^*), f \rangle_H$, $f \in H$, en H ; véase el Teorema 8.1 en Riedle (2011), para una caracterización más amplia de esta clase de proceso de Wiener fuertemente cilíndricos.

1.2.4. El proceso Ornstein-Uhlenbeck en espacios de Hilbert

Sea H un espacio de Hilbert real separable. Consideremos el proceso de Ornstein-Uhlenbeck (O-U), definido por

$$X(t, f) = \exp(tA)(f) + \int_0^t \exp((t-s)A)dW(s), \quad t \geq 0, \quad f \in H, \quad (1.8)$$

donde W es un proceso de Wiener fuertemente cilíndrico en H , inducido por un proceso de Wiener H -valuado; véase el Teorema 8.1 en Riedle (2011), con operador de covarianza $Q : H^* \rightarrow H$, tal que, para $0 \leq s \leq t$,

$$W(t) - W(s) \sim \mathcal{N}(0, (t-s)Q).$$

Se asumen las siguientes condiciones, véase Da Prato y Zabczyk (1992); Fuhrman (1995):

- (i) El operador A es el generador infinitesimal de un semigrupo fuertemente continuo, $\exp(tA)$, $t \geq 0$, de operadores lineales en H satisfaciendo que $\|\exp(tA)\| \leq M \exp(-\omega t)$, $t \geq 0$, para algún M y $\omega > 0$.
- (ii) Q es un operador continuo, lineal, autoadjunto y no negativo en H .
- (iii) Considerando, para $f \in H$,

$$Q_t(f) = \int_0^t \exp(sA)Q \exp(sA^*)(f)ds \quad (1.9)$$

se tiene que $\sup_{t>0} \text{Tr}(Q_t) < \infty$, donde Tr denota la traza.

La integral estocástica (1.8) toma sus valores en H , y se entiende en el sentido de Itô; véase, por ejemplo, Sección 8.1 de Kopp (2011). La variable aleatoria H -valuada $X(t, f)$ tiene distribución gaussiana de media $\exp(tA)(f)$, y operador de covarianza Q_t para cada $t > 0$. La medida gaussiana asociada se denota como $\mu_{\exp(tA)f, Q_t}$. El semigrupo de transición de O-U, P_t , viene entonces dado por

$$(P_t\phi)(f) = \int_H \phi(y) \mu_{\exp(tA)f, Q_t}(dy), \quad t > 0, \quad f \in H,$$

para cada función real-valuada medible acotada ϕ en H . El operador $Q_\infty(f) = \int_0^\infty \exp(tA)Q \exp(tA^*)(f)dt$, $f \in H$, pertenece a la clase de operadores traza bajo la Condición (i), y la medida gaussiana en H , μ_{0, Q_∞} , es la única medida

invariante para P_t . La forma explícita de la densidad de probabilidad y el espacio donde se encuentra, así como la derivada Radon-Nikodym de μ_{0,Q_∞} , se obtienen, bajo ciertas condiciones, en Fuhrman (1995).

1.2.5. Procesos ARH(1)

Se introduce a continuación la ecuación de estados y suposiciones básicas, que definen la familia de procesos ARH(1), véase Bosq (2000). Nótese que esta familia de procesos Hilbert-valuados interviene en el cálculo del predictor funcional de la log-intensidad aleatoria que dirige la clase de procesos de Cox que se considera en los capítulos 3 y 7.

Definición 1.2. *Sea H un espacio de Hilbert real separable. Una secuencia $Y = \{Y_n, n \in \mathbb{Z}\}$ de variables aleatorias H -valuadas en un espacio de probabilidad básico (Ω, \mathcal{A}, P) se denomina un proceso hilbertiano autorregresivo de orden uno, asociado a $(\eta, \varepsilon, \rho)$, si es estacionario y satisface que*

$$X_n = Y_n - \eta = \rho(Y_{n-1} - \eta) + \varepsilon_n = \rho(X_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z}, \quad (1.10)$$

donde la esperanza $\eta = E[Y_n] \in H$ es constante, $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ es un ruido blanco Hilbert-valuado en sentido fuerte (es decir, una secuencia estacionaria de variables aleatorias independientes H -valuadas con $E\|\varepsilon_n\|_H^2 = \sigma^2 < \infty$, para cada $n \in \mathbb{Z}$), y $\rho \in \mathcal{L}(H)$, siendo $\mathcal{L}(H)$ el espacio de operadores lineales acotados en H . Para cada $n \in \mathbb{Z}$, ε_n y X_{n-1} se asume que son incorreladas.

Si existe un valor positivo $j_0 \geq 1$ tal que $\|\rho^{j_0}\|_{\mathcal{L}(H)} < 1$, entonces, el proceso ARH(1) $X = \{X_n, n \in \mathbb{Z}\}$ en (1.10) es estándar, y existe una única solución estacionaria para la ecuación (1.10); véase la Proposición 10.5, p. 248, en Bosq (2000).

Los operadores de autocovarianza y covarianza cruzada vienen dados respectivamente por

$$\begin{aligned} C_X &= E[X_n \otimes X_n] = E[X_0 \otimes X_0], \quad n \in \mathbb{Z}, \\ D_X &= E[X_n \otimes X_{n+1}] = E[X_0 \otimes X_1], \quad n \in \mathbb{Z}, \end{aligned} \quad (1.11)$$

donde $h \otimes g$ denota el producto tensorial de las funciones h y g en H . Es decir,

para $f, g \in H$,

$$f \otimes g(h) = f \langle g, h \rangle_H, \quad \forall h \in H,$$

define un operador Hilbert-Schmidt en H . Se asume que el operador C_X pertenece a la clase de operadores traza. En particular, $E\|X_n\|_H^2 < \infty$, para todo $n \in \mathbb{Z}$. Equivalentemente, para cualquier entero n , la variable X_n es un elemento aleatorio en H . Por tanto, $P(X_n \in H) = 1$, para todo $n \in \mathbb{Z}$.

1.3. Preliminares para el Capítulo 4 y 5

Definición 1.3. (Ruiz-Medina, 2011). Un proceso funcional espacial $X = \{X_{i,j}, (i,j) \in \mathbb{Z}^2\}$, con valores en un espacio separable de Hilbert, \mathcal{H} , se dice que es un proceso autorregresivo espacial de orden uno, evaluado en un espacio de Hilbert H , SARH(1), si es estacionario y cumple la siguiente ecuación:

$$X_{i,j} = R + L_1(X_{i-1,j}) + L_2(X_{i,j-1}) + L_3(X_{i-1,j-1}) + \epsilon_{i,j}, \quad (1.12)$$

donde $R \in \mathcal{H}$ y $L_l \in \mathcal{L}(\mathcal{H})$, para $l = 1, 2, 3$, y $\mathcal{L}(\mathcal{H})$ denota el espacio de operadores lineales acotados en \mathcal{H} . Aquí, $\epsilon = \{\epsilon_{i,j}, (i,j) \in \mathbb{Z}^2\}$ es el proceso de innovación espacial, que se supone que es un proceso espacial funcional débilmente dependiente, más específicamente, se trata de una diferencia de martingala biparamétrica (donde el espacio de parámetros es bidimensional, es decir, un proceso espacial), con $E\|\epsilon_{i,j}\|_{\mathcal{H}}^2 = \sigma^2$, independientemente de la ubicación espacial (i,j) , y $E[\epsilon_{i,j} \otimes \epsilon_{i,j}] = R_{\epsilon_{0,0}}$ también es independiente de (i,j) .

La siguiente suposición está formulada para establecer las condiciones necesarias para probar la existencia como una solución estacionaria de (1.12).

Suposición 1.1. (Ruiz-Medina, 2011). Para $l = 1, 2, 3$, se asume que el operador $L_l \in \mathcal{L}(\mathcal{H})$ admite una descomposición espectral en términos de la secuencia de autovalores $\lambda_{kl}, k \in \mathbb{N}$, y de los sistemas biortonormales de los autovectores izquierdos $\{\psi_k, k \in \mathbb{N}\}$ y autovectores derechos $\{\phi_k, k \in \mathbb{N}\}$, definiendo bases duales Riesz de \mathcal{H} y \mathcal{H}^* , y de forma que se satisfacen las siguientes ecuaciones:

$$\begin{aligned} L_l(\psi_k) &= \lambda_{kl}\psi_k, & k \in \mathbb{N} \\ L_l^*(\phi_k) &= \lambda_{kl}\phi_k, & k \in \mathbb{N} \end{aligned}$$

donde L_l^* denota al adjunto de L_l .

Bajo esta suposición, para $l = 1, 2, 3$, L_l admite la siguiente representación espectral

$$L_l(g)(f) = \sum_{k \in \mathbb{N}} \lambda_{kl} \psi_k(f) \phi_k(g), \quad \forall f, g \in \mathcal{H} \quad (1.13)$$

Nota 1.2. Las siguientes proposiciones proporcionan un conjunto de condiciones adecuadas sobre los parámetros funcionales L_l , $l = 1, 2, 3$, bajo la Suposición 1.1, que garantizan la existencia de una solución estacionaria para (1.12) en la forma (1.14) desarrollada en la Proposición 1.4.

Proposición 1.3. (Ruiz-Medina, 2011). Bajo la Suposición 1.1, para cada $k \in \mathbb{N}$, ninguna de las raíces de $\phi(z_1, z_2) = 0 = 1 - \lambda_{K1}z_1 - \lambda_{K2}z_2 - \lambda_{K3}z_1z_2$ se encuentran dentro del polidisco unitario cerrado ($|z_1| \leq 1, |z_2| \leq 1$) si y solo si, para cada $k \in \mathbb{N}$,

i). $|\lambda_{kl}| < 1$, para $l = 1, 2, 3$,

ii). $(1 + \lambda_{k1}^2 - \lambda_{k2}^2 - \lambda_{k3}^2)^2 - 4(\lambda_{k1} + \lambda_{k2}\lambda_{k3})^2 > 0$,

iii). $(1 - \lambda_{k2}^2) > |\lambda_{k1} + \lambda_{k2}\lambda_{k3}|$.

Bajo i), ii) y iii), la solución a la ecuación (1.12) es estacionaria.

Nota 1.3. En el caso donde L_l , $l = 1, 2, 3$, sean semidefinidos positivos, la Condición i) se puede sustituir por la existencia de un entero, $j_0 \geq 1$, tal que:

$$\|L^{j_0}\|_{\mathcal{L}_\infty(\mathcal{S}^3)} = \max\{\|L_1^{j_0}\|_{\mathcal{L}(\mathcal{S})}, \|L_2^{j_0}\|_{\mathcal{L}(\mathcal{S})}, \|L_3^{j_0}\|_{\mathcal{L}(\mathcal{S})}\} < 1,$$

lo que equivale a la existencia de un entero $j_0 \geq 1$ tal que $\|L_l^{j_0}\|_{\mathcal{L}(\mathcal{S})} < 1$, para $l = 1, 2, 3$. De hecho, podemos considerar un vector, $\mathbf{j}_0 = (j_0^1, j_0^2, j_0^3)$, tal que

$$\|L^{j_0}\|_{\mathcal{L}_\infty(\mathcal{S}^3)} = \max\{\|L_1^{j_0^1}\|_{\mathcal{L}(\mathcal{S})}, \|L_2^{j_0^2}\|_{\mathcal{L}(\mathcal{S})}, \|L_3^{j_0^3}\|_{\mathcal{L}(\mathcal{S})}\} < 1.$$

Proposición 1.4. (Ruiz-Medina, 2011). Sea X un proceso SARH(1), satisfaciendo la Suposición 1.1. Supongamos que se cumplen las condiciones consideradas en la Proposición 1.3. Entonces, (1.12) admite una solución estacionaria única dada por

$$X_{i,j} = R + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!} L_1^k L_2^l L_3^r (\epsilon_{i-k-r, j-l-r}) \quad (1.14)$$

donde L_l , $l = 1, 2, 3$, se han definido en (1.12).

Nota 1.4. Bajo la Suposición 1.1, si L_l es compacto, autoadjunto, y operador positivo para $l = 1, 2, 3$, entonces, en particular, $\phi_k = \psi_k$, para $k \geq 1$, y la proyección

de la ecuación de estados SARH(1), (1.12), en $\{\phi_k, k \geq 1\}$, nos lleva al siguiente sistema diagonal de ecuaciones de dimensión infinita:

$$X_{i,j}^p = \lambda_p(L_1)X_{i-1,j}^p + \lambda_p(L_2)X_{i,j-1}^p + \lambda_p(L_3)X_{i-1,j-1}^p + \epsilon_{i,j}^p, \quad p \geq 1, \quad (1.15)$$

donde $X_{i,j}^p = \langle X_{i,j}, \phi_p \rangle_{\mathcal{H}}$, y $\epsilon_p(i, j) = \langle \epsilon_{i,j}, \phi_p \rangle_{\mathcal{H}}$.

Nota 1.5. Bajo la suposición más restrictiva de que L_l , con $l = 1, 2, 3$, son operadores de Hilbert-Schmidt (positivos y autoadjuntos, como antes), se pueden obtener resultados similares al desarrollo posterior para una base ortonormal arbitraria, $\{\varphi_k\}_{k \geq 1}$ de \mathcal{H} , más allá de la restricción de un sistema común de autovectores para L_l , $l = 1, 2, 3$, en la Suposición 1.1, ya que podemos considerar las siguientes expansiones en serie de L_l , $l = 1, 2, 3$, en \mathcal{H} ,

$$L_l = \sum_{k=1}^{\infty} \sum_{q=1}^{\infty} L_l(\varphi_q)(\varphi_k)\varphi_k \otimes \varphi_q, \quad l = 1, 2, 3.$$

.

1.3.1. Estimación de momentos del proceso autorregresivo espacial de orden uno evaluado en un espacio de Hilbert

Siguiendo Ruiz-Medina (2012), a continuación, se consideran las bases de autofunciones biortogonales que diagonalizan los parámetros del proceso SARH(1) infinito-dimensional.

A partir de la ecuación (1.12), los operadores L_i , $i = 1, 2, 3$, satisfacen el siguiente sistema:

$$\begin{cases} R_{1,0} = L_1 R_{0,0} + L_2 R_{1,1} + L_3 R_{0,1} \\ R_{0,1} = L_1 R_{1,1} + L_2 R_{0,0} + L_3 R_{1,0} \\ R_{1,1} = L_1 R_{0,1} + L_2 R_{1,0} + L_3 R_{0,0} \end{cases}$$

Este sistema de ecuaciones lineales se aproxima, en la práctica, en términos

de los operadores de covarianza empíricos que se definen como:

$$\begin{aligned}\hat{R}_{0,0} &= \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J Z_{i,j} \otimes Z_{i,j}; & \hat{R}_{1,1} &= \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} Z_{i+1,j+1} \otimes Z_{i,j}; \\ \hat{R}_{0,1} &= \frac{1}{I(J-1)} \sum_{i=1}^I \sum_{j=1}^{J-1} Z_{i,j+1} \otimes Z_{i,j}; & \hat{R}_{1,0} &= \frac{1}{(I-1)J} \sum_{i=1}^{I-1} \sum_{j=1}^J Z_{i+1,j} \otimes Z_{i,j};\end{aligned}$$

donde $\{Z_{i,j} = X_{i,j} - R, (i, j) \in (1, \dots, I) \times (1, \dots, J)\}$ define la estructura de dependencia espacial de los datos funcionales y satisfacen la ecuación (1.12).

Para cada $p, q \in \{0, 1\}$, se denota por $\{c_{m,n}(R_{p,q}), m, n \in \mathbb{N}\}$ a los coeficientes de Fourier de $R_{p,q}$, con respecto a las bases de autovectores definidos en la Suposición 1.1, es decir, $c_{m,n}(R_{p,q}) = R_{p,q}(\phi_n)(\psi_m)$, $m, n \in \mathbb{N}$. Así pues, para cada $p, q \in \{0, 1\}$, la siguiente representación basada en *kernels* es válida, véase Ruiz-Medina (2012):

$$R_{p,q}(f)(g) = \sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} c_{m,n}(R_{p,q}) \psi_m(f) \phi_n(g) \quad \forall f, g \in \mathcal{H}$$

Proposición 1.5. (Ruiz-Medina, 2012). *Bajo la Suposición 1.1, los autovalores $\{\lambda_{kl}, k \in \mathbb{N}, l = 1, 2, 3\}$ se pueden estimar a partir de*

$$\begin{cases} c_{m,n}(\hat{R}_{1,0}) = c_{m,n}(\hat{R}_{0,0}) \hat{\lambda}_{k1} + c_{m,n}(\hat{R}_{1,1}) \hat{\lambda}_{k2} + c_{m,n}(\hat{R}_{0,1}) \hat{\lambda}_{k3} & m, n \in \mathbb{N} \\ c_{m,n}(\hat{R}_{0,1}) = c_{m,n}(\hat{R}_{1,1}) \hat{\lambda}_{k1} + c_{m,n}(\hat{R}_{0,0}) \hat{\lambda}_{k2} + c_{m,n}(\hat{R}_{1,0}) \hat{\lambda}_{k3} & m, n \in \mathbb{N} \\ c_{m,n}(\hat{R}_{1,1}) = c_{m,n}(\hat{R}_{0,1}) \hat{\lambda}_{k1} + c_{m,n}(\hat{R}_{1,0}) \hat{\lambda}_{k2} + c_{m,n}(\hat{R}_{0,0}) \hat{\lambda}_{k3} & m, n \in \mathbb{N} \end{cases}$$

Entonces, los estimadores componente a componente de L_i , $i = 1, 2, 3$, basados en el método de los momentos, se pueden calcular a partir de

$$\hat{L}_l(f)(g) = \sum_{k \in \mathbb{N}} \hat{\lambda}_{kl} \psi_k(f) \phi_k(g), \quad \forall f, g \in \mathcal{H}, \quad l = 1, 2, 3,$$

En consecuencia, se obtiene el siguiente predictor plug-in funcional resultante:

$$\hat{Z}_{i+1,j+1} = \hat{L}_1(Z_{i,j+1}) + \hat{L}_2(Z_{i+1,j}) + \hat{L}_3(Z_{i,j}) + \epsilon_{i,j}, \quad (1.16)$$

Nota 1.6. (Ruiz-Medina, 2012). *En la práctica, para un orden de truncamiento M , se calcula la estimación paramétrica como proyección de los operadores empí-*

ricos de autocovarianza:

$$\begin{pmatrix} \hat{\Lambda}(L_1) \\ \hat{\Lambda}(L_2) \\ \hat{\Lambda}(L_3) \end{pmatrix} = \begin{pmatrix} \mathbf{C}(\hat{R}_{0,0}) & \mathbf{C}(\hat{R}_{1,1}) & \mathbf{C}(\hat{R}_{0,1}) \\ \mathbf{C}(\hat{R}_{1,1}) & \mathbf{C}(\hat{R}_{0,0}) & \mathbf{C}(\hat{R}_{1,0}) \\ \mathbf{C}(\hat{R}_{0,1}) & \mathbf{C}(\hat{R}_{1,0}) & \mathbf{C}(\hat{R}_{0,0}) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{C}(\hat{R}_{1,0}) \\ \mathbf{C}(\hat{R}_{0,1}) \\ \mathbf{C}(\hat{R}_{1,1}) \end{pmatrix}$$

donde, para $p, q \in \{0, 1\}$, $\mathbf{C}(\hat{R}_{p,q})$ es una matriz $M \times M$ de entradas $\left(\hat{R}_{p,q}(\phi_m)(\psi_n) \right)_{m,n=1,\dots,M}$ y, para $l = 1, 2, 3$, $\hat{\Lambda}(L_l)$ es una matriz diagonal, donde $\left(\hat{\lambda}_{kl} \right)_{k=1,\dots,M}$ son sus entradas.

METODOLOGÍA Y RESULTADOS

En esta parte se detallan los principales ingredientes y aportaciones de la presente tesis según se describe en los capítulos 2-5, que intervienen en los resultados teóricos y metodológicos derivados en toda la tesis. Asimismo, se introducen los estudios de simulación y aplicaciones con datos reales desarrollados para ilustrar los resultados obtenidos y su implementación práctica.

Capítulo 2

Análisis de las propiedades asintóticas de estimadores espaciales de mínimo contraste basados en el periodograma *tapered*

En este capítulo, se estudian las propiedades asintóticas de los estimadores de mínimo contraste en procesos espaciales a partir del periodograma *tapered*. Se estudia el funcional de contraste de Ibragimov así como el de Whittle, para campos aleatorios gaussianos con dependencia de corto y largo rango. Los resultados derivados en este capítulo sobre consistencia y normalidad asintótica se han publicado en el trabajo titulado “Asymptotic properties of parameter estimates for random fields with tapered data”, que aparece en la revista “Electronic Journal of Statistics”, Vol. 11, pp 3332–3367, (Alomari et al., 2017). Asimismo, se describe el estudio de simulación desarrollado, en el contexto de procesos espaciales autorregresivos fraccionarios, en relación con los polinomios de Gegenbauer.

Sea $Y = \{Y_{\mathbf{t}}, \mathbf{t} \in \mathbb{Z}^d, d > 1\}$ un campo aleatorio homogéneo gaussiano real-valuado centrado y medible, con densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, para $\boldsymbol{\lambda} \in \mathbb{T} = (-\pi, \pi]^d$ y $\boldsymbol{\theta} \in \Theta$, siendo Θ un subespacio compacto de \mathbb{R}^q , $q > 1$. Asumimos que θ_0 , el verdadero valor del parámetro, pertenece al interior de Θ . Se asume también que la parametrización es propia, o equivalentemente, se considera la

correspondiente condición de indentificabilidad del modelo paramétrico, es decir, $\theta \neq \theta'$ implica $f(\lambda, \theta) \neq f(\lambda, \theta')$ sobre un conjunto de medida de Lebesgue positiva.

A continuación, en la Sección 2.1, se exponen los principales resultados sobre consistencia y normalidad asintótica de los correspondientes estimadores de mínimo contraste. Las condiciones utilizadas abarcan el caso de campos aleatorios con dependencia de largo rango, con un modelo particular de densidades espectrales con singularidades que se factorizan. También se dan unas condiciones sobre las funciones tapered para asegurar la normalidad asintótica de los estimadores de mínimo contraste propuestos, dando lugar a un resultado de normalidad asintótica, para las dimensiones $d = 1, 2, 3$.

La Sección 2.2 contiene las demostraciones de los resultados de la Sección 2.1 sobre las propiedades asintóticas de los estimadores de mínimo contraste aportados.

En la Sección 2.3, se ilustran los resultados derivados mediante un estudio de simulación, en el contexto de los procesos autorregresivos espaciales y los campos aleatorios espaciales de Gegenbauer. Finalmente, en la Sección 2.4 se presenta el teorema central del límite para funcionales espectrales de campos aleatorios gaussianos, a partir de datos tapered

2.1. Condiciones para el análisis de propiedades asintóticas

En los resultados derivados en las secciones subsiguientes, se asumen las siguientes condiciones, que definen la familia paramétrica de densidades espectrales espaciales analizada.

B1. Sea $Y(\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^d$, un campo aleatorio gaussiano, estacionario, medible¹, real-valuado, centrado y con densidad espectral $f(\lambda, \theta)$, donde $\lambda \in \mathbb{T} =$

¹Un campo aleatorio $Y(\omega, \mathbf{t}) : \Omega \times T \rightarrow \mathbb{R}^m$, con $T \in \mathcal{B}(\mathbb{R}^d)$, se dice medible si para cualquier conjunto $A \in \mathcal{B}(\mathbb{R}^m)$, $\{(\omega, \mathbf{t}) : Y(\omega, \mathbf{t}) \in A\} \in \mathcal{A} \times \mathcal{B}(T)$. En general, un campo aleatorio se puede considerar simplemente como un proceso estocástico definido sobre un espacio paramétrico de dimensión $d \geq 1$, que usualmente toma valores en un espacio euclidiano. Si el espacio paramétrico $T \subset \mathbb{R}^d$, se dice un campo aleatorio d -dimensional. Es habitual exigir que los campos aleatorios sean separables por las buenas propiedades que se desprenden de dicha condición y para evitar problemas de medición. Esta condición nos dice que existe un subconjunto numera-

$(-\pi, \pi]^d$, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$, y Θ es un conjunto compacto. Se asume que $\boldsymbol{\theta}_0 \in \text{int}(\Theta)$, donde $\boldsymbol{\theta}_0$ es el verdadero valor del vector paramétrico $\boldsymbol{\theta}$.

B2. Si $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ entonces $f(\boldsymbol{\lambda}, \boldsymbol{\theta}_1) \neq f(\boldsymbol{\lambda}, \boldsymbol{\theta}_2)$ para casi todo $\boldsymbol{\lambda} \in \mathbb{T}$ con respecto a la medida Lebesgue.

Restringimos nuestro estudio a los campos con $d \leq 3$. Nótese que los resultados sobre la consistencia de los estimadores se mantendrán para el caso general, $d \geq 1$. Sin embargo, para la normalidad asintótica se le impondrá la restricción $d \leq 3$, ya que sólo para estas dimensiones podemos controlar el sesgo con la ayuda del *tapering*.

Nota 2.1. En lo que sigue, para la diferenciabilidad con respecto a $\boldsymbol{\theta}$, nos referimos a la diferenciación en el interior de Θ .

Para definir el funcional de Ibragimov, vamos a utilizar una función peso que satisfaga la siguiente hipótesis.

B3. Existe una función no negativa $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \mathbb{T}$, tal que

1. $w(\boldsymbol{\lambda})$ es simétrica, es decir que $w(\boldsymbol{\lambda}) = w(-\boldsymbol{\lambda})$;
2. $w(\boldsymbol{\lambda})f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \in L_1(\mathbb{T})$, para todo $\boldsymbol{\theta} \in \Theta$.

Bajo esta condición, tenemos que

$$\sigma^2(\boldsymbol{\theta}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta})w(\boldsymbol{\lambda}) d\boldsymbol{\lambda} \quad (2.1)$$

y representa la densidad espectral en la forma:

$$f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \sigma^2(\boldsymbol{\theta})\psi(\boldsymbol{\lambda}, \boldsymbol{\theta}). \quad (2.2)$$

Para la función $\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ se tiene que

$$\int_{\mathbb{T}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})w(\boldsymbol{\lambda})d\boldsymbol{\lambda} = 1, \quad (2.3)$$

adicionalmente suponemos:

ble fijo $D \subset T$, tal que, con probabilidad 1,

$$\sup_{t \in V \cap D} Y(t) = \sup_{t \in V} Y(t), \quad \inf_{t \in V \cap D} Y(t) = \inf_{t \in V} Y(t), \text{ para todo conjunto abierto } V,$$

(Adler y Taylor, 2007; Azaïs y Wschebor, 2009; Leonenko, 1999).

B4. Las derivadas $\nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ existen y se verifican las condiciones que permiten intercambiar la diferenciación bajo el signo de la integral en la ecuación (2.3), es decir que

$$\nabla_{\boldsymbol{\theta}} \int_{\mathbb{T}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) d\boldsymbol{\lambda} = \int_{\mathbb{T}} \nabla_{\boldsymbol{\theta}} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) d\boldsymbol{\lambda} = 0.$$

Supongamos que existen las funciones $\alpha_i : \Theta \rightarrow (0, 1)$, $i = 1, \dots, d$, tales que cumplen las condiciones siguientes (con la elección apropiada de una función peso $w(\boldsymbol{\lambda})$). Para simplificar la notación en la formulación de las condiciones siguientes, omitiremos el argumento $\boldsymbol{\theta}$ en las funciones $\alpha_i(\boldsymbol{\theta})$ y las escribiremos simplemente como α_i .

B5. Para todo $\boldsymbol{\theta} \in \Theta$, la densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^d |\lambda_i|^{-\alpha_i})$ cuando $\lambda_i \rightarrow 0$, y $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ está acotada para $\delta \leq |\boldsymbol{\lambda}| \leq \pi$, con $\delta > 0$; Así pues, $w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^d |\lambda_i|^{\alpha_i})$ cuando $\lambda_i \rightarrow 0$.

B6. Existe una función $v(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \mathbb{T}$, tal que

1. la función $h(\boldsymbol{\lambda}, \boldsymbol{\theta}) = v(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ es uniformemente continua en $\mathbb{T} \times \Theta$;
2. $w(\boldsymbol{\lambda})/v(\boldsymbol{\lambda}) = O(\prod_{i=1}^d |\lambda_i|^{\alpha_i})$ cuando $\lambda_i \rightarrow 0$.

B7. La función $\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ tiene segunda derivada acotada en Θ y

1. $w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^d |\lambda_i|^{\alpha_i})$ cuando $\lambda_i \rightarrow 0$, para todo i, j , $\boldsymbol{\theta} \in \Theta$.
2. $w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = O(\prod_{i=1}^d |\lambda_i|^{\alpha_i})$ cuando $\lambda_i \rightarrow 0$, para todo i , $\boldsymbol{\theta} \in \Theta$.
3. Las derivadas de segundo orden, $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$, con $i, j = 1, \dots, q$, son continuas en $\boldsymbol{\theta}$.

B8. Las matrices $\mathbf{S}(\boldsymbol{\theta}) = (s_{ij}(\boldsymbol{\theta}))_{i,j=1,\dots,q}$ y $\mathbf{A}(\boldsymbol{\theta}) = (a_{ij}(\boldsymbol{\theta}))_{i,j=1,\dots,q}$ son definidas

positivas donde, para $i, j = 1, \dots, q$,

$$\begin{aligned}
 (s_{ij}(\boldsymbol{\theta})) &= \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \\
 &= \sigma^2(\boldsymbol{\theta}) \int_{\mathbb{T}} w(\boldsymbol{\lambda}) \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right. \\
 &\quad \left. - \frac{1}{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})} \frac{\partial}{\partial \theta_i} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right] d\boldsymbol{\lambda}, \\
 (a_{ij}(\boldsymbol{\theta})) &= 2(2\pi)^d \int_{\mathbb{T}} f^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \\
 &= 2(2\pi)^d (\sigma^2(\boldsymbol{\theta}))^2 \int_{\mathbb{T}} w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}.
 \end{aligned}$$

H2. La función taper $\tilde{h}(t)$ es Lipschitz continua en $[-1, 1]$, y $\tilde{h}(-1) = \tilde{h}(1) = 0$.

B9. La densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$, la función $w(\boldsymbol{\lambda})$, y la función $\varphi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$, son tales que una de las siguientes condiciones se cumple:

- (i) φ tiene segunda derivada acotada;
- (ii) la convolución $g(\mathbf{u}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda} + \mathbf{u}) d\boldsymbol{\lambda}$ tiene segunda derivada acotada en cero.

Nota 2.2. En las condiciones **B5–B7**, se prescribe el comportamiento de la densidad espectral y alguna de sus derivadas, en el punto de singularidad. Estas condiciones aparecen de manera natural y son análogas, por ejemplo, a las que se utilizan en Fox y Taqqu (1986) para el caso de los estimadores de Whittle, con modificaciones evidentes, ya que usamos otro funcional para construir los estimadores. De las suposiciones **B6–B7** se desprende de forma clara que la función peso $w(\boldsymbol{\lambda})$ para el funcional de contraste U_T se construye para compensar posibles singularidades de la densidad espectral. En algunos casos, esta función peso no es necesaria, por ejemplo, cuando la densidad espectral está acotada.

Nota 2.3. Las condiciones **B7** y **B8** se utilizan para lograr una convergencia apropiada del sesgo a 0.

Nota 2.4. Un ejemplo de un campo aleatorio con dependencia de largo rango para el cual se satisfacen todas las condiciones anteriores (con la elección apropiada de

la función peso $w(\boldsymbol{\lambda})$, es la solución a la siguiente ecuación:

$$(1 - B_1)^{d_1}(1 - B_2)^{d_2}Y(t_1, t_2) = \varepsilon(t_1, t_2)$$

, siendo $\varepsilon(t_1, t_2)$ un ruido blanco bidimensional, y siendo B_i un operador de retardos para la i -ésima coordenada, $i = 1, 2$. Este ejemplo de autorregresión espacial fraccionaria, de forma más general, se considera en detalle en la Sección 2.3, junto a otros dos ejemplos de campos aleatorios de Gegenbauer. También se presentan las funciones peso $w(\boldsymbol{\lambda})$, apropiadas para estos modelos. Obsérvese que la estimación de los campos que obedecen al modelo autorregresivo espacial fraccionario (basada en el funcional de Whittle) se ha considerado en Boissy et al. (2005), sin embargo, solo se ha establecido la consistencia de los estimadores. Dentro del enfoque del presente capítulo, se pueden construir estimaciones consistentes y asintóticamente normales para este modelo (véase la Sección 2.3.1).

Consideramos el siguiente campo aleatorio de contraste basado en el periodograma tapered definido en los preliminares:

$$U_T(\boldsymbol{\theta}) = - \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \quad (2.4)$$

y el estimador de mínimo contraste

$$\hat{\boldsymbol{\theta}}_T = \arg \min_{\boldsymbol{\theta} \in \Theta} U_T(\boldsymbol{\theta}).$$

Así mismo, se definen las siguientes funciones:

$$K(\boldsymbol{\theta}_0; \boldsymbol{\theta}) = - \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \frac{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta})}{\psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)} d\boldsymbol{\lambda} \quad (2.5)$$

y

$$U(\boldsymbol{\theta}) = - \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}. \quad (2.6)$$

La propiedad de mínimo contraste de la función $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ se indica en el Teorema 2.1, así como las condiciones bajo las cuales $U_T(\boldsymbol{\theta})$ es el campo aleatorio de contraste relacionado con $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$.

2.1.1. Teoremas sobre la consistencia y la normalidad asintótica de los estimadores

Teorema 2.1. *Asumimos que las condiciones B1–B3, B5–B6, y H1 se satisfacen. Entonces, la función $K(\theta_0; \theta)$ definida por la ecuación (2.5) es la función de contraste para el campo aleatorio de contraste $U_T(\theta)$ definido por la ecuación (2.4). Además, el estimador de mínimo contraste $\hat{\theta}_T$,*

$$\hat{\theta}_T = \arg \min_{\theta \in \Theta} U_T(\theta), \quad (2.7)$$

es un estimador consistente del vector paramétrico θ . Esto es, existe una convergencia en P_0 probabilidad:

$$\hat{\theta}_T \xrightarrow{P_0} \theta_0, \quad \text{cuando } T \rightarrow \infty.$$

El estimador $\hat{\sigma}_T^2 \xrightarrow{P_0} \sigma^2(\theta_0)$, cuando $T \rightarrow \infty$, donde $\hat{\sigma}_T^2$ es un estimador del parámetro $\sigma^2(\theta_0)$ que viene dado por

$$\hat{\sigma}_T^2 = \int_{\mathbb{T}} I_T^h(\lambda) w(\lambda) d\lambda.$$

La demostración se incluye en la Sección 2.2.

Teorema 2.2. *Asumimos que $d \leq 3$, y que las condiciones B1–B9 y H1–H2 se satisfacen. Entonces, el estimador de mínimo contraste, definido en la ecuación (2.7), es asintóticamente normal. Esto es, cuando $T \rightarrow \infty$*

$$T^{d/2}(\hat{\theta}_T - \theta_0) \xrightarrow{D} \mathcal{N}_q(0, e(h)\mathbf{S}^{-1}(\theta_0)\mathbf{A}(\theta_0)\mathbf{S}^{-1}(\theta_0)), \quad (2.8)$$

donde las entradas de las matrices $\mathbf{S}(\theta) = (s_{ij}(\theta))$, y $\mathbf{A}(\theta) = (a_{ij}(\theta))$ están definidas en la Condición B8, $e(h)$ viene dado por la siguiente ecuación,

$$e(h) = \left(\int (\tilde{h}(t))^4 dt \left(\int (\tilde{h}(t))^2 dt \right)^{-2} \right)^d, \quad (2.9)$$

y $\mathcal{N}_q(\cdot, \cdot)$ denota la distribución de Gauss multivariante q -dimensional.

La demostración se incluye en la Sección 2.2.

Nota 2.5. *En el caso de tiempo discreto, usualmente se toma $L_T = [1, T]^d$, como el dominio sobre el cual se observa el campo. Estos resultados obtenidos siguen*

siendo válidos para un dominio de este tipo. Así, solamente tenemos que ajustar los supuestos para una función taper $\tilde{h}(t)$. Es decir, la Suposición **H1** debe modificarse de la siguiente manera: $\tilde{h}(t)$ es una función medible positiva de variación acotada con soporte en $[0, 1]$ y $h(0) = 0$, $h(1 - v) = h(v)$, para $0 \leq v \leq \frac{1}{2}$.

Nota 2.6. Un ejemplo de una función taper $\tilde{h}(t)$ que satisface las condiciones asumidas es

$$\tilde{h}(t) = \frac{1}{2}(1 + \cos(4\pi t)), \quad t \in [-1, 1].$$

Esto es simplemente una modificación de la conocida campana de coseno (o la función taper de Tukey-Hanning):

$$\tilde{h}(t) = \frac{1}{2}(1 - \cos(2\pi t)), \quad t \in [0, 1],$$

adecuada para el dominio $L_T = [1, T]^d$.

Nota 2.7. Para establecer la consistencia de un estimador de mínimo contraste $\hat{\theta}_T$, que corresponde a un funcional $U_T(\theta)$, se siguen los siguientes razonamientos habituales: se debe comprobar que se cumple la convergencia en probabilidad (1.2) y, a continuación, debido al Teorema 3.4.1 en Guyon (1995), basta con probar que la convergencia (1.2) se mantiene uniformemente con respecto a θ .

El enfoque habitual para establecer la normalidad asintótica del estimador $\hat{\theta}_T$ es considerar la siguiente relación:

$$\nabla_{\theta} U_T(\hat{\theta}_T) = \nabla_{\theta} U_T(\theta_0) + \nabla_{\theta} \nabla'_{\theta} U_T(\theta_T^*) (\hat{\theta}_T - \theta_0), \quad |\theta_T^* - \theta_0| < |\hat{\theta}_T - \theta_0|,$$

y entonces evaluar el comportamiento asintótico de $\nabla_{\theta} U_T(\theta_0)$ y $\nabla_{\theta} \nabla'_{\theta} U_T(\theta_T^*)$.

Por lo tanto, se obtendrá la consistencia de los funcionales espectrales empíricos de la forma:

$$J_T^h(\varphi) = \int_{\mathbb{T}} I_T^h(\lambda) \varphi(\lambda) w(\lambda) d\lambda,$$

donde $I_T^h(\lambda)$ es el periodograma basado en datos tapered. En particular, para la consecución de los resultados de normalidad asintótica, es importante establecer las condiciones que garantizan la tasa apropiada de convergencia a cero del sesgo, es decir, la siguiente relación a tener en cuenta:

$$T^{d/2} \left(E J_T^h(\varphi) - \int_{\mathbb{T}} f(\lambda) \varphi(\lambda) w(\lambda) d\lambda \right) \rightarrow 0 \quad \text{cuando } T \rightarrow \infty. \quad (2.10)$$

En el anterior Teorema 2.2, las condiciones que ayudan a controlar el sesgo son H1–H2 y B9.

Nota 2.8. Obsérvese que los resultados anteriores sobre estimadores de Ibragimov² se expresaron bajo las condiciones de integrabilidad de la densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ y la función $w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$, así como de funcionales que involucran a dichas funciones (2.10). Cabe destacar también que la investigación del sesgo para los estimadores de funcionales espectrales, en el ajuste no paramétrico, fue presentada en Anh et al. (2007a); Sakhno (2007, 2014).

Las condiciones asumidas para los resultados de consistencia y normalidad asintótica, se formulan de forma explícita, para facilitar su verificación y aplicación en la práctica.

Nota 2.9. Como se desprende de los resultados de Guyon, véase, por ejemplo, Guyon (1995), resumidos en la siguiente sección: para controlar el sesgo de los estimadores de mínimo contraste de Ibragimov en el caso de campos aleatorios con dependencia de corto rango se pueden usar las siguientes condiciones: la función taper $\tilde{h}(t)$ pertenece a $C^2[-1, 1]$; la densidad espectral $f(\boldsymbol{\lambda}, \boldsymbol{\theta})$ y la función $\varphi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ pertenecen a $C^2(\mathbb{T})$.

Nota 2.10. Con respecto a la Condición B8: si en lugar de la Condición 3 en B7 utilizamos la Condición 3': las derivadas de segundo orden, $\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta})$, $i, j = 1, \dots, q$, son continuas en ambos $(\boldsymbol{\lambda}, \boldsymbol{\theta})$, $\lambda \neq 0$; entonces se puede simplemente utilizar el caso no degenerado $\mathbf{S}(\boldsymbol{\theta})$ en B8. De hecho, dado que el punto $\boldsymbol{\theta}_0$ es el punto máximo para el funcional $-U(\boldsymbol{\theta}) = \int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}$, entonces, bajo la última Condición 3', la matriz $\nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}}(-U_T(\boldsymbol{\theta}_0))$ tiene que ser semidefinida negativa, y suponiendo que $\mathbf{S}(\boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}}(U_T(\boldsymbol{\theta}_0))$ sea no degenerada, se puede concluir que $\nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}}(-U_T(\boldsymbol{\theta}_0))$ es definida negativa y, por lo tanto, $\mathbf{S}(\boldsymbol{\theta}_0) = \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}}(U_T(\boldsymbol{\theta}_0))$ es definida positiva, véase, por ejemplo, los comentarios después de la Condición N8 en Ivanov y Prikhodko (2015).

2.2. Demostraciones

Las demostraciones se basan en los resultados sobre propiedades asintóticas de los funcionales $J_T(\varphi)$ obtenidos mediante la evaluación de sus cumulantes.

²Véase, por ejemplo, Anh et al. (2004a,b, 2007b,c); Avram et al. (2010a).

Estos cumulantes se pueden representar en forma de algunas integrales que implican densidades espectrales, funciones peso, y *kernels* de tipo Fejér. Aunque estos *kernels* son diferentes para los casos tapered y no tapered, todas las técnicas para las demostraciones funcionan de manera similar en ambos casos (ver más detalles en la Sección 2.4).

Se presentan, en esta sección, los pasos principales para las demostraciones y para establecer el Lema 2.1, que permite controlar el sesgo según se requiere para la derivación del Teorema 2.2.

La herramienta principal que utilizamos para obtener la normalidad asintótica de nuestros estimadores es el Teorema 2.4 (ver Sección 2.4). Este teorema da condiciones suficientes para la normalidad asintótica de $J_T(\varphi)$. También necesitamos condiciones para la convergencia en probabilidad de los funcionales $J_T(\varphi)$ que se discute en la siguiente observación.

Nota 2.11. Para establecer la convergencia en probabilidad

$$J_T(\varphi) = \int_{\mathbb{T}} I_T^h(\lambda) \varphi(\lambda) d\lambda \xrightarrow{P} J(\varphi) = \int_{\mathbb{T}} f(\lambda) \varphi(\lambda) d\lambda, \quad (2.11)$$

es suficiente con ver que:

$$(i) \int_{\mathbb{T}} (EI_T^h(\lambda) - f(\lambda)) \varphi(\lambda) d\lambda \rightarrow 0;$$

$$(ii) \int_{\mathbb{T}} (I_T^h(\lambda) - EI_T^h(\lambda)) \varphi(\lambda) d\lambda \xrightarrow{P} 0.$$

La convergencia en probabilidad de (ii) se cumplirá si $Var J_T(\varphi) = E(\int_{\mathbb{T}} (I_T^h(\lambda) - EI_T^h(\lambda)) \varphi(\lambda) d\lambda)^2 \rightarrow 0$. Por lo tanto, para que se cumpla (2.11) podemos utilizar condiciones que garanticen la convergencia (i) y la convergencia de la varianza del funcional normalizado $T^{d/2}(J_T(\varphi) - EJ_T(\varphi))$ a un límite finito. En particular, la convergencia (2.11) se cumple bajo las condiciones asumidas sobre $f(\lambda)$ y $\varphi(\lambda)$ impuestas en el Teorema 2.4.

Demostración del Teorema 2.1: En vista de la Nota 2.11, teniendo en cuenta la expresión para $U_T(\boldsymbol{\theta})$, se puede concluir que la convergencia

$$\lim_{T \rightarrow \infty} U_T(\boldsymbol{\theta}) - U_T(\boldsymbol{\theta}_0) \xrightarrow{P} U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}_0) = K(\boldsymbol{\theta}_0; \boldsymbol{\theta})$$

se mantiene bajo la Condición B5.

Además, bajo las condiciones B5 y B6, la convergencia anterior se mantiene uniformemente con respecto a $\boldsymbol{\theta} \in \Theta$. De hecho, denotando como $\eta(\varepsilon)$ al

módulo de continuidad de la función $h(\lambda, \theta)$, se puede decir que:

$$\sup \{|U_T(\theta_1) - U_T(\theta_2)|, \theta_1, \theta_2 \in \Theta, |\theta_1 - \theta_2| \leq \varepsilon\} \leq \eta(\varepsilon) \int_{\mathbb{T}} I_T^h(\lambda) \frac{w(\lambda)}{v(\lambda)} d\lambda,$$

y la integral del lado derecho de la desigualdad anterior está asintóticamente acotada en probabilidad, bajo las condiciones **B5** y **B6**.

Por lo tanto, en vista del Teorema 3.4.1 de Guyon (1995), concluimos que el estimador $\hat{\theta}_T$ es consistente: $\hat{\theta}_T \xrightarrow{P_0} \theta_0$, cuando $T \rightarrow \infty$.

La propiedad de mínimo contraste para $K(\theta_0; \theta)$ se deduce de la desigualdad de Jensen:

$$\begin{aligned} -K(\theta_0; \theta) &= \int_{\mathbb{T}} f(\lambda, \theta_0) w(\lambda) \log \frac{\psi(\lambda, \theta)}{\psi(\lambda, \theta_0)} d\lambda \\ &= \sigma^2(\theta_0) \int_{\mathbb{T}} \psi(\lambda, \theta_0) w(\lambda) \log \frac{\psi(\lambda, \theta)}{\psi(\lambda, \theta_0)} d\lambda \\ &\leq \sigma^2(\theta_0) \log \int_{\mathbb{T}} w(\lambda) \psi(\lambda, \theta) d\lambda = 0, \end{aligned}$$

por lo tanto, $K(\theta_0; \theta) \geq 0$ y, además, $K(\theta_0; \theta) > 0$ si $\psi(\lambda, \theta_0) \not\equiv \psi(\lambda, \theta)$, para $\theta_0 \neq \theta$, casi seguramente, respecto a la medida de Lebesgue.

La siguiente convergencia,

$$\hat{\sigma}_T^2 = \int_{\mathbb{T}} I_T^h(\lambda) w(\lambda) d\lambda \xrightarrow{P_0} \sigma^2(\theta_0) = \int_{\mathbb{T}} f(\lambda, \theta_0) w(\lambda) d\lambda \quad (2.12)$$

se puede establecer con el uso de los mismos argumentos utilizados en la convergencia de $U_T(\theta) \xrightarrow{P_0} U(\theta)$, según se ha indicado anteriormente. También se puede aplicar nuevamente la observación dada en la Nota 2.11, esto es, para establecer la expresión dada en (2.12) podemos utilizar el Teorema 2.4. Se puede ver que, bajo la Condición **B5**, $w(\lambda)$ compensa la singularidad de $f(\lambda, \theta_0)$ y, por lo tanto, así se llega a la convergencia (2.12). \square

Demostración del Teorema 2.2: Aplicando el teorema del valor medio, se tiene que

$$\nabla_{\theta} U_T(\hat{\theta}_T) = \nabla_{\theta} U_T(\theta_0) + \nabla_{\theta} \nabla'_{\theta} U_T(\theta_T^*)(\hat{\theta}_T - \theta_0),$$

donde $|\boldsymbol{\theta}_T^* - \boldsymbol{\theta}_0| < |\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0|$ y

$$\begin{aligned}\nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}) &= - \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\theta}} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \\ &= \left\{ - \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \right\}_{i=1, \dots, q}, \\ \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}) &= - \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \\ &= \left\{ - \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) w(\boldsymbol{\lambda}) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \right\}_{i, j=1, \dots, q}.\end{aligned}$$

De las condiciones asumidas en la formulación del estimador de mínimo contraste, se deduce que, para un N suficientemente grande,

$$\nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) = - \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_T^*) (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0). \quad (2.13)$$

Por lo tanto, queda por demostrar que

- (i) $\nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}) \rightarrow \mathbf{S}(\boldsymbol{\theta}_0)$ en P_0 -probabilidad,
- (ii) $T^{d/2} \nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} N_q(0, e(h) \mathbf{A}(\boldsymbol{\theta}_0))$,

donde $\mathbf{A}(\boldsymbol{\theta})$ y $\mathbf{S}(\boldsymbol{\theta})$ vienen dados en la Condición **B8**. Entonces, por el lema de Slutsky, la relación (2.8) es una consecuencia de (2.13) y (i)–(ii).

Obsérvese que al cumplirse la Condición **B4**,

$$\int_{\mathbb{T}} f(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) w(\boldsymbol{\lambda}) \nabla_{\boldsymbol{\theta}} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0) d\boldsymbol{\lambda} = 0,$$

por lo tanto, $\nabla_{\boldsymbol{\theta}} U_T(\boldsymbol{\theta}_0) = (J_T(\varphi_i) - J(\varphi_i))_{i=1, \dots, q}$, y la convergencia (ii) se cumplirá si se produce la siguiente convergencia

$$T^{d/2} (J_T(\varphi_i) - E J_T(\varphi_i))_{i=1, \dots, q} \xrightarrow{\mathcal{D}} N_q(0, e(h) \mathbf{A}(\boldsymbol{\theta}_0)) \quad (2.14)$$

y

$$T^{d/2} (E J_T(\varphi_i) - J(\varphi_i)) \rightarrow 0, \quad i = 1, \dots, q, \quad (2.15)$$

donde $\varphi_i(\boldsymbol{\lambda}) = w(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \log \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}_0)$. La convergencia (2.14) se cumple bajo las condiciones **B5** y **B7(2)**, en vista del Teorema 2.4; la convergencia (i) se cumple bajo las condiciones **B5**, **B6**, **B7(1)**, **B7(3)**, teniendo en cuenta la Nota 2.11 (y teniendo en cuenta la consistencia de $\hat{\boldsymbol{\theta}}_T$). La convergencia (2.15) se cum-

plirá bajo las condiciones **H1–H2** y **B9**, en vista del Lema 2.1, que se indica a continuación. Obsérvese que en los trabajos anteriores, la convergencia (2.15) se impuso como hipótesis, véase, por ejemplo Anh et al. (2007b); Avram et al. (2010b).

□

Lema 2.1. *Sea la función taper $\tilde{h}(t)$ tal que satisface los supuestos **H1** y **H2**. Supongamos además que una de las siguientes condiciones se cumple:*

- (i) *f tiene segunda derivada acotada y $\varphi \in L_1(\mathbb{T})$;*
- (ii) *φ tiene segunda derivada acotada;*
- (iii) *la convolución $g(u) = \int f(\boldsymbol{\lambda})\varphi(\boldsymbol{\lambda} + \mathbf{u}) d\boldsymbol{\lambda}$ tiene segunda derivada acotada en cero.*

Entonces, cuando $T \rightarrow \infty$,

$$EJ_T(\varphi) - J(\varphi) = O(T^{-2}). \quad (2.16)$$

Demostración:

El análogo de este lema para los campos paramétricos continuos se obtuvo en Sakhno (2007); los resultados sobre la evaluación del sesgo para los funcionales espectrales de órdenes superiores en contextos discretos y continuos se establecieron en Anh et al. (2007b) y Avram et al. (2010b). La presente demostración utiliza las ideas de estos artículos. Tras realizar algunos cálculos sencillos, el sesgo de $J_T(\varphi)$ se puede expresar como sigue:

$$\begin{aligned} E[J_T(\varphi) - J(\varphi)] &= \int \int \varphi(\boldsymbol{\lambda})f(\mathbf{u}) \Phi_{2,T}^h(\mathbf{u} - \boldsymbol{\lambda}) d\mathbf{u}d\boldsymbol{\lambda} - \int \varphi(\boldsymbol{\lambda})f(\boldsymbol{\lambda})d\boldsymbol{\lambda} \\ &= \int \int \varphi(\boldsymbol{\lambda}) (f(\boldsymbol{\lambda} + \mathbf{u}) - f(\boldsymbol{\lambda})) \Phi_{2,T}^h(\mathbf{u}) d\mathbf{u}d\boldsymbol{\lambda} \end{aligned} \quad (2.17)$$

$$= \int \int f(\boldsymbol{\lambda}) (\varphi(\boldsymbol{\lambda} + \mathbf{u}) - \varphi(\boldsymbol{\lambda})) \Phi_{2,T}^h(\boldsymbol{\lambda}) d\mathbf{u}d\boldsymbol{\lambda} \quad (2.18)$$

$$= \int (g(\mathbf{u}) - g(\mathbf{0})) \Phi_{2,T}^h(\mathbf{u}) d\mathbf{u}, \quad (2.19)$$

donde $\Phi_{2,T}^h(\mathbf{u}) = [(2\pi)^d H_{2,T}(\mathbf{0})]^{-1} |H_{1,T}(\mathbf{u})|^2$, es una función par que se integra a 1 (más específicamente, este es el *kernel* de tipo Fejér; véase la Sección 2.4), con $H_{1,T}(\mathbf{u})$ que viene dado en (1.4) para $k = 1$, y denotando $g(\mathbf{u}) = \int f(\boldsymbol{\lambda})\varphi(\boldsymbol{\lambda} + \mathbf{u}) d\boldsymbol{\lambda}$.

Por lo tanto, para evaluar el sesgo necesitamos analizar el comportamiento asintótico de las expresiones (2.17)–(2.19). Aquí, se pueden aplicar argumentos estándar si imponemos condiciones de regularidad sobre las funciones f , φ , o, más generalmente, sobre su convolución g .

Consideremos la expresión (2.17) y la Condición (i).

Por el teorema de Taylor y la suposición de que $f(\lambda)$ tiene segunda derivada acotada en \mathbb{T} ,

$$\left| f(\lambda + \mathbf{u}) - f(\lambda) - \sum_{i=1}^d u_i \frac{\partial f(\lambda)}{\partial \lambda_i} \right| \leq \text{const} \cdot \sum_{i=1}^d |u_i|^2.$$

Consideremos la expresión

$$\begin{aligned} & \sum_{i=1}^d \int_{\mathbb{T}} \varphi(\lambda) \frac{\partial f(\lambda)}{\partial \lambda_i} \int_{\mathbb{T}} u_i \frac{|H_{1,T}(\mathbf{u})|^2}{(2\pi)^d H_{2,T}(\mathbf{0})} d\mathbf{u} d\lambda \\ &= \sum_{i=1}^d \int_{\mathbb{T}} \varphi(\lambda) \frac{\partial f(\lambda)}{\partial \lambda_i} \int_{\mathbb{T}} u_i \frac{\prod_{i=1}^d |\tilde{H}_{1,T}(u_i)|^2}{(2\pi)^d H_{2,T}(\mathbf{0})} d\mathbf{u} d\lambda. \end{aligned} \quad (2.20)$$

Como $\left| \tilde{H}_{1,T}(u_i) \right|^2 \cdot (2\pi \tilde{H}_{2,T}(0))^{-1}$ es un *kernel* de tipo Fejér e integra 1, siguiendo lo planteado en Dahlhaus (1983), las integrales dadas en (2.20) se reducen a la expresión

$$\int_{-\pi}^{\pi} u_i \frac{|\tilde{H}_{1,T}(u_i)|^2}{2\pi \tilde{H}_{2,T}(0)} du_i,$$

la cual es igual a cero, ya que $|\tilde{H}_{1,T}(u_i)|^2$ es una función par. Por lo tanto, (2.20) es igual a cero y

$$|EJ_T(\varphi) - J(\varphi)| \leq \text{const} \cdot \sum_{i=1}^d \int_{\mathbb{T}} |\varphi(\lambda)| \int_{\mathbb{T}} |u_i|^2 \prod_{i=1}^d \frac{|\tilde{H}_{1,T}(u_i)|^2}{2\pi \tilde{H}_{2,T}(0)} d\mathbf{u} d\lambda. \quad (2.21)$$

El orden de las integrales anteriores en T se reduce al orden de las integrales

$$\int_{-\pi}^{\pi} |u_i|^2 \frac{|\tilde{H}_{1,T}(u_i)|^2}{2\pi \tilde{H}_{2,T}(0)} du_i, \text{ para } i = 1, \dots, d.$$

Observamos que

$$\tilde{H}_{2,T}(0) \sim T \int \tilde{h}^2(t) dt \quad (2.22)$$

cuando $T \rightarrow \infty$.

A continuación, se procede como en Brillinger (1981), obteniéndose:

$$\tilde{H}_{1,T}(u) = \sum_t \tilde{h}(t/T) e^{-iut} = - \sum_t \Delta^t(u) \left(\tilde{h}((t+1)/T) - \tilde{h}(t/T) \right), \quad (2.23)$$

donde $\Delta^t(u)$ involucra $(\sin(u/2))^{-1}$ multiplicado por un término acotado respecto a T y $\Delta^0(u) \equiv 1$. Para más detalles se puede consultar, por ejemplo, Dahlhaus (1983), o, en concreto, la demostración del Corolario 7.2.1 dada en Brillinger (1981). Dado que la función tapered $\tilde{h}(t)$, es Lipschitz continua, podemos afirmar la siguiente estimación:

$$\int_{-\pi}^{\pi} u^2 \left| \tilde{H}_{k,T}(u) \right|^2 du \leq const \sum_t \left| \tilde{h}((t+1)/T) - \tilde{h}(t/T) \right|^2 \leq const \frac{1}{T}. \quad (2.24)$$

Argumentos similares se aplican en Robinson (2007), ver Teorema 3 y ecuación (3.21). Por lo tanto, cada término en (2.21) está acotado por $const \cdot T^{-2}$, lo que nos da el resultado asintótico (2.16) bajo la Condición (i). \square

2.3. Ejemplos numéricos

Esta sección ilustra los resultados obtenidos e investiga numéricamente las propiedades de los estimadores de mínimo contraste considerados. Primero, consideramos la familia de procesos autorregresivos fraccionarios espaciales introducidos en Boissy et al. (2005). Los estudios de simulación también sugieren que resultados similares son válidos para campos aleatorios de Gegenbauer, véase Espejo et al. (2015), y el enfoque es aplicable a modelos más generales.

En los siguientes ejemplos se utilizan modelos de campos aleatorios con densidades espectrales de la forma

$$f(\lambda, \theta) = \frac{\sigma_\varepsilon^2}{(2\pi)^2} \frac{|1 - 2u_1 e^{-i\lambda_1} + e^{-2i\lambda_1}|^{-2d_1} |1 - 2u_2 e^{-i\lambda_2} + e^{-2i\lambda_2}|^{-2d_2}}{|\phi(\exp(-i\lambda_1), \exp(-i\lambda_2), \alpha, \beta)|^2}, \quad (2.25)$$

donde $\phi(z_1, z_2, \alpha, \beta) = (1 - \alpha z_1)(1 - \beta z_2)$.

En el Ejemplo 2.3.1, los valores de los parámetros u_1 y u_2 son $u_1 = u_2 = 1$. En el Ejemplo 2.3.3, $\alpha = \beta = 0$ y $\phi(z_1, z_2, 0, 0) \equiv 1$. En el Ejemplo 2.3.2 se considera el caso general, donde las singularidades están localizadas fuera del origen, y suponemos nuevamente que $\alpha = \beta = 0$.

En todos los casos se aprecia que el enfoque de mínimo contraste proporciona muy buenas tasas de convergencia y normalidad asintótica.

En todos los ejemplos que vienen a continuación suponemos que los campos aleatorios $Y(\mathbf{t})$ se observan en la rejilla $\{(t_1, t_2) : t_1, t_2 = 1, \dots, T\}$. Así, utilizamos la ventana de observación a partir de la Nota 2.5. Obsérvese que en nuestro caso $\mathbb{T} = [-\pi, \pi]$.

Elegimos la función *taper* de la forma $h_T(t_1, t_2) = \tilde{h}(t_1/T) \tilde{h}(t_2/T)$, con $\tilde{h}(t) = \frac{1}{2}(1 - \cos(2\pi t))$, $t \in [0, 1]$, y $w(\boldsymbol{\lambda}) = w(\lambda_1, \lambda_2) = |\lambda_1|^{2a_1} |\lambda_2|^{2a_2}$, $a_i > 1/2$, $i = 1, 2$. Entonces, $w(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in [-\pi, \pi]^2$, es una función que satisface la Suposición B3. La función *taper*, $h(\cdot)$, también satisface las condiciones de la Nota 2.5.

2.3.1. Procesos autorregresivos fraccionarios espaciales

Primeramente, partimos de la familia de procesos autorregresivos fraccionarios espaciales introducidos en Boissy et al. (2005).

Sea $Y_{\mathbf{t}}$, $\mathbf{t} = (t_1, t_2) \in \mathbb{Z}^2$, el proceso espacial que satisface el siguiente modelo autorregresivo fraccionario:

$$\phi(B_1, B_2, \alpha, \beta) \nabla_1^{d_1} \nabla_2^{d_2} Y_{t_1, t_2} = \epsilon_{t_1, t_2}, \quad (t_1, t_2) \in \mathbb{Z}^2, \quad (2.26)$$

donde $d_i \in (-1/2, 1/2)$, $i = 1, 2$, $\phi(B_1, B_2, \alpha, \beta) Y_{t_1, t_2} = Y_{t_1, t_2} - \alpha Y_{t_1-1, t_2} - \beta Y_{t_1, t_2-1} + \alpha\beta Y_{t_1-1, t_2-1}$, y $\nabla_1^{d_1} \nabla_2^{d_2} = (1 - B_1)^{d_1} (1 - B_2)^{d_2}$, con B_i denotando el operador de retardos para la coordenada t_i , $i = 1, 2$, es decir, $B_1 Y_{t_1, t_2} = Y_{t_1-1, t_2}$, y $B_2 Y_{t_1, t_2} = Y_{t_1, t_2-1}$.

La densidad espectral del proceso definido por (2.26) viene dada por

$$f_Y(\boldsymbol{\lambda}) = \frac{\sigma^2}{4\pi^2} \frac{|1 - \exp(-i\lambda_1)|^{-2d_1} |1 - \exp(-i\lambda_2)|^{-2d_2}}{|\phi(\exp(-i\lambda_1), \exp(-i\lambda_2), \alpha, \beta)|^2}, \quad \boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in [-\pi, \pi]^2,$$

véase Boissy et al. (2005).

Supongamos que los valores de los parámetros α y β son conocidos o han sido estimados previamente. Entonces, estamos interesados en estimar el vector paramétrico $(d_1, d_2) \in (0, 1/2)^2$. Esto significa que $\boldsymbol{\theta} = (\theta_1, \theta_2) = (d_1, d_2)$ y $\boldsymbol{\theta} \in \Theta = (0, 1/2)^2$.

Mediante cálculos directos es fácil comprobar que se cumplen todas las condiciones de consistencia y normalidad asintótica del estimador $\hat{\boldsymbol{\theta}}_T$. En particular, para la Condición **B6**, podemos usar la función $v(\boldsymbol{\lambda}) = v(\lambda_1, \lambda_2) = |\lambda_1|^{2\beta} |\lambda_2|^{2\beta}$, con $\beta \in (0, 1/2)$. Para verificar la definición positiva de las matrices $S(\boldsymbol{\theta})$ y $A(\boldsymbol{\theta})$, como se requiere en la Condición **B8**, podemos seguir análogamente lo considerado de forma muy detallada en Espejo et al. (2015), ver verificación de la Condición A8.

Así, tenemos la consistencia del estimador y la normalidad asintótica

$$T^{d/2}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{D} \mathcal{N}_q(0, e(h)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)A(\boldsymbol{\theta}_0)\mathbf{S}^{-1}(\boldsymbol{\theta}_0)), \quad (2.27)$$

donde, para la función *taper* escogida, tenemos $e(h) \approx 3.78$. Los elementos de las matrices $\mathbf{S}(\boldsymbol{\theta})$ y $\mathbf{A}(\boldsymbol{\theta})$ se calculan de forma análoga a pp. 671 y 672, en Espejo et al. (2015). En particular, nótese que

$$\begin{aligned} \frac{\partial}{\partial \theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= -2 \log \left| 2 \operatorname{sen} \frac{\lambda_i}{2} \right| f(\boldsymbol{\lambda}, \boldsymbol{\theta}), \\ \frac{\partial}{\partial \theta_i} \sigma^2(\boldsymbol{\theta}) &= -2 \int_{[-\pi, \pi]^2} \log \left| 2 \operatorname{sen} \frac{\lambda_i}{2} \right| w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}. \end{aligned}$$

Así se obtiene,

$$\begin{aligned} s_{ij} &= 3 \int_{[-\pi, \pi]^2} \frac{w(\boldsymbol{\lambda})}{\sigma^2(\boldsymbol{\theta})} \left[\frac{\partial}{\partial \theta_j} \sigma^2(\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial \theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right] d\boldsymbol{\lambda} \\ &\quad - \int_{[-\pi, \pi]^2} \frac{w(\boldsymbol{\lambda})}{f(\boldsymbol{\lambda}, \boldsymbol{\theta})} \left[\frac{\partial}{\partial \theta_i} \sigma^2(\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial \theta_j} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right] d\boldsymbol{\lambda} \\ &= \frac{3}{\sigma^2(\boldsymbol{\theta})} \left[\frac{\partial}{\partial \theta_j} \sigma^2(\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial \theta_i} \sigma^2(\boldsymbol{\theta}) \right] \\ &\quad - 4 \int_{[-\pi, \pi]^2} \log \left| 2 \operatorname{sen} \frac{\lambda_i}{2} \right| \log \left| 2 \operatorname{sen} \frac{\lambda_j}{2} \right| w(\boldsymbol{\lambda}) f(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}, \end{aligned}$$

y

$$\begin{aligned}
 a_{ij} &= 8\pi^2 \sigma^4(\boldsymbol{\theta}) \int_{[-\pi, \pi]^2} w^2(\boldsymbol{\lambda}) \frac{\partial}{\partial \theta_i} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} \psi(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda} \\
 &= 8\pi^2 \int_{[-\pi, \pi]^2} w^2(\boldsymbol{\lambda}) \left(\left[\frac{\partial}{\partial \theta_i} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right] \sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial \theta_i} \sigma^2(\boldsymbol{\theta}) \right] f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \\
 &\quad \times \left(\left[\frac{\partial}{\partial \theta_j} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right] \sigma^2(\boldsymbol{\theta}) - \left[\frac{\partial}{\partial \theta_j} \sigma^2(\boldsymbol{\theta}) \right] f(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) d\boldsymbol{\lambda}.
 \end{aligned}$$

Por lo tanto, podemos considerar $a_{ij} = S_1 - S_2(i, j) - S_2(j, i) + S_3$, donde

$$\begin{aligned}
 S_1 &= 32\pi^2 \sigma^4(\boldsymbol{\theta}) \int_{[-\pi, \pi]^2} \log \left| 2 \operatorname{sen} \frac{\lambda_i}{2} \right| \log \left| 2 \operatorname{sen} \frac{\lambda_j}{2} \right| w^2(\boldsymbol{\lambda}) f^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}, \\
 S_2(i, j) &= 16\pi^2 \sigma^2(\boldsymbol{\theta}) \left[\frac{\partial}{\partial \theta_j} \sigma^2(\boldsymbol{\theta}) \right] \int_{[-\pi, \pi]^2} \log \left| 2 \operatorname{sen} \frac{\lambda_i}{2} \right| w^2(\boldsymbol{\lambda}) f^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}, \\
 S_3 &= 8\pi^2 \left[\frac{\partial}{\partial \theta_i} \sigma^2(\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial \theta_j} \sigma^2(\boldsymbol{\theta}) \right] \int_{[-\pi, \pi]^2} w^2(\boldsymbol{\lambda}) f^2(\boldsymbol{\lambda}, \boldsymbol{\theta}) d\boldsymbol{\lambda}.
 \end{aligned}$$

En lo que sigue, se proporcionan resultados numéricos para el modelo autorregresivo fraccionario espacial (2.26). Se generaron 100 realizaciones de Y_{t_1, t_2} para cada T en la secuencia de tamaño muestral creciente $T=10, 30, 50, 70, 90, 110, 130, 150$. Para los parámetros utilizados en las simulaciones se usaron los siguientes valores: $d_1=0.2$, $d_2=0.3$, $\alpha=0.1$, $\beta=0.2$, y $\sigma_\varepsilon^2=1$.

El operador $(1-B_1)^{d_1}(1-B_2)^{d_2}$ de la ecuación (2.26) fue ampliado en una serie doble de potencias con respecto a cada B_i , $i = 1, 2$. Las realizaciones de Y se aproximaron usando las sumas truncadas con potencias sin exceder 30 términos para cada B_i . El periodograma I_T^h fue calculado con $h_T(t_1, t_2) = \tilde{h}(t_1/T)\tilde{h}(t_2/T)$, donde $\tilde{h}(t) = \frac{1}{2}(1 - \cos(2\pi t))$, $t \in [0, 1]$; y también con $\omega(\lambda_1, \lambda_2)$ dados anteriormente.

Los argumentos de minimización $\hat{\boldsymbol{\theta}}_T$ del funcional $U_T(\boldsymbol{\theta})$ en (2.4) se encontraron numéricamente para cada simulación.

Para cada T , la Tabla 2.1 muestra los valores estimados \hat{d}_1 y \hat{d}_2 de los parámetros, y los errores cuadrados medios (ECM) de \hat{d}_1 y \hat{d}_2 . Para cada valor de T la Figura 2.1 produce un diagrama de cajas de valores estimados \hat{d}_1 y \hat{d}_2 basado en 100 realizaciones. Se ve claramente que los valores estimados están centrados en los valores verdaderos $d_1=0.2$ y $d_2=0.3$.

Se realizó el gráfico Q-Q de normalidad de \hat{d}_1 y \hat{d}_2 para valores grandes

Tabla 2.1: Media, desviación estándar y ECM de \hat{d}_1 y \hat{d}_2

T	\hat{d}_1	\hat{d}_2	ECM de \hat{d}_1	ECM de \hat{d}_2
10	0.227 (0.122)	0.403 (0.141)	0.02	0.02
30	0.219 (0.054)	0.361 (0.072)	0.003	0.01
50	0.206 (0.034)	0.346 (0.054)	0.0012	0.03
70	0.199 (0.027)	0.326 (0.034)	0.001	0.0012
90	0.205 (0.029)	0.314 (0.025)	0.001	0.001
110	0.199 (0.026)	0.311 (0.022)	0.001	0.001
130	0.197 (0.027)	0.310 (0.025)	0.001	0.001
150	0.199 (0.022)	0.307 (0.022)	0.0005	0.001

de T y así verificar la normalidad asintótica de los estimadores. La Figura 2.2 demuestra que las distribuciones empíricas de \hat{d}_1 y \hat{d}_2 coinciden con la distribuciones normales teóricas.

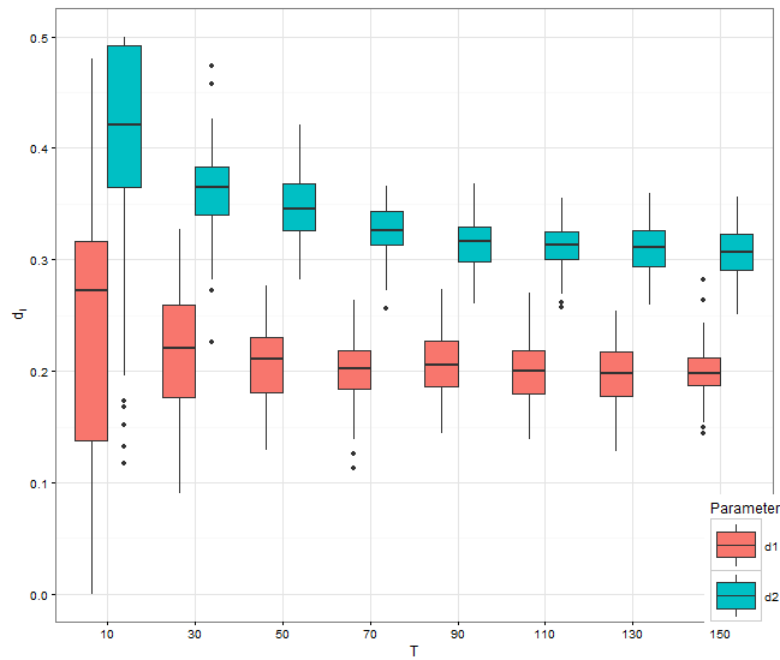


Figura 2.1: Diagrama de cajas de valores muestrales de \hat{d}_1 y \hat{d}_2

Para cada T , se usaron 100 valores simulados para estimar $P_0(|\hat{\theta}_T - \theta_0| < \varepsilon)$. El gráfico de las probabilidades muestrales $P_0(|\hat{\theta}_T - \theta_0| < \varepsilon)$ en la Figura 2.3 confirma la convergencia de $\hat{\theta}_T = (\hat{d}_1, \hat{d}_2)$ a $\theta_0 = (0.2, 0.3)$ en probabilidad.

Finalmente, se investigan las propiedades asintóticas de $\hat{\sigma}_T$. De manera similar a los casos anteriores, para cada T se construyen los diagramas de cajas y el gráfico de las probabilidades muestrales para 100 simulaciones. Las figuras

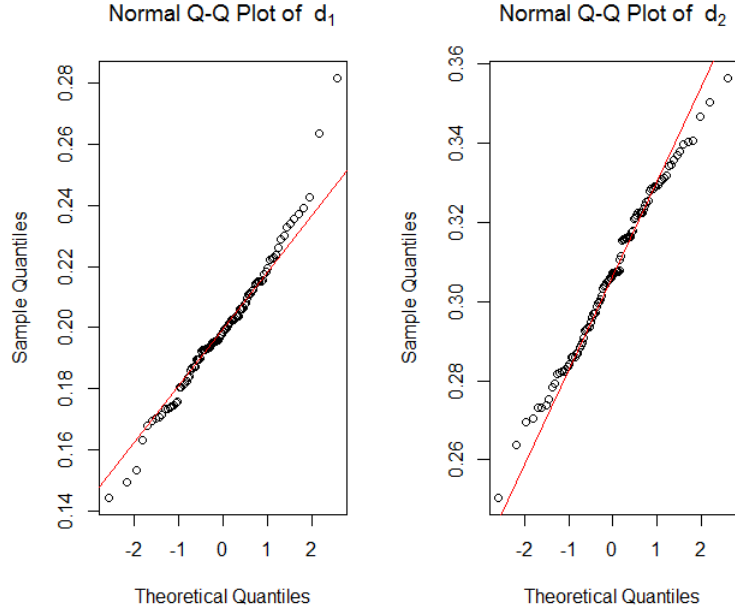


Figura 2.2: Gráfico Q-Q de normalidad de \hat{d}_1 y \hat{d}_2

2.4 y 2.5 respaldan la convergencia en probabilidad $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon) \rightarrow 1$, cuando T aumenta.

2.3.2. Campos aleatorios de Gegenbauer espaciales

Esta sección ilustra la técnica de estimación de mínimo contraste para el caso de campos aleatorios de Gegenbauer.

Sea Y_{t_1, t_2} un campo aleatorio discreto que satisface que

$$\Delta_{u_1}^{d_1} \Delta_{u_2}^{d_2} Y_{t_1, t_2} = (I - 2u_1 B_1 + B_1^2)^{d_1} (I - 2u_2 B_2 + B_2^2)^{d_2} Y_{t_1, t_2} = \epsilon_{t_1, t_2},$$

donde ϵ_{t_1, t_2} , con $(t_1, t_2) \in \mathbb{Z}^2$, es un campo de ruido blanco de media cero y varianza común $E[\epsilon_{t_1, t_2}^2] = \sigma_\epsilon^2$, y, para $j = 1, 2$,

$$\begin{aligned} \Delta_{u_j}^{d_j} &= (I - 2u_j B_j + B_j^2)^{d_j} \\ &= (1 - 2 \cos \nu_j B_j + B_j^2)^{d_j} \\ &= [(1 - \exp(i\nu_j) B_j)(1 - \exp(-i\nu_j) B_j)]^{d_j}. \end{aligned}$$

Como antes, B_j , $j = 1, 2$, denota el operador de retardos para cada coordenada

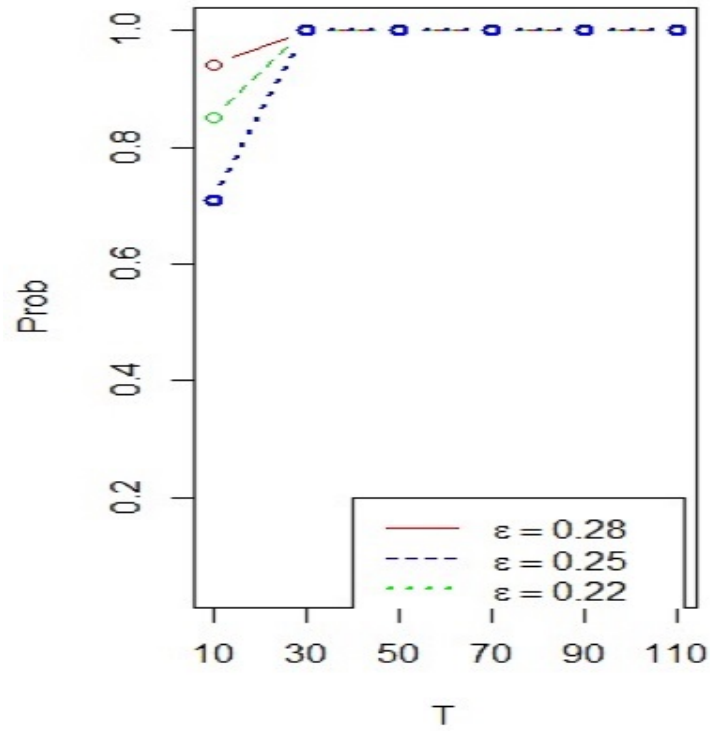


Figura 2.3: Probabilidades muestrales $P_0(|\hat{\theta}_T - \theta_0| < \epsilon)$

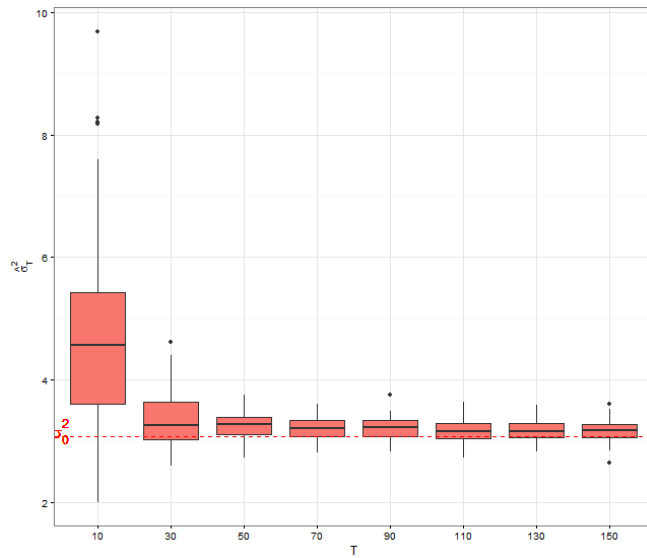


Figura 2.4: Diagrama de cajas de valores muestrales de $\hat{\sigma}_T^2$

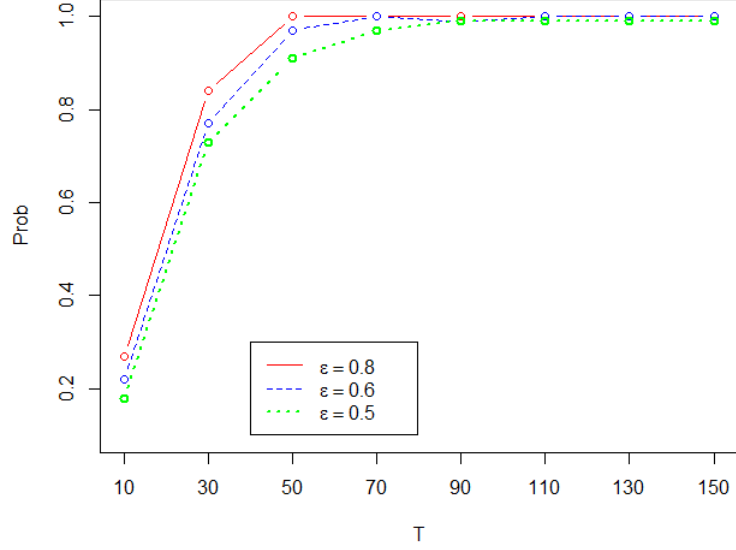


Figura 2.5: Probabilidades muestrales $P_0(|\hat{\sigma}_T^2 - \sigma^2(\theta)| < \epsilon)$

espacial, y $u = \cos \nu$, es decir, $\nu = \arccos(u)$.

Existe la siguiente representación de un campo aleatorio espacial Gegenbauer estacionario, véase Espejo et al. (2015),

$$Y_{t_1, t_2} = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} C_{n_1}^{(d_1)}(u_1) C_{n_2}^{(d_2)}(u_2) \epsilon_{t_1 - n_1, t_2 - n_2}, \quad (2.28)$$

donde $d_i \neq 0$, y $|u_i| \leq 1$, $i = 1, 2$.

El polinomio de Gegenbauer $C_n^{(d)}(u)$ viene dado por

$$C_n^{(d)}(u) = \sum_{k=0}^{[n/2]} (-1)^k \frac{(2u)^{n-2k} \Gamma(d-k+n)}{k!(n-2k)! \Gamma(d)}.$$

La densidad espectral f del campo aleatorio Gegenbauer espacial se define como

$$\begin{aligned} f(\boldsymbol{\lambda}, \boldsymbol{\theta}) &= \frac{\sigma_\epsilon^2}{(2\pi)^2} |1 - 2u_1 e^{-i\lambda_1} + e^{-2i\lambda_1}|^{-2d_1} |1 - 2u_2 e^{-i\lambda_2} + e^{-2i\lambda_2}|^{-2d_2} \\ &= \frac{\sigma_\epsilon^2}{(2\pi)^2} |2 \cos(\lambda_1) - 2u_1|^{-2d_1} |2 \cos(\lambda_2) - 2u_2|^{-2d_2}. \end{aligned}$$

Para la secuencia creciente de tamaños de muestra, $T=10, 30, 50, 70, 90, 110$, se generaron 200 repeticiones de campo aleatorio en (2.28), utilizando los valores de los parámetros $d_1=0.2, d_2=0.3, u_1=0.4, u_2=0.3$, y $\sigma_\epsilon^2=1$.

Análogamente al ejemplo de la Sección 2.3.1, se aproximaron realizaciones de Y_{t_1, t_2} por sumas truncadas con 100 términos en (2.28). El periodograma I_T^h se calculó con $h_T(t_1, t_2)=\tilde{h}(t_1/T)\tilde{h}(t_2/T)$, $\tilde{h}(t) = \frac{1}{2}(1 - \cos(2\pi t))$, $t \in [0, 1]$. Se utilizó la función peso $\omega(\lambda_1, \lambda_2) = |2 \cos(\lambda_1) - 2u_1|^2 |2 \cos(\lambda_2) - 2u_2|^2$.

El análisis, las gráficas y las conclusiones sobre la ilustración de los resultados obtenidos son análogas al ejemplo de la Sección 2.3.1. Para cada simulación, se encontraron los argumentos de minimización $\hat{\theta}_T$ del funcional $U_T(\theta)$ en (2.4), ver Figura 2.6 y Tabla 2.2, en relación con los valores muestrales y errores cuadráticos medios (ECM), respectivamente.

Tabla 2.2: Media, desviación estándar y ECM de \hat{d}_1 y \hat{d}_2

T	\hat{d}_1	\hat{d}_2	ECM de \hat{d}_1	ECM de \hat{d}_2
10	0.264 (0.186)	0.334 (0.181)	0.04	0.035
30	0.220 (0.115)	0.346 (0.130)	0.013	0.02
50	0.208 (0.078)	0.314 (0.085)	0.01	0.01
70	0.205 (0.063)	0.313 (0.085)	0.004	0.005
90	0.204 (0.063)	0.298 (0.064)	0.003	0.003
110	0.193 (0.059)	0.304 (0.060)	0.003	0.003

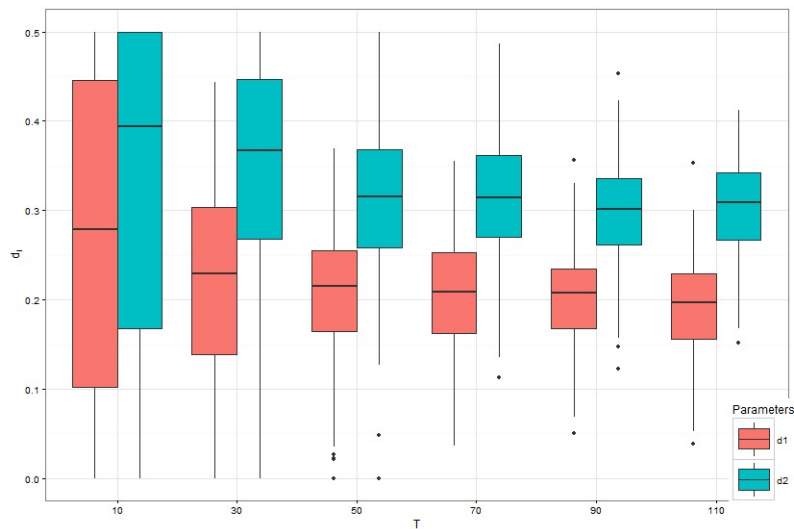


Figura 2.6: Diagrama de cajas de valores muestrales de \hat{d}_1 y \hat{d}_2

Los gráficos Q-Q de normalidad de \hat{d}_1 y \hat{d}_2 en la Figura 2.7 coinciden con las distribuciones normales teóricas. La gráfica de probabilidades muestrales $P_0(|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$ en la Figura 2.8 también confirma la convergencia en probabilidad. Las figuras 2.9 y 2.10 respaldan la convergencia en probabilidad $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon) \rightarrow 1$ cuando T aumenta.

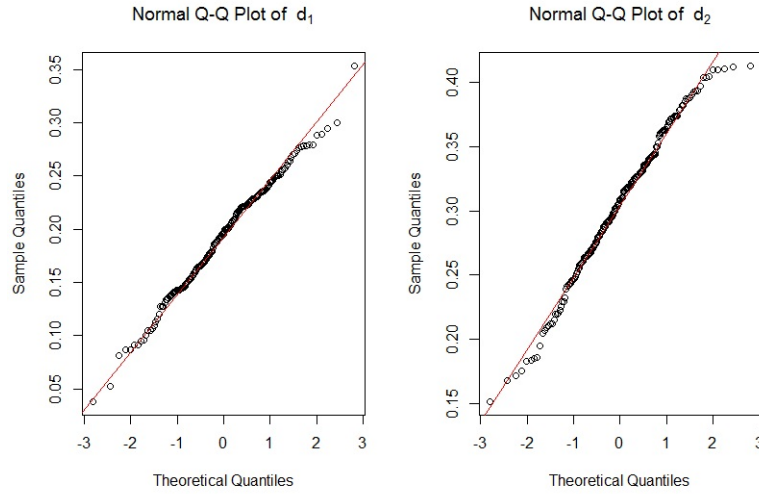


Figura 2.7: Gráficos Q-Q de normalidad de \hat{d}_1 y \hat{d}_2

Resultados similares también se obtuvieron usando la función peso $\omega(\lambda_1, \lambda_2) = (|\lambda_1| - \arccos(u_1))^2 (|\lambda_2| - \arccos(u_2))^2$.

2.3.3. Campos aleatorios espaciales de Gegenbauer: singularidades en el origen

Finalmente, en el tercer ejemplo, se considera el caso de densidades espectrales con singularidades en el origen y la generación de $\phi(B_1, B_2, \alpha, \beta) = I$, es decir, $\alpha = \beta = 0$.

En este ejemplo se trabaja de nuevo con las generaciones de campos aleatorios espaciales que satisfacen (2.28). Sus realizaciones fueron simuladas para los siguientes valores de los parámetros: $u_i=1$, con $i = 1, 2$, $\sigma_\epsilon^2=1$, $d_1 = 0.2$ y $d_2 = 0.3$. Las realizaciones del campo aleatorio espacial Y fueron generadas a partir de sumas truncadas con 100 términos en la ecuación (2.28): El periodograma I_T^h se calculó con $h_T(t_1, t_2)=\tilde{h}(t_1/T)\tilde{h}(t_2/T)$, $\tilde{h}(t) = \frac{1}{2}(1 - \cos(2\pi t))$, $t \in [0, 1]$. La función peso utilizada fue la siguiente: $\omega(\lambda_1, \lambda_2) =$

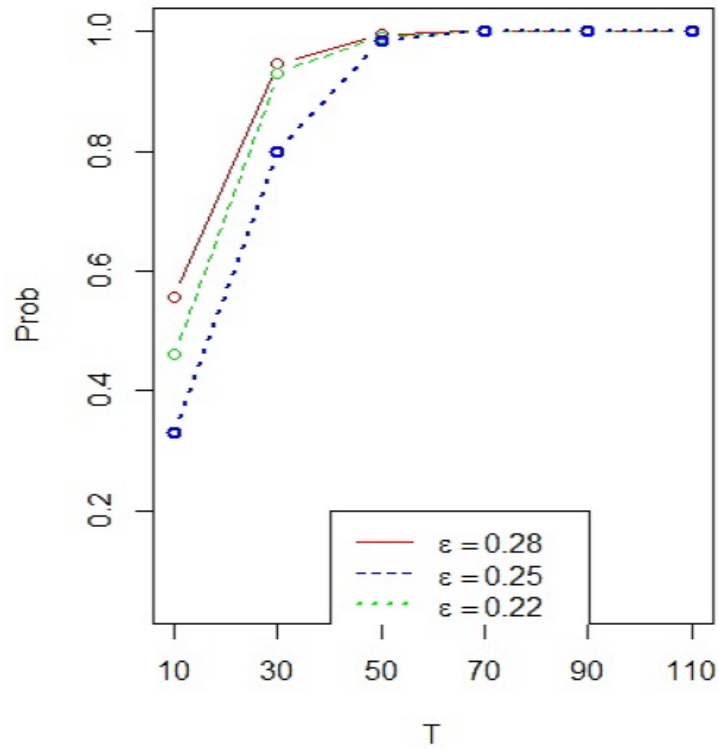


Figura 2.8: Probabilidades muestrales $P_0(|\hat{\theta}_T - \theta_0| < \epsilon)$

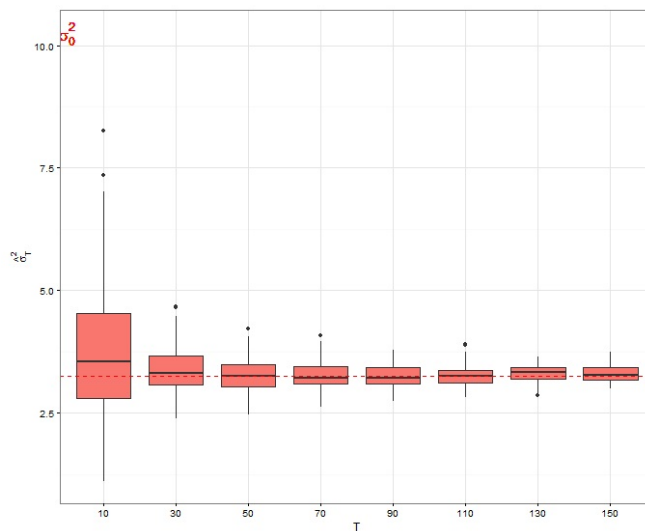


Figura 2.9: Diagramas de valores muestrales de $\hat{\sigma}_T^2$

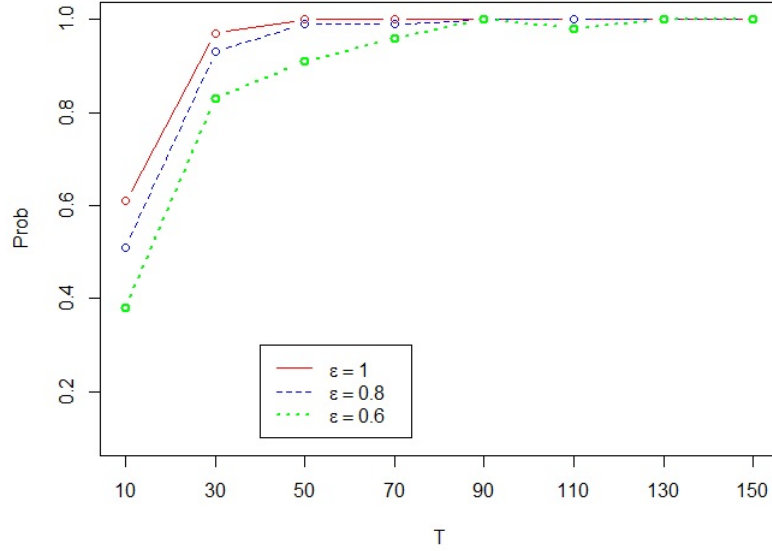


Figura 2.10: Probabilidades muestrales $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon)$

$(|\lambda_1| - \arccos(u_1))^2(|\lambda_2| - \arccos(u_2))^2 = |\lambda_1|^2|\lambda_2|^2$. Resultados similares se obtuvieron también para $\omega(\lambda_1, \lambda_2) = |2 \cos(\lambda_1) - 2u_1|^2|2 \cos(\lambda_2) - 2u_2|^2$.

El análisis, las gráficas y las son análogas al ejemplo de la Sección 2.3.1. Los argumentos de minimización $\hat{\boldsymbol{\theta}}_T$ del funcional $U_T(\boldsymbol{\theta})$ en (2.4) se encontraron numéricamente para cada simulación (ver Figura 2.11 y Tabla 2.3).

Tabla 2.3: Media, desviación estándar y ECM de \hat{d}_1 y \hat{d}_2

T	\hat{d}_1	\hat{d}_2	ECM de \hat{d}_1	ECM de \hat{d}_2
10	0.284 (0.211)	0.361 (0.184)	0.04	0.034
30	0.219 (0.141)	0.356 (0.130)	0.02	0.02
50	0.218 (0.105)	0.323 (0.102)	0.01	0.01
70	0.203 (0.092)	0.301 (0.098)	0.01	0.01
90	0.206 (0.091)	0.304 (0.089)	0.01	0.01
110	0.190 (0.096)	0.299 (0.082)	0.01	0.01

Los gráficos Q-Q de normalidad de \hat{d}_1 y \hat{d}_2 , en la Figura 2.12, coinciden con las distribuciones normales teóricas. El gráfico de las probabilidades muestrales $P_0(|\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0| < \varepsilon)$, en la Figura 2.13, también confirma la convergencia en probabilidad. La Tabla 2.3 muestra el error cuadrático medio de \hat{d}_1 y \hat{d}_2 , calculado para diferentes valores de T . Las figuras 2.14 y 2.15 respaldan

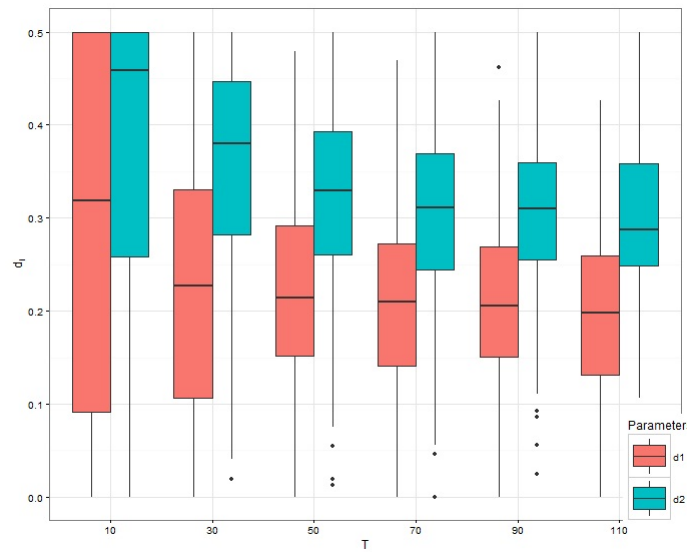


Figura 2.11: Diagrama de cajas de los valores muestrales de \hat{d}_1 y \hat{d}_2 con $u_i = 1$, $i = \{1, 2\}$

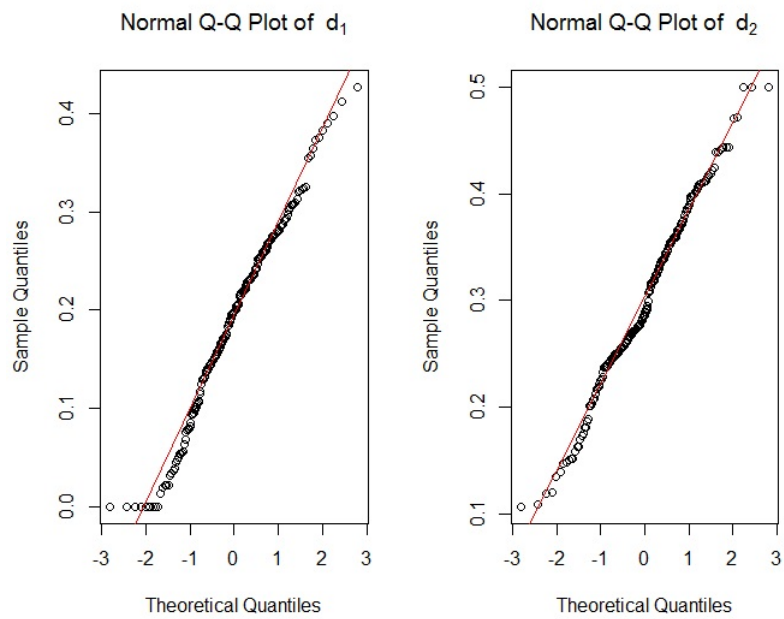


Figura 2.12: Gráfico Q-Q de normalidad de \hat{d}_1 y \hat{d}_2 con $u_i = 1$, $i = \{1, 2\}$

la convergencia en probabilidad, $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon) \rightarrow 1$, cuando T aumenta. Resultados similares se encontraron con la función peso: $\omega(\lambda_1, \lambda_2) = |2 \cos(\lambda_1) - 2u_1|^2 |2 \cos(\lambda_2) - 2u_2|^2$.

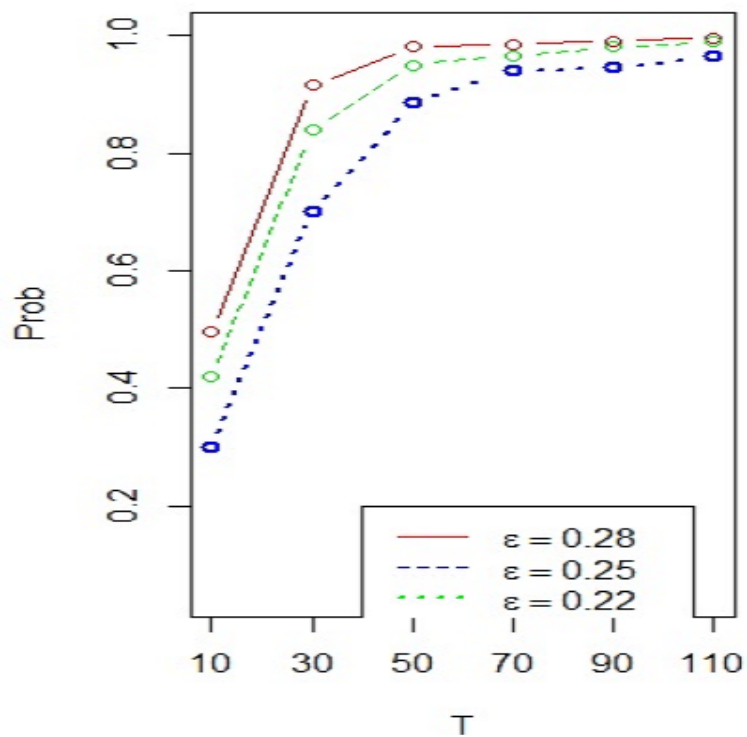


Figura 2.13: Probabilidades muestrales $P_0(|\hat{\theta}_T - \theta_0| < \epsilon)$

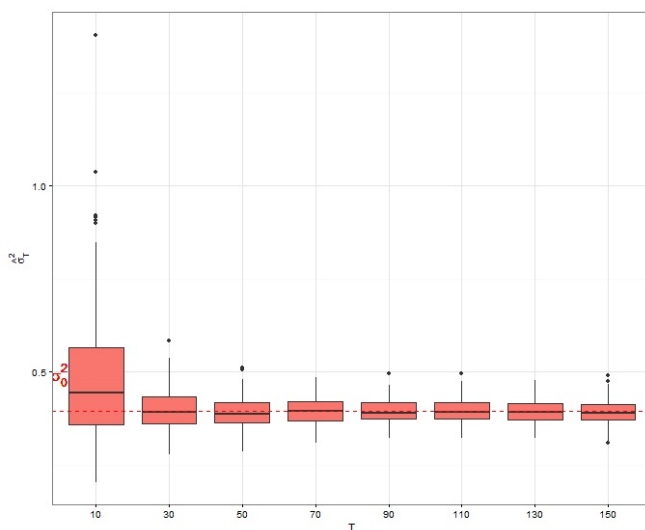


Figura 2.14: Diagrama de cajas de valores muestrales de $\hat{\sigma}_T^2$.

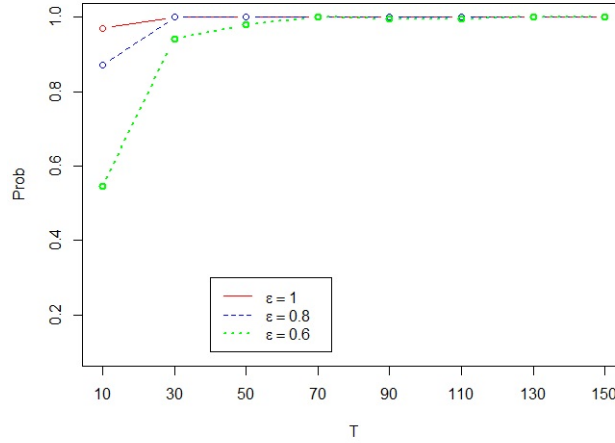


Figura 2.15: Probabilidades muestrales $P_0(|\hat{\sigma}_T^2 - \sigma^2(\boldsymbol{\theta})| < \varepsilon)$

2.4. Teoremas centrales del límite

Uno de los enfoques clásicos para llegar al teorema central del límite para los funcionales espectrales, $J_T(\varphi) = \int_{\mathbb{T}} I_T(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$, consiste en calcular y evaluar sus cumulantes. Se presentan aquí algunos detalles de este enfoque, y los resultados correspondientes para el caso de campos aleatorios gaussianos y datos tapered, según se ha considerado en el desarrollo del presente capítulo. Para establecer el teorema central del límite (TCL) para el funcional normalizado $\tilde{J}_T(\varphi) = T^{d/2}(J_T(\varphi) - E J_T(\varphi))$ bastará con proporcionar condiciones para la convergencia al límite finito del cumulante de segundo orden de $\tilde{J}_T(\varphi)$, y proporcionar condiciones para la convergencia a cero del resto de cumulantes de órdenes superiores.

A lo largo de esta sección se omitirá el exponente 'h' y se tendrá en cuenta simplemente $I_T(\boldsymbol{\lambda})$ para denotar el periodograma tapered y $J_T(\varphi)$ para denotar el correspondiente funcional espectral. Solamente la Proposición 2.1 se refiere al caso de datos sin el tapering.

El cumulante general de orden k -ésimo se puede representar de la forma:

$$\begin{aligned}
 c_k(\tilde{J}_T(\varphi)) &= T^{kd/2} 2^{k-1} (k-1)! [H_{2,T}(\mathbf{0})]^{-k} \int_{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_{2k} \in \mathbb{T}^{2k}} f(\boldsymbol{\lambda}_1) \varphi(\boldsymbol{\lambda}_2) \\
 &\quad \times f(\boldsymbol{\lambda}_3) \varphi(\boldsymbol{\lambda}_4) \dots f(\boldsymbol{\lambda}_{2k-1}) \varphi(\boldsymbol{\lambda}_{2k}) H_{1,T}(\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1) H_{1,T}(\boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_2) \dots \\
 &\quad \times H_{1,T}(\boldsymbol{\lambda}_{2k} - \boldsymbol{\lambda}_{2k-1}) H_{1,T}(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_{2k}) d\boldsymbol{\lambda}_1 \dots d\boldsymbol{\lambda}_{2k} \\
 &= T^{kd/2} 2^{k-1} (k-1)! \int_{\mathbf{u}_1, \dots, \mathbf{u}_{2k-1} \in \mathbb{T}^{2k-1}} \int_{\boldsymbol{\lambda} \in \mathbb{T}} f(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda} + \mathbf{u}_1) \\
 &\quad \times f(\boldsymbol{\lambda} + \mathbf{u}_1 + \mathbf{u}_2) \varphi(\boldsymbol{\lambda} + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3) \dots \varphi\left(\boldsymbol{\lambda} + \sum_{i=1}^{2k-1} \mathbf{u}_i\right) \\
 &\quad \times \prod_{i=1}^{2k-1} H_{1,T}(\mathbf{u}_i) H_{1,T}\left(-\sum_{i=1}^{2k-1} \mathbf{u}_i\right) d\boldsymbol{\lambda} d\mathbf{u}_1 \dots d\mathbf{u}_{2k-1},
 \end{aligned}$$

donde $H_{k,T}(\boldsymbol{\lambda}) = \sum_{\mathbf{t} \in L_T} h_T^k(\mathbf{t}) e^{-i(\boldsymbol{\lambda}, \mathbf{t})}$, y $h_T(\mathbf{t}) = h(\mathbf{t}/T)$.

Más detalles de los cálculos de los cumulantes de funcionales espectrales se pueden encontrar, por ejemplo, en Anh et al. (2004a); Avram et al. (2010b); Bentkus (1976), para el caso en que no se utilizan datos tapered; y en Anh et al. (2004b); Dahlhaus (1983); Dahlhaus y Künsch (1987), para el caso en el que se utilicen datos tapered. Los cálculos se basan en la denominada fórmula del producto para cumulantes que da la expresión para cumulantes de productos de variables aleatorias en términos de cumulantes de las variables individuales. En el caso gaussiano, la fórmula mencionada se reduce a una fórmula mucho más simplificada.

Obsérvese que las funciones

$$\Phi_{k,T}^h(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_{k-1}) := \frac{1}{(2\pi)^{d(k-1)} H_{k,T}(\mathbf{0})} \prod_{j=1}^{k-1} H_{1,T}(\boldsymbol{\lambda}_j) H_{1,T}\left(-\sum_{j=1}^{k-1} \boldsymbol{\lambda}_j\right) \quad (2.29)$$

son kernels multidimensionales de tipo Fejér sobre \mathbb{T}^{k-1} , o identidades aproximadas para la convolución.

Para el caso de datos sin el tapering, en Bentkus (1972, 1976) se se obtienen los resultados correspondientes para $h(t) \equiv 1$. Bajo este supuesto, cuando la función taper se factoriza (como se define en el Apartado 1.1.3), y el dominio

de observación es un cubo $L_T = [-T, T]^d$, los resultados se obtienen de forma directa a partir del correspondiente resultado dado en Dahlhaus (1983) para dimensión $d = 1$.

La propiedad del *kernel*, $\Phi_{k,T}^h(\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_{k-1})$, implica que el

$$\begin{aligned} & \lim_{T \rightarrow \infty} \int_{\mathbb{T}^{k-1}} G(\mathbf{u}_1 - \mathbf{v}_1, \dots, \mathbf{u}_{k-1} - \mathbf{v}_{k-1}) \Phi_{k,T}^h(\mathbf{u}_1, \dots, \mathbf{u}_{k-1}) d\mathbf{u}_1 \dots d\mathbf{u}_{k-1} \\ &= G(\mathbf{v}_1, \dots, \mathbf{v}_{k-1}), \end{aligned} \quad (2.30)$$

siempre que la función G sea acotada y continua en el punto $(\mathbf{v}_1, \dots, \mathbf{v}_{k-1})$.

En particular, se tiene,

$$\begin{aligned} \text{Var}(\tilde{J}_T(\varphi)) &= 2T^d (2\pi)^{3d} H_{4,T}(\mathbf{0}) [H_{2,T}(\mathbf{0})]^{-2} \int_{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \in \mathbb{T}^3} \int_{\boldsymbol{\lambda} \in \mathbb{T}} f(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda} + \mathbf{u}_1) \\ &\quad \times f(\boldsymbol{\lambda} + \mathbf{u}_1 + \mathbf{u}_2) \varphi(\boldsymbol{\lambda} + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3) \Phi_{4,T}^h(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) d\boldsymbol{\lambda} d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3 \quad (2.31) \\ &= 2T^d (2\pi)^{3d} H_{4,T}(\mathbf{0}) [H_{2,T}(\mathbf{0})]^{-2} \int_{(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4) \in \mathbb{T}^4} f(\boldsymbol{\lambda}_1) \varphi(\boldsymbol{\lambda}_2) f(\boldsymbol{\lambda}_3) \varphi(\boldsymbol{\lambda}_4) \\ &\quad \times \Phi_{4,T}^h(\boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_4 - \boldsymbol{\lambda}_3) d\boldsymbol{\lambda}_1 d\boldsymbol{\lambda}_2 d\boldsymbol{\lambda}_3 d\boldsymbol{\lambda}_4. \end{aligned}$$

El análisis asintótico de las expresiones de los cumulantes, basado en la Propiedad (2.30), permite afirmar el siguiente resultado de normalidad asintótica para el funcional $J_T(\varphi) = J_T^h(\varphi) = \int_{\mathbb{T}} I_T^h(\boldsymbol{\lambda}) \varphi(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$ en el caso de datos tapered.

Teorema 2.3. *Sea $X(\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^d$, un campo aleatorio gaussiano de media cero con densidad espectral $f(\boldsymbol{\lambda}) \in L_p$ y $\varphi(\boldsymbol{\lambda}) \in L_q$, donde $\frac{1}{p} + \frac{1}{q} \leq \frac{1}{2}$. Entonces,*

$$T^{d/2}(J_T(\varphi) - EJ_T(\varphi)) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \text{ cuando } T \rightarrow \infty, \quad (2.32)$$

donde

$$\sigma^2 = 2(2\pi)^{3d} e(h) \int_{\mathbb{T}} f^2(\boldsymbol{\lambda}) \varphi^2(\boldsymbol{\lambda}) d\boldsymbol{\lambda}, \quad (2.33)$$

y $e(h)$ está definido en (2.9).

La convergencia de la varianza (2.31) a σ^2 se puede obtener siguiendo las siguientes consideraciones:

(i) Bajo las condiciones de integrabilidad de f y φ , impuestas en el teorema, la integral dada respecto de λ en la expresión (2.31) es una función continua y acotada en $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, que podemos llamar $G(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, por lo tanto, la integral general en la ecuación (2.31) se puede escribir como

$$\int G(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \Phi_{4,T}^h(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) d\mathbf{u}_1 d\mathbf{u}_2 d\mathbf{u}_3.$$

Entonces, debido a la propiedad del Kernel (2.30), el límite de esta integral es igual a $G(\mathbf{0}, \mathbf{0}, \mathbf{0}) = \int_{\mathbb{T}} f^2(\lambda) \varphi^2(\lambda) d\lambda$;

(ii) El factor de normalización, $T^d H_{4,T}(\mathbf{0}) [H_{2,T}(\mathbf{0})]^{-2}$, converge a $e(h)$ debido a su comportamiento asintótico, es decir, $H_{k,T}(\mathbf{0}) \sim T^d \left(\int \tilde{h}^k(t) dt \right)^d$ cuando $T \rightarrow \infty$.

Siguiendo los mismos argumentos que en Avram et al. (2010a,b), las condiciones de integrabilidad en f y φ implican también convergencia a cero de todos los cumulantes de órdenes $k \geq 3$. Obsérvese que, bajo las suposiciones consideradas en la función tapered, las siguientes estimaciones para las normas de $H_{k,T}(\lambda)$ se cumplen: $\|H_{k,T}(\lambda)\|_p \leq CT^{d(1-\frac{1}{p})}$, $p > 1$, lo cual es una consecuencia del Lema 1 de Dahlhaus (1983), que se pueden aplicar en las estimaciones análogas para el caso de datos sin el tapering. Por lo tanto, todas las demostraciones se pueden deducir de forma análoga.

Del Teorema 2.3 y del Lema 2.1 (véase la Sección 2.2), se deduce el siguiente corolario.

Corolario 2.1. *Bajo las condiciones del Teorema 2.3 y del Lema 2.1 se tiene,*

$$T^{d/2} (J_T(\varphi) - J(\varphi)) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \text{ cuando } T \rightarrow \infty, \quad (2.34)$$

donde σ^2 viene dada por la ecuación (2.33).

Obsérvese que el reciente artículo de Ginovyan et al. (2014), Sección 4.2.1, presenta una colección de condiciones suficientes clásicas para que se cumpla el TCL para formas cuadráticas, $Q_T = T^d J_T(\varphi)$, en el caso de procesos gaussianos en tiempo discreto y continuo.

El Teorema 2.3 es una generalización de las afirmaciones dadas en el Teorema 4.7-(C) desarrollado en Ginovyan et al. (2014), véase también Avram (1988), para el caso de los campos (en Z^d) y para los datos tapered.

Para el caso de procesos gaussianos con tiempo discreto y densidades espectrales con posibles singularidades, se obtuvieron los siguientes resultados en Fox y Taqqu (1987), sin utilización de tapering, es decir, $h(t) \equiv 1$.

Proposición 2.1. Sea $X(t)$, $t \in \mathbb{Z}$, un proceso gaussiano con media cero y densidad espectral $f(\lambda)$.

I. Supongamos que las siguientes condiciones se cumplen:

- (i) Los conjuntos de discontinuidades de las funciones $f(\lambda)$ y $\varphi(\lambda)$ tienen medida de Lebesgue cero, y estas funciones están acotadas en el intervalo $[\delta, \pi]$ para todo $\delta > 0$;
- (ii) Existe $\alpha < 1$ y $\beta < 1$ tales que para $\alpha + \beta < 1/2$, y para cada $\delta > 0$, $f(\lambda) = O(|\lambda|^{-\alpha-\delta})$ y $\varphi(\lambda) = O(|\lambda|^{-\beta-\delta})$ cuando $\lambda \rightarrow 0$.

Entonces,

$$T^{1/2}(J_T(\varphi) - EJ_T(\varphi)) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \text{ cuando } T \rightarrow \infty, \quad (2.35)$$

$$\text{donde } \sigma^2 = 16\pi^3 \int_{-\pi}^{\pi} \varphi^2(\lambda) f^2(\lambda) d\lambda.$$

- II. Si existe $\alpha < 1$ y $\beta < 1$ tales que $\alpha + \beta < 1/2$, y $f(\lambda) = |\lambda|^{-\alpha} L_1(\lambda)$ y $\varphi(\lambda) = |\lambda|^{-\beta} L_2(\lambda)$ cuando $\lambda \rightarrow 0$, donde L_1 y L_2 son funciones que varían lentamente en cero, entonces, la Condición (ii) se cumple

Presentamos aquí una extensión de la Proposición 2.1 para el caso de campos gaussianos y periodogramas tapered en la siguiente forma.

Teorema 2.4. Sea $X(\mathbf{t})$, $\mathbf{t} \in \mathbb{Z}^d$, un campo aleatorio gaussiano de media cero con densidad espectral $f(\boldsymbol{\lambda})$ tal que para algún $0 < \alpha_i < 1$, $i = 1, \dots, d$, $f(\boldsymbol{\lambda}) = O(\prod_{i=1}^d |\lambda_i|^{-\alpha_i})$ cuando $\lambda_i \rightarrow 0$, y $\varphi(\boldsymbol{\lambda}) = O(\prod_{i=1}^d |\lambda_i|^{\alpha_i})$ cuando $\lambda_i \rightarrow 0$. Los conjuntos de puntos de discontinuidad de funciones $f(\boldsymbol{\lambda})$ y $\varphi(\boldsymbol{\lambda})$ tienen medida Lebesgue cero, y estas funciones están acotadas, para $\delta \leq |\boldsymbol{\lambda}| \leq \pi$ y para todo $\delta > 0$. Entonces,

$$T^{d/2}(J_T(\varphi) - EJ_T(\varphi)) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \text{ cuando } T \rightarrow \infty, \quad (2.36)$$

donde σ^2 se ha introducido en el Teorema 2.3.

A partir del Teorema 2.4 y el Lema 2.1 (véase la Sección 2.2) se obtiene el siguiente corolario.

Corolario 2.2. Supongamos que se cumplen las condiciones del Teorema 2.4 y del Lema 2.1. Entonces,

$$T^{d/2}(J_T(\varphi) - J(\varphi)) \xrightarrow{D} \mathcal{N}(0, \sigma^2) \text{ cuando } T \rightarrow \infty, \quad (2.37)$$

donde σ^2 viene dada por la ecuación (2.33).

Nota 2.12. La Proposición 2.1 se establece para las funciones $f(\lambda)$ y $\varphi(\lambda)$, que tienen singularidades en el punto $\lambda = 0$. Sin embargo, se puede demostrar, ver Ginovyan et al. (2014), que la elección $\lambda = 0$ no es esencial y la Proposición 2.1 se cumple si la singularidad se localiza en cualquier otro punto $\lambda_0 \in [-\pi, \pi]$. La observación análoga es verdadera para el resultado indicado, en el caso de campos aleatorios (con forma multiplicativa de singularidades), en el Teorema 2.4. De hecho, el Teorema 2.4 sigue siendo válido si, en lugar del origen ($\lambda_i = 0$, $i = 1, \dots, d$), la singularidad de la forma prescrita en el teorema tiene lugar en cualquier otro punto.

Demostración del Teorema 2.4: Para la demostración, utilizamos una idea del artículo Heyde y Gay (1993). Primeramente, consideramos el caso $d = 1$. Se considera el proceso filtrado

$$Y(t) = \nabla^{\alpha/2} X(t),$$

donde $\nabla = 1 - B$, B es el operador de retardos ($BX(t) = X(t - 1)$), y $\nabla^{\alpha/2} = (1 - B)^{\alpha/2} := \sum_{j=0}^{\infty} C_j^{\alpha/2} (-B)^j$. Entonces, el proceso $Y(t)$ tiene por densidad espectral $f_Y(\lambda) = (2 \sin |\frac{\lambda}{2}|)^{\alpha} f_X(\lambda)$, ya que $Y(t)$ se obtiene de $X(t)$ usando el filtro con función de transferencia $D(i\lambda) = (1 - e^{i\lambda})^{\alpha/2}$, y $|D(i\lambda)|^2 = (2 \sin |\frac{\lambda}{2}|)^{\alpha}$.

Sea $\psi(\lambda) = \varphi(\lambda)/(2 \sin |\frac{\lambda}{2}|)^d$, y consideremos el funcional

$$\tilde{J}_T^Y(\psi) = \int_{-\pi}^{\pi} \psi(\lambda) I_T^Y(\lambda) d\lambda - E \int_{-\pi}^{\pi} \psi(\lambda) I_T^Y(\lambda) d\lambda,$$

donde $I_T^Y(\lambda) = \frac{1}{2\pi T} |\sum_{t \in L_T} h_T(t) e^{i\lambda t} Y(t)|^2$ es el periodograma tapered que corresponde a $\{Y(t), t \in L_T\}$.

Dado que la densidad espectral $f_Y(\lambda)$ del proceso $Y(t)$ y la función $\psi(\lambda)$ satisfacen las condiciones del Teorema 2.3, para el funcional $\tilde{J}_T^Y(\psi)$ se tiene la siguiente convergencia cuando $T \rightarrow \infty$,

$$T^{1/2} \tilde{J}_T^Y(\psi) \xrightarrow{D} \mathcal{N}(0, \sigma^2), \quad (2.38)$$

donde

$$\sigma^2 = 16\pi^3 e(h) \int_{-\pi}^{\pi} \psi^2(\lambda) f_Y^2(\lambda) d\lambda = 16\pi^3 e(h) \int_{-\pi}^{\pi} \varphi^2(\lambda) f_X^2(\lambda) d\lambda. \quad (2.39)$$

Por lo tanto, para probar la afirmación del teorema, basta con demostrar que

$$\lim_{T \rightarrow \infty} TE |\tilde{J}_T^X(\varphi) - \tilde{J}_T^Y(\psi)|^2 = 0, \quad (2.40)$$

donde

$$\tilde{J}_T^X(\varphi) = \int_{-\pi}^{\pi} \varphi(\lambda) I_T^X(\lambda) d\lambda - E \int_{-\pi}^{\pi} \varphi(\lambda) I_T^X(\lambda) d\lambda.$$

Consideremos

$$TE |\tilde{J}_T^X(\varphi) - \tilde{J}_T^Y(\psi)|^2 = TE |\tilde{J}_T^X(\varphi)|^2 + TE |\tilde{J}_T^Y(\psi)|^2 - 2TE \tilde{J}_T^X(\varphi) \tilde{J}_T^Y(\psi). \quad (2.41)$$

Para el funcional que corresponde al proceso $Y(t)$ se tiene fácilmente la convergencia $TE |\tilde{J}_T^Y(\psi)|^2 \rightarrow \sigma^2$ cuando $T \rightarrow \infty$.

Vamos a demostrar que bajo las condiciones del teorema $TE |\tilde{J}_T^X(\varphi)|^2$ y $TE \tilde{J}_T^X(\varphi) \tilde{J}_T^Y(\psi)$ también tienden a σ^2 cuando $T \rightarrow \infty$, y, por lo tanto, la convergencia (2.40) se cumple.

Usando (2.31) se puede considerar:

$$\begin{aligned} TE |\tilde{J}_T^X(\varphi)|^2 &= 2T^d (2\pi)^{3d} H_{4,T}(0) [H_{2,T}(0)]^{-2} \int_{[-\pi, \pi]^4} f_X(\lambda_1) \varphi(\lambda_2) f_X(\lambda_3) \varphi(\lambda_4) \\ &\quad \times \Phi_{4,T}^h(\lambda_2 - \lambda_1, \lambda_3 - \lambda_2, \lambda_4 - \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4. \end{aligned} \quad (2.42)$$

$$\begin{aligned} TE |\tilde{J}_T^Y(\psi)|^2 &= 2T^d (2\pi)^{3d} H_{4,T}(0) [H_{2,T}(0)]^{-2} \int_{[-\pi, \pi]^4} f_Y(\lambda_1) \psi(\lambda_2) f_Y(\lambda_3) \psi(\lambda_4) \\ &\quad \times \Phi_{4,T}^h(\lambda_2 - \lambda_1, \lambda_3 - \lambda_2, \lambda_4 - \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4. \end{aligned} \quad (2.43)$$

$$\begin{aligned} TE |\tilde{J}_T^X(\varphi) \tilde{J}_T^Y(\psi)| &= 2T^d (2\pi)^{3d} H_{4,T}(0) [H_{2,T}(0)]^{-2} \\ &\quad \times \int_{[-\pi, \pi]^4} f_{XY}(\lambda_1) \varphi(\lambda_2) f_{XY}(\lambda_3) \psi(\lambda_4) \\ &\quad \times \Phi_{4,T}^h(\lambda_2 - \lambda_1, \lambda_3 - \lambda_2, \lambda_4 - \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4, \end{aligned} \quad (2.44)$$

donde por f_{XY} hemos denotado la densidad espectral cruzada de los procesos

$X(t)$ e $Y(t)$.

Dado que $Y(t)$ se obtiene a partir de $X(t)$ con el uso del filtrado de la función de transferencia $D(i\lambda)$, entonces, $f_{XY}(\lambda) = D(-i\lambda)f_X(\lambda)$, o $f_{XY}(\lambda) = D(-i\lambda)\frac{f_Y(\lambda)}{|D(i\lambda)|^2}$, recordemos también que $\varphi(\lambda) = \psi(\lambda)|D(i\lambda)|^2$, $f_X(\lambda) = \frac{f_Y(\lambda)}{|D(i\lambda)|^2}$.

Si definimos la medida μ_T en $[-\pi, \pi]^4$ como

$$\mu_T(E) = \int_E \Phi_{4,T}^h(\lambda_2 - \lambda_1, \lambda_3 - \lambda_2, \lambda_4 - \lambda_3) d\lambda_1 d\lambda_2 d\lambda_3 d\lambda_4, \text{ para } E \subset [-\pi, \pi]^4,$$

entonces, μ_T converge débilmente a la medida μ que se concentra en la diagonal $D = \{y_1 = y_2 = y_3 = y_4\}$, y satisface: $\mu(\{y : a \leq y_1 = y_2 = y_3 = y_4 \leq b\}) = b - a$. Para el caso de datos sin *tapering*, esto fue considerado en Fox y Taqqu (1987). El resultado se mantiene para el caso de datos con el *tapering* y también en nuestro caso, bajo las condiciones consideradas. Esto se puede demostrar, siguiendo los mismos argumentos que en Fox y Taqqu (1987), considerando los coeficientes de Fourier de las medidas μ y μ_T . La convergencia de los coeficientes de Fourier de μ_T a los de μ se obtiene aplicando las propiedades del núcleo (2.30).

A continuación, observamos que la siguiente estimación se cumple para $H_{1,T}$: $|H_{1,T}(\lambda)| \leq \text{const} \cdot l_T(\lambda)$, donde $l_T(u)$ denota la extensión 2π -periódica de la función $l_T^*(u)$, que se define como: $l_T^*(u) = T$ para $|u| \leq \frac{1}{T}$, y $l_T^*(u) = \frac{1}{|u|}$ para $\frac{1}{T} < |u| \leq \pi$. A partir de este punto se puede obtener la convergencia de (2.42), (2.43), y (2.44), a σ^2 , definida por (2.39), siguiendo las consideraciones correspondientes de la demostración en Fox y Taqqu (1987). De forma más precisa, no se necesitan sus argumentos de forma completa, sino simplemente usar los argumentos para el cumulante de segundo orden, y las partes a) de sus proposiciones 6.1 y 6.2 que funcionan para nuestro caso.

La demostración para el caso $d = 1$ puede extenderse directamente para $d > 1$ en el caso considerado, cuando las singularidades de la densidad espectral se factorizan como se describe en la formulación del teorema, y para la función *taper* factorizada, de modo que las integrales pueden ser partidas como d -tuplas de integrales, las cuales aparecen cuando $d = 1$, y, por lo tanto, los razonamientos correspondientes se pueden conservar. Téngase en cuenta que

el campo filtrado se introduce como

$$\begin{aligned} Y(\mathbf{t}) = Y(t_1, \dots, t_d) &= \nabla_1^{\alpha_1/2} \dots \nabla_d^{\alpha_d/2} X(\mathbf{t}) \\ &= \sum_{k_1=0}^{\infty} \dots \sum_{k_d=0}^{\infty} \prod_{i=1}^d C_{k_i}^{\alpha_i/2} X(t_1 - k_1, \dots, t_d - k_d), \end{aligned}$$

y se tiene la densidad espectral

$$f_Y(\lambda_1, \dots, \lambda_d) = \left(\prod_{i=1}^d \left| 2 \operatorname{sen} \frac{\lambda_i}{2} \right|^{\alpha_i} \right) f_X(\lambda_1, \dots, \lambda_d).$$

□

Capítulo 3

Procesos de Cox log-gaussianos temporales dirigidos mediante un proceso de Ornstein-Uhlenbeck Hilbert-valuado

Los resultados de este capítulo se recogen en el artículo titulado “Log-Gaussian Cox processes in infinite-dimensional spaces”, publicado en “*Theory of Probability and Mathematical Statistics*”, Vol. 95, pp 157–177, (Torres et al., 2016).

Este capítulo introduce una nueva familia de Procesos de Cox en el tiempo, dirigidos por un proceso de Ornstein-Uhlenbeck (O-U) evaluado en un espacio de Hilbert separable. La estimación paramétrica y predicción funcional se realiza a partir de una aproximación en tiempo discreto de dicha intensidad mediante un proceso autorregresivo hilbertiano de orden 1, ARH(1). La predicción de los valores funcionales de la intensidad aleatoria se obtiene mediante la implementación del filtrado de Kalman. Finalmente, se calcula el predictor *plug-in* asociado para el proceso de Cox log-gaussiano.

Se supone que la transformada logarítmica de la intensidad se controla mediante un proceso de difusión Hilbert-valuado, esto es, el proceso de O-U; véase, por ejemplo, Duffie y Kan (1996) y Fuhrman (1995), donde los procesos de Feller en una o d dimensiones son incorporados como modelos de intensidad. En este sentido, la familia de procesos introducidos en este capítulo extiende los modelos estudiados en Basu y Dassios (2002); Kozachenko y Pogorilyak (2008);

Wei et al. (2002).

El esquema de este capítulo es el que sigue. En la Sección 3.1, se aplican métodos de proyección numérica, bajo condiciones adecuadas, para una representación espectral diagonal de la ecuación de integral que define el proceso O-U en espacios de Hilbert. La estimación paramétrica y la predicción del proceso O-U Hilbert-valuado, a partir de su aproximación en tiempo discreto en términos de un proceso ARH(1), se desarrolla en la Sección 3.3. La definición y predicción del proceso temporal de Cox log-gaussiano en espacios de Hilbert reales separables se proporciona en la Sección 3.4. Los resultados obtenidos se aplican a la estimación y predicción espacio-temporal, en la Sección 3.5. Para ilustrar los resultados obtenidos, se lleva a cabo un estudio de simulación condicionada, en la Sección 3.6.

3.1. Proyección del proceso O-U Hilbert-valuado

Sea $X = \{X_t, t \geq 0\}$ un proceso Hilbert-valuado que satisface la ecuación (1.8), así como las condiciones (i)–(iii) formuladas en el Apartado 1.2.4. Aquí, también asumimos la siguiente condición adicional:

Suposición 3.1. *Los operadores A , en la ecuación (1.8), y Q , el operador de covarianza que caracteriza la distribución de Wiener fuertemente cilíndrica W , en dicha ecuación, son operadores positivos, admitiendo una descomposición espectral diagonal, con respecto a una base ortonormal común de autovectores $\{\phi_j\}_{j \geq 1}$ en H . Es decir,*

$$A = \sum_{j=1}^{\infty} \lambda_j(A) \phi_j \otimes \phi_j \quad (3.1)$$

$$Q = \sum_{j=1}^{\infty} \lambda_j(Q) \phi_j \otimes \phi_j. \quad (3.2)$$

Nota 3.1. *La Suposición 3.1 se cumple, en particular, cuando los operadores A y Q están relacionados por una función continua; véase Dautray y Lions (1985), pp. 119–140. Las bases incondicionadas, como las bases wavelet, también permiten la representación de las series espectrales diagonales, por ejemplo, de los kernels distributivos de operadores de Calderón-Zygmund; ver, por ejemplo, Heil y Walnut (2006), pp. 664–665.*

En particular,

$$\exp(tA)(f) = \sum_{j=1}^{\infty} \exp(-\lambda_j(A)t) \langle f, \phi_j \rangle_H \phi_j, \quad \forall f \in H. \quad (3.3)$$

Obsérvese que Q en la ecuación (1.8) representa el operador de covarianzas del proceso de Wiener fuertemente cilíndrico W , ya que, como se comentó en el Apartado 1.2.4, se asumen las condiciones que aseguran que el proceso de Wiener fuertemente cilíndrico W en la ecuación (1.8) sea un proceso de Wiener Hilbert-valuado; ver Teorema 8.1 en Riedle (2011).

Bajo la Condición 3.1, a partir de los teoremas 7.1 y 8.1 que aparecen en Riedle (2011), de las ecuaciones(3.1)–(3.3), y teniendo en cuenta que

$$\{\psi_j = \lambda_j^{1/2}(Q)\phi_j\}_{j \geq 1} \quad (3.4)$$

define una base ortonormal del espacio de Hilbert con núcleo reproductor de Q , W admite la siguiente representación:

$$W(t)(g^*) = \sum_{j=1}^{\infty} \langle i_Q(\psi_j), g \rangle_H B_j(t), \quad \forall g^* \in U^* = [\overline{Q^{1/2}(H)}]^{\|\cdot\|_Q}, \quad (3.5)$$

donde U^* denota el espacio dual de $U = \overline{Q^{1/2}(H)}^{\|\cdot\|_Q}$, el cierre de $Q^{1/2}(H)$, con respecto a la norma $\|\cdot\|_Q$, inducida por Q , es decir, $\|f\|_Q = Q(f)(f) = \langle Q(f^*), f \rangle_H$, para todo $f \in H$, con, como anteriormente, $f^* \in H^*$ denotando el elemento dual de f , dado por el teorema de representación de Riesz. Aquí, $\{B_j(t), t \geq 0\}_{j \geq 1}$ son procesos de Wiener estándar real-valuados independientes y, como antes, i_Q es la aplicación de inclusión continua que introduce el rango de Q en H .

Teorema 3.1. *Bajo las condiciones (i)–(iii) asumidas en el Apartado 1.2.4, y la Condición 3.1, para cada $j \geq 1$, $\{Y_p(t) = \langle X(t, \phi_j), \phi_p \rangle_H, t \in \mathbb{R}_+\}_{p \geq 1}$, introducido en la ecuación (1.8) para $f = \phi_p$, define una secuencia de procesos O-U real-valuados independientes, con respectivos parámetros $\{\lambda_p(A)\}_{p \geq 1}$ y $\{\lambda_p^{1/2}(Q)\}_{p \geq 1}$, tales que $Y_p(0) \equiv 1$, para cada $p \geq 1$.*

Demostración:

Bajo la Condición 3.1, a partir de las ecuaciones (3.3) y (3.5), la proyección de la ecuación (1.8) en el sistema de autovectores ortonormales $\{\phi_j\}_{j \geq 1}$ nos lleva

a las siguientes identidades: Para cada $t > 0$, $p \geq 1$, y para cualquier $j \geq 1$,

$$\begin{aligned}
 & \langle X(t, \phi_j), \phi_p \rangle_H \\
 &= \langle \exp(tA)(\phi_j), \phi_p \rangle_H + \left\langle \int_0^t \exp((t-s)A) dW(s), \phi_p \right\rangle_H \\
 &= \sum_{k=1}^{\infty} \exp(-\lambda_k(A)t) \langle \phi_k, \phi_j \rangle_H \langle \phi_k, \phi_p \rangle_H \\
 &\quad + \sum_{k=1}^{\infty} \langle \phi_k, \phi_p \rangle_H \sum_{l=1}^{\infty} \int_0^t \exp(-\lambda_k(A)(t-s)) \langle \phi_k, i_Q(\psi_l) \rangle_H B_l(ds) \\
 &= \exp(-\lambda_p(A)t) + \lambda_p^{1/2}(Q) \int_0^t \exp(-\lambda_p(A)(t-s)) B_p(ds),
 \end{aligned} \tag{3.6}$$

donde la integral se interpreta como una integral estocástica uniparamétrica de Wiener-Itô con respecto a la medida de Wiener $B_p(ds)$, para cada $p \geq 1$. La última identidad de la ecuación (3.6) significa que, para cada $p \geq 1$, y para todo $j \geq 1$, $\{Y_p(t) = \langle X(t, \phi_j), \phi_p \rangle_H, t \in \mathbb{R}_+\}$ es un proceso O-U real-valuado que satisface la ecuación de Langevin:

$$dY_p(t) = -\lambda_p(A)Y_p(t)dt + \lambda_p^{1/2}(Q)B_p(dt). \tag{3.7}$$

Esto es, para cada $p \geq 1$, $\{Y_p(t), t \in \mathbb{R}_+\}$ es un proceso O-U comenzando en $t = 0$, con condición inicial $Y_p(0) \equiv 1$, y parámetros $\lambda_p(A)$ y $\lambda_p^{1/2}(Q)$. De hecho, $\{Y_p\}_{p \geq 1}$ es una secuencia de procesos O-U real-valuados independientes. \square

3.2. Aproximación ARH(1) en tiempo discreto del proceso O-U Hilbert-valuado

En la aproximación en tiempo discreto de la ecuación (1.8) se puede encontrar varias dificultades relacionadas con la definición de un adecuado concepto de derivada (por ejemplo, derivada total como derivada Fréchet, derivada direccional como derivada Gâteaux, H -derivada, etc.). Esta es la razón por la que se considera el operador $\gamma = I - \rho$, y el proceso de innovación ε_n definido mediante $\varepsilon_n = \sigma(W(n) - W(n-1))$, para todo $n \in \mathbb{N}_*$, siendo W un proceso de Wiener H -valuado y fuertemente cilíndrico, que lleva a reescribir la ecuación

(1.10) como sigue:

$$X_{t+1} - X_t = -\gamma X_t + \sigma(W_{t+1} - W_t), \quad t = 1, 2, \dots \quad (3.8)$$

El análogo a esta ecuación en tiempo continuo es la ecuación diferencial estocástica

$$dX_t = \gamma X_t dt + \sigma dW_t, \quad t \geq 0.$$

Formalmente, la solución única a esta ecuación se da en (1.8). Bajo las suposiciones realizadas sobre los operadores que definen la ecuación (1.8), la proyección de dicha ecuación en la base ortonormal de autovectores, permite aplicar la aproximación de un proceso O-U real-valuado, en términos de un proceso AR(1), que a su vez, nos proporciona la definición de un proceso ARH(1) proyectado. Se obtiene, para cada $t = 1, 2, \dots$, y $p \geq 1$,

$$Y_p(t+1) - Y_p(t) = -\lambda_p(A)Y_p(t) + \lambda_p^{1/2}(Q)(B_p(t+1) - B_p(t)), \quad (3.9)$$

donde, para cada $\tau > 0$, se puede ahora considerar

$$Y_{p,\tau}(n+1) = \gamma_\tau Y_{p,\tau}(n) + \sqrt{\tau} \varepsilon_n, \quad (3.10)$$

con $\gamma_\tau = 1 - \lambda_p(A)\tau$. Entonces, para cada $t > 0$, y $p \geq 1$, se tiene que

$$\begin{aligned} Y_{p,\tau}([t/\tau]) &= \gamma_\tau^{[t/\tau]} Y_p(0) + \lambda_p^{1/2}(Q) \sqrt{\tau} \gamma_\tau^{[t/\tau]-1} \int_0^{[t/\tau]} \gamma_\tau^{-[s]} dB_p(s) \\ &\sim \exp(-\lambda_p(A)t) + \lambda_p^{1/2}(Q) \exp(-\lambda_p(A)t) \int_0^t \exp(\lambda_p(A)s) dB_p(s), \end{aligned} \quad (3.11)$$

donde, en el último paso se utiliza la propiedad de escala del movimiento browniano $\sqrt{\tau} B_p(s) \sim B_p(\tau s)$. Obsérvese que, $\gamma_\tau \sim (1 - \lambda_p(A))^\tau = \gamma^\tau$, cuando $\tau \rightarrow 0$. Este argumento puede ser expresado de una manera rigurosa para llegar a la convergencia en sentido débil del proceso $Y_{p,\tau}([t/\tau])$ al proceso O-U real-valuado y continuo, cuando $\tau \rightarrow 0$; ver Fouque et al. (2000); Nelson (1990), para más detalles. La conclusión es que si el intervalo entre observaciones sucesivas se lleva al cero, y los coeficientes se ajustan adecuadamente, entonces, cada elemento (3.9)–(3.10) de la secuencia obtenida de procesos AR(1) con-

verge al proceso de O-U real-valuado y continuo, véase las ecuaciones (3.7) y (3.11).

Por otro lado, bajo la siguiente hipótesis, la secuencia de proceso AR(1) introducida en (3.10) también define un proceso ARH(1). Esta es la razón por la que en esta sección nos referimos a la aproximación ARH(1) en tiempo discreto del proceso O-U H -valuado.

Suposición 3.2. *El operador de autocorrelación ρ , en la ecuación (1.10), admite la siguiente representación espectral diagonal, en sentido débil:*

Para todo $f \in H$,

$$\rho(f) = \sum_{k=1}^{\infty} \lambda_k(\rho) \langle \phi_k, f \rangle_H \phi_k, \quad (3.12)$$

donde la serie (3.12) es finita para cualquier $f \in H$, y, como antes, $\{\phi_j\}_{j \geq 1}$ denota el sistema de autovectores del operador de autocovarianza C_X dado en (1.11).

A partir de las ecuaciones (3.9)–(3.11), la Suposición 3.1 implica la Suposición 3.2, cuando el proceso considerado ARH(1) se construye como la aproximación en tiempo discreto de un proceso O-U H -valuado, es decir, cuando ambos procesos están caracterizados por una secuencia de procesos ARH(1) independientes tal como se da en las ecuaciones (3.9)–(3.10), y la ecuación (3.13) que aparece en el siguiente resultado:

Lema 3.1. *Bajo la Condición 3.2, la proyección de la ecuación ARH(1), (1.10), en $\{\phi_j\}_{j \geq 1}$ nos lleva a la siguiente expresión: para cada $p \geq 1$*

$$\begin{aligned} X_p(n) &= \langle X_n, \phi_p \rangle_H = \lambda_p(\rho) \langle X_{n-1}, \phi_p \rangle_H + \langle \varepsilon_n, \phi_p \rangle_H \\ &= \lambda_p(\rho) X_p(n-1) + \varepsilon_p(n), \quad n \in \mathbb{Z}, \end{aligned} \quad (3.13)$$

donde $\{X_p(n), n \in \mathbb{Z}\}$ es un proceso AR(1).

Demostración:

Es una consecuencia inmediata de la representación espectral diagonal en sentido débil de ρ en la ecuación (3.12), así como de la representación espectral

$$C_X(h) = \sum_{k=1}^{\infty} \lambda_k(C_X) \langle \phi_k, h \rangle_H \phi_k, \quad \forall h \in H,$$

la cual se cumple a partir de la propiedad del operador traza C_X . □

3.3. Estimación paramétrica y predicción del proceso O-U Hilbert-valuado

El objetivo de esta sección es la formulación de los respectivos estimadores paramétricos componente a componente de los operadores A y Q , que caracterizan la distribución del proceso de O-U H -valuado en la ecuación (1.8).

En el supuesto de que se conozcan los autovectores del operador de autocovarianza del proceso ARH(1), C_X , en la ecuación (1.10), se consideran los siguientes estimadores de componentes para ρ y $R_\varepsilon = E[\varepsilon_1 \otimes \varepsilon_1] = E[\varepsilon_n \otimes \varepsilon_n]$, para cada $n \in \mathbb{N}_*$, basados en una muestra de tamaño T ; véase, por ejemplo, Ruiz-Medina et al. (2016):

$$\begin{aligned} \hat{\rho}_{k_T} &= \sum_{k=1}^{k_T} \hat{\lambda}_{T,k}(\rho) \phi_k \otimes \phi_k, & \hat{\lambda}_{T,k}(\rho) &= \frac{\frac{1}{T-1} \sum_{i=0}^{T-2} X_k(i) X_k(i+1)}{\frac{1}{T} \sum_{i=0}^{T-1} X_k^2(i)}, \\ & k = 1, \dots, k_T, \\ \hat{R}_\varepsilon &= \sum_{k=1}^{k_T} \hat{\lambda}_{T,k}(R_\varepsilon) \phi_k \otimes \phi_k, & \hat{\lambda}_{T,k}(R_\varepsilon) &= \frac{1}{T} \left(\sum_{i=0}^{T-1} X_k^2(i) \right) \left(1 - [\hat{\lambda}_{T,k}(\rho)]^2 \right), \\ & k = 1, \dots, k_T, \end{aligned} \tag{3.14}$$

donde $X_k(i)$ satisface la ecuación (3.13), para $i = 0, \dots, T-1$, y para cada $k = 1, \dots, k_T$. El orden de truncamiento, k_T , debe satisfacer condiciones adecuadas de acuerdo a la magnitud relativa de $\lambda_{k_T}(C_X)$ y T en relación con asegurar la consistencia, ya que $k_T \rightarrow \infty$ cuando $T \rightarrow \infty$, y $k_T/T \rightarrow 0$ cuando $T \rightarrow \infty$.

Por último, bajo las condiciones 3.1–3.2, a partir de las ecuaciones (3.9) y (3.13), los operadores A y Q , de las ecuaciones (3.1) y (3.2), caracterizando la distribución del proceso O-U H -valuado en (1.8), se pueden aproximar como sigue:

$$\hat{A}_T = \sum_{k=1}^{k_T} (1 - \hat{\lambda}_{T,k}(\rho)) \phi_k \otimes \phi_k \tag{3.15}$$

$$\hat{Q}_T = \sum_{k=1}^{k_T} \hat{\lambda}_{T,k}(R_\varepsilon) \phi_k \otimes \phi_k, \tag{3.16}$$

donde, en este caso, $\{\phi_j\}_{j \geq 1}$ denota al sistema común de autovectores de A y Q en las ecuaciones (3.1) y (3.2), y en las ecuaciones (3.12) y (3.13), considerando $\rho = I - A$. Nótese que, bajo la Condición 3.1, $\{\phi_j\}_{j \geq 1}$ coincide también con el conocido sistema de autovectores que interviene en la representación en serie (3.5) del proceso de Wiener H -valuado, W .

A partir de las ecuaciones (1.8), (3.15) y (3.16), se puede considerar el siguiente predictor *plug-in* del proceso de O-U H -valuado:

$$\widehat{X}_T(t, f) = \exp(t\widehat{A}_T)(f) + \int_0^t \exp((t-s)\widehat{A}_T)d\widehat{W}_T(s), \quad t \geq 0, \quad f \in H, \quad (3.17)$$

donde $\widehat{W}_T(t) - \widehat{W}_T(s) \sim \mathcal{N}(\mathbf{0}, (t-s)\widehat{Q}_T)$, es decir, $\widehat{W}_T(t) - \widehat{W}_T(s)$ tiene distribución de probabilidad gaussiana $\mu_{\mathbf{0}, (t-s)\widehat{Q}_T}$ en H , con media cero y operador de covarianza $(t-s)\widehat{Q}_T$, con \widehat{Q}_T dado en (3.16).

3.4. Proceso de Cox log-gaussiano en espacios de dimensión infinita

Esta sección presenta los principales elementos y resultados teóricos involucrados en la introducción, estimación de parámetros y predicción de procesos de Cox con intensidad log-gaussiana, definida a partir del proceso O-U H -valuado.

En primer lugar, se extraen condiciones suficientes para la definición de $\{\mathcal{C}(t), t \in \mathbb{R}_+\} = \{\mathcal{C}_p(t), p \geq 1, t \in \mathbb{R}_+\}$ como un proceso de Poisson doblemente estocástico ℓ^2 -valuado, con una intensidad aleatoria ℓ^2 -valuada $\{\lambda_t = \exp(Y_p(t)), p \geq 1, t \in \mathbb{R}_+\}$, donde Y_p satisface las ecuaciones (3.6) y (3.7), para cada $p \geq 1$. Para calcular la varianza de $\Lambda_p(t) = \int_0^t \exp(Y_p(s))ds$, en la Proposición 3.1 que sigue, se considera el sistema de polinomios de Hermite.

Es bien sabido que los polinomios de Hermite forman un sistema ortogonal completo del espacio de Hilbert $L_2(\mathbb{R}, \varphi(u)du)$, el espacio de funciones de cuadrado integrable con respecto a la densidad normal estándar φ . Dichos polinomios se definen como sigue:

$$H_k(u) = (-1)^k e^{\frac{u^2}{2}} \frac{d^k}{du^k} e^{-\frac{u^2}{2}}, \quad k = 0, 1, \dots$$

En particular, para un proceso gaussiano de media cero Y , para $k \geq 1$,

$$\begin{aligned} \mathbb{E} [H_k(Y(\mathbf{x}))] &= 0, \\ \mathbb{E} [H_k(Y(\mathbf{x})) H_m(Y(\mathbf{y}))] &= \delta_{m,k} m! (\mathbb{E}[Y(\mathbf{x})Y(\mathbf{y})])^m; \end{aligned} \quad (3.18)$$

véase, por ejemplo, Peccati y Taqqu (2011).

Proposición 3.1. *Asumamos que*

$$\lambda_p(Q) < 2\lambda_p(A), \quad p \geq 1. \quad (3.19)$$

Entonces, para cada $p \geq 1$, considerando la covarianza condicionada de Y_p , y asumiendo que t es suficientemente grande, para asegurar que la diferencia con la covarianza incondicionada del proceso de O-U real-valuado Y_p es despreciable, se cumple la siguiente identidad:

$$\begin{aligned} \text{Var} \left(\int_0^t \exp(Y_p(s)) ds \right) \\ = \sum_{k=1}^{\infty} \frac{C_k^2}{k!} \int_0^t \int_0^t \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \exp(-\lambda_p(A)|u-v|) \right]^k dudv < \infty, \end{aligned} \quad (3.20)$$

donde $C_k = \int_{\mathbb{R}} \exp(u) H_k(u) \varphi(u) du$, $k \geq 1$, y $G(u) = \exp(u) = \sum_{k=1}^{\infty} \frac{C_k}{k!} H_k(u)$, para todo $u \in \mathbb{R}$, en $L^2(\mathbb{R}, \varphi(u) du)$.

Demostración:

A la ecuación (3.20) se llega directamente a partir de (3.18), y de la definición de Y_p en la ecuación (3.7), para cada $p \geq 1$. Además, $G(u) = \exp(u) \in L^2(\mathbb{R}, \varphi(u) du)$ implica

$$\|G\|_{L^2(\mathbb{R}, \varphi(u) du)}^2 = \sum_{k=1}^{\infty} \left[\frac{C_k}{k!} \right]^2 < \infty. \quad (3.21)$$

Teniendo en cuenta la desigualdad de Cauchy-Schwarz, la ecuación (3.21), y la

Condición (3.19), nos lleva a que $\left[\frac{\lambda_p(Q)}{2\lambda_p(A)}\right]^2 < 1$, con lo que se obtiene:

$$\begin{aligned}
 & \sum_{k=1}^{\infty} \frac{C_k^2}{k!} \int_0^t \int_0^t \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \exp(-\lambda_p(A)|u-v|) \right]^k dudv \\
 & \leq \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k^2}{k!} \right]^2} \sqrt{\sum_{k=1}^{\infty} \left[\int_0^t \int_0^t \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \exp(-\lambda_p(A)|u-v|) \right]^k dudv \right]^2} \\
 & \leq \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k^2}{[k!]^2} \right]^2} \sqrt{\sum_{k=1}^{\infty} \left[\int_0^t \int_0^t \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \exp(-\lambda_p(A)|u-v|) \right]^k dudv \right]^2} \\
 & \leq \tilde{K} \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k}{k!} \right]^2} t^2 \sqrt{\sum_{k=1}^{\infty} \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \right]^{2k}} \\
 & = \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k}{k!} \right]^2} \frac{t^2}{\sqrt{1 - \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \right]^2}} < \infty, \tag{3.22}
 \end{aligned}$$

donde \tilde{K} es una constante positiva, sabiendo que (3.21) implica que la secuencia $\left\{ \left[\frac{C_k}{k!} \right]^2 \right\}_{k \geq 1}$ converge a cero y, en particular, para k suficientemente grande, se tiene que $\left[\frac{C_k}{k!} \right]^4 < \left[\frac{C_k}{k!} \right]^2 < 1$. \square

En el desarrollo subsiguiente se asume la siguiente condición:

Suposición 3.3. Se asume el siguiente comportamiento asintótico.

$$\frac{2\lambda_p(A)}{\sqrt{[2\lambda_p(A)]^2 - [\lambda_p(Q)]^2}} = \mathcal{O}(p^{-\alpha}), \quad p \rightarrow \infty, \quad \alpha > 1.$$

Proposición 3.2. Bajo las condiciones (i)–(iii) formuladas en el Apartado 1.2.4, considerando también que las condiciones 3.1, 3.3, y la ecuación (3.19) se cumplen, se tiene, para cada $t \in \mathbb{R}_+$,

$$\sum_{p=1}^{\infty} C_p^2(t) < \infty, \quad \text{c. s.}$$

Demostración:

A partir de la ecuación (3.22), bajo la Condición (3.19), y las condiciones 3.1 y

3.3,

$$\begin{aligned}
 E \left[\sum_{p=1}^{\infty} C_p^2(t) \right] &= \sum_{p=1}^{\infty} E [C_p^2(t)] \\
 &\leq \tilde{K} \sum_{p=1}^{\infty} \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k}{k!} \right]^2} \frac{t^2}{\sqrt{1 - \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \right]^2}} \\
 &= \tilde{K} t^2 \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k}{k!} \right]^2} \sum_{p=1}^{\infty} \left[1 - \left[\frac{\lambda_p(Q)}{2\lambda_p(A)} \right]^2 \right]^{-1/2} \\
 &= \tilde{K} t^2 \sqrt{\sum_{k=1}^{\infty} \left[\frac{C_k}{k!} \right]^2} \sum_{p=1}^{\infty} \frac{2\lambda_p(A)}{\sqrt{[2\lambda_p(A)]^2 - [\lambda_p(Q)]^2}} < \infty.
 \end{aligned}$$

□

A partir de la Proposición 3.2, $\mathcal{C} = \{C_p(t), p \geq 1, t \in \mathbb{R}^+\}$ es un proceso de Poisson doblemente estocástico ℓ^2 -valuado. Con un razonamiento similar al caso de los procesos de Poisson homogéneos ℓ^2 -valuados, comentado en el Apartado 1.2.2, se llega a la existencia de una medida de conteo σ -finita $\mu_{\mathcal{C}}$ asociada a tal proceso, con respecto a la cual, en particular, se puede definir la probabilidad de $\{C_p(T), p \geq 1\}$ para cada $T > 0$. Además, para cada $p \geq 1$, y para cualquier $0 < t_1 < t_2$, se cumple la siguiente identidad, dados los valores $Y_p(\omega, s)$, $\omega \in \Omega$, y $t_1 \leq s \leq t_2$, con, como antes, Y_p satisfaciendo (3.7) para cada $p \geq 1$,

$$\begin{aligned}
 \Pr(C_p(t_2) - C_p(t_1) = k | Y_p(\omega, s); t_1 \leq s \leq t_2) \\
 = \frac{\exp\left(-\int_{t_1}^{t_2} \exp(Y_p(\omega, s)) ds\right) \left[\int_{t_1}^{t_2} \exp(Y_p(\omega, s)) ds\right]^k}{k!}.
 \end{aligned}$$

Por lo tanto, bajo la condición de que $\{\mathbf{Y}(t), t \in \mathbb{R}_+\} = \{Y_p(t), p \geq 1, t \in \mathbb{R}_+\}$, $\{\mathcal{C}(t), t \in \mathbb{R}_+\} = \{C_p(t), p \geq 1, t \in \mathbb{R}_+\}$ define una secuencia de procesos de Poisson no homogéneos independientes, con lo que se tiene, para $\mathbf{k} = (k_1, \dots, k_n, \dots) \in \mathbb{N}^\infty$,

$$\begin{aligned}
 \Pr(\mathbf{C}(t_1) - \mathbf{C}(t_2) = \mathbf{k} | \mathbf{y}_{t_1, t_2}) &= \Pr(\mathbf{C}(t_1) - \mathbf{C}(t_2) = \mathbf{k} | \mathbf{Y}(\omega, s) = \mathbf{y}(s); t_1 \leq s \leq t_2) \\
 &= \prod_{p=1}^{p_0(\omega, \mathbf{y}_{t_1, t_2}, t_1, t_2)} \frac{\exp\left(-\int_{t_1}^{t_2} \exp(Y_p(\omega, s)) ds\right) \left[\int_{t_1}^{t_2} \exp(Y_p(\omega, s)) ds\right]^{k_p}}{(k_p)!}.
 \end{aligned}$$

Además, dadas las observaciones $\mathbf{y}_T = \{Y_p(\omega, t) = y_p(t), p \geq 1, 0 \leq t \leq T\}$, para cierto $\omega \in \Omega$, con, como antes, $Y_p(t) = \langle X(t, \phi), \phi_p \rangle_H$, satisfaciendo la ecuación (3.7), para cada $p \geq 1$, y cualquier $j \geq 1$, se puede calcular, a partir de (3.17), ver secciones 3.2 y 3.3, el predictor *plug-in*

$$\begin{aligned}
 \widehat{Y}_p(T+h) &= \left\langle \widehat{X}_T(T+h, \phi_j), \phi_p \right\rangle_H \\
 &= \left\langle \exp((T+h)\widehat{A}_T)(\phi_j) + \int_0^{T+h} \exp((t-s)\widehat{A}_T) d\widehat{W}_T(s), \phi_p \right\rangle_H,
 \end{aligned} \tag{3.23}$$

para $h \in \mathbb{R}_+$. Así, para cualquier $h > 0$, se formula el siguiente predictor ℓ^2 -valuado del proceso de Cox log-gaussiano \mathbf{C} :

$$E_{\widehat{\mathbf{Y}}_T(\omega)}(\mathbf{C}(T+h) | \mathbf{C}(\omega, T)) = \prod_{p=1}^{p_0(\omega, \widehat{\mathbf{Y}}_T(\omega), T, T+h)} \int_T^{T+h} \exp\left(\widehat{Y}_p(\omega, s)\right) ds + \mathbf{C}(\omega, T),$$

donde $\widehat{\mathbf{Y}}_T(\omega) = \{\widehat{Y}_p(\omega, s), p \geq 1, T \leq s \leq T+h\}$, siendo $\widehat{Y}_p(\cdot)$, tal y como se ha definido en (3.23), para cada $p \geq 1$, el predictor de la intensidad aleatoria, calculado a partir de las observaciones en el intervalo $[0, T]$, es decir,

$$\begin{aligned}
 \{\mathbf{Y}(\omega, s), 0 \leq s \leq T\} &= \{Y_p(\omega, s), p \geq 1, 0 \leq s \leq T\} \\
 &= \{\langle X(\omega, s, \phi), \phi_p \rangle_H, p \geq 1, 0 \leq s \leq T\} \\
 &= \{y_p(s), p \geq 1, 0 \leq s \leq T\}.
 \end{aligned}$$

3.5. Proceso de Cox log-gaussiano espacio-temporal en el mapeo de enfermedades

Los resultados dados en la Sección 3.4 se aplican aquí al caso particular donde $H = L^2(D)$, con $D \subseteq \mathbb{R}^2$. De hecho, la motivación reside en el problema de la estimación funcional y predicción de mapas de enfermedades. En la siguiente sección, en el estudio de simulación condicionada realizado, nos centraremos en la estimación y predicción de los mapas de riesgo de cáncer de mama, próstata y encéfalo en España.

3.5.1. Observación de la tasa aleatoria funcional de incidencia de la enfermedad

Sea $\{\mathcal{C}_t, t \in \mathbb{R}_+\}$ un proceso de Cox log-gaussiano con intensidad aleatoria $L^2(D)$ -valuada,

$$\{\lambda_t = E_t \exp(\beta_t + X_t), t \in \mathbb{R}_+\}, \quad (3.24)$$

siendo X_t un proceso O-U $L^2(D)$ -valuado, que satisface las condiciones asumidas en la sección anterior, y siendo E_t , para cada $t \in \mathbb{R}_+$, una función determinista $L^2(D)$ -valuada, proporcionando el mapa de los casos esperados de la enfermedad en las subregiones de D analizadas, en ese tiempo t . Por tanto, E_t tiene valores funcionales positivos en D , y β_t es un parámetro de efecto fijo $L^2(D)$ -valuado, para cada $t \in \mathbb{R}_+$, ver [Ruiz-Medina \(2016\)](#). En particular, para cada $t \in \mathbb{R}_+$, y para todo i , con $\mathbf{x}^i = (x_1^i, x_2^i)$ representando la localización espacial de la subregión i analizada en D ,

$$\ln(\lambda_t(\mathbf{x}^i)) = \ln(E_t(\mathbf{x}^i)) + \beta_t(\mathbf{x}^i) + X_t(\mathbf{x}^i), \quad i = 1, \dots, N. \quad (3.25)$$

Nuestro modelo de observación $L^2(D)$ -valuado se obtiene de la ecuación (3.25) añadiendo un ruido de observación, como sigue: Para $i = 1, \dots, N$,

$$Z_t(\mathbf{x}^i) = \ln(\lambda_t(\mathbf{x}^i)) + \epsilon_t(\mathbf{x}^i) = \ln(E_t(\mathbf{x}^i)) + \beta_t(\mathbf{x}^i) + X_t(\mathbf{x}^i) + \epsilon_t(\mathbf{x}^i),$$

donde, para cada $t \in \mathbb{R}_+$, $Z_t, \lambda_t, E_t, \epsilon_t, \beta_t, X_t \in L^2(D)$. En particular, $\{\epsilon_t, t \in \mathbb{R}_+\}$ es un proceso gaussiano $L^2(D)$ -valuado de media cero con componentes independientes y operador diagonal de autocovarianza.

Para cualquier $h > 0$, la estimación funcional de un mapa de riesgo de enfermedad en el tiempo $T+h$, basada en observaciones de intensidad funcional hasta el tiempo T , se construye a partir de

$$\frac{\Phi(\widehat{\mathcal{C}(T+h)})(\mathbf{x}^i)}{E_{T+h}(\mathbf{x}^i)}, \quad i = 1, \dots, N,$$

donde Φ denota la inversa del operador de proyección en el sistema de auto-vectores conocidos $\{\phi_j\}_{j \geq 1}$ de Q y A , y

$$\widehat{\mathcal{C}(T+h)} = \prod_{p=1}^{p_0(\omega, \widehat{\mathbf{Y}}_T(\omega), T, T+h)} \int_T^{T+h} \widehat{\lambda}_p(\omega, s) ds + \mathcal{C}(\omega, T), \quad (3.26)$$

con, para todo $p \geq 1$,

$$\{\widehat{\lambda}_p(\omega, s)\} = \langle E_s, \phi_p(\cdot) \rangle_H \exp \left(\left\langle \widehat{\beta}_s(\omega, \cdot), \phi_p(\cdot) \right\rangle_H + \left\langle \widehat{X}_T(\omega, s, \phi), \phi_p \right\rangle_H \right),$$

para $\omega \in \Omega$ denotando, como antes, el punto de muestreo asociado con la realización observada de la muestra aleatoria funcional. Aquí, para cada $s \in (T, T+h]$, $\widehat{\beta}_s$ es una estimación funcional no paramétrica de β_s (véase la ecuación (3.29) posterior), y \widehat{X}_T se define como en la ecuación (3.17), basada en observaciones funcionales hasta el tiempo T .

3.6. Estudio de simulación condicionada

Se generan tres muestras funcionales, bajo los supuestos 3.1, 3.2 y 3.3, condicionadas a los registros disponibles sobre casos esperados de cáncer de mama, próstata y encéfalo, en las provincias de España. Además, siguiendo las expresiones (3.24)–(3.25), nuestro modelo de observación se define a partir de la siguiente ecuación, extraída de considerar el desarrollo de Taylor en el ln de la función $\frac{C_t(\mathbf{x}^i)}{E_t(\mathbf{x}^i)}$ ¹ en el valor observado $\lambda_t(\mathbf{x}^i)$, para cada $t = 1, \dots, T$, y para

¹En el caso de estudio de mortalidad, esta función se conoce como Razón de Mortalidad Estandarizada (RME) y se trata de un estimador del Riesgo Relativo (RR), véase, por ejemplo, Bivand et al. (2013), entre otros. Consideremos $O_{t,i}$ y $E_{t,i}$ como el número de muertes observadas y esperadas (respectivamente) causadas por una enfermedad en el año $t = 1, \dots, T$,

cualquier $i = 1, \dots, N$,

$$\begin{aligned}
 \tilde{Z}_t(\mathbf{x}^i) &= \ln \left(\frac{C_t(\mathbf{x}^i)}{E_t(\mathbf{x}^i)} \right) = \ln(C_t(\mathbf{x}^i)) - \ln(E_t(\mathbf{x}^i)) \\
 &= \ln(\lambda_t(\mathbf{x}^i)) + \frac{C_t(\mathbf{x}^i) - \lambda_t(\mathbf{x}^i)}{\lambda_t(\mathbf{x}^i)} - \ln(E_t(\mathbf{x}^i)) \\
 &= \ln(E_t(\mathbf{x}^i)) + \beta_t(\mathbf{x}^i) + X_t(\mathbf{x}^i) - \ln(E_t(\mathbf{x}^i)) + \frac{C_t(\mathbf{x}^i) - \lambda_t(\mathbf{x}^i)}{\lambda_t(\mathbf{x}^i)} \\
 &= \beta_t(\mathbf{x}^i) + X_t(\mathbf{x}^i) + \frac{C_t(\mathbf{x}^i) - \lambda_t(\mathbf{x}^i)}{\lambda_t(\mathbf{x}^i)} = \beta_t(\mathbf{x}^i) + X_t(\mathbf{x}^i) + \xi_t(\mathbf{x}^i),
 \end{aligned} \tag{3.27}$$

donde, a partir de la aproximación gaussiana de la distribución de Poisson, para cada $t = 1, \dots, T$, y para cualquier $i = 1, \dots, N$,

$$\xi_t(\mathbf{x}^i) = \left\{ \frac{C_t(\mathbf{x}^i) - \lambda_t(\mathbf{x}^i)}{\lambda_t(\mathbf{x}^i)}, i = 1, \dots, N, t = 1, \dots, T \right\} \tag{3.28}$$

se distribuyen aproximadamente como variables aleatorias normales independientes con respectivas varianzas

$$\left\{ \frac{1}{\lambda_t(\mathbf{x}^i)}, i = 1, \dots, N, t = 1, \dots, T \right\}.$$

Como antes, se adopta el enfoque ARH(1) para obtener la estimación paramétrica funcional del proceso de O-U $L^2(D)$ -valuado, X_t , siendo D la península ibérica (véase la Sección 3.3 y, en particular, las secciones 3.2 y 3.3). En la última identidad de (3.27), para cada $t \in \mathbb{R}_+$, se estima el parámetro β_t , $L^2(D)$ -valuado, en un marco no paramétrico, a partir de la siguiente ecuación:

$$\hat{\beta}_t = \Phi(\Phi^*(\hat{Q}_t + \hat{C}_{\xi_t})^{-1}\Phi)^{-1}\Phi^*(\hat{Q}_t + \hat{C}_{\xi})^{-1}\tilde{Z}_t, \tag{3.29}$$

donde, como antes, \tilde{Z}_t se define a partir de (3.27), Φ^* denota el operador de proyección en el sistema de autovectores conocido $\{\phi_j\}_{j \geq 1}$ de Q y A , que también coincide con el sistema de autovectores común del operador de autocoprovincia $i = 1, \dots, N$. Entonces, la razón de mortalidad estandarizada se define como

$$RME_{t,i} = \frac{O_{t,i}}{E_{t,i}},$$

con $t = 1, \dots, T, i = 1, \dots, N$.

relación ρ , y el operador de autocovarianza C_X del proceso ARH(1), proporcionando la aproximación en tiempo discreto del proceso de O-U H -valuado. Aquí, \widehat{Q}_t denota el estimador componente a componente del operador de autocovarianza Q_t , en la ecuación (1.9), de X_t , calculado a partir de los estimadores introducidos en (3.15) y (3.16) de los operadores A y Q , respectivamente. Por último, \widehat{C}_{ξ_t} denota la aproximación empírica finito-dimensional del operador de autocovarianza C_{ξ_t} del proceso $L^2(D)$ -valuado ξ_t , introducido en la ecuación (3.28), basado en los valores observados $\left\{ \frac{1}{\lambda_t(\mathbf{x}^i)}, i = 1, \dots, N, t = 1, \dots, T \right\}$. En la práctica, se adoptan reglas de truncamiento que dependen del tamaño T de la muestra funcional, siguiendo los criterios reflejados en Bosq (2000); Bosq y Blanke (2007).

El filtrado de Kalman se aplica para obtener la predicción de los valores funcionales del proceso ARH(1) en los tiempos $t = T + h$, $h > 0$, aproximándose al correspondiente proceso de O-U H -valuado. En concreto, para un orden de truncamiento dado $M(T) < T$ adecuado, considerando la aproximación diagonal (3.13) de la ecuación de estados ARH(1), se formula el siguiente modelo de dimensión finita: Para $t = 1, \dots, T$,

$$\mathbf{X}_{M \times 1}(t) = \mathbf{\Lambda}_{M \times M}(\rho) \mathbf{X}_{M \times 1}(t-1) + \boldsymbol{\varepsilon}_{M \times 1}(t), \quad (3.30)$$

donde, para cada $t = 1, \dots, T$, $\mathbf{X}_{M \times 1}(t)$ es un vector $M \times 1$ con entradas $X_p(t) = \langle X_t, \phi_p \rangle_H$, $p = 1, \dots, M$; $\mathbf{\Lambda}_{M \times M}(\rho)$ es una matriz diagonal $M \times M$ con entradas $\lambda_p(\rho)$, $p = 1, \dots, M$; y $\boldsymbol{\varepsilon}_{M \times 1}(t)$ es un vector $M \times 1$ con entradas $\varepsilon_p(t) = \langle \varepsilon_t, \phi_p \rangle_H$, $p = 1, \dots, M$. Las ecuaciones para el filtrado de Kalman se implementan, a partir de (3.27) y (3.30), como sigue:

$$\widehat{\mathbf{X}}(t | t) = \widehat{\mathbf{X}}(t | t-1) + \mathbf{K}_t \left(Z_t - \mathbf{\Phi}_M \widehat{\mathbf{X}}(t | t-1) \right)$$

donde $\widehat{\mathbf{X}}(t | t) = E(\mathbf{X}(t) | Z_t, \dots, Z_1)$ es la estimación de la proyección actualizada (a-posteriori) de los coeficientes aleatorios $X_p(t)$, $p = 1, \dots, M$, en el tiempo t , y $\widehat{\mathbf{X}}(t | t-1) = E(\mathbf{X}(t) | Z_{t-1}, \dots, Z_1)$ es la estimación de la proyección (a-priori) correspondiente, previamente calculada, para el tiempo t . Aquí, Z_t se obtiene a partir de los valores observados (3.27) restando la estimación no paramétrica (3.29) en ese momento t . El operador \mathbf{K}_t denota al *operador de ganancias*, reflejando el suavizado realizado sobre datos funcionales, dado por

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \Phi_M^* (\mathbf{C}_{\xi_t} + \Phi_M \mathbf{P}_{t|t-1} \Phi_M^*)^{-1}$$

donde Φ_M^* y \mathbf{C}_{ξ_t} denotan respectivamente el operador de proyección en ϕ_1, \dots, ϕ_M , y el operador de autocovarianza de ξ_t , el ruido de observación. Además,

$$\mathbf{P}_{t|t-1} = \text{Cov}(\mathbf{X}(t) | Z_{t-1}, \dots, Z_1) = \Lambda_{M \times M}(\rho) \mathbf{P}_{t-1|t-1} \Lambda_{M \times M}(\rho) + \mathbf{R}_{\varepsilon_t},$$

siendo $\mathbf{R}_{\varepsilon_t}$ la matriz de covarianzas del vector aleatorio $\varepsilon_{M \times 1}(t)$, con las componentes $\varepsilon_p(t) = \langle \varepsilon_t, \phi_p \rangle_H$, $p = 1, \dots, M$, y

$$\mathbf{P}_{t|t} = E \left((\mathbf{X}(t) - \widehat{\mathbf{X}}(t | t)) (\mathbf{X}(t) - \widehat{\mathbf{X}}(t | t))^* \right) = \mathbf{P}_{t|t-1} - \mathbf{K}_t \Phi_M \mathbf{P}_{t|t-1}.$$

Así, la aproximación de dimensión finita del predictor del filtrado de Kalman, para cada tiempo t de interés, viene dada por

$$\widehat{X}_t = \Phi_M \widehat{\mathbf{X}}(t | t-1) = \Phi_M \Lambda(\rho) \widehat{\mathbf{X}}(t-1 | t-1). \quad (3.31)$$

Los valores iniciales considerados son

$$\begin{aligned} \widehat{\mathbf{X}}(0 | 0) &= \mathbf{0} \\ P_{0|0} &= E [\mathbf{X}_{M \times 1}(0) [\mathbf{X}_{M \times 1}(0)]^T], \end{aligned}$$

con $\mathbf{X}_{M \times 1}(0) = (X_1(0), \dots, X_M(0))^T = (\langle X_0, \phi_1 \rangle_H, \dots, \langle X_0, \phi_M \rangle_H)^T$.

En la práctica, $\Lambda_{M \times M}(\rho)$ se estima a partir de la matriz diagonal con entradas $\widehat{\lambda}_{T,k}(\rho)$, $k = 1, \dots, M$, definida en (3.14), y la matriz $\mathbf{R}_{\varepsilon_t}$ de dimensión $M \times M$ se aproxima a partir de la matriz diagonal con entradas $\widehat{\lambda}_{T,k}(R_\varepsilon)$, $k = 1, \dots, M$, también definida en (3.14).

La predicción funcional de los mapas de riesgo relativo, en cada momento t de interés, se obtiene entonces a partir de

$$\widehat{r}_t = \exp(\widehat{\beta}_t + \widehat{X}_t), \quad (3.32)$$

con $\widehat{\beta}_t$ como viene en (3.29), y \widehat{X}_t calculada a partir de las anteriores ecuaciones del filtrado de Kalman, en términos de los residuos proyectados, asociada con la estimación del parámetro de efecto fijo $L^2(D)$ -valuado β_t , para cada

tiempo t de interés. En la Tabla 3.1 se muestra el promedio espacial de los errores cuadráticos medios empíricos para los datos generados de los tres tipos de cáncer (de mama, próstata y encéfalo), para cada uno de los años estudiados, calculados a partir de 100 realizaciones del predictor de Kalman (3.31), y del estimador no paramétrico de efectos fijos (3.29). Se puede observar que los datos de cáncer de mama y próstata se generan y se estiman mejor a partir del enfoque de intensidad ARH(1) log-gaussiano, ya que el error es mucho más pequeño que para el caso de cáncer de encéfalo. Esto puede ser debido a varias razones. La incidencia del cáncer de encéfalo es menor en la población, con lo que hay provincias con pocos casos observados y esperados de esta enfermedad, lo que puede provocar un mayor error cuadrático medio. También puede ser debido a las propias características del tipo de cáncer.

3.6.1. Algoritmo de simulación condicionada para el enfoque de intensidad log-gaussiana con aproximación ARH(1)

Mostramos a continuación, el algoritmo seguido para el estudio de simulación. Para dicho estudio se han utilizado los datos reales con el número de muertes por año y provincia española de tres tipos de cáncer distintos durante un periodo de observación en torno a los 30 años.

Paso 1. Partimos de los registros disponibles del logaritmo del Riesgo Relativo (RR) de mortalidad asociado al cáncer de mama, próstata y encéfalo, según se define en la ecuación (3.27), primera línea. A continuación se procede a la eliminación de la tendencia lineal.

Paso 2. Se calcula una estimación funcional truncada del parámetro de efectos fijos en el modelo de observación considerado en (3.27), $\hat{\beta}_t$, a partir de los datos del Paso 1.

Paso 3. Se generan valores funcionales aleatorios del proceso ARH(1) en los tiempos $t = 1, \dots, T$.

Paso 4. Se generan valores funcionales aleatorios para el modelo de observación definido en la ecuación (3.27), para $t = 1, \dots, T$, y se evalúan en los 50 nodos espaciales correspondientes a las provincias analizadas.

Paso 5. Se calcula la exponencial de los datos generados en el Paso 4.

Tabla 3.1: Promedio espacial de los errores cuadráticos medios empíricos (ECMEs) por año, obtenido a partir de 100 generaciones de la muestra funcional de tamaño T del proceso de intensidad ARH(1), para los datos generados de los tres tipos de cáncer

Año	Cáncer de próstata	Cáncer de mama	Cáncer de encéfalo
1	4.587e-002	4.197e-002	9.337e-001
2	4.150e-002	3.516e-002	6.712e-001
3	3.410e-002	3.001e-002	3.264e-001
4	3.586e-002	3.089e-002	4.651e-001
5	3.227e-002	2.688e-002	3.823e-001
6	3.216e-002	3.058e-002	1.766e-001
7	3.391e-002	2.797e-002	4.192e-001
8	2.891e-002	3.304e-002	3.921e-001
9	3.128e-002	2.886e-002	3.182e-001
10	3.096e-002	3.127e-002	1.744e-001
11	3.070e-002	3.137e-002	2.977e-001
12	2.989e-002	3.248e-002	2.250e-001
13	3.052e-002	3.263e-002	6.456e-001
14	2.809e-002	3.462e-002	2.350e-001
15	2.866e-002	3.489e-002	1.209e+000
16	2.814e-002	3.419e-002	2.742e-001
17	2.627e-002	3.397e-002	3.685e-001
18	2.858e-002	3.629e-002	1.244e-001
19	3.101e-002	3.637e-002	2.864e-001
20	3.556e-002	3.677e-002	1.857e-001
21	3.528e-002	3.497e-002	6.199e-001
22	3.715e-002	3.167e-002	3.545e-001
23	3.337e-002	3.344e-002	2.227e-001
24	3.432e-002	2.605e-002	1.223e+000
25	3.405e-002	3.184e-002	1.396e-001
26	3.171e-002	2.634e-002	
27	3.165e-002	3.126e-002	
28	3.156e-002	2.438e-002	
29	2.998e-002	2.771e-002	
30	2.892e-002	2.809e-002	
31	2.593e-002	2.387e-002	
32	2.671e-002		
33	2.459e-002		
34	2.206e-002		

Paso 6. Se calculan los valores de la intensidad aleatoria, de acuerdo al modelo considerado en la ecuación (3.24).

Paso 7. Se calculan los valores generados de la razón de mortalidad estandarizada a partir de los valores del Paso 6 y los registros disponibles de casos esperados de mortalidad por padecer la enfermedad.

Los correspondientes mapas con el riesgo de incidencia de enfermedad se representan seguidamente (véase las figuras 3.1–3.3). Dichas figuras muestran año por año los correspondientes mapas de riesgo de mortalidad según el tipo de cáncer bajo estudio (34 años para el caso de cáncer de próstata, desde 1975 a 2008; 31 años para cáncer de mama, desde 1975 a 2005; y 25 años para cáncer de encéfalo, desde 1986 a 2010). Para cada uno de los tres tipos de cáncer analizados se obtienen las gráficas para la población en los años indicados, diferenciando el riesgo simulado o generado a partir de una realización y el riesgo estimado a partir de la aproximación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana.

3.6.2. Intensidad espacio-temporal con dependencia de largo rango

En este apartado se considera el caso en que la intensidad aleatoria, tal y como aparece en (3.24), $\{\lambda_t, t \in \mathbb{R}_+\}$, del proceso de Cox log-gaussiano estudiado se define a partir de un proceso gaussiano espacio-temporal con Dependencia de Largo Rango (DLR), satisfaciendo la siguiente ecuación integral, en el sentido de media cuadrática:

$$X_t(\mathbf{x}^i) \underset{\text{m.c.}}{=} \int_{\mathbb{R}_+} \int_D K(t, s, \mathbf{x}^i, \mathbf{y}, \boldsymbol{\theta}) V(s, \mathbf{y}) ds d\mathbf{y}, \quad (3.33)$$

para $i = 1, \dots, N$, y $t \in \mathbb{R}_+$. Aquí,

$$K(t, s, \mathbf{x}^i, \mathbf{y}, \boldsymbol{\theta}) = |t - s|^{-1+\nu} |x_1^i - y_1|^{-1+\varpi_1} |x_2^i - y_2|^{-1+\varpi_2},$$

con $\boldsymbol{\theta} = (\nu, \varpi_1, \varpi_2) \in (0, 1)^3$, ver Frías y Ruiz-Medina (2011). El proceso V es un proceso gaussiano regular de media cero tal que la integral (3.33) es finita en el sentido de media cuadrática. Los datos se generan, considerando el modelo de observación (3.27), pero reemplazando X_t por el proceso dado en (3.33). El parámetro estimado β_t , $L^2(D)$ -valuado, se calcula a partir de la ecuación (3.29). A continuación, se calcula el predictor de mínimos cuadrados, a partir

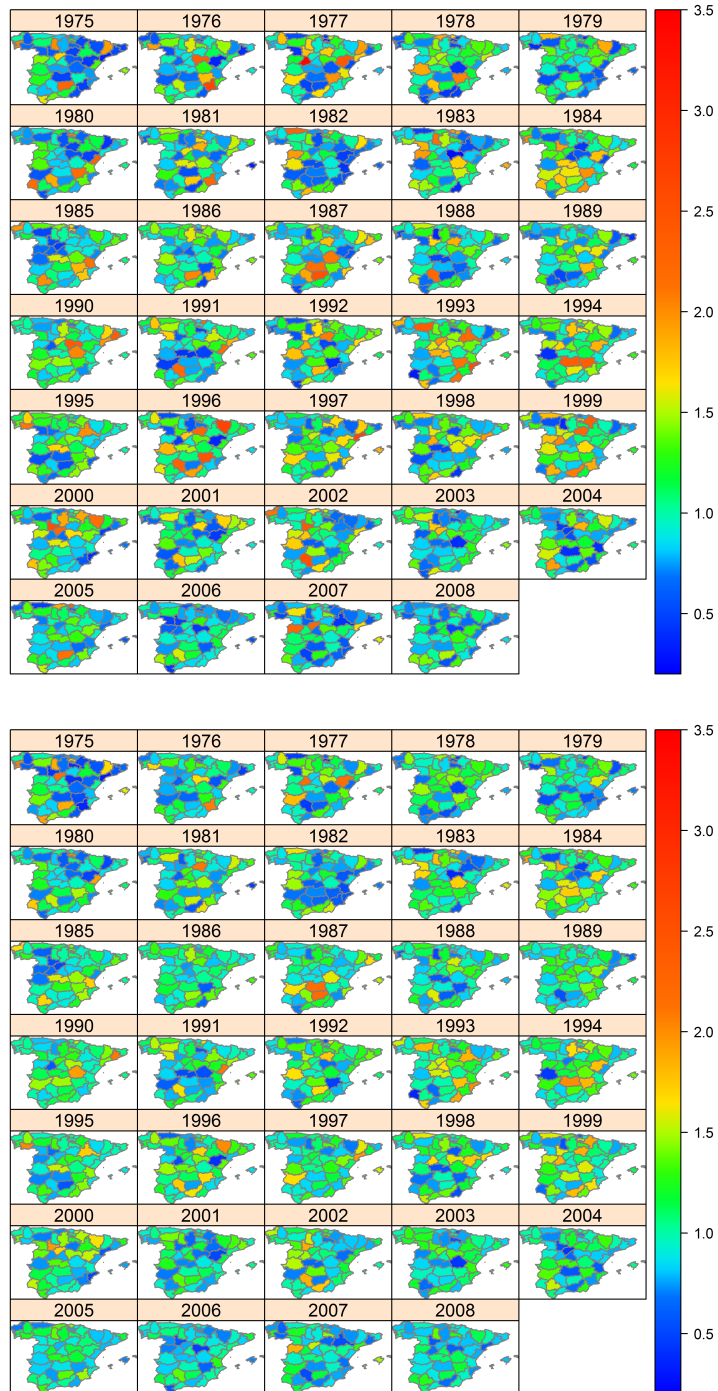


Figura 3.1: Mapas de riesgo relativo de mortalidad por cáncer de próstata: para una generación (figura superior) y su estimación (figura inferior), basada en la aproximación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana

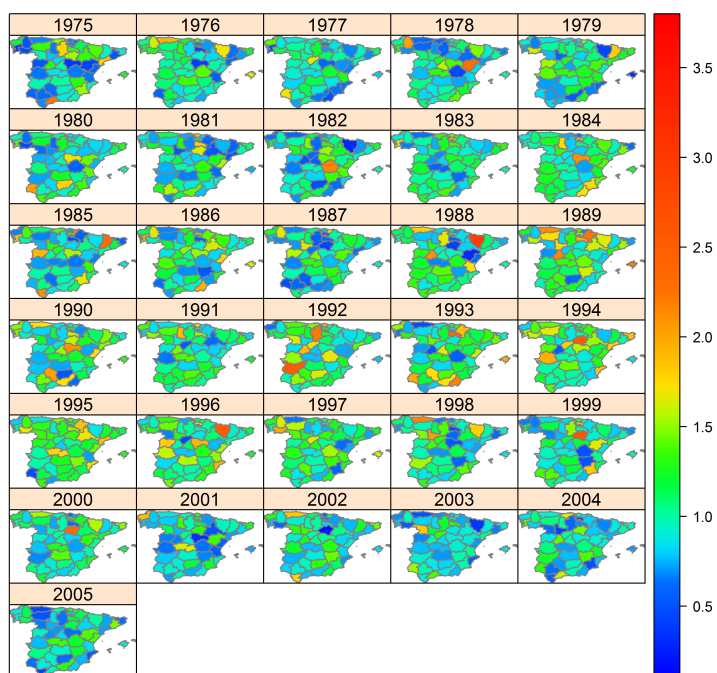
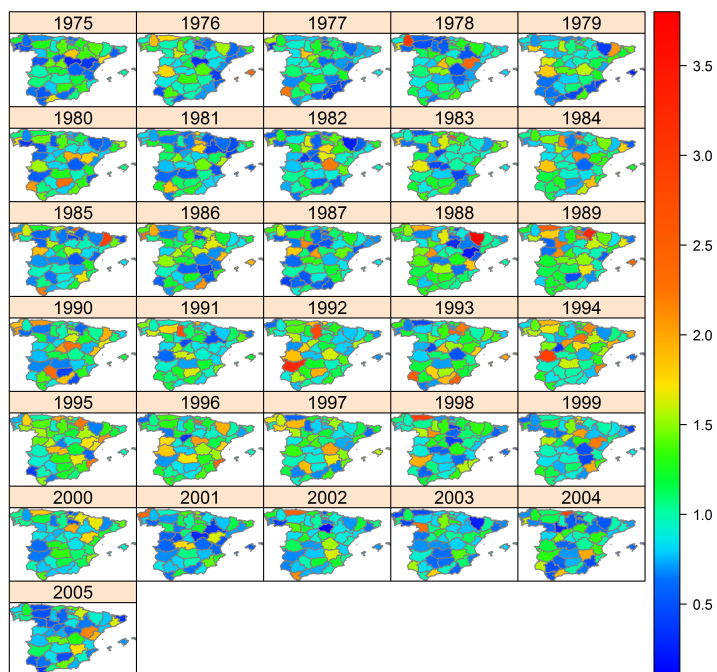


Figura 3.2: Mapas de riesgo relativo de mortalidad por cáncer de mama: para una generación (figura superior) y su estimación (figura inferior), basada en la aproximación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana

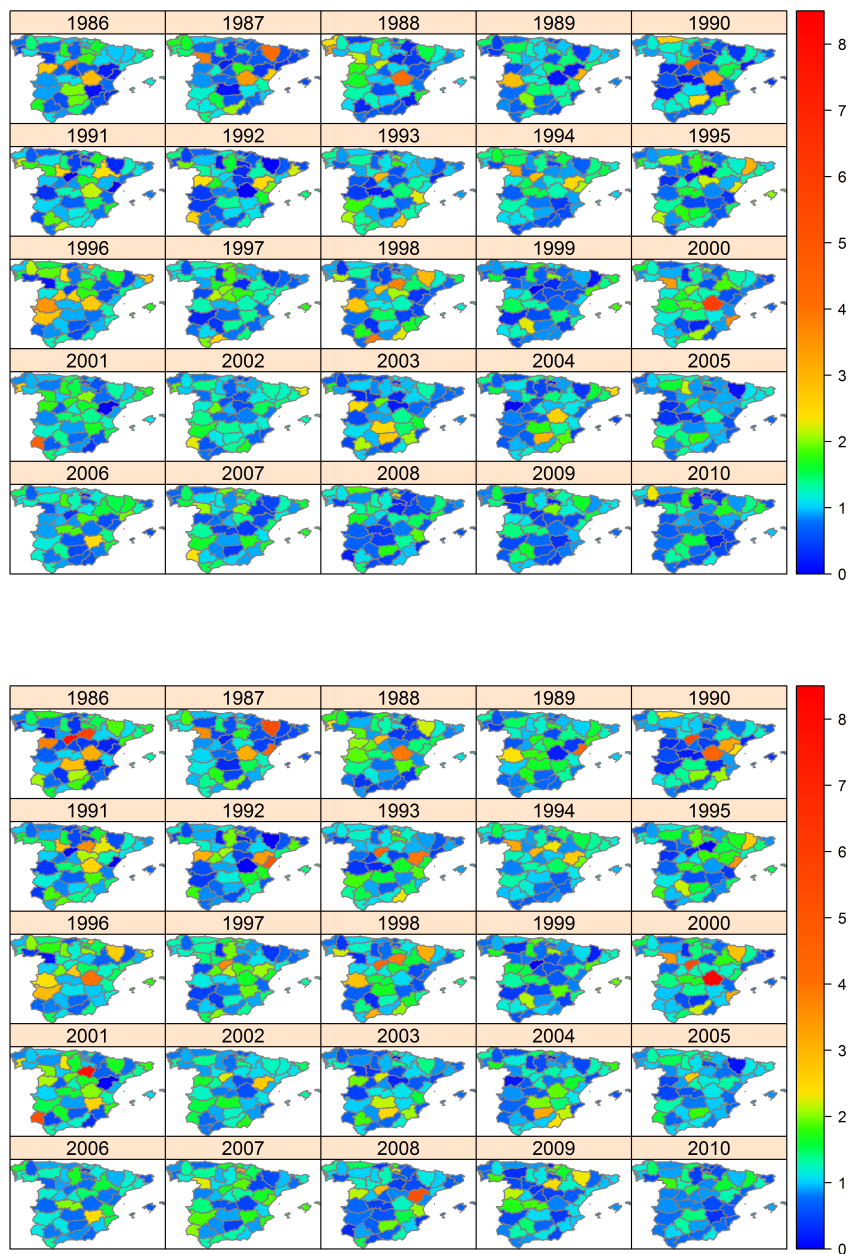


Figura 3.3: Mapas de riesgo relativo de mortalidad por cáncer de encéfalo: para una generación (figura superior) y su estimación (figura inferior), basada en la aproximación en tiempo discreto ARH(1) del modelo de intensidad log-gaussiana

de los residuos asociados como sigue:

$$\widehat{X}_t = R_{X_t(\widetilde{Z}_t - \widehat{\beta}_t)} R_{(\widetilde{Z}_t - \widehat{\beta}_t)(\widetilde{Z}_t - \widehat{\beta}_t)}^{-1} (\widetilde{Z}_t - \widehat{\beta}_t), \quad \forall t \in \mathbb{R}_+, \quad (3.34)$$

donde

$$\begin{aligned} R_{X_t(\widetilde{Z}_t - \widehat{\beta}_t)} &= E[X_t \otimes \widetilde{Z}_t - \widehat{\beta}_t] \\ R_{(\widetilde{Z}_t - \widehat{\beta}_t)(\widetilde{Z}_t - \widehat{\beta}_t)}^{-1} &= \left[E[(\widetilde{Z}_t - \widehat{\beta}_t) \otimes (\widetilde{Z}_t - \widehat{\beta}_t)] \right]^{-1}, \end{aligned} \quad (3.35)$$

definiéndose \widetilde{Z}_t como en la ecuación (3.27), y X_t como en la ecuación (3.33). Aquí, como es usual, \otimes denota el producto tensorial de las funciones en $L^2(D)$. El parámetro $\theta = (\nu, \varpi_1, \varpi_2)$, que caracteriza los operadores integrales $R_{X_t(\widetilde{Z}_t - \widehat{\beta}_t)}$ y $R_{(\widetilde{Z}_t - \widehat{\beta}_t)(\widetilde{Z}_t - \widehat{\beta}_t)}^{-1}$, se ha estimado a partir de los registros originales de cáncer de mama, próstata y encéfalo, en las provincias de España, aplicando la metodología basada en wavelets propuesta en Frías y Ruiz-Medina (2011); véase también Frías et al. (2013). La presencia de dependencia de largo rango en el tiempo (memoria larga) es más evidente que en el espacio (correlación fuerte) para los tres tipos de datos de cáncer analizados (véase la Tabla 3.2, donde la primera columna se corresponde con la estimación del parámetro de dependencia de largo rango para el tiempo, y las columnas 2 y 3 para el espacio).

Tabla 3.2: Estimaciones paramétricas con DLR

	$\hat{\nu}$	ϖ_1	ϖ_2
Cáncer de próstata	0.6997	0.4103	0.5483
Cáncer de mama	0.6667	0.4441	0.4356
Cáncer de encéfalo	0.6883	0.4555	0.55435

El promedio espacial de los errores cuadráticos medios empíricos, para cada uno de los años estudiados, calculado a partir de 100 realizaciones del predictor de mínimos cuadrados (3.34), y el estimador no paramétrico de efectos fijos (3.29), se muestran en la Tabla 3.3, para los tres tipos de datos de cáncer (próstata, mama y encéfalo), generados de forma condicionada. Se puede observar que las estimaciones obtenidas a partir de los datos de cáncer de encéfalo, bajo un modelo DLR de log-intensidad gaussiana, mejoran las obtenidas bajo dependencia ARH(1) de corto rango en el tiempo. Sin embargo, se observa un mejor

ajuste bajo el modelo ARH(1), a partir de los registros de cáncer de mama y próstata (ver tablas 3.1 y 3.3).

Algoritmo de simulación condicionada para el enfoque log-gaussiano con DLR

Mostramos a continuación, el algoritmo seguido para el estudio de simulación siguiendo un enfoque log-gaussiano con dependencia de largo rango. Como en el algoritmo anterior, para dicho estudio se han utilizado los datos reales con el número de muertes por año y provincia española de tres tipos de cáncer distintos en un periodo de observación en torno a los 30 años.

Paso 1. Partimos de los registros disponibles sobre el logaritmo del riesgo relativo de mortalidad asociado al cáncer de próstata, mama o encéfalo, según se define en la ecuación (3.27), primera línea.

Paso 2. A partir de los datos cargados en el Paso 1, basándonos en el modelo (3.33) se estiman los parámetros de dependencia de rango largo (DLR) ν , ϖ_1 y ϖ_2 , utilizando el método de regresión log-wavelet introducido en Frías y Ruiz-Medina (2011), ver también Frías et al. (2013), este método se basa en el comportamiento del logaritmo de la Transformada Wavelet en un entorno de las frecuencias cercanas a cero. Las estimaciones obtenidas aparecen en la Tabla 3.2.

Paso 3. Se aproxima la ecuación (3.33) mediante el método de Monte Carlo, obteniendo una muestra de valores funcionales aleatorios del proceso X_t definido en (3.33) en los tiempos $t = 1, \dots, T$, y evaluados en los 50 nodos espaciales correspondientes a las provincias analizadas.

Paso 4. Se realiza una estimación funcional truncada del parámetro de efectos fijos del modelo de observación definido en la ecuación (3.27), $\hat{\beta}_t$, a partir de los datos obtenidos en el Paso 1, tras la eliminación de la tendencia lineal.

Paso 5. Se generan valores funcionales aleatorios para el modelo de observación definido en (3.27), pero reemplazando X_t por el proceso dado en (3.33). Estos valores funcionales generados se evalúan en los 50 nodos espaciales correspondientes a las provincias analizadas. Se calcula la exponencial a estos datos.

Tabla 3.3: Promedio espacial de los errores cuadráticos medios empíricos (ECMEs) por año, obtenidos a partir de 100 generaciones de la muestra funcional de tamaño T del modelo de intensidad aleatoria con DLR, para los tres tipos de datos de cáncer generados

Año	Cáncer de próstata	Cáncer de mama	Cáncer de encéfalo
1	2.715e-001	4.238e+000	2.539e-001
2	2.492e+001	5.445e-002	7.854e-002
3	2.742e+000	2.530e-002	6.705e-002
4	6.400e+000	6.835e-002	6.445e-002
5	8.250e-001	2.399e-001	1.089e-001
6	4.614e-002	1.549e+001	9.699e-002
7	4.589e-001	3.204e-002	2.633e-001
8	5.280e-001	7.784e-002	9.773e-002
9	4.348e-001	5.195e-002	7.650e-002
10	5.723e-002	3.386e+000	6.673e-002
11	1.919e-002	1.168e-001	7.677e-002
12	1.427e+000	2.654e-002	8.259e-002
13	1.610e-002	2.723e-002	7.083e-002
14	3.094e+000	4.607e-002	5.450e-002
15	1.495e-002	2.165e+000	8.851e-002
16	3.012e-001	6.044e-002	4.365e-002
17	4.050e-002	4.107e-002	5.732e-002
18	1.409e-002	1.848e-002	5.499e-002
19	2.185e-001	1.178e+000	3.053e-002
20	2.088e-002	2.170e-001	3.951e-002
21	1.027e+000	4.749e-002	2.638e-002
22	1.574e-002	4.337e-002	5.691e-002
23	9.894e-002	6.459e-002	2.416e-002
24	1.254e-002	1.791e+000	4.403e-002
25	1.391e-001	3.138e-002	2.506e-002
26	2.559e-001	3.077e-002	
27	1.532e-002	2.021e-002	
28	4.820e-002	2.019e-002	
29	2.951e-002	2.623e-001	
30	1.298e-001	2.447e-002	
31	3.505e-002	1.050e-002	
32	3.021e-002		
33	7.394e-002		
34	5.161e-002		

Paso 6. Se calculan los valores de la intensidad aleatoria, de acuerdo al modelo considerado en la ecuación (3.24).

Paso 7. Para cada tiempo, $t = 1, \dots, T$, y provincia, $i = 1, \dots, 50$, se genera una muestra de valores aleatorios de una distribución de Poisson con parámetro, el elemento (t, i) de la matriz obtenida en el Paso 6.

Paso 8. Se calculan los valores generados de la razón de mortalidad estandarizada a partir de los valores del Paso 6 y los registros disponibles de casos esperados de mortalidad por padecer la enfermedad.

Seguidamente, se representan los correspondientes mapas con el riesgo relativo de mortalidad, año por año, según el tipo de cáncer bajo estudio (34 años para el caso de cáncer de próstata, desde 1975 a 2008; 31 años para cáncer de mama, desde 1975 a 2005; y 25 años para cáncer de encéfalo, desde 1986 a 2010), ver figuras 3.4–3.6. Dichas figuras muestran, para cada uno de los tres tipos de cáncer, las gráficas para la población con los años indicados, diferenciando el riesgo observado simulado o generado a partir de una realización y el riesgo estimado a partir del modelo de intensidad log-gaussiana con DLR.

Al considerar las gráficas de la aproximación ARH(1), figuras 3.1–3.3, con las del modelo de intensidad log-gaussiana con DLR (figuras 3.4–3.6), teniendo en cuenta que en cada caso estamos utilizando la simulación del riesgo de mortalidad por enfermedad para contrastarlo con las estimaciones de dicho riesgo, podemos también comparar los dos tipos de estimación realizados (ver también figuras 3.7 y 3.8).

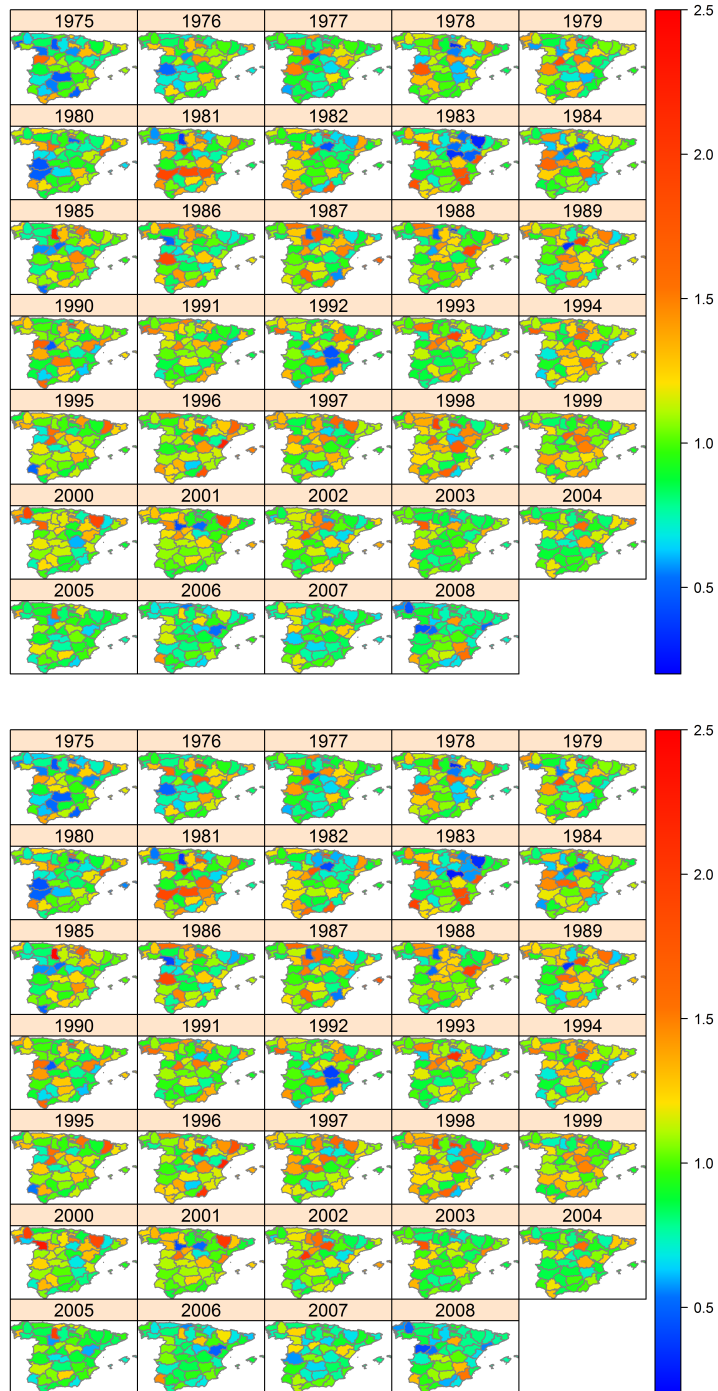


Figura 3.4: Mapas de riesgo relativo de mortalidad por cáncer de próstata: para una generación (figura superior) y su estimación (figura inferior), basada en el modelo de intensidad log-gaussiana con DLR

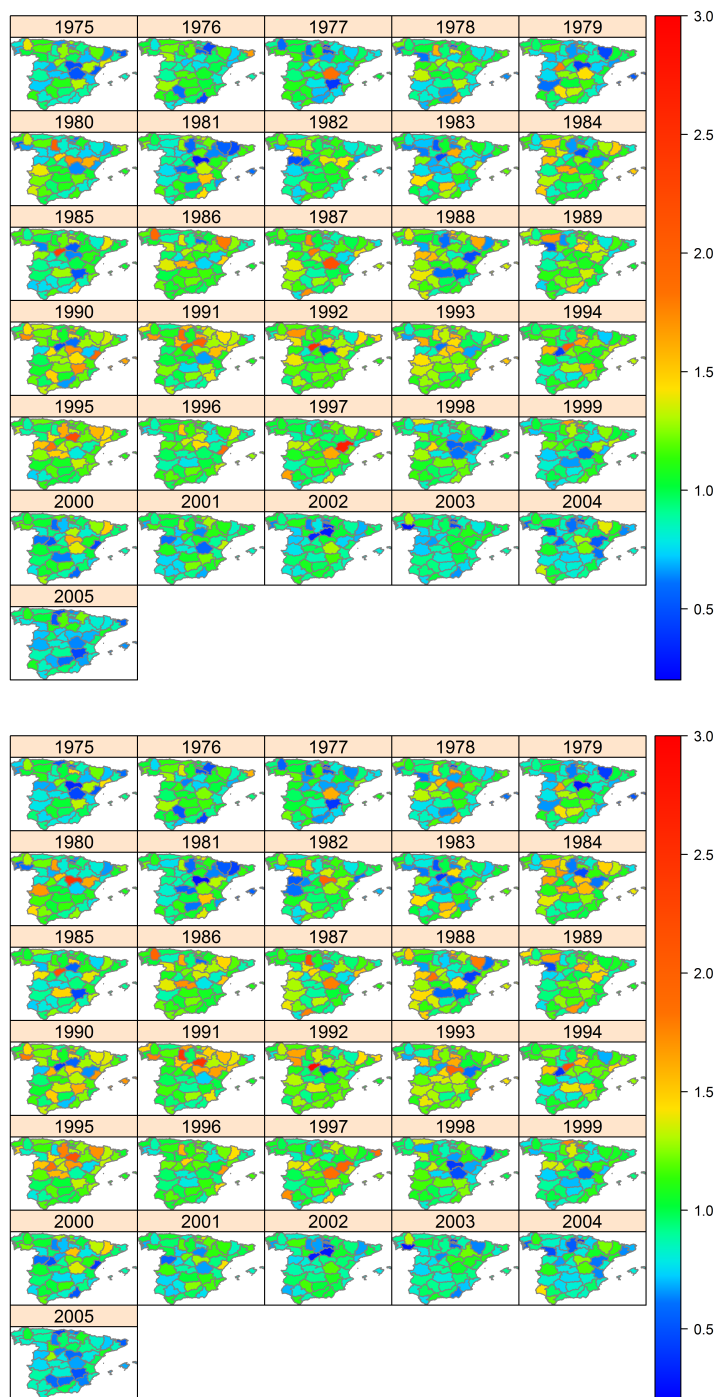


Figura 3.5: Mapas de riesgo relativo de mortalidad por cáncer de mama: para una generación (figura superior) y su estimación (figura inferior), basada en el modelo de intensidad log-gaussiana con DLR

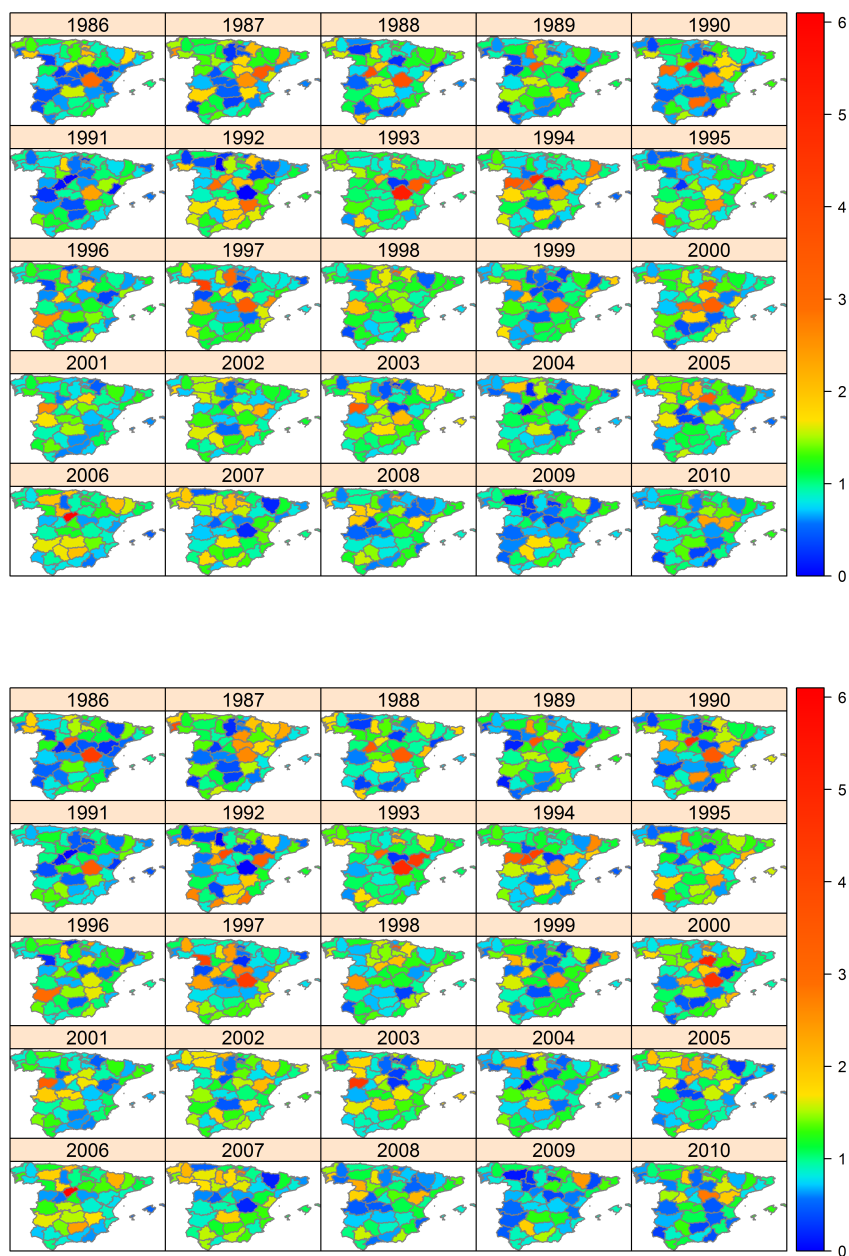


Figura 3.6: Mapas de riesgo relativo de mortalidad por cáncer de encéfalo: para una generación (figura superior) y su estimación (figura inferior), basada en el modelo de intensidad log-gaussiana con DLR

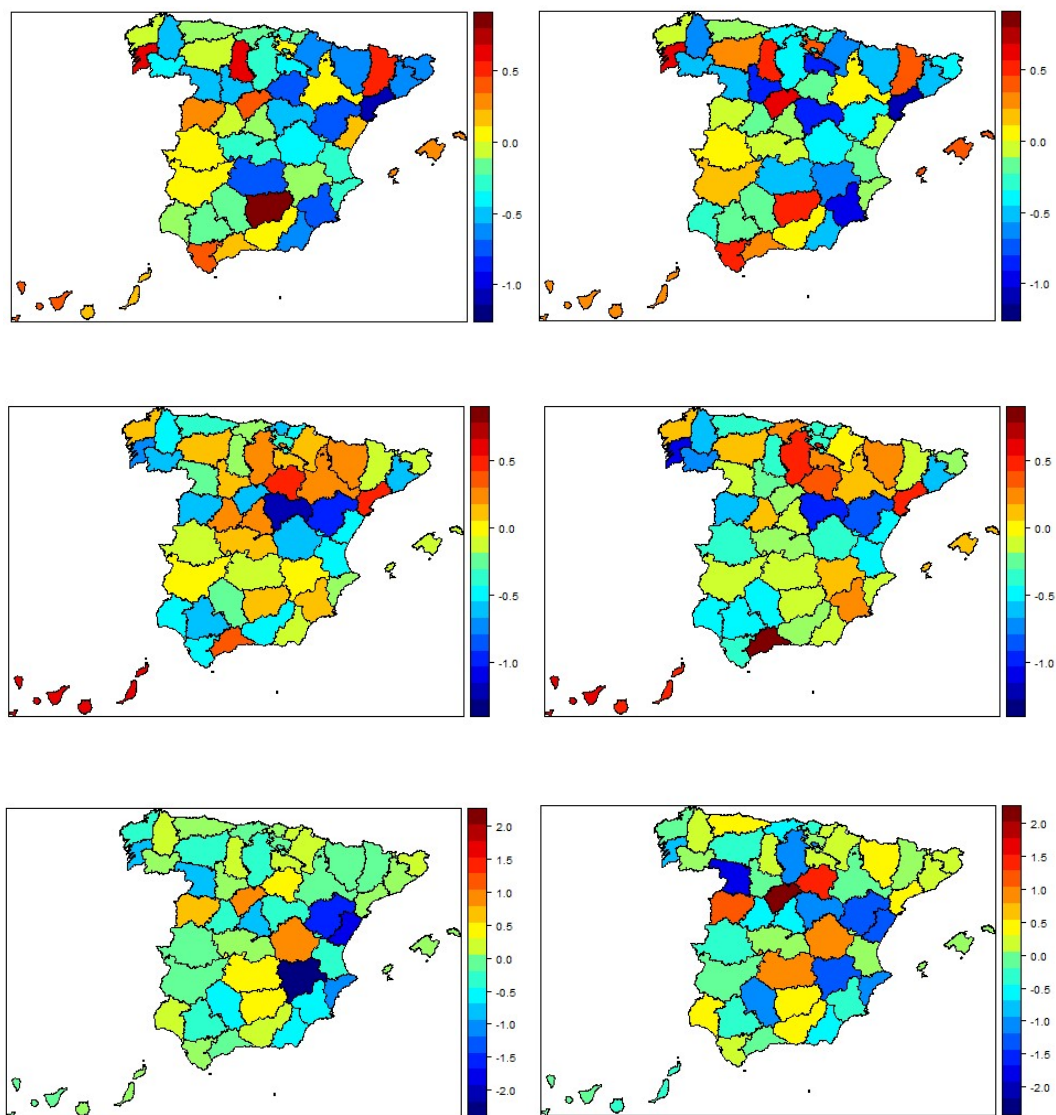


Figura 3.7: Enfoque O-U log-gaussiano. Mapas de España con el log-riesgo de mortalidad por tipo de cáncer: de próstata (año 1975), de mama (año 1975), y de encéfalo (año 1986); en base a 1a generación de un proceso O-U (de la parte superior izquierda a la inferior izquierda) y sus respectivas estimaciones con el filtrado de Kalman (de la parte superior derecha a la inferior derecha)

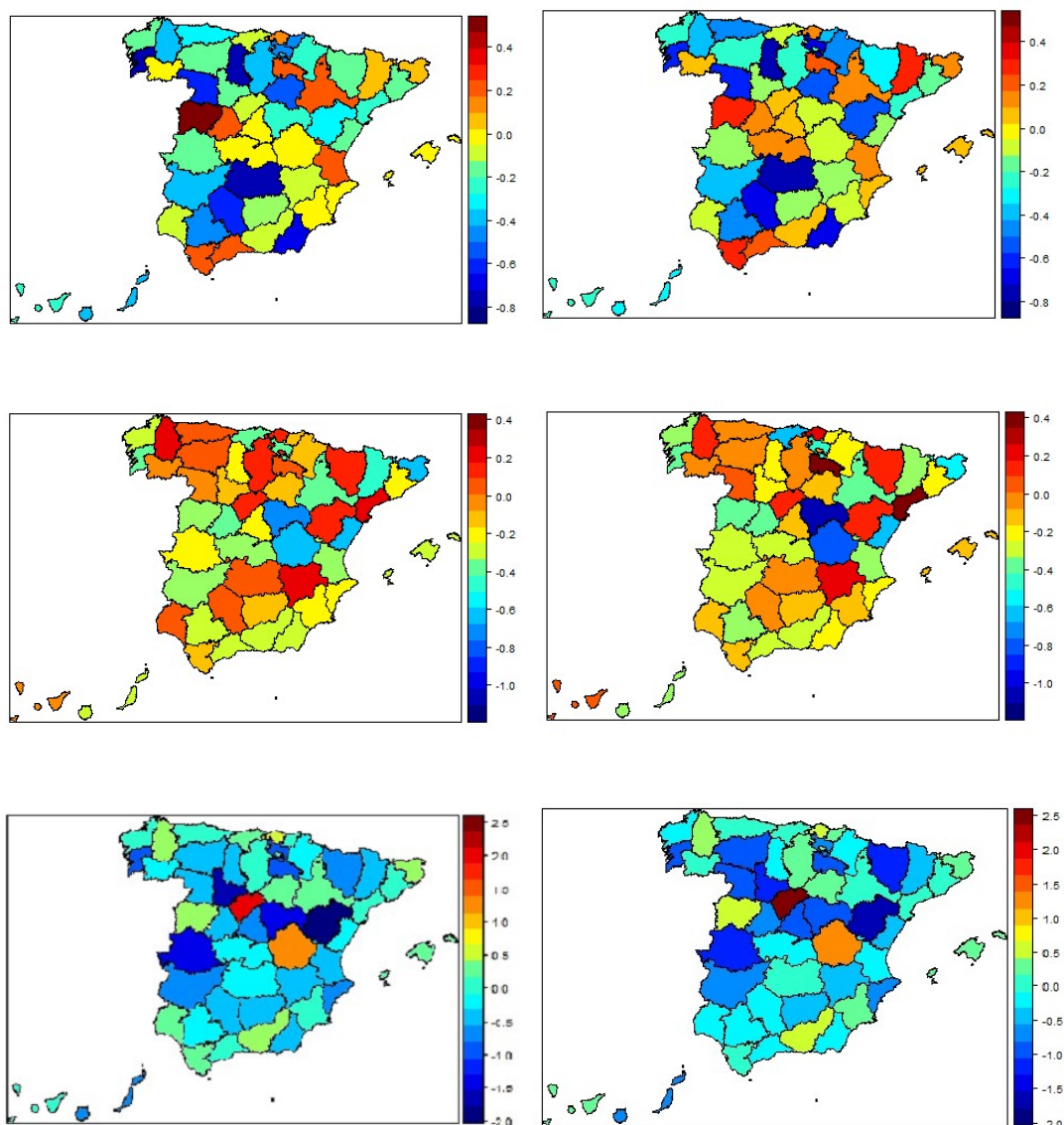


Figura 3.8: Enfoque log-gaussiano con DLR. Mapas de España con el log-riesgo de mortalidad por tipo de cáncer: de próstata (año 1975), de mama (año 1975), y de encéfalo (año 1986); en base a 1 generación del proceso de Cox con intensidad log-gaussiana con DLR (de la parte superior izquierda a la inferior izquierda) y sus respectivas estimaciones de mínimos cuadrados de DLR (de la parte superior derecha a la inferior derecha)

Capítulo 4

Procesos de Cox espaciales en un marco infinito-dimensional

Este capítulo surge como continuación del Capítulo 3, donde se ha introducido una familia de procesos de Cox log-gaussianos temporales valuados en ℓ^2 . En lo siguiente, se presenta una nueva clase de procesos espaciales de Cox, dirigidos por una log-intensidad aleatoria infinito-dimensional. Se considera la definición en sentido débil de los valores funcionales de la log-intensidad. Nos referimos al modelo de campo aleatorio de log-intensidad resultante generalizada como proceso de log-riesgo generalizado. Las distribuciones de probabilidad marginal condicionada de la medida aleatoria de recuento asociada tienen una distribución de Poisson. Como es habitual, en el caso log-Gaussiano, la estructura de correlación funcional espacial de la log-intensidad aleatoria caracteriza los momentos de orden superior.

Bajo la estacionariedad en el espacio, en este capítulo se adopta un marco paramétrico en el dominio espectral funcional espacial. En el contexto de la metodología de estimación de Whittle, los estadísticos funcionales de segundo orden se aproximan en términos de un funcional espectral, basado en el periodograma y los operadores de densidad espectral. La función de pérdida mide la proximidad de la estructura de correlación funcional espacial empírica de los datos, reflejada por el operador periodograma, y los elementos de la familia de operadores de densidad espectral paramétrica contrastados. Se demuestra la consistencia en sentido fuerte del estimador paramétrico formulado de la densidad del producto de segundo orden.

Para ilustrar los resultados, se realiza un estudio de simulación en el mar-

co de un proceso espacial autorregresivo hilbertiano, proceso, $SAR_{\mathcal{H}}(1)$, véase Ruiz-Medina (2011). Aplicamos nuestro enfoque a la estimación paramétrica de los autovalores y autovectores de los operadores de autocorrelación implicados. Estos estimadores paramétricos permiten la aproximación del operador de densidad espectral. Se ilustra la consistencia en sentido fuerte y en media-cuadrada. En particular, en el primer ejemplo numérico, se estima el hiperparámetro que define el soporte de los autovectores de los operadores de autocorrelación implicados. El problema de estimar los hiperparámetros de escala y localización, que caracterizan los autovalores de los operadores de autocorrelación, se aborda en un segundo ejemplo numérico. La metodología de estimación también se implementa a partir de datos reales. En concreto, analizamos la mortalidad por enfermedades respiratorias en 48 provincias españolas durante el periodo 1980-2015. El Instituto Nacional de Estadística de España proporcionó los datos, que consisten en 432 registros mensuales de mortalidad por enfermedades respiratorias en las 48 provincias españolas.

El esquema del capítulo es el siguiente. En la Sección 4.1 se introduce la clase de procesos espaciales de Cox, dirigidos por una log-intensidad espacial Hilbert-valuada. En la Sección 4.2 se obtiene la estimación paramétrica fuertemente consistente de su estructura de correlación funcional espacial en el dominio espectral. Un estudio de simulación ilustra en la Sección 4.3 las propiedades de la metodología de estimación espacial funcional introducida. Este enfoque se valida en la Sección 4.4, utilizando datos reales. El capítulo termina con una discusión final.

Los resultados desarrollados en este capítulo se recogen en el artículo titulado: “Spatial Cox Processes in an Infinite-Dimensional Framework”, actualmente sometido (arXiv:1811.11139), cuyos co-autores son M. P. Frías, A. Torres-Signes, M. D. Ruiz-Medina, J. Mateu, (Frías et al., 2020).

4.1. Procesos de Cox dirigidos por una log-intensidad aleatoria \mathcal{H} -valuada espacial

Como en los capítulos anteriores, (Ω, \mathcal{A}, P) denota el espacio de probabilidad básico sobre el que se definen las variables aleatorias en este capítulo. En el desarrollo posterior, se omite la dependencia de $\omega \in \Omega$ de todas las aplicaciones

medibles, desde (Ω, \mathcal{A}, P) a la recta real o compleja, así como desde (Ω, \mathcal{A}, P) a un espacio de funciones. Sólo reflejamos dicha dependencia la primera vez, cuando se introducen distribuciones de probabilidad condicionadas a la observación de valores muestrales. Normalmente, el subíndice denota la localización del elemento de una familia de operadores o variables aleatorias dada, y el argumento se refiere al elemento ubicado en el conjunto que define su soporte.

Sea \mathcal{H} un espacio de Hilbert real separable de funciones, denotamos por $\mathcal{H} + i\mathcal{H}$ su versión compleja. En lo que sigue, $\langle \cdot, \cdot \rangle$ y $\| \cdot \|$ denotan respectivamente el producto interno y la norma en el espacio $\mathcal{H} + i\mathcal{H}$. La misma notación se utilizará para el producto interno y la norma de \mathcal{H} , considerado como un subespacio de $\mathcal{H} + i\mathcal{H}$. A efectos prácticos nos referimos a $\mathcal{H} = L^2(\mathcal{T})$, como el espacio de las funciones de cuadrado integrable en el intervalo de tiempo \mathcal{T} . En esta sección B representa un conjunto acotado de Borel $B \in \mathcal{B}^d$.

Se denota por $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$ a un proceso espacial con valores en \mathcal{H} . Al aplicar el Teorema de representación de Riesz, podemos definir un funcional aleatorio $X_{\mathbf{z}}$, sobre \mathcal{H} utilizando la identidad

$$X_{\mathbf{z}}(\varphi) = \langle \varkappa_{\mathbf{z}}, \varphi \rangle, \quad \forall \varphi \in \mathcal{H}, \quad \mathbf{z} \in \mathbb{R}^d.$$

Desde un punto de vista técnico, esta definición ayuda a introducir diferentes conceptos relacionados con la distribución de probabilidad, y la extrapolación espacial de variables aleatorias \mathcal{H} -valuadas.

Nota 4.1. Obsérvese que bajo la suposición

$$\sum_{p=1}^{\infty} E |\langle \varkappa_{\mathbf{z}}, \phi_p \rangle|^2 = \sum_{p=1}^{\infty} E |X_{\mathbf{z}}(\phi_p)|^2 < \infty, \quad \mathbf{z} \in \mathbb{R}^d, \quad (4.1)$$

para cualquier base ortonormal $\{\phi_p, p \geq 1\}$ de \mathcal{H} , $\varkappa_{\mathbf{z}}$ es un elemento aleatorio en \mathcal{H} , es decir, $P[\varkappa_{\mathbf{z}} \in \mathcal{H}] = 1$, para cualquier $\mathbf{z} \in \mathbb{R}^d$, véase Ledoux y Talagrand (1991). Por lo tanto, a lo largo de este capítulo aparecen tanto identidades en sentido fuerte (puntuales) como en sentido débil (funcionales sobre \mathcal{H}). Por ejemplo, en la ecuación (4.2) siguiente, se considera la definición puntual de $\varkappa_{\mathbf{z}} \in \mathcal{H}$, mientras que en la ecuación (4.3), se aplica su definición en sentido débil.

Consideremos una intensidad aleatoria funcional espacial $\Lambda = \{\Lambda_{\mathbf{z}}(\cdot), \mathbf{z} \in$

\mathbb{R}^d definida como

$$\Lambda_{\mathbf{z}}(t) = \exp(\varkappa_{\mathbf{z}}(t)) = \sum_{k=1}^{\infty} \frac{[\varkappa_{\mathbf{z}}(t)]^k}{k!}, \quad \forall t \in \mathcal{T}, \quad \mathbf{z} \in \mathbb{R}^d, \quad (4.2)$$

donde, $\varkappa_{\mathbf{z}}(t)$ denota el valor puntual de la aplicación $t \rightarrow \varkappa_{\mathbf{z}}(t)$, $t \in \mathcal{T}$, ya que $\varkappa_{\mathbf{z}} \in \mathcal{H} = L^2(\mathcal{T})$, casi seguro. Obsérvese que $\varkappa_{\mathbf{z}} : (\Omega, \mathcal{A}, P) \rightarrow \mathcal{H}$ define una función medible, para cualquier $\mathbf{z} \in \mathbb{R}^d$. Por construcción, para un $\omega \in \Omega$, $\ln(\lambda_{\mathbf{z}})(t) := \varkappa_{\mathbf{z}}(\omega, t) = \mathcal{X}_{\mathbf{z}}(t)$, para cada $t \in \mathcal{T}$, y $\mathbf{z} \in \mathbb{R}^d$. Las realizaciones $\lambda = \{\lambda_{\mathbf{z}}(\cdot), \mathbf{z} \in \mathbb{R}^d\}$ de $\Lambda = \{\Lambda_{\mathbf{z}}(\cdot), \mathbf{z} \in \mathbb{R}^d\}$ son, entonces, casi seguramente (c. s.) positivas. Como anteriormente se ha indicado, omitimos la dependencia de $\omega \in \Omega$ de las realizaciones $\{\varkappa_{\mathbf{z}}(\omega, \cdot), \mathbf{z} \in \mathbb{R}^d, \omega \in \Omega\}$ de la log-intensidad espacial \mathcal{H} -valuada $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$.

Se asume la siguiente suposición.

Suposición 4.1. Para cualquier $B \in \mathcal{B}^d$,

$$\int_B \exp(X_{\mathbf{z}}(\varphi)) d\mathbf{z} = \int_B \exp(\langle \varkappa_{\mathbf{z}}, \varphi \rangle) d\mathbf{z} < \infty, \quad \forall \varphi \in \mathcal{H}, \quad (4.3)$$

en la norma del espacio $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, dado por $\|X\|_{\mathcal{L}^2(\Omega, \mathcal{A}, P)}^2 = E[X^2]$, para cualquier variable aleatoria de segundo orden con media cero X en (Ω, \mathcal{A}, P) .

Nota 4.2. De la Suposición 4.1, la familia de procesos de Cox introducida tiene momentos de segundo orden finitos.

La ecuación (4.3) es válida si el proceso $\{\varkappa_{\mathbf{z}}(t), t \in \mathcal{T}, \mathbf{z} \in \mathbb{R}^d\}$ es continuo en media cuadrática con respecto a ambos argumentos $t \in \mathcal{T}$, y $\mathbf{z} \in \mathbb{R}^d$, lo que equivale a suponer que el kernel $r_{\mathbf{z}, \mathbf{y}}(t, s) = E[\varkappa_{\mathbf{z}}(t)\varkappa_{\mathbf{y}}(s)]$, que define el operador de covarianza $\mathcal{R}_{\mathbf{z}, \mathbf{y}} = E[\varkappa_{\mathbf{z}} \otimes \varkappa_{\mathbf{y}}]$ es continuo en $t, s \in \mathcal{T}$; $\mathbf{z}, \mathbf{y} \in \mathbb{R}^d$. La expresión $\varkappa_{\mathbf{z}} \otimes \varkappa_{\mathbf{y}}$ denota el producto tensorial de dos elementos aleatorios en \mathcal{H} . Por lo tanto, dicha expresión define un operador de Hilbert-Schmidt en \mathcal{H} casi seguramente.

Consideramos ahora $N : (\Omega, \mathcal{A}, \mathcal{P}) \times \mathcal{B}^d \rightarrow \mathbb{N}$, una medida de recuento aleatoria sobre la σ -álgebra de Borel \mathcal{B}^d de \mathbb{R}^d . Es decir, para cada $\omega \in \Omega$, fijo, $N(\omega, B)$ con $B \in \mathcal{B}^d$, define una medida de recuento sobre los conjuntos de \mathcal{B}^d , mientras que, para cada conjunto $B \in \mathcal{B}^d$, $N(\omega, B)$ con $\omega \in \Omega$, define una variable aleatoria de valores enteros sobre (Ω, \mathcal{A}, P) . Suponemos que, para cualquier $\varphi \in \mathcal{H}$, dada la realización espacial $x(\varphi) = \{x_{\mathbf{z}}(\varphi) = \langle \mathcal{X}_{\mathbf{z}}, \varphi \rangle, \mathbf{z} \in \mathbb{R}^d\}$ de $X(\varphi)$, definida a partir de las curvas muestrales $\mathcal{X}_{\mathbf{z}} \in \mathcal{H}$, espacialmente

distribuidas sobre $\mathbf{z} \in \mathbb{R}^d$, la distribución de probabilidad condicionada del número de sucesos aleatorios $N(B)$ que ocurren en un conjunto B , sigue una distribución de probabilidad de Poisson de media $\int_B \exp(x_{\mathbf{z}}(\varphi)) d\mathbf{z}$, para cada conjunto B .

Nota 4.3. Se considera la terminología función de contraste en un sentido amplio, a partir del Teorema de representación de Riesz. Obsérvese que, por lo general, las técnicas de FDA se aplican tras la interpolación y suavizado de los datos. Alternativamente, cuando se analizan sistemas altamente singulares, el espacio $C_0^\infty(\mathcal{T})$ de funciones infinitamente diferenciables con soporte compacto en \mathcal{T} se suele considerar como espacio de la función de contraste, permitiendo la regularización del sistema.

4.1.1. Predicción por mínimos cuadrados y momentos condicionados marginales de segundo orden

Se introducen brevemente los momentos de primer y segundo orden de las distribuciones de probabilidad condicionadas marginales de nuestra medida aleatoria $\{N(B); |B| < \infty, B \in \mathcal{B}^d\}$, donde $|B| = \int_B d\mathbf{z}$.

Para $B \in \mathcal{B}^d$ y $\varphi \in \mathcal{H}$, a partir de la distribución condicionada de Poisson, $N(B)$ dado $x_{\mathbf{z}}(\varphi)$, $\mathbf{z} \in \mathbb{R}^d$, se obtiene

$$\begin{aligned} f_{N(B)/x_{\mathbf{z}}(\varphi), \mathbf{z} \in \mathbb{R}^d}(t) &:= E [\exp(tN(B)) / x_{\mathbf{z}}(\varphi), \mathbf{z} \in \mathbb{R}^d] \\ &= \exp \left(\left[\int_B \exp(x_{\mathbf{z}}(\varphi)) d\mathbf{z} \right] (\exp(t) - 1) \right), \quad t \in \mathbb{R}, \quad \forall \varphi \in \mathcal{H}. \end{aligned} \quad (4.4)$$

Como es habitual, el predictor mínimo cuadrático $\widehat{N(B)}$ de $N(B)$, se calcula a partir de la primera derivada de (4.4) en $t = 0$. Por lo tanto,

$$\widehat{N(B)} = E [N(B) / x_{\mathbf{z}}(\varphi), \mathbf{z} \in \mathbb{R}^d] = \int_B \exp(x_{\mathbf{z}}(\varphi)) d\mathbf{z}, \quad (4.5)$$

que coincide con la varianza condicionada. La correspondiente fórmula de descomposición de la varianza se calcula a partir de las siguientes identidades,

obtenidas de la segunda derivada de (4.4) en $t = 0$ aplicando (4.5)

$$\begin{aligned} E [\text{Var} (N(B)/X_{\mathbf{z}}(\varphi), \mathbf{z} \in \mathbb{R}^d)] &= \int_B E [\exp (X_{\mathbf{z}}(\varphi))] d\mathbf{z} \\ \text{Var} (E [N(B)/X_{\mathbf{z}}(\varphi), \mathbf{z} \in \mathbb{R}^d]) &= \int_{B \times B} E [\exp (X_{\mathbf{z}}(\varphi) + X_{\mathbf{y}}(\varphi))] d\mathbf{z}d\mathbf{y} \\ &\quad - \int_{B \times B} E [\exp (X_{\mathbf{z}}(\varphi))] E [\exp (X_{\mathbf{y}}(\varphi))] d\mathbf{z}d\mathbf{y}, \end{aligned} \quad (4.6)$$

donde, a partir de (4.5)

$$E_{\varphi} [N(B)] = \int_B E [\exp (X_{\mathbf{z}}(\varphi))] d\mathbf{z}, \quad (4.7)$$

para $B \in \mathcal{B}^d$ y para cualquier $\varphi \in \mathcal{H}$. Así, a partir de (4.6) y (4.7), obtenemos

$$\begin{aligned} \text{Var}_{\varphi} (N(B)) &= \int_B E [\exp (X_{\mathbf{z}}(\varphi))] d\mathbf{z} \\ &\quad + \int_{B \times B} E [\exp (X_{\mathbf{y}}(\varphi) + X_{\mathbf{z}}(\varphi))] d\mathbf{y}d\mathbf{z} \\ &\quad - \int_{B \times B} E [\exp (X_{\mathbf{y}}(\varphi))] E [\exp (X_{\mathbf{z}}(\varphi))] d\mathbf{y}d\mathbf{z}. \end{aligned} \quad (4.8)$$

En el caso estacionario espacial log-gaussiano, las ecuaciones (4.7)–(4.8) se pueden reescribir, en términos de los momentos de segundo orden de una log-intensidad \mathcal{H} -valuada espacial de media cero, $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$. Para cada $\varphi \in \mathcal{H}$ y conjunto $B \in \mathcal{B}^d$,

$$\begin{aligned} E_{\varphi} [N(B)] &= \exp \left(\frac{\mathcal{R}_0(\varphi)(\varphi)}{2} \right) |B| \\ \text{Var}_{\varphi} (N(B)) &= \exp (\mathcal{R}_0(\varphi)(\varphi)) \int_{B \times B} \exp \left(\frac{\mathcal{R}_{\mathbf{z}-\mathbf{y}}(\varphi)(\varphi) + \mathcal{R}_{\mathbf{y}-\mathbf{z}}(\varphi)(\varphi)}{2} \right) d\mathbf{z}d\mathbf{y} \\ &\quad + |B| \exp \left(\frac{\mathcal{R}_0(\varphi)(\varphi)}{2} \right) \left[1 - |B| \exp \left(\frac{\mathcal{R}_0(\varphi)(\varphi)}{2} \right) \right], \end{aligned} \quad (4.9)$$

donde, como antes, $|B|$ denota la medida de Lebesgue de B , y $\mathcal{R}_{\mathbf{z}-\mathbf{y}}$ es el operador de covarianza espacial de $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$ dado por

$$\begin{aligned} \mathcal{R}_{\mathbf{z}-\mathbf{y}}(f)(g) &= E [X_{\mathbf{z}}(g)X_{\mathbf{y}}(f)] = E [\langle \varkappa_{\mathbf{z}}, g \rangle \langle \varkappa_{\mathbf{y}}, f \rangle] \\ &= E \langle X_{\mathbf{z}} \otimes X_{\mathbf{y}}(f), g \rangle, \quad \forall f, g \in \mathcal{H}, \quad \mathbf{z}, \mathbf{y} \in \mathbb{R}^d. \end{aligned} \quad (4.10)$$

La densidad producto de orden n se calcula entonces a partir de la intensidad y la densidad producto de segundo orden como sigue: para $\varphi \in \mathcal{H}$,

$$\begin{aligned} \rho_\varphi^{(n)}(\mathbf{z}_1, \dots, \mathbf{z}_n) &= E \left[\prod_{i=1}^n \exp(x_{\mathbf{z}_i}(\varphi)) \right] = E \left[\exp \left(\sum_{i=1}^n x_{\mathbf{z}_i}(\varphi) \right) \right] \\ &= [\rho_\varphi]^n \exp \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \mathcal{R}_{\mathbf{z}_i - \mathbf{z}_j}(\varphi)(\varphi) \right), \quad \mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{R}^d, \end{aligned} \quad (4.11)$$

donde la intensidad funcional espacial ρ_φ se define como

$$\rho_\varphi = \rho_\varphi^{(1)}(\mathbf{z}) = \exp \left(\frac{\mathcal{R}_0(\varphi)(\varphi)}{2} \right), \quad \forall \mathbf{z} \in \mathbb{R}^d, \quad \varphi \in \mathcal{H}. \quad (4.12)$$

En el caso isotrópico, para $i, j = 1, \dots, n$, el funcional de correlación par espacial se define a partir de la identidad

$$g_\varphi(\mathbf{z}_j - \mathbf{z}_i) = \frac{\rho_\varphi^{(2)}(\mathbf{z}_i - \mathbf{z}_j)}{[\rho_\varphi]^2} = \exp(\mathcal{R}_{\mathbf{z}_i - \mathbf{z}_j}(\varphi)(\varphi)), \quad \forall \varphi \in \mathcal{H}. \quad (4.13)$$

4.2. Estimación de la estructura de correlación espacial de la log-intensidad infinito-dimensional

En esta sección, asumiendo estacionariedad en el espacio, se propone un estimador paramétrico fuertemente consistente, basado en el operador periodograma, inspirado en el funcional de Whittle, para el operador de densidad espectral (véase, por ejemplo, Panaretos y Tavakoli (2013b), en el marco no paramétrico). Restringimos nuestra atención al caso $\{\nu_{\mathbf{z}}, \mathbf{z} \in \mathbb{Z}^d\}$, es decir, la log-intensidad \mathcal{H} -valuada se observa en una rejilla espacial regular. Desarrollamos las herramientas de inferencia funcional espacial en el dominio espectral, considerando el intervalo $[-\pi, \pi]^d$, como el conjunto de índices de nuestra familia de operadores de densidad espectral $\{\mathcal{F}_\xi, \xi \in [-\pi, \pi]^d\}$, dada por

$$\mathcal{F}_\xi(h)(g) = \frac{1}{(2\pi)^d} \sum_{\mathbf{z} \in \mathbb{Z}^d} \exp \left(-i \sum_{j=1}^d \xi_j z_j \right) \mathcal{R}_{\mathbf{z}}(h)(g), \quad \xi \in [-\pi, \pi]^d, \quad (4.14)$$

para todo $h, g \in \mathcal{H} + i\mathcal{H}$. Asumimos que la serie $\frac{1}{(2\pi)^d} \sum_{\mathbf{z} \in \mathbb{Z}^d} \exp\left(-i \sum_{j=1}^d \xi_j z_j\right) \mathcal{R}_{\mathbf{z}}$ converge en la norma de la traza, véase, por ejemplo, Panaretos y Tavakoli (2013b). Como antes, $\mathcal{H} + i\mathcal{H}$ denota la versión complejo-valuada del espacio de Hilbert real separable \mathcal{H} , y el operador de covarianza $\mathcal{R}_{\mathbf{z}}$ se ha introducido en la ecuación (4.10). Asumimos la siguiente condición

Suposición 4.2. Para cualquier $\boldsymbol{\xi} \in [-\pi, \pi]^d$, $\mathcal{F}_{\boldsymbol{\xi}}$ es un operador traza en $\mathcal{H} + i\mathcal{H}$. Consideremos $\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{Z}^d$, con $z_i \in [-\mathcal{N}_i + 1, \mathcal{N}_i - 1]$, $i = 1, \dots, d$, y $\mathcal{N} = \mathcal{N}_1 \times \dots \times \mathcal{N}_d$. Para una base ortonormal $\{\psi_k, k \geq 1\}$ de \mathcal{H} , se asume que, cuando $\mathcal{N} \rightarrow \infty$,

$$\sum_{k, l \geq 1} \left\| \left[\frac{1}{\mathcal{N}} \left[\sum_{y_1=1}^{\mathcal{N}_1 - z_1} \dots \sum_{y_d=1}^{\mathcal{N}_d - z_d} X_{\mathbf{y}}(\psi_k) X_{\mathbf{y}+\mathbf{z}}(\psi_l) \right] - \mathcal{R}_{\mathbf{z}}(\psi_k)(\psi_l) \right] \right\|^2 \rightarrow c. s. 0,$$

donde se ha utilizado la notación $\mathbf{y} = (y_1, \dots, y_d) \in \prod_{i=1}^d [1, \mathcal{N}_i - z_i] \cap \mathbb{Z}^d$.

Esta condición proporciona la ergodicidad, para los momentos empíricos de orden dos, con respecto a la norma del operador de Hilbert-Schmidt $\|\cdot\|_{S(\mathcal{H})}$ del proceso de log-intensidad funcional espacial $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{Z}^d\}$.

Nota 4.4. Como es habitual, la consistencia fuerte de una variable aleatoria evaluada en \mathcal{H} se define a partir de la convergencia c. s. en la norma $\|\cdot\|_{\mathcal{H}}$. Este hecho se aplica en la Suposición 4.2, considerando el espacio de Hilbert $S(\mathcal{H})$ de los operadores de Hilbert-Schmidt sobre \mathcal{H} . Así, la Suposición 4.2 define la consistencia fuerte de los momentos de segundo orden funcionales empíricos de $\varkappa_{\mathbf{z}}$ con respecto a la norma $S(\mathcal{H})$.

Nota 4.5. Si la norma del operador de Hilbert-Schmidt $\|\mathcal{F}_{\boldsymbol{\xi}}\|_{S(\mathcal{H})}$ de la familia de operadores de densidad espectral $\{\mathcal{F}_{\boldsymbol{\xi}}, \boldsymbol{\xi} \in [-\pi, \pi]^d\}$ es de cuadrado integrable en $[-\pi, \pi]^d$, entonces, la Suposición 4.2 se cumple. En particular, si la familia de operadores de densidad espectral es continua en el dominio de la frecuencia, con respecto a la norma del operador Hilbert-Schmidt, es decir, $\|\mathcal{F}_{\boldsymbol{\xi}} - \mathcal{F}_{\boldsymbol{\xi}_0}\|_{S(\mathcal{H})} \rightarrow 0$, $\boldsymbol{\xi} \rightarrow \boldsymbol{\xi}_0$, entonces, la Suposición 4.2 se cumple.

Consideremos ahora una familia paramétrica de operadores de densidad espectral

$$\{\mathcal{F}_{\boldsymbol{\xi}, \theta}, \boldsymbol{\xi} \in [-\pi, \pi]^d, \theta \in \Theta \subset \mathbb{R}^q\},$$

donde el espacio paramétrico compacto $\Theta \subset \mathbb{R}^q$, $q \geq 1$, es tal que

$$\left\| \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_1} \mathcal{F}_{\xi, \theta_2}^{-1} d\xi \right\|_{\mathcal{L}(\mathcal{H} + i\mathcal{H})} = \left\| \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_2}^{-1} \mathcal{F}_{\xi, \theta_1} d\xi \right\|_{\mathcal{L}(\mathcal{H} + i\mathcal{H})} < \infty, \quad (4.15)$$

para cada $\theta_1, \theta_2 \in \Theta$. Aquí, $\|\cdot\|_{\mathcal{L}(\mathcal{H} + i\mathcal{H})}$ denota la norma en el espacio de los operadores lineales acotados en $\mathcal{H} + i\mathcal{H}$.

Obsérvese que los ejemplos 4.1 y 4.2 de la Sección 4.3 proporcionan un escenario adecuado en el que se satisface la Condición (4.15) en el contexto lineal funcional. En concreto, en el escenario uniparamétrico del Ejemplo 4.1, θ caracteriza el soporte de los autovectores de los operadores de autocorrelación, mientras que en el Ejemplo 4.2, las componentes de θ son parámetros de localización y escala que intervienen en la definición de los autovalores paramétricos. En particular, el enfoque presentado extiende al marco infinito-dimensional, la modelización lineal paramétrica adoptada en Hannan (1973) en el caso real-valorado, donde también se parametrizan los coeficientes de correlación.

Restringiendo nuestra atención al marco paramétrico lineal funcional, se consideran los siguientes cuatro supuestos en la obtención del Teorema 4.1 siguiente. Específicamente, para $\{\varphi_k, k \geq 1\}$, una base ortonormal arbitraria de $\mathcal{H} + i\mathcal{H}$, supongamos las siguientes condiciones:

Condición 4.3. Para $\theta, \theta' \in \Theta$, con $\theta \neq \theta'$, el conjunto $\{\xi \in [-\pi, \pi]^d; \mathcal{F}_{\xi, \theta} \neq \mathcal{F}_{\xi, \theta'}\}$ tiene una medida positiva de Lebesgue. Por lo tanto, valores diferentes de θ corresponden a diferentes estructuras de dependencia funcional espacial.

Condición 4.4. El operador $\mathcal{F}_{\xi, \theta}$ es continuo en $\xi \in [-\pi, \pi]^d$, y $\theta \in \Theta$, con respecto a la norma $\mathcal{L}(\mathcal{H} + i\mathcal{H})$. Es decir, $\|\mathcal{F}_{\xi, \theta} - \mathcal{F}_{\xi^*, \theta^*}\|_{\mathcal{L}(\mathcal{H} + i\mathcal{H})} \rightarrow 0$, $(\xi, \theta) \rightarrow (\xi^*, \theta^*)$, para todo $(\xi^*, \theta^*) \in [-\pi, \pi]^d \times \Theta$.

Condición 4.5. Se asume la sumabilidad espacial del cuadrado de la norma del operador de Hilbert–Schmidt de los operadores lineales, involucrados en la representación funcional de medias móviles de orden infinito, y la normalización en varianza componente a componente de las proyecciones independientes de las innovaciones de ruido blanco. Por lo tanto,

$$\int_{[-\pi, \pi]^d} \log((2\pi)^d \mathcal{F}_{\xi, \theta}(\varphi_k)(\varphi_k)) d\xi = 0, \quad \forall k \geq 1, \theta \in \Theta. \quad (4.16)$$

Condición 4.6. La suma de Cesaro de las series de Fourier de $\mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k)$ converge uniformemente en $\xi \in [-\pi, \pi]^d$, $k \geq 1$, y $\theta \in \Theta$.

Nota 4.6. Como se demuestra en el siguiente Teorema 4.1, las condiciones 4.3–4.6 proporcionan un escenario adecuado para la estimación consistente en sentido fuerte en el dominio espectral funcional espacial del modelo paramétrico adoptado para los autoelementos del operador de correlación espacial. Véanse también los ejemplos 4.1 y 4.2 de la Sección 4.3.

Nota 4.7. La Condición 4.5 implica, en particular, la invertibilidad del modelo lineal en términos de una secuencia de operadores lineales aplicados a un proceso de innovación de ruido blanco generalizado. En este contexto funcional lineal, la Condición 4.5 significa que las innovaciones son curvas independientes e idénticamente distribuidas con un operador de autocovarianza identidad. Por lo tanto, sus proyecciones sobre una base ortonormal de \mathcal{H} tienen varianzas unidad. El comportamiento singular local que muestra este modelo de campo aleatorio de innovaciones generalizado se regulariza mediante la convolución con una secuencia de operadores lineales adecuada (que define la representación funcional de medias móviles de orden infinito de la log-intensidad). En particular, debe asumirse la sumabilidad espacial del cuadrado de la norma del operador de Hilbert–Schmidt de esta secuencia. Esta condición es equivalente a la integrabilidad en el dominio de frecuencias de la norma de la traza del operador de densidad espectral.

Sea $\tilde{\varkappa}_{\xi}^{\mathcal{N}}$ la transformada de Fourier discreta funcional de los datos, dada por la siguiente identidad en la norma de $\mathcal{H} + i\mathcal{H}$,

$$\tilde{\varkappa}_{\xi}^{\mathcal{N}} = \frac{1}{\sqrt{\mathcal{N}(2\pi)^d}} \sum_{y_1=1}^{\mathcal{N}_1} \cdots \sum_{y_d=1}^{\mathcal{N}_d} \exp\left(-i \sum_{j=1}^d \xi_j y_j\right) \varkappa_{\mathbf{y}}, \quad \xi \in [-\pi, \pi]^d, \quad \mathcal{N} = \prod_{i=1}^d \mathcal{N}_i, \quad (4.17)$$

donde hemos aplicado que $\tilde{\varkappa}_{\xi}^{\mathcal{N}} \in \mathcal{H} + i\mathcal{H}$ casi seguramente, para cualquier $\xi \in [-\pi, \pi]^d$, dada la propiedad de la traza del operador de autocovarianza de $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$, véase la ecuación (4.1) en la Nota 4.1. El operador periodograma

se define entonces como el operador empírico

$$\begin{aligned} \mathcal{I}_{\xi}^{\mathcal{N}}(g)(h) &= \frac{1}{\mathcal{N}(2\pi^d)} \sum_{y_1=1}^{\mathcal{N}_1} \cdots \sum_{y_d=1}^{\mathcal{N}_d} \sum_{z_1=1}^{\mathcal{N}_1} \cdots \sum_{z_d=1}^{\mathcal{N}_d} X_{y_1, \dots, y_d}^{\mathcal{N}}(h) \\ &\quad \times X_{z_1, \dots, z_d}^{\mathcal{N}}(g) \exp\left(-i \sum_{j=1}^d \xi_j (y_j - z_j)\right), \end{aligned} \quad (4.18)$$

para $g, h \in \mathcal{H} + i\mathcal{H}$, y $\xi \in [-\pi, \pi]^d$. Obsérvese que a partir de (4.17), la propiedad de traza de $\mathcal{I}_{\xi}^{\mathcal{N}}$ se deduce directamente de la identidad de Parseval en $\mathcal{H} + i\mathcal{H}$, teniendo en cuenta que $\mathcal{I}_{\xi}^{\mathcal{N}}$ es no negativo.

Bajo la Suposición 4.1, y las condiciones formuladas anteriormente, se considera el siguiente estimador paramétrico, basado en una muestra funcional de tamaño \mathcal{N} ,

$$\hat{\theta}_{\mathcal{N}} = \arg \min_{\theta \in \Theta} [\sigma_{\mathcal{N}}(\theta)], \quad (4.19)$$

donde $\sigma_{\mathcal{N}}(\theta)$ denota la función de pérdida empírica

$$\begin{aligned} \sigma_{\mathcal{N}}(\theta) &= \sup_{k \geq 1} \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} |\mathcal{F}_{\xi, \theta}^{-1} \mathcal{I}_{\xi}^{\mathcal{N}}(\varphi_k)(\varphi_k)| d\xi \\ &= \sup_{k \geq 1} \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} |\mathcal{I}_{\xi}^{\mathcal{N}} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k)| d\xi, \quad \forall \theta \in \Theta, \end{aligned} \quad (4.20)$$

cuyas propiedades asintóticas serán analizadas en el siguiente Teorema 4.1.

Consideramos el correspondiente estimador paramétrico del operador de covarianza espacial se define como sigue:

$$\hat{\mathcal{R}}_{\mathbf{z}, \theta}(h)(g) = \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \exp\left(i \sum_{j=1}^d z_j \xi_j\right) \mathcal{F}_{\xi, \hat{\theta}_{\mathcal{N}}}(h)(g) d\xi, \quad (4.21)$$

para $h, g \in \mathcal{H}$, y $\mathbf{z} \in \mathbb{Z}^d$. Además, en el caso isotrópico log-gaussiano, se obtiene el siguiente estimador paramétrico de la función de correlación par

$$\begin{aligned} \widehat{g}_{\varphi, \theta}(\mathbf{z}) &= \exp\left(\widehat{\mathcal{R}}_{\mathbf{z}, \theta}(\varphi)(\varphi)\right) \\ &= \exp\left(\mathcal{R}_{\mathbf{z}, \hat{\theta}_{\mathcal{N}}}(\varphi)(\varphi)\right), \quad \forall \varphi \in \mathcal{H}, \mathbf{z} \in \mathbb{R}^d. \end{aligned}$$

En este sentido, y aunque el siguiente resultado se obtiene más allá del supuesto log-Gaussiano, recordemos que, bajo el escenario log-Gaussiano de Cox la consistencia fuerte demostrada en el Teorema 4.1 también se traslada al estimador paramétrico de la densidad producto

$$\rho_{\varphi, \hat{\theta}_N}^{(n)}(\mathbf{z}_1, \dots, \mathbf{z}_n) = [\rho_{\varphi, \hat{\theta}_N}]^n \exp\left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \mathcal{R}_{\mathbf{z}_i - \mathbf{z}_j, \hat{\theta}_N}(\varphi)(\varphi)\right),$$

para $\mathbf{z}_1, \dots, \mathbf{z}_n \in \mathbb{Z}^d$, y $\varphi \in \mathcal{H}$, con

$$\rho_{\varphi, \hat{\theta}_N} = \widehat{\rho_{\varphi, \theta}} = \widehat{\rho_{\varphi, \theta}^{(1)}}(\mathbf{z}) = \exp\left(\frac{\mathcal{R}_{0, \hat{\theta}_N}(\varphi)(\varphi)}{2}\right), \quad \forall \mathbf{z} \in \mathbb{Z}^d. \quad (4.22)$$

Teorema 4.1. *Sea $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{Z}^d\}$ un campo aleatorio de log-intensidad lineal \mathcal{H} -valuado espacial estacionario de media cero. Denotamos $\theta_0 \in \text{Int}(\Theta)$ al verdadero valor del parámetro, que se encuentra en el interior $\text{Int}(\Theta)$ de Θ . Bajo las suposiciones 4.1–4.2, y las condiciones (4.15) y 4.3–4.6, el estimador (4.19)–(4.20) satisface*

$$\hat{\theta}_N \rightarrow_{\text{c. s.}} \theta_0, \quad \mathcal{N} \rightarrow \infty, \quad (4.23)$$

donde $\rightarrow_{\text{c. s.}}$ significa convergencia casi segura.

Demostración:

Consideramos aquí una versión compacta de la demostración. Algunos detalles técnicos adicionales se pueden encontrar en la Nota 4.8. En concreto, la demostración de este resultado se desprende de la hipótesis 4.2 que implica, para cualquier base ortonormal, $\{\varphi_k, k \geq 1\}$ de $\mathcal{H} + i\mathcal{H}$, y para $\mathbf{z} = (z_1, \dots, z_d) \in \mathbb{Z}^d$ con $z_i \in [-\mathcal{N}_i + 1, \mathcal{N}_i - 1]$, $i = 1, \dots, d$,

$$\sum_{k, l \geq 1} |\mathcal{C}(\mathbf{z}, k, l) - \mathcal{R}_{\mathbf{z}, \theta_0}(\varphi_k)(\varphi_l)|^2 \rightarrow_{\text{c. s.}} 0, \quad \mathcal{N} \rightarrow \infty, \quad (4.24)$$

donde los coeficientes de Fourier empíricos $\mathcal{C}(\mathbf{z}, k, l)$ del operador periodograma

vienen dados por

$$\begin{aligned} \mathcal{C}(\mathbf{z}, k, l) &= \int_{[-\pi, \pi]^d} \exp\left(i \sum_{j=1}^d \xi_j z_j\right) \tilde{X}_{\xi}^{\mathcal{N}}(\varphi_k) \tilde{X}_{-\xi}^{\mathcal{N}}(\varphi_l) d\xi \\ &= \frac{1}{\mathcal{N}} \left[\sum_{y_1=1}^{\mathcal{N}_1-z_1} \cdots \sum_{y_d=1}^{\mathcal{N}_d-z_d} X_{\mathbf{y}}(\varphi_k) X_{\mathbf{y}+\mathbf{z}}(\varphi_l) \right], \quad |z_i| < \mathcal{N}_i, \quad i \in \{1, \dots, d\}, \\ \mathcal{C}(\mathbf{z}, k, l) &= 0, \quad |z_i| \geq \mathcal{N}_i, \quad \text{para algún } i \in \{1, \dots, d\}, \end{aligned} \quad (4.25)$$

con $\mathbf{y} = (y_1, \dots, y_d) \in \prod_{i=1}^d [1, \mathcal{N}_i - z_i] \cap \mathbb{Z}^d$. Aquí, $\tilde{X}_{\xi}^{\mathcal{N}}(\varphi) = \langle \tilde{z}_{\xi}^{\mathcal{N}}, \varphi \rangle$, para cada $\varphi \in \mathcal{H} + i\mathcal{H}$.

Bajo las condiciones 4.5 y 4.6, cuando $M \rightarrow \infty$, con $M = \prod_{j=1}^d M_j$, se tiene la convergencia de

$$\begin{aligned} q_{\xi, \theta}^M(\varphi_k)(\varphi_k) &= \frac{1}{(2\pi)^d} \sum_{z_j \in [-M_j+1, M_j-1]; j=1, \dots, d} \exp\left(-i \sum_{j=1}^d \xi_j z_j\right) \\ &\quad \times \prod_{j=1}^d \left(1 - \frac{z_j}{M_j}\right) g(z_1, \dots, z_d, \theta, k) \\ &= \int_{[-\pi, \pi]^d} F_M(\xi - \varpi) \mathcal{F}_{\varpi, \theta}^{-1}(\varphi_k)(\varphi_k) d\varpi, \end{aligned} \quad (4.26)$$

a $\mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k)$, uniformemente en $k \geq 1$, $\xi \in [-\pi, \pi]^d$, y $\theta \in \Theta$. Aquí, para cada $\mathbf{z} = (z_1, \dots, z_d) \in \prod_{j=1}^d [-M_j + 1, M_j - 1]$, $g(\mathbf{z}, \theta, k) = g(z_1, \dots, z_d, \theta, k)$ denota el \mathbf{z} -ésimo coeficiente de Fourier de $\mathcal{F}_{\varpi, \theta}^{-1}(\varphi_k)(\varphi_k)$, véase la ecuación (4.34) posterior.

Para un M suficientemente grande, aplicando la desigualdad de Hölder, a partir de la Suposición 4.2 y las condiciones (4.3)–(4.6) se obtiene la convergencia c. s. a cero, cuando $\mathcal{N} \rightarrow \infty$, para cualquier $\theta \in \Theta$, de

$$\begin{aligned} &\lim_{\mathcal{N} \rightarrow \infty} \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| q_{\xi, \theta}^M \left[\mathcal{I}_{\xi}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k) \right] \right| d\xi \\ &\leq_{c.s.} \lim_{\mathcal{N} \rightarrow \infty} \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| q_{\xi, \theta}^M \left[\mathcal{I}_{\xi}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k) \right] \right|^2 d\xi \\ &= \lim_{\mathcal{N} \rightarrow \infty} \sup_{k \geq 1} \frac{(2\pi)^d}{\mathcal{N}} \sum_{\mathbf{z}} \left| q_{\xi, \theta}^M \mathcal{I}_{\xi, \mathbf{z}}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - q_{\xi, \theta}^M \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k) \right|^2 \stackrel{c.s.}{=} 0, \end{aligned} \quad (4.27)$$

véase Hannan (1973) y la Nota 4.8 para más detalles.

Aplicando la desigualdad triangular, bajo las condiciones (4.5)–(4.6),

$$\begin{aligned}
 & \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| \mathcal{I}_{\xi}^{(\mathcal{N})} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) \right| d\xi \\
 & \leq_{c.s.} \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| \mathcal{I}_{\xi}^{(\mathcal{N})} [\mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) - q_{\xi, \theta}^M(\varphi_k)(\varphi_k)] \right| d\xi \\
 & \quad + \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| q_{\xi, \theta}^M [\mathcal{I}_{\xi}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k)] \right| d\xi \\
 & \quad + \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| \mathcal{F}_{\xi, \theta_0} [q_{\xi, \theta}^M(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k)] \right| d\xi \\
 & \leq_{c.s.} [2\pi]^d \varepsilon(M) \sup_{k \geq 1} \mathcal{C}(0, k, k) \\
 & \quad + \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| q_{\xi, \theta}^M [\mathcal{I}_{\xi}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k)] \right| d\xi \\
 & \quad + \varepsilon(M) \sup_{k \geq 1} \mathcal{R}_{0, \theta_0}(\varphi_k)(\varphi_k) \\
 & = S_1(\mathcal{N}, M) + S_2(\mathcal{N}, M) + S_3(M), \quad \theta \in \Theta. \tag{4.28}
 \end{aligned}$$

donde el parámetro M denota, como antes, el número de términos implicados en la suma parcial de Cesaro de la serie de Fourier de $\mathcal{F}_{\xi, \theta}^{-1}$, y $\varepsilon(M)$ es el término residual correspondiente. Obsérvese que, bajo la Condición 4.6, la serie de Fourier de $\mathcal{F}_{\xi, \theta}^{-1}$ es sumable por Cesaro en norma $\mathcal{L}(\mathcal{H} + i\mathcal{H})$, uniformemente en $(\xi, \theta) \in [-\pi, \pi]^d \times \Theta$.

Cuando $\mathcal{N} \rightarrow \infty$, aplicando la convergencia c. s. a cero de la ecuación (4.27),

$$S_2(\mathcal{N}, M) = \sup_{k \geq 1} \int_{[-\pi, \pi]^d} \left| q_{\xi, \theta}^M [\mathcal{I}_{\xi}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k)] \right| d\xi \rightarrow_{c.s.} 0.$$

La sumabilidad de Cesaro de la serie de Fourier de $\mathcal{F}_{\xi, \theta}^{-1}$ en norma $\mathcal{L}(\mathcal{H} + i\mathcal{H})$, uniformemente en $(\xi, \theta) \in [-\pi, \pi]^d \times \Theta$ en la Condición 4.6 implica que, cuando $M \rightarrow \infty$, $\varepsilon(M) \rightarrow 0$. Finalmente, bajo la Suposición 4.2, $\sup_{k \geq 1} \mathcal{C}(0, k, k) \rightarrow$

$\sup_{k \geq 1} \mathcal{R}_{\mathbf{0}, \theta_0}(\varphi_k)(\varphi_k)$, cuando $\mathcal{N} \rightarrow \infty$. Por lo tanto, se obtiene

$$\begin{aligned} S_1(\mathcal{N}, M) &\rightarrow_{c.s.} [2\pi]^d \varepsilon(M) \sup_{k \geq 1} \mathcal{R}_{\mathbf{0}, \theta_0}(\varphi_k)(\varphi_k), \quad \mathcal{N} \rightarrow \infty \\ \text{y } \varepsilon(M) \sup_{k \geq 1} \mathcal{R}_{\mathbf{0}, \theta_0}(\varphi_k)(\varphi_k) &\rightarrow 0, \quad M \rightarrow \infty \\ S_2(\mathcal{N}, M) &\rightarrow_{c.s.} 0 \quad \mathcal{N} \rightarrow \infty \\ S_3(M) = \varepsilon(M) \sup_{k \geq 1} \mathcal{R}_{\mathbf{0}, \theta_0}(\varphi_k)(\varphi_k) &\rightarrow 0, \quad S_3(M) \rightarrow 0, \quad M \rightarrow \infty. \end{aligned} \quad (4.29)$$

A partir de las ecuaciones (4.28)–(4.29), se obtiene, para $\theta \in \Theta$,

$$\begin{aligned} \sigma_{\mathcal{N}}(\theta) &\rightarrow_{c. s.} \sup_{k \geq 1} \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) d\xi \\ &= \sup_{k \geq 1} \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta}^{-1} \mathcal{F}_{\xi, \theta_0}(\varphi_k)(\varphi_k) d\xi, \quad \mathcal{N} \rightarrow \infty. \end{aligned} \quad (4.30)$$

Bajo la Condición 4.5,

$$\sup_{k \geq 1} \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) d\xi > 1, \quad \theta \neq \theta_0, \quad (4.31)$$

ya que, según se deduce de la Nota 4.7, para cada $k \geq 1$, $\mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k)$ define el espectro de un proceso lineal estacionario cuya varianza de predicción en un paso es la unidad. Específicamente, la varianza de ese proceso viene dada por $\int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) d\xi$, que debe ser mayor que la varianza de predicción en un paso, a menos que el espectro sea constante para cada $k \geq 1$. Este hecho no se cumple porque significa que la composición $\mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}$ de los operadores $\mathcal{F}_{\xi, \theta_0}$ y $\mathcal{F}_{\xi, \theta}^{-1}$ coincide con el operador identidad en $\mathcal{H} + i\mathcal{H}$, que no es posible en vista de la condición de identificabilidad, 4.3, ya que $\theta \neq \theta_0$ en (4.31). Por lo tanto,

$$\inf_{\theta \in \Theta} \sup_{k \geq 1} \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1}(\varphi_k)(\varphi_k) d\xi = 1. \quad (4.32)$$

A partir de (4.32), la versión teórica de la función de pérdida empírica alcanza el mínimo en $\theta = \theta_0$. En cuanto a la consistencia en sentido fuerte del estimador paramétrico obtenido se desprende, de este hecho y de las ecuaciones (4.30) y (4.31); véase la Nota 4.8 para más detalles. \square

Nota 4.8 (Detalles para la demostración del Teorema 4.1).

En el desarrollo de la demostración del Teorema 4.1, la expansión empírica truncada del operador periodograma

$$\begin{aligned} \mathcal{I}_{\xi}^{\mathcal{N}}(\varphi_k)(\varphi_l) &= \tilde{X}_{\xi}^{\mathcal{N}}(\varphi_k)\tilde{X}_{-\xi}^{\mathcal{N}}(\varphi_l) \\ &= \frac{1}{(2\pi)^d} \sum_{\mathbf{z} \in \prod_{i=1}^d [-\mathcal{N}_i+1, \mathcal{N}_i-1]} \mathcal{C}(\mathbf{z}, k, l) \exp\left(-i \sum_{j=1}^d \xi_j z_j\right), \quad k, l \geq 1, \\ \mathbf{z} &= (z_1, \dots, z_d) \in \mathbb{Z}^d, \quad z_i \in [-\mathcal{N}_i + 1, \mathcal{N}_i - 1], \quad i = 1, \dots, d, \end{aligned} \quad (4.33)$$

se ha aplicado en primer lugar, en términos de los coeficientes de Fourier, introducidos en (4.25). Al mismo tiempo, también se considera la serie de Fourier truncada (4.26) definida a partir de los coeficientes de Fourier

$$\begin{aligned} g(\mathbf{z}, \theta, k) &= \frac{1}{(2\pi)^d} \int_{[-\pi, \pi]^d} \exp\left(i \sum_{j=1}^d \varpi_j z_j\right) \mathcal{F}_{\varpi, \theta}^{-1}(\varphi_k)(\varphi_k) d\varpi, \\ \mathbf{z} &= (z_1, \dots, z_d) \in \mathbb{Z}^d, \quad \theta \in \Theta, \quad k \geq 1, \end{aligned} \quad (4.34)$$

de la inversa $\mathcal{F}_{\varpi, \theta}^{-1}$ del operador de densidad espectral. Este truncamiento se expresa de forma equivalente en la ecuación (4.26) en términos del kernel de Féjer multidimensional,

$$F_M(\xi) = \frac{1}{\prod_{j=1}^d M_j} \sum_{y_1=1}^{M_1} \cdots \sum_{y_d=1}^{M_d} \sum_{x_1=1}^{M_1} \cdots \sum_{x_d=1}^{M_d} \exp\left(-i \sum_{j=1}^d (y_j - x_j) \xi_j\right). \quad (4.35)$$

Las ecuaciones (4.33)–(4.35) se aplican en la obtención de la convergencia a cero de las ecuaciones (4.27) y (4.28).

Específicamente, sea $\xi_{z_j} = \frac{2\pi z_j}{\mathcal{N}_j}$, con $z_j \in \mathbb{Z}$ tal que $-\frac{\mathcal{N}_j}{2} < z_j \leq \left[\frac{\mathcal{N}_j}{2}\right]$, para $j = 1, \dots, d$, siendo $\mathcal{N} = \prod_{j=1}^d \mathcal{N}_j$ el tamaño de la muestra funcional. Aplicando la identidad de Parseval en $L^2([-\pi, \pi]^d)$, para un M fijado grande, y $\theta \in \Theta$,

$$\begin{aligned} &\frac{(2\pi)^d}{\mathcal{N}} \sum_{\mathbf{z}_j} \left| q_{\xi_{\mathbf{z}_j}, \theta}^M \mathcal{I}_{\xi_{\mathbf{z}_j}}^{(\mathcal{N})}(\varphi_k)(\varphi_k) - q_{\xi_{\mathbf{z}_j}, \theta}^M \mathcal{F}_{\xi_{\mathbf{z}_j}, \theta_0}(\varphi_k)(\varphi_k) \right|^2 \\ &= \frac{(2\pi)^d}{\mathcal{N}} \sum_{\mathbf{u}_j \in [-M_j+1, M_j-1], j=1, \dots, d} \left| \prod_{j=1}^d \left(1 - \frac{|u_j|}{M_j}\right) g(\mathbf{u}, \theta, k) \right|^2 \\ &\quad \times \left| [\mathcal{C}(\mathbf{u}, k, k) + \mathcal{K}(\mathcal{N})] - \mathcal{R}_{\mathbf{u}, \theta_0}(\varphi_k)(\varphi_k) \right|^2, \end{aligned} \quad (4.36)$$

donde $\mathcal{K}(\mathcal{N}) \rightarrow 0$, cuando $\mathcal{N} \rightarrow \infty$ (Hannan, 1973). La Suposición 4.2 implica entonces la convergencia c. s. a cero de (4.36), cuando $\mathcal{N} \rightarrow \infty$, uniformemente en $k \geq 1$, en vista de la ecuación (4.24).

A continuación, se procede a detallar la obtención de la consistencia fuerte de $\hat{\theta}_{\mathcal{N}}$, a partir de las ecuaciones (4.30) y (4.31). Específicamente, si, cuando $\mathcal{N} \rightarrow \infty$, $\hat{\theta}_{\mathcal{N}}$ no converge c. s. a θ_0 , existe un $\hat{\theta}_{\mathcal{N}_m}$ que converge a $\theta' \neq \theta_0$, cuando $m \rightarrow \infty$ ($\mathcal{N}_m \rightarrow \infty$), tal que $\theta' \in \Theta$. A partir de (4.30), para una constante positiva, $\eta > 0$, el límite inferior $\underline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\hat{\theta}_{\mathcal{N}_m})$ satisface c. s.

$$\underline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\hat{\theta}_{\mathcal{N}_m}) \geq \sup_{k \geq 1} \frac{1}{[2\pi]^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} [\mathcal{F}_{\xi, \theta'} + \eta I_{\mathcal{H} + i\mathcal{H}}]^{-1} (\varphi_k)(\varphi_k) d\xi,$$

donde $I_{\mathcal{H} + i\mathcal{H}}$ denota el operador identidad en $\mathcal{H} + i\mathcal{H}$. A partir de la ecuación (4.31), para un η suficientemente pequeño, $\underline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\hat{\theta}_{\mathcal{N}_m}) > 1$, c. s. Por otra parte, para cada $\theta \in \Theta$, $\overline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\hat{\theta}_{\mathcal{N}_m}) \leq \overline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\theta)$, c. s. Por lo tanto, para cada $\theta \in \Theta$,

$$\overline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\hat{\theta}_{\mathcal{N}_m}) \leq \sup_{k \geq 1} \frac{1}{[2\pi]^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1} (\varphi_k)(\varphi_k) d\xi,$$

que implica c. s. la desigualdad

$$\overline{\lim}_{\mathcal{N}_m \rightarrow \infty} \sigma_{\mathcal{N}_m}(\hat{\theta}_{\mathcal{N}_m}) \leq \inf_{\theta \in \Theta} \sup_{k \geq 1} \frac{1}{[2\pi]^d} \int_{[-\pi, \pi]^d} \mathcal{F}_{\xi, \theta_0} \mathcal{F}_{\xi, \theta}^{-1} (\varphi_k)(\varphi_k) d\xi = 1,$$

teniendo en cuenta (4.32), lo que lleva a una contradicción. Por lo tanto, $\theta' = \theta_0 = \lim_{\mathcal{N} \rightarrow \infty} \hat{\theta}_{\mathcal{N}}$, c. s., como se quería demostrar.

4.3. Estudio de simulación

Esta sección ilustra la consistencia fuerte obtenida del estimador paramétrico (4.19)–(4.20), así como su consistencia en media cuadrática. Restringimos nuestra atención al caso en que el campo aleatorio de log-intensidad es un proceso SAR $\mathcal{H}(1)$, véase Ruiz-Medina (2011). Como se comentó en la introducción del presente capítulo, en los ejemplos numéricos que siguen se aborda el problema de la estimación de los hiperparámetros, que caracterizan los espectros puntuales y los autovectores de los operadores de correlación espacial, que también intervienen en la parametrización de nuestra familia de operadores de

densidad espectral, véase, por ejemplo, la ecuación (4.40) más adelante.

Sea $\mathcal{H} = L^2(\mathcal{T})$, como antes, el espacio de las funciones de cuadrado integrable en el intervalo $\mathcal{T} \subseteq \mathbb{R}_+$. Consideremos el caso $d = 2$, y supongamos que el campo aleatorio de log-intensidad \mathcal{H} -valuado $\{\varkappa_{\mathbf{z}}, \mathbf{z} \in \mathbb{Z}^2\}$ satisface la ecuación de estados SARH(1)

$$\varkappa_{i,j} = L_1(\varkappa_{i-1,j}) + L_2(\varkappa_{i,j-1}) + L_3(\varkappa_{i-1,j-1}) + \epsilon_{i,j}, \quad (i,j) \in \mathbb{Z}^2, \quad (4.37)$$

donde $\epsilon = \{\epsilon_{i,j}, (i,j) \in \mathbb{Z}^2\}$ denota un proceso de innovación espacial generalizado de media cero que satisface, para cualquier nodo espacial (i,j) , $E|\epsilon_{i,j}(\varphi_k)|^2 = \sigma_k^2 = 1$, para $k \geq 1$, siendo $\{\varphi_k, k \geq 1\}$ una base ortonormal de \mathcal{H} . Por lo tanto, $\mathcal{R}_{i,j}^\epsilon = E(\epsilon_{i+k,j+l} \otimes \epsilon_{k,l}) = E(\epsilon_{i,j} \otimes \epsilon_{0,0}) = I_{\mathcal{H}}$, para cada $(i,j), (k,l) \in \mathbb{Z}^2$, con $I_{\mathcal{H}}$ denotando el operador identidad en \mathcal{H} .

En lo siguiente, trabajamos bajo el supuesto de que $\{\epsilon_{i,j}, (i,j) \in \mathbb{Z}^2\}$ es un ruido blanco gaussiano en \mathcal{H} . En concreto, $\{\epsilon_{i,j}(\varphi_k), k \geq 1, (i,j) \in \mathbb{Z}^2\}$ es una familia de variables aleatorias gaussianas independientes tipificadas \mathcal{H} -valuadas. Para cada $(i,j) \in \mathbb{Z}^2$, $\epsilon_{i,j}$ está incorrelado con $\varkappa_{i,j}$. Los operadores L_q , $q = 1, 2, 3$, son operadores lineales acotados en \mathcal{H} , que satisfacen las condiciones dadas en las proposiciones 3 y 4 en Ruiz-Medina (2011), que aseguran la existencia de una única solución estacionaria para (4.37), admitiendo una representación funcional espacial de medias móviles de orden infinito, SMAH(∞). Nótese que se ha asumido la sumabilidad en el espacio del cuadrado de la norma del operador de Hilbert-Schmidt de los operadores involucrados en la representación SMAH(∞).

En los ejemplos expuestos a continuación, L_i , $i = 1, 2, 3$, se introduce en términos de los autovectores $\{\phi_k, k \geq 1\}$ del operador laplaciano negativo de Dirichlet $(-\Delta)_{\mathcal{T}}$ sobre un intervalo \mathcal{T} ,

$$\phi_k(t) = \text{sen} \left(\frac{\pi t k}{T} \right), \quad t \in \mathcal{T} = [0, T], \quad k \geq 1. \quad (4.38)$$

Aplicando el teorema espectral sobre el cálculo espectral para operadores auto-adjuntos en un espacio separable de Hilbert, podemos caracterizar el espectro puntual puro de cualquier función continua de $(-\Delta)_{\mathcal{T}}$, aplicando dicha función a los autovalores de $(-\Delta)_{\mathcal{T}}$, véase, por ejemplo, Dautray y Lions (1985). Este es el contexto adoptado en los ejemplos 4.1 y 4.2 siguientes para la introducción de los operadores L_i , $i = 1, 2, 3$. Por otra parte, en el caso de considerar que los

valores funcionales de los procesos de log-intensidad \mathcal{H} -valuados tienen soporte sobre toda la recta real, estos modelos pueden extenderse considerando la transformada de Fourier en sentido débil de los potenciales de Riesz y Bessel, véase, por ejemplo, Triebel (1978).

Ejemplo 4.1. En este ejemplo, se aborda el problema de la estimación del parámetro $\theta \in \Theta$ que caracteriza el soporte $\mathcal{T}(\theta) = [0, \theta]$ de los autovectores (4.38) que definen $L_{i,\theta}$, $i = 1, 2, 3$, dados por

$$\begin{aligned} L_{1,\theta} &= \Phi_{M_1(\theta)} \mathcal{K}_1(\theta, \pi) (-\Delta)_{\mathcal{T}(\theta)}^{-\beta_1}, & L_{2,\theta} &= \Phi_{M_2(\theta)} \mathcal{K}_2(\theta, \pi) (-\Delta)_{\mathcal{T}(\theta)}^{-\beta_2}, \\ L_{3,\theta} &= -L_{1,\theta} L_{2,\theta} = -L_{2,\theta} L_{1,\theta}, \end{aligned} \quad (4.39)$$

donde $\Phi_{M_i(\theta)}$ denota al operador de proyección sobre el subespacio generado por los autovectores $\phi_1, \dots, \phi_{M_i(\theta)}$ de $(-\Delta)_{\mathcal{T}(\theta)}$, siendo $M_i(\theta) \geq 1$ una función acotada en $\theta \in \Theta$, para $i = 1, 2$. Aquí, $\beta_i, \mathcal{K}_i \in \mathbb{R}_+$ son conocidos, y \mathcal{K}_i es continua en $\theta \in \Theta$, para $i = 1, 2$. Aplicando el Teorema espectral sobre el cálculo espectral para operadores autoadjuntos en un espacio separable de Hilbert, véase, por ejemplo, Dautray y Lions (1985), los autovalores $\{\lambda_{k,i}(\theta), k \geq 1\}$ del operador de autocorrelación $L_{i,\theta}$ vienen dados, para $i = 1, 2, 3$, por

$$\begin{aligned} \lambda_{k,1}(\theta) &= 1_{[1, M_1(\theta)]}(k) \mathcal{K}_1(\theta, \pi) [\lambda_k((-\Delta)_{\mathcal{T}(\theta)})]^{-\beta_1} \\ \lambda_{k,2}(\theta) &= 1_{[1, M_2(\theta)]}(k) \mathcal{K}_2(\theta, \pi) [\lambda_k((-\Delta)_{\mathcal{T}(\theta)})]^{-\beta_2} \\ \lambda_{k,3}(\theta) &= -\lambda_{k,1}(\theta) \lambda_{k,2}(\theta), \quad k \geq 1, \quad \theta \in \Theta, \end{aligned} \quad (4.40)$$

donde $1_{[1, M_i(\theta)]}(k)$ denota la función indicadora del conjunto $[1, M_i(\theta)] \subset \mathbb{N}$, con, como antes, $M_i(\theta) \geq 1$, para $\theta \in \Theta$, y $i = 1, 2$. Para $k \geq 1$, $\lambda_k((-\Delta)_{\mathcal{T}(\theta)})$ denota el k -ésimo autovalor de $(-\Delta)_{\mathcal{T}(\theta)}$. Consideramos el caso particular $M_1(\theta) = M_2(\theta) = M(\theta)$, y asumimos que $\lambda_{k,i}(\theta)$, $k \geq 1$, $i = 1, 2, 3$, son tales que las raíces del polinomio $P_{k,\theta}(z_1, z_2) = \sum_{i=0}^1 \sum_{j=0}^1 p_{k,i,j,\theta} z_1^i z_2^j$, están fuera del polidisco unitario (es decir, el producto de dos discos unitarios abiertos en el plano complejo). Aquí $p_{k,0,0,\theta} = 1$, $p_{k,1,0,\theta} = \lambda_{k,1}(\theta)$, $p_{k,0,1,\theta} = \lambda_{k,2}(\theta)$, y $p_{k,1,1,\theta} = -\lambda_{k,1}(\theta) \lambda_{k,2}(\theta)$, $k \geq 1$. A partir de las ecuaciones (4.37)–(4.40), calculamos el operador de densidad

espectral

$$\begin{aligned} & \mathcal{F}_{\xi, \theta}(\phi_{k, \theta})(\phi_{k, \theta}) \\ &= \frac{1}{|1 - \lambda_{k,1}(\theta) \exp(i\xi_1) - \lambda_{k,2}(\theta) \exp(i\xi_2) - \lambda_{k,3}(\theta) \exp(i(\xi_1 + \xi_2))|^2}, \end{aligned} \quad (4.41)$$

$$\begin{aligned} L_{q, \theta}(g)(h) &= \sum_{k=1}^{\infty} \lambda_{k,q}(\theta) \langle \phi_{k, \theta}, h \rangle \langle \phi_{k, \theta}, g \rangle, \quad \forall g, h \in \mathcal{H}, \quad q = 1, 2, 3, \\ \xi &\in [-\pi, \pi]^2, \quad k \in [1, M(\theta)], \quad \theta \in \Theta. \end{aligned}$$

Bajo las condiciones asumidas sobre los parámetros del modelo (4.40), la Suposición 4.2 se satisface en nuestro marco SARH(1), como se deduce del Corolario 2.3 en (Bosq, 2000), aplicando resultados similares a los obtenidos en el Corolario 4.1 y el Teorema 4.8 de dicha monografía. La condición de identificabilidad 4.3 se deduce de la caracterización de los autovectores y autovalores del operador laplaciano negativo de Dirichlet a partir del soporte $\mathcal{T}(\theta)$, definido unívocamente por el parámetro θ . Las condiciones (4.15), 4.4 y 4.6 se obtienen a partir de las ecuaciones (4.40)–(4.41), bajo las restricciones consideradas en los coeficientes paramétricos $\{\lambda_{k,i}(\theta), k \in [1, M(\theta)], \theta \in \Theta, i = 1, 2\}$ definiendo la familia de polinomios $\{P_{k,\theta}, k \in [1, M(\theta)], \theta \in \Theta\}$. La Condición 4.5 también se cumple ya que $\sigma_k^2 = 1$, para cada $k \geq 1$.

La consistencia fuerte del predictor paramétrico plug-in SARH(1)

$$\widehat{\mathcal{X}}_{i,j}^{\mathcal{N}}(\theta) = L_{1, \widehat{\theta}_{\mathcal{N}}} \mathcal{X}_{i-1,j} + L_{2, \widehat{\theta}_{\mathcal{N}}} \mathcal{X}_{i,j-1} + L_{3, \widehat{\theta}_{\mathcal{N}}} \mathcal{X}_{i-1,j-1}, \quad \forall (i, j) \in \mathbb{Z}^2, \quad (4.42)$$

se obtiene entonces bajo condiciones similares a las asumidas en Bosq (2000) para el caso ARH(1).

Los resultados de la estimación se muestran en la Tabla 4.1 y en la Figura 4.1, donde se ha considerado $\theta_0 = 1$, $\mathcal{K}_1(\theta, \pi) = \frac{\theta^{2-11/10}}{\pi^{2-11/10}}$, $\mathcal{K}_2(\theta, \pi) = \frac{\theta^{2-12/10}}{\pi^{2-12/10}}$, $\beta_1 = \frac{11}{20}$, $\beta_2 = \frac{12}{20}$, $M_1(\theta) = M_2(\theta) = [10(\theta)]$, y $\theta \in \Theta = [0.7, 4]$, con $[\cdot]$ denotando la función parte entera. Se han probado los tamaños de muestra funcional espacial $\mathcal{N} = 40000, 62500, 90000, 122500, 160000, 202500, 250000, 302500$. Específicamente, la Figura 4.1, en la parte izquierda, muestra los valores muestrales de $\widehat{\theta}_{\mathcal{N}}$, basados en 100 generaciones de cada muestra funcional. Los errores cuadráticos medios empíricos también se han calculado en la Tabla 4.1 y la Figura 4.1, en la parte derecha.

Tabla 4.1: Estimaciones de los parámetros de localización. Tamaños de muestra, media, desviación típica, y errores cuadrático medio empíricos (M.S.E.), a partir de 100 generaciones, $\theta_0 = 1$

\mathcal{N}	$\overline{\hat{\theta}_{\mathcal{N}}}$	$\sigma(\hat{\theta}_{\mathcal{N}})$	M.S.E
40000	0.9867	0.1391	0.0193
62500	1.0066	0.0972	0.0094
90000	0.9993	0.0857	0.0073
122500	1.0008	0.0810	0.0065
160000	1.0034	0.0716	0.0051
202500	0.9893	0.0536	0.0030
250000	0.9995	0.0449	0.0020
302500	1.0060	0.0436	0.0019

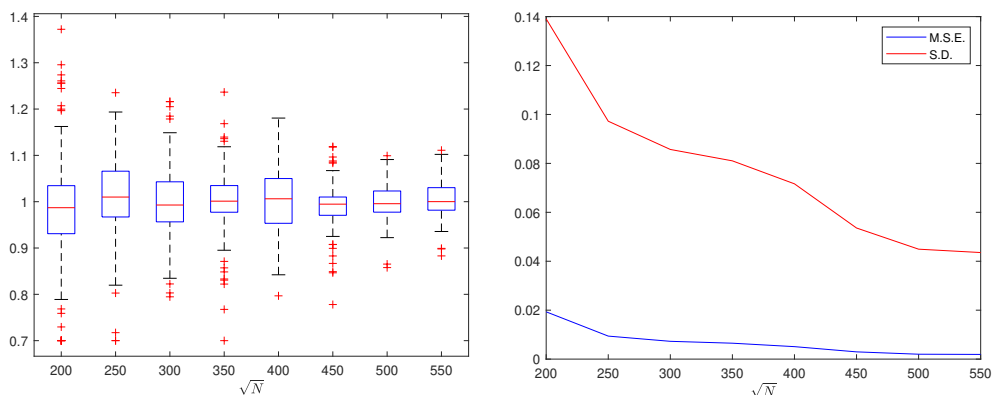


Figura 4.1: Diagramas de cajas de los valores muestrales de $\hat{\theta}_{\mathcal{N}}$, para $\theta_0 = 1$ basados en 100 generaciones (a la izquierda). A la derecha, el error cuadrático medio empírico (M.S.E), línea azul, y desviación típica (S.D.), línea roja continua

Ejemplo 4.2. Consideramos ahora la siguiente familia paramétrica de operadores de autocorrelación L_{q,θ_q} , $q = 1, 2$,

$$\begin{aligned}
 L_{1,\theta_1} &= \Phi_M \theta_{1,1} \left[\theta_{1,2} + (1/\pi)(-\Delta)_{[0,1]}^{1/2} \right]^{-1} \\
 L_{2,\theta_2} &= \Phi_M \theta_{2,1} \left[\theta_{2,2} + (1/\pi)(-\Delta)_{[0,1]}^{1/2} \right]^{-1}, \quad (\theta_1, \theta_2) \in \Theta \\
 L_{3,\theta_3} &= -L_{1,\theta_1} L_{2,\theta_2} = -L_{2,\theta_2} L_{1,\theta_1} \\
 \phi_k(t) &= \text{sen}(\pi tk), \quad t \in [0, 1], \quad k \geq 1 \\
 \lambda_{k,q}(\theta_q) &= 1_{[1,M]}(k) \frac{\theta_{q,1}}{k + \theta_{q,2}}, \quad q = 1, 2, \\
 \lambda_{k,3}(\theta_3) &= -\lambda_{k,1}(\theta_1) \lambda_{k,2}(\theta_2), \quad k \geq 1
 \end{aligned} \tag{4.43}$$

donde, como antes, $1_{[1,M]}(k)$ es la función indicadora del conjunto $[1, M] \subset \mathbb{N}$, Φ_M , $M \geq 1$, denota el operador de proyección sobre el subespacio generado por los autovectores ϕ_1, \dots, ϕ_M del operador laplaciano negativo de Dirichlet $(-\Delta)_{[0,1]}$ en el intervalo $[0, 1]$. Aquí, $\{\lambda_{k,1}(\theta_1), \lambda_{k,2}(\theta_2), k \geq 1, (\theta_1, \theta_2) \in \Theta\}$ es la familia paramétrica de autovalores de los operadores de autocorrelación $\{L_{1,\theta_1}, L_{2,\theta_2}, (\theta_1, \theta_2) \in \Theta\}$.

Bajo condiciones similares a las del Ejemplo 4.1 sobre estas familias paramétricas de autovalores de $\{L_{1,\theta_1}, L_{2,\theta_2}, (\theta_1, \theta_2) \in \Theta\}$, se cumplen las suposiciones 4.1–4.2, y las condiciones (4.15) y 4.3–4.6. Por lo tanto, la consistencia fuerte del estimador paramétrico $\hat{\theta}_{\mathcal{N}}$ se deduce del Teorema 4.1. En particular, se han considerado los valores de los parámetros de escala $\theta_{1,1} = 1$, $\theta_{2,1} = 1.5$, y los valores de los parámetros de localización $\theta_{1,2} = 1.6$, y $\theta_{2,2} = 1.2$, que se encuentran en el interior de $\Theta = ([0.7, 1.3] \times [1.3, 1.9]) \times ([1.2, 1.8] \times [0.9, 1.5])$.

Véanse los resultados numéricos mostrados en los gráficos ubicados en la parte derecha de las figuras 4.2–4.5, a partir de las 100 generaciones de cada una de las muestras funcionales espaciales de tamaños $\mathcal{N} = 40000, 62500, 90000, 122500, 160000, 202500, 250000, 302500$. El error cuadrático medio empírico también se ha calculado para estos tamaños de muestra (ver los gráficos ubicados en la parte derecha de las figuras 4.2–4.5 y la Tabla 4.2). Se observa una buena aproximación a partir de la técnica de estimación propuesta, para los tamaños de muestra funcional espacial considerados, y para el valor del parámetro de truncamiento elegido, $M = 10$. Obsérvese que este valor del parámetro de truncamiento proporciona la dimensión umbral para asegurar un comportamiento estable del

estimador vectorial paramétrico, con respecto a los datos funcionales espaciales de entrada, y el tamaño de paso de discretización temporal.

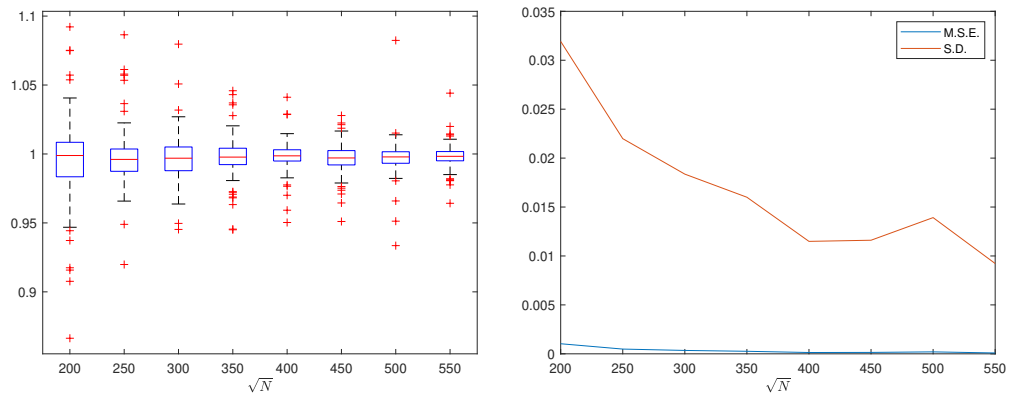


Figura 4.2: Diagramas de cajas de los valores muestrales de $\hat{\theta}_{1,1,\mathcal{N}}$, para $\theta_{1,1,0} = 1$, basados en 100 generaciones (a la izquierda). A la derecha, el error cuadrático medio empírico (M.S.E), línea azul, y desviación típica (S.D.), línea roja continua

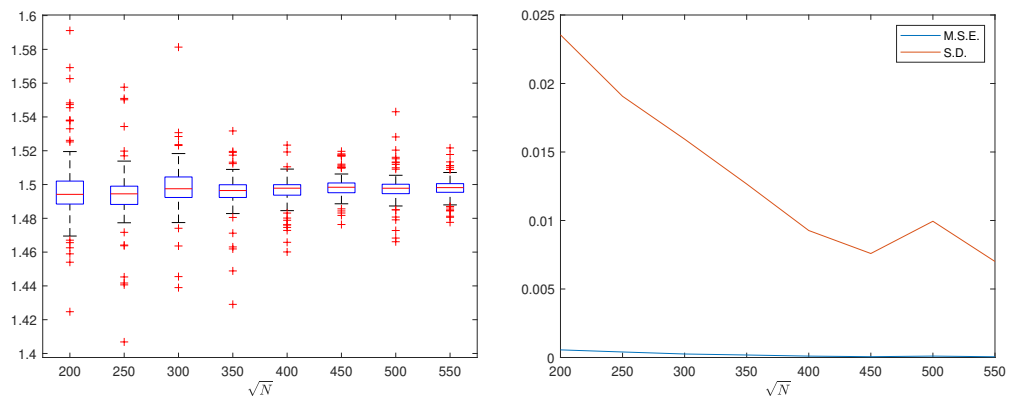


Figura 4.3: Diagramas de cajas de los valores muestrales de $\hat{\theta}_{2,1,\mathcal{N}}$, para $\theta_{2,1,0} = 1.5$, basados en 100 generaciones (a la izquierda). A la derecha, el error cuadrático medio empírico (M.S.E), línea azul, y desviación típica (S.D.), línea roja discontinua

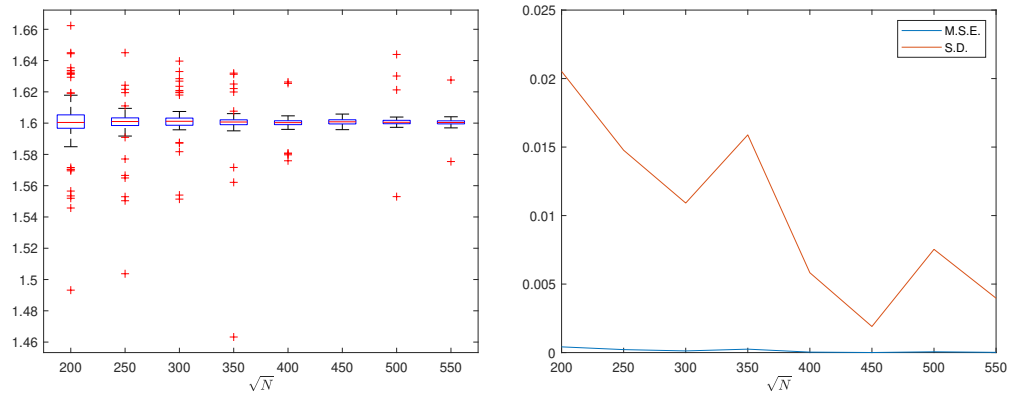


Figura 4.4: Diagramas de cajas de los valores muestrales de $\hat{\theta}_{1,2,\mathcal{N}}$, para $\theta_{1,2,0} = 1.6$, basados en 100 generaciones (a la izquierda). A la derecha, el error cuadrático medio empírico (M.S.E), línea azul, y desviación típica (S.D.), línea roja continua

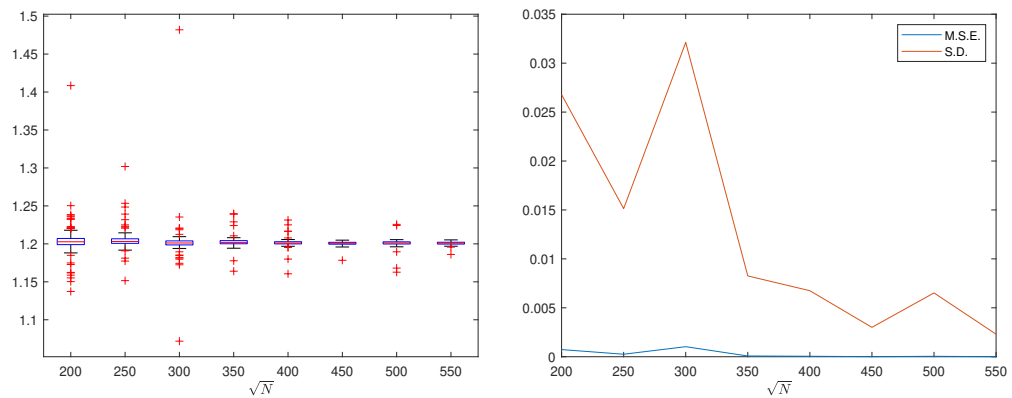


Figura 4.5: Diagramas de cajas de los valores muestrales de $\hat{\theta}_{2,2,\mathcal{N}}$, para $\theta_{2,2,0} = 1.2$, basados en 100 generaciones (a la izquierda). A la derecha, el error cuadrático medio empírico (M.S.E), línea azul, y desviación típica (S.D.), línea roja discontinua

Tabla 4.2: Estimaciones de los parámetros de escala y localización. Tamaños de muestra, media, desviación típica, y errores cuadrático medio empíricos (M.S.E.), a partir de 100 generaciones, $\theta_{1,1,0} = 1$, $\theta_{2,1,0} = 1.5$, $\theta_{1,2,0} = 1.6$, y $\theta_{2,2,0} = 1.2$

N	$\overline{\hat{\theta}_{1,1,N}}$	$\sigma(\hat{\theta}_{1,1,N})$	M.S.E.
40000	0.9955	0.0319	0.00103
62500	0.9975	0.0220	0.00048
90000	0.9965	0.0184	0.00035
122500	0.9978	0.0160	0.00026
160000	0.9984	0.0115	0.00013
202500	0.9969	0.0116	0.00014
250000	0.9973	0.0139	0.00020
302500	0.9983	0.0092	0.00009

\mathcal{N}	$\overline{\hat{\theta}_{2,1,\mathcal{N}}}$	$\sigma(\hat{\theta}_{2,1,\mathcal{N}})$	M.S.E.
40000	1.4972	0.0236	0.00056
62500	1.4934	0.0191	0.00040
90000	1.4980	0.0159	0.00026
122500	1.4951	0.0127	0.00018
160000	1.4959	0.0093	0.00010
202500	1.4989	0.0076	0.00006
250000	1.4977	0.0099	0.00010
302500	1.4979	0.0070	0.00005

\mathcal{N}	$\overline{\hat{\theta}_{1,2,\mathcal{N}}}$	$\sigma(\hat{\theta}_{1,2,\mathcal{N}})$	M.S.E.
40000	1.6003	0.0205	0.00042
62500	1.5991	0.0148	0.00022
90000	1.6018	0.0109	0.00012
122500	1.5998	0.0159	0.00025
160000	1.6001	0.0058	0.00003
202500	1.6008	0.0019	0.00000
250000	1.6011	0.0075	0.00006
302500	1.6005	0.0040	0.00002

\mathcal{N}	$\overline{\hat{\theta}_{2,2,\mathcal{N}}}$	$\sigma(\hat{\theta}_{2,2,\mathcal{N}})$	M.S.E.
40000	1.2046	0.0268	0.00073
62500	1.2054	0.0151	0.00026
90000	1.2024	0.0321	0.00103
122500	1.2027	0.0083	0.00007
160000	1.2015	0.0067	0.00005
202500	1.2005	0.0030	0.00001
250000	1.2008	0.0065	0.00004
302500	1.2008	0.0023	0.00001

4.4. Ejemplo con datos reales

Los datos son facilitados por el Instituto Nacional de Estadística de España, y consisten en 432 registros mensuales sobre el número de casos de mortalidad por enfermedades respiratorias en 48 provincias españolas durante el período 1980–2015. Se consigue un ajuste del modelo espacial heterogéneo de tendencia polinómica de grado 10, tras un preprocesamiento adecuado de nuestro conjunto de datos originales. El análisis de correlación residual funcional espacial se realiza en el dominio espectral, en términos del modelo $SAR\mathcal{H}(1)$ introducido en la sección anterior. En concreto, los principales pasos del algoritmo de estimación implementado son los siguientes.

Paso 1. Se calcula el número acumulado de casos de mortalidad en las 48 provincias españolas, en el período 1980–2015, a partir de los 432 registros mensuales,

Paso 2. Las curvas de mortalidad acumulada se interpolan en 1725 nodos temporales y se suavizan mediante B-splines cúbicos.

Paso 3. Aplicando la distancia inversa ponderada, se realiza la interpolación espacial de nuestros datos curva del Paso 2 a una rejilla regular $\mathcal{N} = 20 \times 20$

Paso 4. La transformación logarítmica de los datos curva en el Paso 3 proporciona los valores funcionales de la intensidad en nuestra rejilla regular espacial, $\mathcal{N} = 20 \times 20$.

Paso 5. Se implementa el ajuste mínimo cuadrático del modelo de tendencia polinomial de grado 10.

Paso 6. Los valores residuales de las curvas de log-intensidad obtenidos en el Paso 5 se proyectan sobre la base ortonormal $\{\sin(\pi pt/1725), p \geq 1\}$ de $L^2([0, 1725])$.

Paso 7. Las proyecciones espaciales empíricas residuales resultantes se normalizan según la Condición 4.5.

Paso 8. La transformada rápida de Fourier bidimensional se aplica a los resultados del Paso 7.

Paso 9. La función de pérdida empírica $\sigma_{\mathcal{N}}(\boldsymbol{\theta})$ en (4.20) se calcula para $\mathcal{N} = 400$.

Paso 10. El modelo (4.41) se ajusta calculando el mínimo de la función multivariante no lineal restringida $\sigma_{\mathcal{N}}(\boldsymbol{\theta})$, con $\mathcal{N} = 400$. Aquí, L_3 no está dado necesariamente por la composición de los operadores L_i , $i = 1, 2$, como en la sección anterior. Se ha utilizado la opción “lbfgs” de la función de MatLab *fmincon*, que es un algoritmo de optimización de la familia de los métodos cuasi-Newton, que implica una estimación de la matriz hessiana inversa para dirigir su búsqueda a través del espacio de variables.

Paso 11. El predictor *plug-in* paramétrico SARH(1) (4.42) se obtiene a partir del Paso 10.

Nuestro ajuste del modelo paramétrico se ha realizado en términos del *kernel* espectral definido por los autovectores $\{\sin(\pi pt/1725), p \geq 1\}$ del operador laplaciano negativo de Dirichlet sobre $L^2([0, 1725])$. La estimación mínimo cuadrática de los valores del proceso de Cox se obtiene a partir del Paso 11 aplicando (4.5) en términos de $\{\sin(\pi pt/1725), p \geq 1\}$, y la correspondiente fórmula de proyección inversa. Las figuras 4.6–4.7 muestran, respectivamente, las proyecciones espaciales originales y estimadas del campo aleatorio de la log-intensidad residual funcional, para el parámetro de truncamiento $M = 10$ (ver resultados de los pasos 6–7 y 10).

En los pasos 9–10, se ha ajustado el siguiente modelo paramétrico

$$\hat{\lambda}_{p,i} = \lambda_{p,i}(\hat{\boldsymbol{\theta}}_{i,N}) = \hat{\theta}_{i,1,N} + |\sin(p\pi/2)|\hat{\theta}_{i,2,N}(p), \quad p = 1, \dots, 10, \quad i = 1, 2, 3, \quad (4.44)$$

para los autovalores de los operadores $L_{i,\hat{\boldsymbol{\theta}}_{i,N}}$, $i = 1, 2, 3$, considerando $M = 10$, a partir de un tamaño muestral funcional $\mathcal{N} = 400$. Los valores de los paráme-

tros ajustados $\widehat{\boldsymbol{\theta}}_{i,\mathcal{N}}$, $i = 1, 2, 3$, se muestran a continuación

$$\begin{aligned}
 &\widehat{\theta}_{1,1,\mathcal{N}} = 0.56; \widehat{\theta}_{1,2,\mathcal{N}}(p) = 0, \quad p = 2, 4, 6, 8, 10; \\
 &\widehat{\theta}_{1,2,\mathcal{N}}(p) = 0.08, \quad p = 1, 3, 5; \widehat{\theta}_{1,2,\mathcal{N}}(p) = 0.189; \quad p = 7, 9; \\
 &\widehat{\theta}_{2,1,\mathcal{N}} = 0.28; \widehat{\theta}_{2,2,\mathcal{N}}(p) = 0, \quad p = 1, 2, 3, 4, 5, 6, 8, 10; \\
 &\widehat{\theta}_{2,2,\mathcal{N}}(7) = 0.0725; \widehat{\theta}_{2,2,\mathcal{N}}(9) = 0.0814; \widehat{\theta}_{3,1,\mathcal{N}} = 0.0033; \\
 &\widehat{\theta}_{3,2,\mathcal{N}}(p) = -0.0052, \quad p = 2, 4, 6; \widehat{\theta}_{3,2,\mathcal{N}}(p) = 0, \quad p = 8, 10; \\
 &\widehat{\theta}_{3,2,\mathcal{N}}(p) = 0.4444, \quad p = 1, 3, 5; \widehat{\theta}_{3,2,\mathcal{N}}(p) = 0.1230, \quad p = 7, 9, \quad \mathcal{N} = 400.
 \end{aligned}
 \tag{4.45}$$

Las correspondientes estimaciones paramétricas $\{\lambda_{p,i}(\widehat{\boldsymbol{\theta}}_{i,\mathcal{N}}), p = 1, \dots, 10\}$, $i = 1, 2, 3$, de los espectros puntuales de $L_{i,\widehat{\boldsymbol{\theta}}_{i,\mathcal{N}}}$, $i = 1, 2, 3$, se muestran en la Tabla 4.3 y en la Figura 4.8. Finalmente, las curvas originales de log-intensidad de enfermedades respiratorias (línea azul discontinua), y sus estimaciones (línea roja), en las 48 provincias españolas, se muestran en las figuras 4.9–4.10. El peor y mejor ajuste del modelo (provincias de Soria y Pontevedra, respectivamente) se amplían en la Figura 4.11. Asimismo, los mapas de riesgo de mortalidad por enfermedades respiratorias originales y estimados, promediados anualmente, se muestran en la Figura 4.12.

La capacidad predictiva funcional espacial de nuestro modelo se evalúa mediante validación cruzada. Retirando las curvas observadas en los nodos que cubren el área espacial correspondiente a la provincia de interés (conjunto de datos funcionales de validación), se ejecutan los pasos 1–11 a partir de las restantes observaciones funcionales, distribuidas espacialmente dentro del resto de las provincias españolas (conjunto de datos funcionales de entrenamiento). En cada iteración se calcula el correspondiente predictor *plug-in* paramétrico $\text{SARH}(1)$, (4.42). Después de las 48 iteraciones, se obtiene la media del error absoluto funcional de validación cruzada, $\text{CVFARE}(t) = \frac{1}{48} \sum_{\mathbf{z} \in \mathcal{R}} \left| \frac{r_{\mathbf{z}}(t) - \widehat{r}_{\mathbf{z}}(t)}{r_{\mathbf{z}}(t)} \right|$, $t = 1, \dots, 1725$, donde \mathcal{R} denota el conjunto de longitudes y latitudes que definen la ubicación espacial de las 48 provincias españolas. La Tabla 4.4 muestra la media anual de los valores puntuales del CVFARE. El valor empírico $\|\text{CVFARE}\|_{L^1([0,1725])} = 0.0016$ calculado a partir de la Tabla 4.4 muestra una buena capacidad de predicción funcional espacial del enfoque presentado más allá del escenario log-gaussiano.

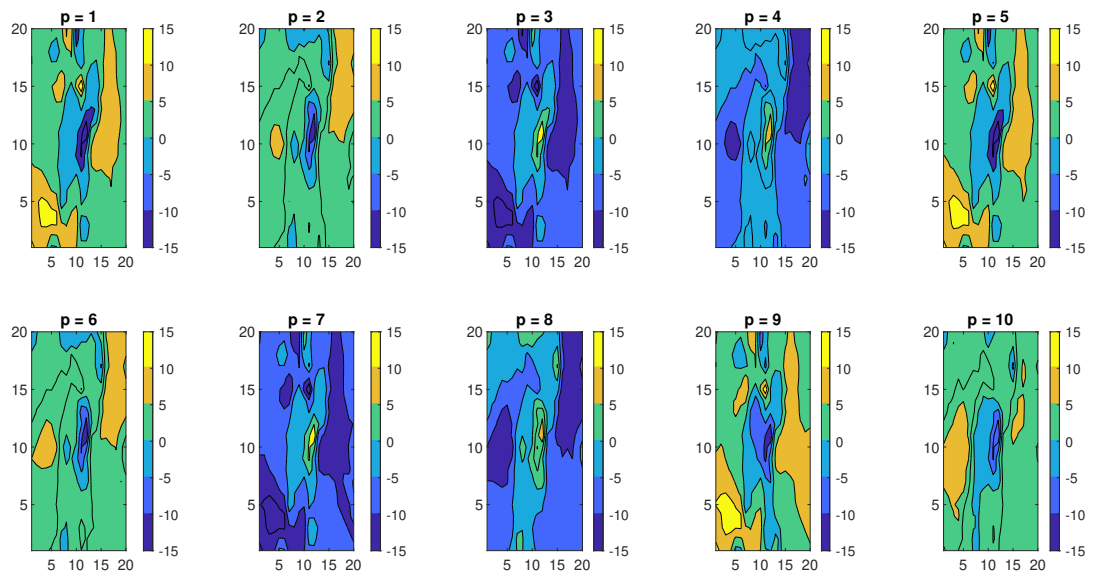


Figura 4.6: Diez proyecciones normalizadas de la log-intensidad residual funcional espacial

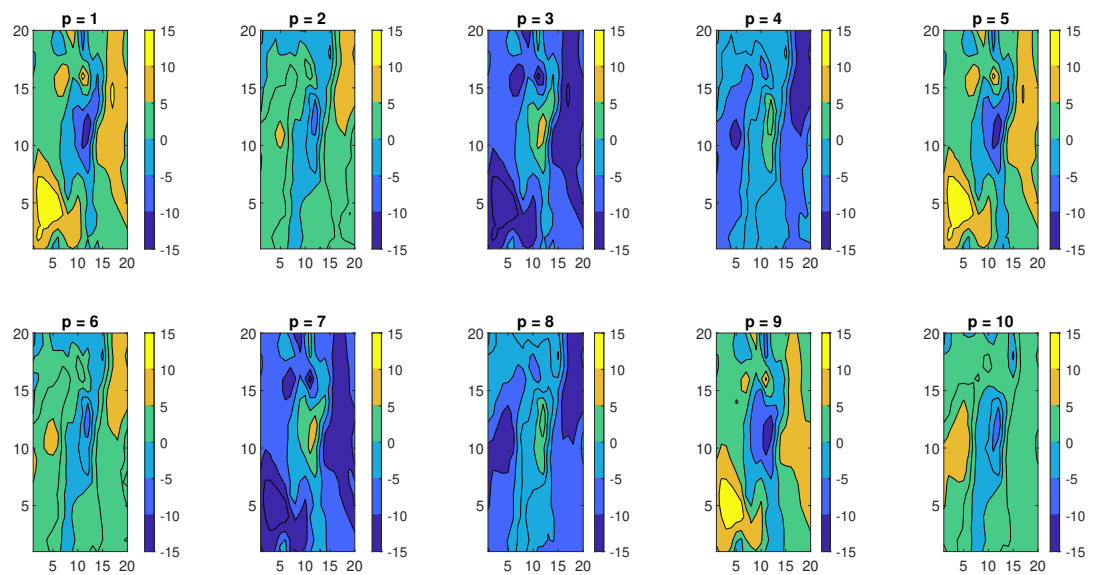


Figura 4.7: Estimaciones de las proyecciones normalizadas de la log-intensidad residual funcional espacial

Tabla 4.3: Estimaciones de espectros puntuales de $L_{i,\hat{\theta}_{i,\mathcal{N}}}$, $i = 1, 2, 3$, para $M = 10, \mathcal{N} = 400$

	$\lambda_{p,1}(\hat{\theta}_{1,\mathcal{N}})$	$\lambda_{p,2}(\hat{\theta}_{2,\mathcal{N}})$	$\lambda_{p,3}(\hat{\theta}_{3,\mathcal{N}})$
$p = 1$	0.6578	0.2801	0.4493
$p = 2$	0.5534	0.2878	-0.0102
$p = 3$	0.6583	0.2812	0.4482
$p = 4$	0.5547	0.2873	-0.0091
$p = 5$	0.6595	0.2841	0.4456
$p = 6$	0.5572	0.2860	-0.0061
$p = 7$	0.7490	0.3525	0.1316
$p = 8$	0.5620	0.2819	0.0021
$p = 9$	0.7507	0.3614	0.1211
$p = 10$	0.5678	0.2634	0.0398

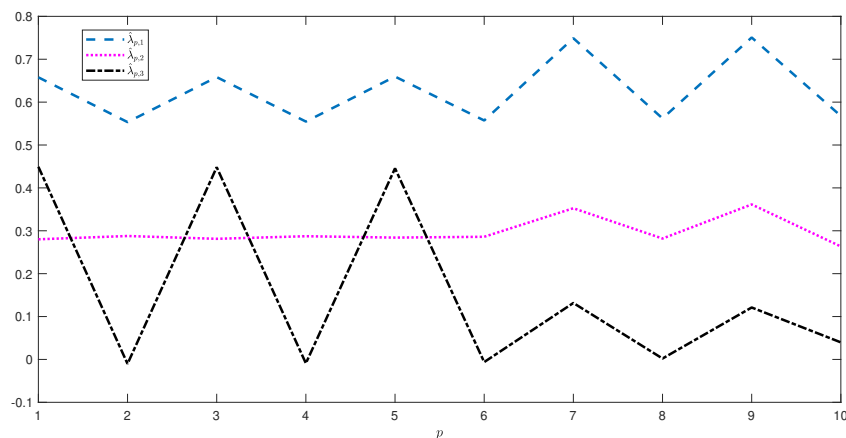


Figura 4.8: Gráfico de espectros puntuales estimados de $L_{i,\hat{\theta}_{i,\mathcal{N}}}$, $i = 1, 2, 3$, para $M = 10$ y $\mathcal{N} = 400$. En el gráfico, se ha utilizado la notación $\hat{\lambda}_{p,i} = \lambda_{p,i}(\hat{\theta}_{i,\mathcal{N}})$, para $p = 1, \dots, 10, i = 1, 2, 3$

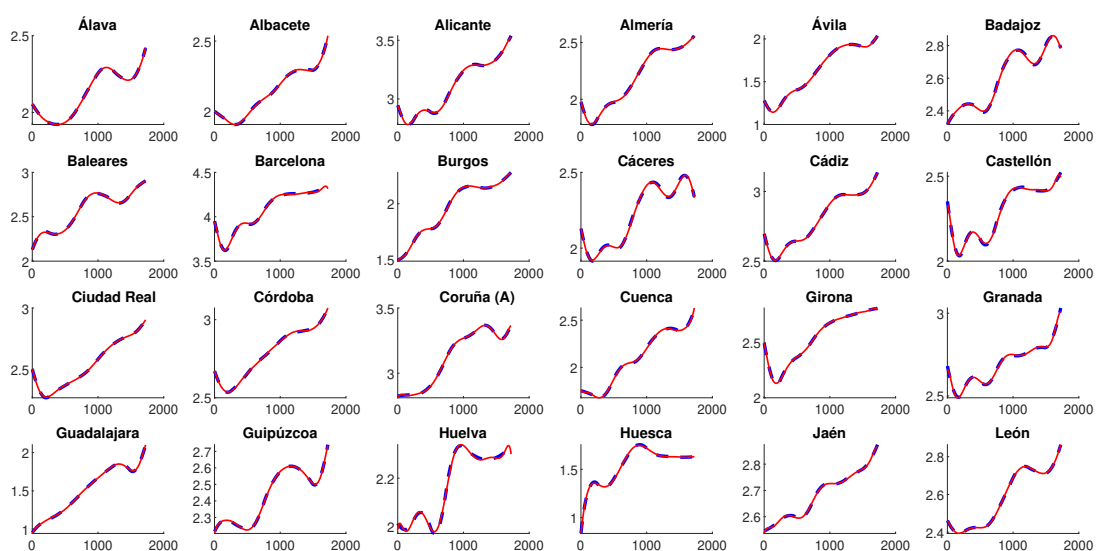


Figura 4.9: Curvas de log-intensidad de mortalidad por enfermedades respiratorias originales (línea azul discontinua), y sus respectivas estimaciones (línea roja), para las 24 primeras provincias españolas por orden alfabético

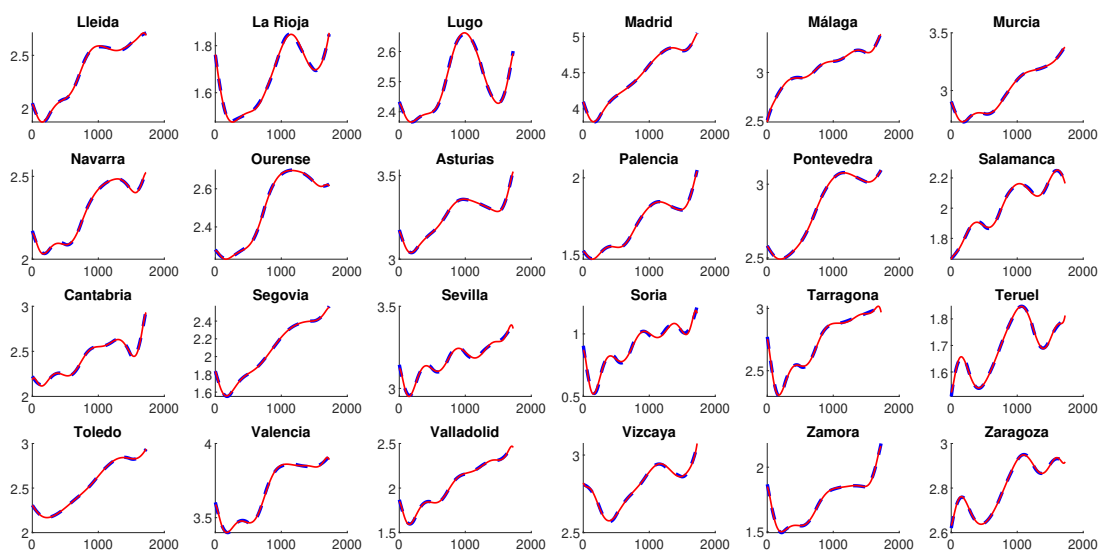


Figura 4.10: Curvas de log-intensidad de mortalidad por enfermedades respiratorias originales (línea azul discontinua), y sus respectivas estimaciones (línea roja), para las 24 últimas provincias españolas por orden alfabético

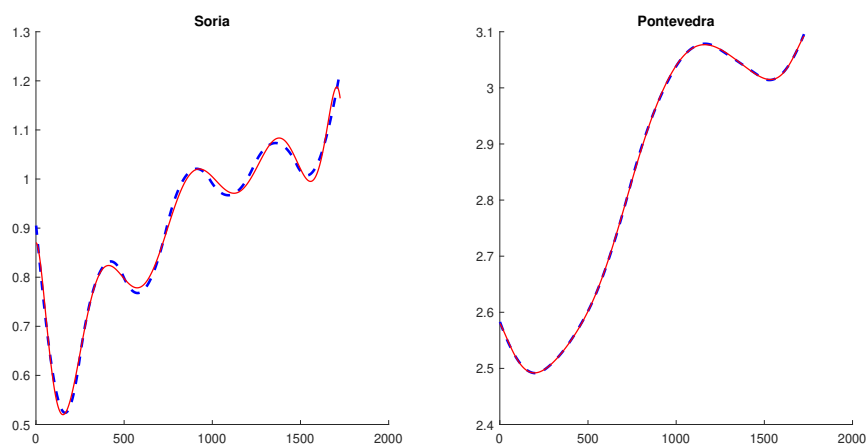


Figura 4.11: Zoom. Provincias españolas con mayor y menor norma $L^1([0, 1725])$ del error relativo absoluto medio funcional asociado (Soria a la izquierda, Pontevedra a la derecha, respectivamente). Las curvas originales aparecen en línea azul discontinua, y las curvas estimadas se representan en línea roja continua

Tabla 4.4: ACVFARE. Promedio puntual anual del CVFARE obtenido a partir de la validación

Año	ACVFARE	Año	ACVFARE	Año	ACVFARE	Año	ACVFARE
1980	0.00301	1989	0.00322	1998	0.00135	2007	0.00007
1981	0.00273	1990	0.00245	1999	0.00036	2008	0.00008
1982	0.00137	1991	0.00073	2000	0.00106	2009	0.00008
1983	0.00295	1992	0.00116	2001	0.00191	2010	0.00053
1984	0.00139	1993	0.00315	2002	0.00134	2011	0.00124
1985	0.00192	1994	0.00357	2003	0.00033	2012	0.00092
1986	0.00323	1995	0.00129	2004	0.00085	2013	0.00159
1987	0.00175	1996	0.00276	2005	0.00100	2014	0.00142
1988	0.00118	1997	0.00244	2006	0.00045	2015	0.00230

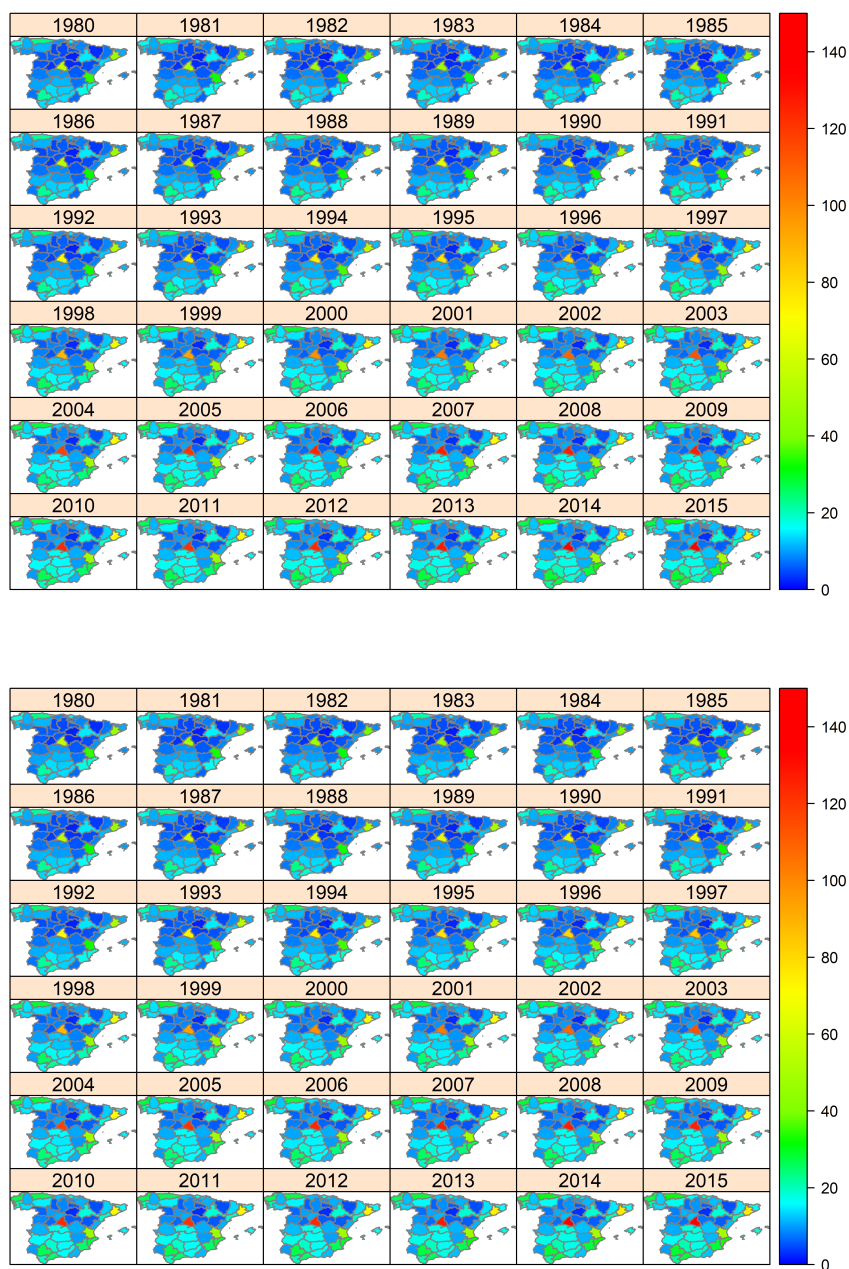


Figura 4.12: Mapas de riesgo de mortalidad por enfermedad respiratoria observada (arriba) y estimada (abajo) promediados anualmente, durante el período 1980–2015

Capítulo 5

Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo

Los datos de recuento y agregados pueden encontrarse generalmente en problemas de incidencia de enfermedades, mortalidad, dinámica de poblaciones o aparición de incendios forestales, que abarcan los campos científicos de la salud ambiental, la ecología, la epidemiología y el medio ambiente atmosférico, por mencionar sólo algunos. En tales casos, la modelización estocástica de los recuentos permite una comprensión más profunda y predicciones precisas para la evaluación y gestión de riesgos, véase [Choi et al. \(2003\)](#); [Christakos \(1992, 2000, 2017\)](#); [Christakos y Hristopulos \(1998\)](#); [Christakos y Olea \(2005\)](#); [Daley y Vere-Jones \(2008\)](#); [Diggle \(2013\)](#); [He et al. \(2020\)](#); [Illian et al. \(2008\)](#), y las referencias en las mismas.

En la mayoría de estos casos, el concepto de datos de procesos puntuales agregados (o datos agregados, para abreviar) se utiliza para referirse a los datos observados discretamente que muy probablemente surgieron a partir de un proceso puntual continuo espacial o espacio-temporal subyacente, véase [Diggle et al. \(2010a\)](#) y [Taylor et al. \(2018\)](#). Estos últimos autores sostienen que es posible ajustar un modelo discreto y obtener una inferencia continua espacial o espacio-temporal a través de la predicción espacial. Remitimos al lector a [Møller y Waagepetersen \(2004\)](#) para obtener material de fondo sobre los procesos espaciales puntuales y los detalles teóricos correspondientes.

Las principales herramientas utilizadas en la introducción de esta nueva cla-

se de procesos de Cox log-gaussianos espaciales multiescalares en espacios ℓ^2 se pueden encontrar en la Sección 5.1. Los resultados teóricos para un análisis multiescalar de la varianza funcional se encuentran en la Sección 5.2. El enfoque de estimación funcional adoptado en el dominio espectral espacial se desarrolla en la Sección 5.3. Posteriormente, en el estudio de simulación (Sección 5.4), se muestran las propiedades asintóticas de los estimadores multiescalares planteados. Finalmente, en la Sección 5.5, se ilustra la aplicación de los contenidos del capítulo sobre un ejemplo con datos reales.

Los resultados desarrollados en este capítulo se recogen en el artículo titulado: “A spatial functional count model for heterogeneity analysis in time”, actualmente aceptado para publicación en *Stochastic Environmental Research and Risk Assessment*, cuyos co-autores son A. Torres-Signes, M. P. Frías, J. Mateu, M. D. Ruiz-Medina, (Torres-Signes et al., 2021a).

5.1. Procesos de Cox log-gaussianos espaciales de dimensión infinita

Sea (Ω, \mathcal{A}, P) el espacio de probabilidad básico, en el que se definen consecuentemente todas las variables aleatorias. Denotamos por \mathcal{H} un espacio arbitrario real separable de funciones de Hilbert, con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, y norma asociada $\|\cdot\|_{\mathcal{H}}$. Sea $\mathbf{X} = \{X_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$ un campo aleatorio gaussiano de media cero estacionario, que toma valores en \mathcal{H} . Por lo tanto, $\sigma^2 = E\|X_{\mathbf{z}}\|_{\mathcal{H}}^2 < \infty$, y $P[X_{\mathbf{z}} \in \mathcal{H}] = 1$, para cada $\mathbf{z} \in \mathbb{R}^d$. Es decir, $X_{\mathbf{z}}$ define un elemento aleatorio en \mathcal{H} , para cada $\mathbf{z} \in \mathbb{R}^d$.

El operador nuclear de covarianza cruzada

$$\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}(f)(g) = E(X_{\mathbf{z}} \otimes X_{\mathbf{y}})(f)(g) = \langle E(X_{\mathbf{z}} \otimes X_{\mathbf{y}})(f), g \rangle_{\mathcal{H}}, f, g \in \mathcal{H}, \quad (5.1)$$

define la estructura de dependencia funcional espacial del campo aleatorio gaussiano infinito-dimensional. Se ha aplicado el teorema de representación de Riesz para definir $\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}(f)(g)$ como el elemento dual de $\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}(f)$ actuando sobre $g \in \mathcal{H}$, para cada $f, g \in \mathcal{H}$. Aquí, estamos restringiendo nuestra atención a la clase de operadores traza o nucleares, es decir, en el espacio $\ell^1(\mathcal{H})$, que

satisfacen

$$\|\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}\|_{\ell^1(\mathcal{H})} = \sum_{j=1}^{\infty} \left\langle \left([\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}]^* \mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}} \right)^{1/2} (\varphi_j), \varphi_j \right\rangle_{\mathcal{H}} < \infty,$$

para cualquier base ortonormal $\{\varphi_j\}_{j \geq 1}$ en \mathcal{H} .

Nota 5.1. Obsérvese que el enfoque presentado se centra en la modelización de la dependencia y la variabilidad funcional espacial (curva) en un marco en \mathcal{H} . Esta es la razón por la que, en nuestras suposiciones iniciales, se considera que se ha eliminado la tendencia del campo espacial infinito-dimensional \mathbf{X} , en virtud de la homogeneidad espacial. Remitimos al lector a la Sección 4 de Bosq y Ruiz-Medina (2014), por ejemplo, donde se adopta un enfoque componente a componente en la estimación de la media funcional de una población gaussiana infinito-dimensional, bajo los marcos clásico y bayesiano.

A partir de (5.1), $\mathcal{R}_0^{\mathbf{X}}$, con kernel $r_0^{\mathbf{X}}$, es un operador traza autoadjunto (simétrico), que satisface

$$\mathcal{R}_0^{\mathbf{X}}(\phi_j) = \lambda_j(\mathcal{R}_0^{\mathbf{X}})\phi_j, \quad j \geq 1, \quad (5.2)$$

donde $\{\phi_j, j \geq 1\}$ denota el sistema ortonormal de autovectores de $\mathcal{R}_0^{\mathbf{X}}$ en \mathcal{H} . Para cada $\mathbf{z} \in \mathbb{R}^d$, $X_{\mathbf{z}}$ admite la siguiente expansión ortogonal en $\mathcal{L}_{\mathcal{H}}^2(\Omega, \mathcal{A}, P)$, véase Angulo y Ruiz-Medina (1997),

$$\begin{aligned} X_{\mathbf{z}} &= \sum_{j=1}^{\infty} \langle X_{\mathbf{z}}, \phi_j \rangle_{\mathcal{H}} \phi_j = \sum_{j=1}^{\infty} X_{\mathbf{z}}(\phi_j) \phi_j \\ r_0^{\mathbf{X}} &=_{\mathcal{H} \otimes \mathcal{H}} \sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j. \end{aligned} \quad (5.3)$$

Es decir,

$$E \left\| X_{\mathbf{z}} - \sum_{j=1}^M \langle X_{\mathbf{z}}, \phi_j \rangle_{\mathcal{H}} \phi_j \right\|_{\mathcal{H}}^2 \rightarrow 0, \quad M \rightarrow \infty,$$

con $E [\langle X_{\mathbf{z}}, \phi_j \rangle_{\mathcal{H}} \langle X_{\mathbf{z}}, \phi_p \rangle_{\mathcal{H}}] = \delta_{j,p} \lambda_j(\mathcal{R}_0^{\mathbf{X}})$, $j \geq 1$, para cada $\mathbf{z} \in \mathbb{R}^d$. Aquí, $\delta_{j,p}$ denota la función delta de Kronecker.

Supongamos que \mathbf{X} es tal que, para cada $\mathbf{z} \in \mathbb{R}^d$, los valores funcionales de $X_{\mathbf{z}}$ casi seguramente (c. s.) tienen su soporte en el intervalo temporal acotado

$\mathcal{T} \subset \mathbb{R}_+$. Se define, para cada $\mathbf{z} \in \mathbb{R}^d$ fijo,

$$\Lambda_{\mathbf{z}}(t) = \exp(X_{\mathbf{z}}(t)) = \sum_{p=0}^{\infty} \frac{C_p}{p!} H_p(X_{\mathbf{z}}(t)), \quad \forall t \in \mathcal{T}, \quad \text{c. s.}, \quad (5.4)$$

donde la última igualdad sigue la expansión del polinomio de Hermite en el espacio $L_2(\mathbb{R}, \varphi(u)du)$, con $\varphi(u) = (1/\sqrt{2\pi}) \exp(-u^2/2)$. Aquí, para cada $p \geq 0$, H_p denota el p -ésimo polinomio de Hermite, y C_p es el coeficiente de función asociado $G(u) = \exp(u)$, por la proyección en el espacio $L_2(\mathbb{R}, \varphi(u)du)$.

La siguiente condición sobre $\Lambda_{\mathbf{z}}$, $\mathbf{z} \in \mathbb{R}^d$, permite la introducción, a partir de (5.4), de nuestro modelo funcional para el conteo espacial de la densidad aleatoria en el sentido de $\mathcal{L}_{\mathcal{H}}^p(\Omega, \mathcal{A}, P)$, $p \geq 1$.

Condición C1. Supongamos que para cualquier conjunto acotado $A \in \mathcal{B}^d$ de la σ -álgebra de Borel \mathcal{B}^d de \mathbb{R}^d , la siguiente integral es finita casi seguramente:

$$\Lambda(A) = \int_A \int_{\mathcal{T}} \Lambda_{\mathbf{z}}(t) dt d\mathbf{z} < \infty, \quad (\text{c. s.}). \quad (5.5)$$

Dadas las observaciones $\Psi_{\mathbf{z}, \omega_0} = \int_{\mathcal{T}} \Lambda_{\mathbf{z}, \omega_0}(t) dt$, $\mathbf{z} \in A \subset \mathbb{R}^d$, para cierto $\omega_0 \in \Omega$, el número de sucesos $\mathcal{C}(A)$, que ocurren durante el periodo \mathcal{T} , en la región A , sigue una distribución de probabilidad de Poisson de media $\Lambda(A)$. Nótese que, el predictor de mínimos cuadrados de $\mathcal{C}(A)$ viene dado por $\Lambda(A)$, introducido en (5.5), para cualquier conjunto de Borel acotado $A \in \mathcal{B}^d$. De (5.4), la ecuación (5.3) lleva a la siguiente expresión de la variación de segundo orden de $\Psi_{\mathbf{z}}$:

$$E[\Psi_{\mathbf{z}}^2] = \sum_{p=0}^{\infty} \int_{\mathcal{T} \times \mathcal{T}} \frac{C_p(t)C_p(s)}{p!} \left[\sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{x}}) \phi_j \otimes \phi_j(t, s) \right]^p dt ds. \quad (5.6)$$

5.2. Análisis de segundo orden espacial a diferentes escalas temporales

Consideremos el caso especial en el que $\mathcal{H} = L^2(\mathcal{T})$, el espacio de funciones de cuadrado integrable en \mathcal{T} .

Teorema 5.1. *Bajo la Condición C1, si $\{\phi_j, j \geq 1\}$, en la ecuación (5.3), están uniformemente acotadas en \mathcal{T} , $\Psi_{\mathbf{z}}$ define una densidad aleatoria de segundo orden*

espacial c. s. integrable localmente absoluta.

Demostración:

De la ecuación (5.6), aplicando la Proposición 4.9 en la p. 92 de Marinucci y Peccati (2011), después de considerar la desigualdad de Cauchy-Schwarz, en términos del producto interno introducido en la Fórmula (4.7) en la p.89 de Marinucci y Peccati (2011), las propiedades de expansión de Hermite nos llevan a

$$\begin{aligned}
 E[\Psi_{\mathbf{z}}^2] &= \sum_{p=0}^{\infty} \int_{\mathcal{T} \times \mathcal{T}} \frac{C_p(t)C_p(s)}{p!} \left[\sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j(t, s) \right]^p dt ds \\
 &\leq \sum_{p=0}^{\infty} \frac{1}{p!} \int_{\mathcal{T} \times \mathcal{T}} \sqrt{E[\Lambda_{\mathbf{z}}(t)]^2 E[\Lambda_{\mathbf{z}}(s)]^2} \sqrt{E[H_p(X_{\mathbf{z}}(t))]^2 E[H_p(X_{\mathbf{z}}(s))]^2} \\
 &\quad \times \left[\sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j(t, t) \sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j(s, s) \right]^{p/2} dt ds \\
 &= \sum_{p=0}^{\infty} \frac{1}{p!} \int_{\mathcal{T} \times \mathcal{T}} \exp(r_0(t, t)/2 + r_0(s, s)/2) [r_0(t, t)r_0(s, s)]^{p/2} \\
 &\quad \times \left[\sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j(t, t) \sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j(s, s) \right]^{p/2} dt ds \\
 &= \sum_{p=0}^{\infty} \frac{1}{p!} \left\{ \int_{\mathcal{T}} \exp(r_0(t, t)/2) \left[\sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \phi_j \otimes \phi_j(t, t) \right]^p dt \right\}^2 \\
 &\leq |\mathcal{T}|^2 \exp(2\mathcal{M}^2 \|\mathcal{R}_0^{\mathbf{X}}\|_{\ell^1(\mathcal{H})}) \left\{ \sum_{p=0}^{\infty} \frac{\mathcal{M}^{2p}}{p!} \left[\sum_{j=1}^{\infty} \lambda_j(\mathcal{R}_0^{\mathbf{X}}) \right]^p \right\}^2 \\
 &= |\mathcal{T}|^2 \exp(4\|\mathcal{R}_0^{\mathbf{X}}\|_{\ell^1(\mathcal{H})} \mathcal{M}^2) < \infty,
 \end{aligned}$$

donde $\mathcal{M} > 0$, es tal que $\sup_{t \in \mathcal{T}} |\phi_j(t)| \leq \mathcal{M}$, para cada $j \geq 1$. □

Sea $\{\psi_{j:k}, k \in \Gamma_j, j \in \mathbb{Z}\}$ una base ortonormal de wavelets, proporcionando un análisis multiresolución de $L^2(\mathcal{T})$, véase, por ejemplo, Ruiz-Medina y Angulo (2002). Para cada $\mathbf{z} \in \mathbb{R}^d$, la secuencia de coeficientes aleatorios gaussianos de media cero $\{X_{\mathbf{z}}(\psi_{j:k}), k \in \Gamma_j, j \in \mathbb{Z}\}$, con $X_{\mathbf{z}}(\psi_{j:k}) = \langle X_{\mathbf{z}}, \psi_{j:k} \rangle_{L^2(\mathcal{T})}$, $k \in \Gamma_j$, $j \in \mathbb{Z}$, tiene covarianza

$$E[X_{\mathbf{z}}(\psi_{j_1:k_1})X_{\mathbf{z}}(\psi_{j_2:k_2})] = \mathcal{R}_0^{\mathbf{X}}(\psi_{j_1:k_1})(\psi_{j_2:k_2}), \quad k \in \Gamma_{j_i}, j_i \in \mathbb{Z}, i = 1, 2,$$

que proporciona un análisis multiescalar de la estructura de dependencia curva en la localización espacial \mathbf{z} , a través del operador de autocovarianza $\mathcal{R}_0^{\mathbf{X}}$. De manera similar, para cada $\mathbf{z}, \mathbf{y} \in \mathbb{R}^d$, un análisis multiescalar es introducido por

$$E[X_{\mathbf{z}}(\psi_{j_1:k_1})X_{\mathbf{y}}(\psi_{j_2:k_2})] = \mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}(\psi_{j_1:k_1})(\psi_{j_2:k_2}), \quad k \in \Gamma_{j_i}, \quad j_i \in \mathbb{Z}, \quad i = 1, 2,$$

sobre la estructura de dependencia cruzada curva entre la localización espacial \mathbf{z} y \mathbf{y} , a través del operador de covarianza cruzada $\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}$. La estructura de covarianza de la secuencia log-gaussiana $\{\exp(X_{\mathbf{z}}(\psi_{j:k})), k \in \Gamma_j, j \in \mathbb{Z}\}$ también muestra un análisis multiescalar en el tiempo, en el espacio $L^2(\mathcal{T})$, de la estructura de dependencia espacial curva del proceso de intensidad infinito-dimensional espacial $\{\Lambda_{\mathbf{z}}(t), t \in \mathcal{T}, \mathbf{z} \in \mathbb{R}^d\}$. Nótese que, de (5.4), aplicando la identidad de Parseval en $[L^2(\mathcal{T})]^{\otimes p}$, $p \geq 1$, y la desigualdad de Cauchy-Schwarz en $L^2(\mathcal{T})$,

$$\begin{aligned} & \sum_{j=1}^{\infty} \sum_{k \in \Gamma_j} E[\exp(X_{\mathbf{z}}(\psi_{j:k})) \exp(X_{\mathbf{y}}(\psi_{j:k}))] \\ &= \sum_{j=1}^{\infty} \sum_{k \in \Gamma_j} \exp\left(\mathcal{R}_0^{\mathbf{X}}(\psi_{j:k})(\psi_{j:k}) + \frac{\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}(\psi_{j:k})(\psi_{j:k}) + \mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}}(\psi_{j:k})(\psi_{j:k})}{2}\right) \\ &= \sum_{j=1}^{\infty} \sum_{k \in \Gamma_j} \sum_{p_1, p_2, p_3} \frac{(1/2)^{p_2+p_3}}{p_1! p_2! p_3!} \sum_{h_1, \dots, h_{p_1}} \sum_{l_1, \dots, l_{p_2}} \sum_{q_1, \dots, q_{p_3}} |\lambda_{h_1}(\mathcal{R}_0^{\mathbf{X}}) \cdots \lambda_{h_{p_1}}(\mathcal{R}_0^{\mathbf{X}})| \\ & \quad \times |\lambda_{l_1}(\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}) \cdots \lambda_{l_{p_2}}(\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}})| |\lambda_{q_1}(\mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}}) \cdots \lambda_{q_{p_3}}(\mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}})| \\ & \quad \times |\phi_{h_1}(\psi_{j:k}) \cdots \phi_{h_{p_1}}(\psi_{j:k})|^2 \\ & \quad \times \left| \psi_{l_1}^{\mathbf{z}-\mathbf{y}}(\psi_{j:k}) \cdots \psi_{l_{p_2}}^{\mathbf{z}-\mathbf{y}}(\psi_{j:k}) \varphi_{l_1}^{\mathbf{z}-\mathbf{y}}(\psi_{j:k}) \cdots \varphi_{l_{p_2}}^{\mathbf{z}-\mathbf{y}}(\psi_{j:k}) \right| \\ & \quad \times \left| \psi_{q_1}^{\mathbf{y}-\mathbf{z}}(\psi_{j:k}) \cdots \psi_{q_{p_3}}^{\mathbf{y}-\mathbf{z}}(\psi_{j:k}) \varphi_{q_1}^{\mathbf{y}-\mathbf{z}}(\psi_{j:k}) \cdots \varphi_{q_{p_3}}^{\mathbf{y}-\mathbf{z}}(\psi_{j:k}) \right| \\ & \leq \sum_{p_1, p_2, p_3} \frac{(1/2)^{p_2+p_3}}{p_1! p_2! p_3!} \left[\sum_{h=1}^{\infty} |\lambda_h(\mathcal{R}_0^{\mathbf{X}})| \right]^{p_1} \left[\sum_{l=1}^{\infty} |\lambda_l(\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}})| \right]^{p_2} \left[\sum_{q=1}^{\infty} |\lambda_q(\mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}})| \right]^{p_3} \\ & = \exp\left(\|\mathcal{R}_0^{\mathbf{X}}\|_{\ell^1(\mathcal{H})} + \frac{1}{2} [\|\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}\|_{\ell^1(\mathcal{H})} + \|\mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}}\|_{\ell^1(\mathcal{H})}]\right) < \infty, \quad \forall \mathbf{z}, \mathbf{y} \in \mathbb{R}^d, \end{aligned} \tag{5.7}$$

lo que implica que la serie

$$\sum_{j=1}^{\infty} \sum_{k \in \Gamma_j} E [\exp (X_{\mathbf{z}}(\psi_{j:k}))]^2 = \sum_{j=1}^{\infty} \sum_{k \in \Gamma_j} \exp (\mathcal{R}_0^{\mathbf{X}}(\psi_{j:k})(\psi_{j:k}))$$

es convergente, para cada $\mathbf{z} \in \mathbb{R}^d$. En (5.7), hemos considerado (5.3), es decir,

$$\mathcal{R}_0^{\mathbf{X}} = \sum_{h=1}^{\infty} \lambda_h (\mathcal{R}_0^{\mathbf{X}}) \phi_h \otimes \phi_h. \quad (5.8)$$

Además, para cualquier $\mathbf{z}, \mathbf{y} \in \mathbb{R}^d$, $\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}$ y $\mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}}$ son los operadores nucleares que admiten una descomposición de valores singulares, dada por

$$\begin{aligned} \mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}} &= \sum_{l=1}^{\infty} \lambda_l (\mathcal{R}_{\mathbf{z}-\mathbf{y}}^{\mathbf{X}}) \psi_l^{\mathbf{z}-\mathbf{y}} \otimes \varphi_l^{\mathbf{z}-\mathbf{y}} \\ \mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}} &= \sum_{q=1}^{\infty} \lambda_q (\mathcal{R}_{\mathbf{y}-\mathbf{z}}^{\mathbf{X}}) \psi_q^{\mathbf{y}-\mathbf{z}} \otimes \varphi_q^{\mathbf{y}-\mathbf{z}}. \end{aligned} \quad (5.9)$$

5.3. Estimación funcional espacial multiresolución en el dominio espectral

En esta sección se presenta el enfoque de estimación funcional espacial adoptado en el dominio espectral espacial siguiendo un marco paramétrico componente a componente multiescalar. En la siguiente sección, se introduce por primera vez el modelo Hilbertiano autoregresivo espacial de orden uno, modelo SAR $\mathcal{H}(1)$, en un marco de modelos espectrales y curvas espaciales.

5.3.1. Una ecuación de espacio de estados curva espacial

Sea $\mathbf{X} = \{X_{\mathbf{z}}, \mathbf{z} \in \mathbb{R}^d\}$ el proceso curva espacial gaussiana introducido en la Sección 5.1. Sin pérdida de generalidad, restringimos nuestra atención aquí al caso $d = 2$, y $\mathcal{H} = L^2(\mathcal{T})$, $\mathcal{T} = [0, 1]$. Supongamos que \mathbf{X} obedece una ecuación de estados hilbertiana autorregresiva espacial (ecuación SAR $\mathcal{H}(1)$), como se da

en Ruiz-Medina (2011). Así,

$$X_{p,q} = Y_{p,q} - R = L_1(X_{p-1,q}) + L_2(X_{p,q-1}) + L_3(X_{p-1,q-1}) + \epsilon_{p,q}, \quad (p, q) \in \mathbb{Z}^2, \quad (5.10)$$

donde $R \in \mathcal{H}$ es la media funcional, que se estima aplicando la metodología propuesta en Bosq y Ruiz-Medina (2014), a partir de una base wavelet ortonormal de soporte compacto $\{\psi_{j:k}, k \in \Gamma_j, j \in \mathbb{Z}\}$ en $L^2([0, 1])$. Los operadores de autocorrelación $L_i, i = 1, 2, 3$, se asumen acotados en $L^2([0, 1])$. Las fluctuaciones aleatorias, introducidas por fuerza externa, se representan en términos del proceso de innovación gaussiano de media cero $L^2([0, 1])$ -valuado $\epsilon = \{\epsilon_{p,q}, (p, q) \in \mathbb{Z}^2\}$. Bajo homogeneidad espacial, este proceso muestra variación funcional constante $E\|\epsilon_{p,q}\|_{L^2([0,1])}^2 = \sigma^2$, a través de las localizaciones espaciales $(p, q) \in \mathbb{Z}^2$. La estructura de dependencia funcional espacial de \mathbf{X} se representa en términos de un operador nuclear de covarianza, dado por $\mathcal{R}_{p,q}^\epsilon = E(\epsilon_{p+k,q+l} \otimes \epsilon_{k,l}) = E(\epsilon_{p,q} \otimes \epsilon_{0,0})$, para cada $(p, q), (k, l) \in \mathbb{Z}^2$. En lo siguiente, trabajaremos bajo el supuesto de que $\{\epsilon_{p,q}, (p, q) \in \mathbb{Z}^2\}$ es un fuerte ruido blanco gaussiano en $L^2([0, 1])$. Por lo tanto, $\mathcal{R}_{p,q}^\epsilon = 0$, para $p \neq q$. En nuestro marco, la ecuación (5.10) se interpreta como la aproximación discreta de un proceso de log-intensidad funcional espacial sobre el espacio continuo, considerando valores constantes dentro de los cuadrantes de la rejilla regular que define la red de observación espacial (ver, por ejemplo, Rathbun y Cressie (1994), en el caso real-valuado). Véase también Ogata y Katsura (1988) sobre la aproximación de funciones spline, para representar la intensidad de primer orden de un marcado proceso puntual de Poisson no homogéneo.

En la implementación de nuestra estimación basada en wavelets, en el dominio espectral, de la estructura de dependencia funcional espacial de $\{\Lambda_{\mathbf{z}}(\cdot), \mathbf{z} \in \mathbb{R}^d\}$, trabajamos bajo las condiciones asumidas en las proposiciones 3 y 4 en Ruiz-Medina (2011), para la existencia de una única solución estacionaria de la ecuación (5.10); además, también consideramos la siguiente suposición.

Condición C2. $\mathcal{R}_{p,q}^{\mathbf{X}}$ es tal que $\sum_{(p,q) \in \mathbb{Z}^2} \|\mathcal{R}_{p,q}^{\mathbf{X}}\|_{l^1(\mathcal{H})} < \infty$.

Bajo la Condición C2, el operador de densidad espectral viene dado por

$$\mathcal{F}_{\omega_1, \omega_2} := \frac{1}{(2\pi)^2} \sum_{(p,q) \in \mathbb{Z}^2} \mathcal{R}_{p,q}^{\mathbf{X}} \exp(-i(p\omega_1 + q\omega_2)), \quad (\omega_1, \omega_2) \in [0, 2\pi) \times [0, 2\pi), \quad (5.11)$$

que es un operador traza autoadjunto no negativo.

Para una muestra funcional dada, de tamaño $N = S_1 \times S_2$, $\{X_{p,q}, p = 1, \dots, S_1, q = 1, \dots, S_2\}$, su Transformada de Fourier Discreta funcional (fDFT) se define como

$$\tilde{X}_{\omega_1, \omega_2}^N(\cdot) := \frac{1}{2\pi\sqrt{N}} \sum_{p=1}^{S_1} \sum_{q=1}^{S_2} X_{p,q}(\cdot) \exp(-i(p\omega_1 + q\omega_2)). \quad (5.12)$$

Esta transformación es lineal, periódica y hermitiana. Bajo condiciones *kernel* cumulantes adecuadas (véase el Teorema 2.2 en Panaretos y Tavakoli (2013b)), la fDFT (5.12) en las frecuencias $\omega_1 := \omega_{1,N} = 0$, $\omega_{2,N} := \omega_2 = \pi$, $\omega_{j,N} \in \left\{ \frac{2\pi}{N}, \dots, \frac{2\pi[(N-1)/2]^-}{N} \right\}$; $\omega_{j,N} \rightarrow \omega_j$, $N \rightarrow \infty$, $j = 3, \dots, J$, converge, cuando $N \rightarrow \infty$, a elementos gaussianos independientes en $L^2([0, 1], \mathbb{R})$, para $j = 1, 2$, y en $L^2([0, 1], \mathbb{C})$, para $j = 3, \dots, J$, con los respectivos operadores de covarianza \mathcal{F}_{ω_j} , $j = 1, \dots, J$, ver la ecuación (5.11).

De una muestra funcional de tamaño N , el operador del periodograma en la frecuencia $(\omega_1, \omega_2) \in [0, 2\pi) \times [0, 2\pi)$ viene dado por

$$\begin{aligned} \mathcal{I}_{\omega_1, \omega_2}^N(\cdot, \cdot) \\ := \sum_{p=1}^{S_1} \sum_{q=1}^{S_2} \sum_{p'=1}^{S_1} \sum_{q'=1}^{S_2} \frac{X_{p,q} \otimes X_{p',q'}(\cdot, \cdot) \exp(-i[(p-p')\omega_1 + (q-q')\omega_2])}{(2\pi)^2 N}, \end{aligned} \quad (5.13)$$

o, equivalentemente por

$$\mathcal{I}_{\omega_1, \omega_2}^N := \tilde{X}_{\omega_1, \omega_2}^N \otimes \overline{\tilde{X}_{\omega_1, \omega_2}^N}. \quad (5.14)$$

Para una base ortonormal dada de wavelets con soporte compacto $\{\psi_{j:k}, k \in$

$\Gamma_j, j \in \mathbb{Z}$ en $L^2(\mathcal{T})$, de las ecuaciones (5.12)–(5.14),

$$\tilde{X}_{\omega_1, \omega_2}^N(\psi_{j:k}) = \frac{1}{2\pi\sqrt{N}} \sum_{p=1}^{S_1} \sum_{q=1}^{S_2} X_{p,q}(\psi_{j:k}) \exp(-i(p\omega_1 + q\omega_2)), \quad k \in \Gamma_j, j \in \mathbb{Z} \quad (5.15)$$

$$\begin{aligned} \mathcal{I}_{\omega_1, \omega_2}^N(\psi_{j_1:k_1})(\psi_{j_2:k_2}) &= \tilde{X}_{\omega_1, \omega_2}^N(\psi_{j_1:k_1}) \overline{\tilde{X}_{\omega_1, \omega_2}^N(\psi_{j_2:k_2})} \\ &= \sum_{p=1}^{S_1} \sum_{q=1}^{S_2} \sum_{p'=1}^{S_1} \sum_{q'=1}^{S_2} \frac{X_{p,q}(\psi_{j_1:k_1}) X_{p',q'}(\psi_{j_2:k_2}) \exp(-i[(p-p')\omega_1 + (q-q')\omega_2])}{(2\pi)^2 N}, \end{aligned} \quad (5.16)$$

para cualquier $k_i \in \Gamma_{j_i}, j_i \in \mathbb{Z}, i = 1, 2$. Por lo tanto, se considera un análisis multiescalar temporal en el dominio espectral espacial.

5.3.2. El enfoque de estimación

A partir de (5.10)–(5.16), se define la secuencia de vectores paramétricos de coeficientes wavelets diagonales

$$\begin{aligned} \boldsymbol{\theta}_{j:k} &= (\theta_{j:k,1}, \theta_{j:k,2}, \theta_{j:k,3}) \\ &= (L_1(\psi_{j:k})(\psi_{j:k}), L_2(\psi_{j:k})(\psi_{j:k}), L_3(\psi_{j:k})(\psi_{j:k})) \in \Theta_{j:k} \subset \Theta, \quad k \in \Gamma_j, j \in \mathbb{Z} \end{aligned} \quad (5.17)$$

y la aproximación multiresolución

$$\{X_{p,q}(\psi_{j:k}), p = 1, \dots, S_1, q = 1, \dots, S_2, k \in \Gamma_j, j \in \mathbb{Z}\}$$

en el tiempo de la información muestral espacial. Obsérvese que, aquí, para cada $k \in \Gamma_j, j \in \mathbb{Z}$, $\Theta_{j:k}$ es finita, y $\Theta = \cup_{j \in \mathbb{Z}} \cup_{k \in \Gamma_j} \Theta_{j:k}$ es un conjunto compacto. Para $k \in \Gamma_j, j \geq 1$, asumimos también que el verdadero valor del parámetro $\boldsymbol{\theta}_{0,j:k}$ siempre se encuentra en el interior de $\Theta_{j:k}$, y nuestro modelo espectral espacial es identificable en el dominio wavelet.

Para cualquier nodo $k \in \Gamma_j$, a nivel de resolución $j \in \mathbb{Z}$, se puede considerar el estimador paramétrico $\hat{\boldsymbol{\theta}}_{N,j:k} = \left(\hat{\theta}_{N,j:k,1}, \hat{\theta}_{N,j:k,2}, \hat{\theta}_{N,j:k,3} \right)$ de $\boldsymbol{\theta}_{j:k} =$

$(\theta_{j:k,1}, \theta_{j:k,2}, \theta_{j:k,3})$, calculado a partir de la función de pérdida

$$\begin{aligned} K_{j:k}(\theta_{0,j:k}, \theta_{j:k}) &:= \int_{[0,2\pi) \times [0,2\pi)} f_{j:k}(\varpi, \theta_{0,j:k}) \eta_{j:k}(\varpi) \log \frac{\Psi_{j:k}(\varpi, \theta_{0,j:k})}{\Psi_{j:k}(\varpi, \theta_{j:k})} d\varpi \\ &= U_{j:k}(\theta_{j:k}) - U_{j:k}(\theta_{0,j:k}), \end{aligned} \quad (5.18)$$

donde $\theta_{0,j:k}$ denota el verdadero valor del parámetro, asociado con el nodo k a escala $j \in \Gamma_j$. La densidad espectral espacial normalizada multiescalar

$$\{\Psi_{j:k}(\varpi, \theta_{j:k}), k \in \Gamma_j, j \in \mathbb{Z}, \varpi \in [0, 2\pi) \times [0, 2\pi)\}$$

se obtiene a partir de las identidades

$$\begin{aligned} f_{j:k}(\varpi, \theta_{j:k}) &= \sigma^2(\theta_{j:k}) \Psi_{j:k}(\varpi, \theta_{j:k}) \\ &= \left[\int_{[0,2\pi) \times [0,2\pi)} f_{j:k}(\varpi, \theta_{j:k}) \eta_{j:k}(\varpi) d\varpi \right] \Psi_{j:k}(\varpi, \theta_{j:k}) \\ f_{j:k}(\varpi, \theta_{j:k}) &= \frac{\sigma_{\epsilon(\psi_{j:k})}^2}{2\pi^2} \left| 1 - \theta_{j:k,1} e^{i\varpi_1} - \theta_{j:k,2} e^{i\varpi_2} - \theta_{j:k,3} e^{i(\varpi_1 + \varpi_2)} \right|^{-2}, \end{aligned} \quad (5.19)$$

para cada $\varpi = (\varpi_1, \varpi_2) \in [0, 2\pi) \times [0, 2\pi)$, siendo $\eta_{j:k}(\varpi)$ una función espacial simétrica no negativa, para cada $k \in \Gamma_j$, y $j \in \mathbb{Z}$, tal que $\eta_{j:k}(\varpi) f_{j:k}(\varpi, \theta_{j:k}) \in L_1([0, 2\pi) \times [0, 2\pi))$, el espacio de funciones integrables absolutas en $[0, 2\pi) \times [0, 2\pi)$, para cada $\theta_{j:k} \in \Theta_{j:k} \subset \Theta$.

Las funciones de pérdida en (5.18) miden la discrepancia, a diferentes niveles de resolución temporal, entre el verdadero modelo paramétrico espectral espacial $\Psi_{j:k}(\varpi, \theta_{0,j:k})$, subyacente a los datos, y los candidatos paramétricos $\Psi_{j:k}(\varpi, \theta_{j:k})$, $\theta_{j:k} \in \Theta_{j:k} \subset \Theta$, en el nodo k , dentro de la escala de variación temporal $j \in \mathbb{Z}$. En la última identidad de la ecuación (5.18), para cada escala $j \in \mathbb{Z}$,

$$U_{j:k}(\theta_{j:k}) := - \int_{[0,2\pi) \times [0,2\pi)} f_{j:k}(\varpi, \theta_{0,j:k}) \eta_{j:k}(\varpi) \log \Psi_{j:k}(\varpi, \theta_{j:k}) d\varpi, \quad k \in \Gamma_j. \quad (5.20)$$

En la práctica, podemos entonces considerar el *funcional multiescalar empírico*

$$\hat{U}_{N,j:k}(\boldsymbol{\theta}_{j:k}) := - \int_{[0,2\pi) \times [0,2\pi)} I_{N,j:k}(\boldsymbol{\varpi}) \eta_{j:k}(\boldsymbol{\varpi}) \log \Psi_{j:k}(\boldsymbol{\varpi}, \boldsymbol{\theta}_{j:k}) d\boldsymbol{\varpi}, \quad (5.21)$$

donde $I_{N,j:k}(\boldsymbol{\varpi}) = \mathcal{I}_{\boldsymbol{\varpi}_1, \boldsymbol{\varpi}_2}^N(\psi_{j:k})(\psi_{j:k})$ denota, como antes, el periodograma multiescalar presentado en (5.16), para $k \in \Gamma_j$, y $j \in \mathbb{Z}$.

Para cada $k \in \Gamma_j$, y $j \in \mathbb{Z}$, $\eta_{j:k}$ debe satisfacer condiciones adecuadas (véase el Teorema 2.1), de tal manera que la función de pérdida (5.18) tenga un mínimo en el verdadero valor del parámetro, para cada nodo en cualquier escala, y se dé el siguiente comportamiento asintótico (véase el Capítulo 2):

$$\hat{U}_{N,j:k}(\boldsymbol{\theta}_{j:k}) - \hat{U}_{N,j:k}(\boldsymbol{\theta}_{0,j:k}) \rightarrow_{P_{0,j:k}} K_{j:k}(\boldsymbol{\theta}_{0,j:k}, \boldsymbol{\theta}_{j:k}), \quad N \rightarrow \infty, \quad (5.22)$$

para cada $\boldsymbol{\theta}_{j:k} \in \Theta_{j:k} \subset \Theta$, donde $P_{0,j:k}$ denota la medida asociada con función de densidad $f_{j:k}(\boldsymbol{\varpi}, \boldsymbol{\theta}_{0,j:k})$, para cada $k \in \Gamma_j$, y $j \in \mathbb{Z}$. Para minimizar la divergencia en (5.18), en la práctica, podemos calcular el mínimo de $\hat{U}_{N,j:k}(\boldsymbol{\theta}_{j:k})$ sobre $\boldsymbol{\theta}_{j:k} \in \Theta_{j:k}$, a través de los diferentes nodos k en cada escala $j \in \mathbb{Z}$. Es decir, consideraremos los estimadores paramétricos multiescalares

$$\hat{\boldsymbol{\theta}}_{N,j:k} = \arg \min_{\boldsymbol{\theta}_{j:k} \in \Theta_{j:k}} \hat{U}_{N,j:k}(\boldsymbol{\theta}_{j:k}), \quad k \in \Gamma_j, \quad j \in \mathbb{Z}. \quad (5.23)$$

El mismo procedimiento de estimación, basado en el periodograma multiescalar en (5.16), se aplica para el resto de los coeficientes en las transformaciones wavelet bidimensionales de los operadores L_i , $i = 1, 2, 3$, incluyendo los coeficientes de la función escala, con respecto a la base $\{\varphi_{j_0, \tilde{k}}, \tilde{k} \in \Upsilon_{j_0}\}$ del espacio $V_0 \subset L^2([0, 1])$. Es decir, se calculan estimadores similares para los

parámetros

$$\begin{aligned}\boldsymbol{\theta}_{j_0:\tilde{k},\tilde{l}} &= \left(\theta_{j_0:\tilde{k},\tilde{l},1}, \theta_{j_0:\tilde{k},\tilde{l},2}, \theta_{j_0:\tilde{k},\tilde{l},3} \right) \\ &= \left(L_1(\varphi_{j_0:\tilde{l}})(\varphi_{j_0:\tilde{k}}), L_2(\varphi_{j_0:\tilde{l}})(\varphi_{j_0:\tilde{k}}), L_3(\varphi_{j_0:\tilde{l}})(\varphi_{j_0:\tilde{k}}) \right) \in \Theta_{j_0:\tilde{k}} \times \Theta_{j_0:\tilde{l}}, \\ \Theta_{j_0:\tilde{k}} \times \Theta_{j_0:\tilde{l}} &\subset \Theta \times \Theta, \quad \tilde{k}, \tilde{l} \in \Upsilon_{j_0}.\end{aligned}$$

$$\begin{aligned}\boldsymbol{\theta}_{j_0:\tilde{k};j:k} &= \left(\theta_{j_0:\tilde{k};j:k,1}, \theta_{j_0:\tilde{k};j:k,2}, \theta_{j_0:\tilde{k};j:k,3} \right) \\ &= \left(L_1(\psi_{j:k})(\varphi_{j_0:\tilde{k}}), L_2(\psi_{j:k})(\varphi_{j_0:\tilde{k}}), L_3(\psi_{j:k})(\varphi_{j_0:\tilde{k}}) \right) \in \Theta_{j_0:\tilde{k}} \times \Theta_{j:k}, \\ \Theta_{j_0:\tilde{k}} \times \Theta_{j:k} &\subset \Theta \times \Theta, \quad \tilde{k} \in \Upsilon_{j_0}; k \in \Gamma_j, j \geq j_0.\end{aligned}$$

$$\begin{aligned}\boldsymbol{\theta}_{j:k;j_0:\tilde{k}} &= \left(\theta_{j:k;j_0:\tilde{k},1}, \theta_{j:k;j_0:\tilde{k},2}, \theta_{j:k;j_0:\tilde{k},3} \right) \\ &= \left(L_1(\varphi_{j_0:\tilde{k}})(\psi_{j:k}), L_2(\varphi_{j_0:\tilde{k}})(\psi_{j:k}), L_3(\varphi_{j_0:\tilde{k}})(\psi_{j:k}) \right) \in \Theta_{j:k} \times \Theta_{j_0:\tilde{k}}, \\ \Theta_{j:k} \times \Theta_{j_0:\tilde{k}} &\subset \Theta \times \Theta, \quad k \in \Gamma_j, j \geq j_0; \tilde{k} \in \Upsilon_{j_0}.\end{aligned}$$

$$\begin{aligned}\boldsymbol{\theta}_{j:k;l:h} &= (\theta_{j:k;l:h,1}, \theta_{j:k;l:h,2}, \theta_{j:k;l:h,3}) \\ &= (L_1(\psi_{j:k})(\psi_{l:h}), L_2(\psi_{j:k})(\psi_{l:h}), L_3(\psi_{j:k})(\psi_{l:h})) \in \Theta_{j:k} \times \Theta_{l:h}, \\ \Theta_{j:k} \times \Theta_{l:h} &\subset \Theta \times \Theta, \quad k \in \Gamma_j, j \geq j_0; h \in \Gamma_l, l \geq j_0; (j, k) \neq (h, l).\end{aligned}$$

(5.24)

Para cualquier localización espacial (p, q) , se calcula el predictor *plug-in* SAR $\ell^2(1)$ multiescalar resultante se calcula, como

$$\begin{aligned}
\widehat{X}_{N,p,q}(\cdot) = & \sum_{\tilde{k} \in \Upsilon_{j_0}} \sum_{\tilde{l} \in \Upsilon_{j_0}} \widehat{\theta}_{N,j_0:\tilde{k},\tilde{l},1} X_{p-1,q}(\varphi_{j_0:\tilde{l}}) \varphi_{j_0:\tilde{k}}(\cdot) \\
& + \sum_{\tilde{k} \in \Upsilon_{j_0}} \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \widehat{\theta}_{N,j_0:\tilde{k};j:k,1} X_{p-1,q}(\psi_{j:k}) \varphi_{j_0:\tilde{k}}(\cdot) \\
& + \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \sum_{\tilde{k} \in \Upsilon_{j_0}} \widehat{\theta}_{N,j:k;j_0:\tilde{k},1} X_{p-1,q}(\varphi_{j_0:\tilde{k}}) \psi_{j:k}(\cdot) \\
& + \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \sum_{l \geq j_0} \sum_{h \in \Gamma_l} \widehat{\theta}_{N,j:k;l:h,1} X_{p-1,q}(\psi_{l:h}) \psi_{j:k}(\cdot) \\
& + \sum_{\tilde{k} \in \Upsilon_{j_0}} \sum_{\tilde{l} \in \Upsilon_{j_0}} \widehat{\theta}_{N,j_0:\tilde{k},\tilde{l},2} X_{p,q-1}(\varphi_{j_0:\tilde{l}}) \varphi_{j_0:\tilde{k}}(\cdot) \\
& + \sum_{\tilde{k} \in \Upsilon_{j_0}} \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \widehat{\theta}_{N,j_0:\tilde{k};j:k,2} X_{p,q-1}(\psi_{j:k}) \varphi_{j_0:\tilde{k}}(\cdot) \\
& + \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \sum_{\tilde{k} \in \Upsilon_{j_0}} \widehat{\theta}_{N,j:k;j_0:\tilde{k},2} X_{p,q-1}(\varphi_{j_0:\tilde{k}}) \psi_{j:k}(\cdot) \\
& + \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \sum_{l \geq j_0} \sum_{h \in \Gamma_l} \widehat{\theta}_{N,j:k;l:h,2} X_{p,q-1}(\psi_{l:h}) \psi_{j:k}(\cdot) \\
& + \sum_{\tilde{k} \in \Upsilon_{j_0}} \sum_{\tilde{l} \in \Upsilon_{j_0}} \widehat{\theta}_{N,j_0:\tilde{k},\tilde{l},3} X_{p-1,q-1}(\varphi_{j_0:\tilde{l}}) \varphi_{j_0:\tilde{k}}(\cdot) \\
& + \sum_{\tilde{k} \in \Upsilon_{j_0}} \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \widehat{\theta}_{N,j_0:\tilde{k};j:k,3} X_{p-1,q-1}(\psi_{j:k}) \varphi_{j_0:\tilde{k}}(\cdot) \\
& + \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \sum_{\tilde{k} \in \Upsilon_{j_0}} \widehat{\theta}_{N,j:k;j_0:\tilde{k},3} X_{p-1,q-1}(\varphi_{j_0:\tilde{k}}) \psi_{j:k}(\cdot) \\
& + \sum_{j \geq j_0} \sum_{k \in \Gamma_j} \sum_{l \geq j_0} \sum_{h \in \Gamma_l} \widehat{\theta}_{N,j:k;l:h,3} X_{p-1,q-1}(\psi_{l:h}) \psi_{j:k}(\cdot).
\end{aligned} \tag{5.25}$$

En la práctica, seleccionamos un número finito D de escalas, de acuerdo con el tamaño del paso de discretización adoptado en el tiempo, en el procedimiento de preprocesamiento involucrado en la construcción de nuestro conjunto de datos curva. Nótese que, como se ha comentado antes, para una escala dada $j \in \{1, \dots, D\}$, el correspondiente número de nodos $k(j)$ es finito.

5.4. Estudio de simulación

Para ilustrar las propiedades asintóticas de los estimadores multiescales formulados, se ha considerado una secuencia de tamaño muestral de curva espacial creciente, $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$. Para nuestra implementación se ha seleccionado el sistema wavelet de Haar, véase, por ejemplo, Daubechies (1992). En particular, consideramos $L_3 = -L_1L_2$, y, como antes, $\mathcal{T} = [0, 1]$. Los operadores L_1 y L_2 se definen en términos de los autovectores comunes,

$$\phi_p(t) = \text{sen}(\pi pt), \quad t \in (0, 1), \quad p \geq 1, \quad (5.26)$$

con $\phi_p(0) = \phi_p(1) = 0$. Los correspondientes sistemas de autovalores $\{\lambda_{pl}, p \geq 1, l = 1, 2\}$ satisfacen las condiciones (i)–(iii) de la Proposición 3 en Ruiz-Medina (2011), para la existencia de una solución estacionaria única de la ecuación SAR $\mathcal{H}(1)$. Nótese que las condiciones asumidas en el Teorema 5.1, y en la Condición C2 también se mantienen, bajo este escenario. En la descomposición ortogonal (5.3), hemos considerado el parámetro de truncamiento $k_N = k_{22500} = [\ln(N)]^- = [\ln(22500)]^- = 10$, donde hemos seleccionado el caso más desfavorable (es decir, el mayor orden de truncamiento correspondiente al tamaño de la muestra funcional $N = 22500$). La Tabla 5.1 muestra los $k_N = 10$ autovalores $\{\lambda_{pi}, p = 1, \dots, 10\}$ de los operadores $L_i, i = 1, 2$.

Tabla 5.1: Autovalores $\lambda_{p1}, \lambda_{p2}, p = 1, \dots, 10$

	λ_{p1}^0	λ_{p2}^0
$p = 1$	0.300	0.500
$p = 2$	0.270	0.470
$p = 3$	0.230	0.430
$p = 4$	0.200	0.400
$p = 5$	0.170	0.370
$p = 6$	0.130	0.330
$p = 7$	0.100	0.300
$p = 8$	0.030	0.230
$p = 9$	0.010	0.200
$p = 10$	0.005	0.150.

Las propiedades muestrales a gran escala (el borrador) de \mathbf{X} se obtienen por

su proyección en el espacio $V_0 \subset L^2([0, 1])$, generado por las funciones de escala

$$\left\{ \varphi_{j_0: \tilde{k}}, \tilde{k} \in \Upsilon_{j_0} \right\}$$

en la escala más burda j_0 . La variabilidad local de la muestra (*detalles*) de \mathbf{X} se reproduce a diferentes niveles de resolución, por su proyección en los subespacios $W_j \subset L^2([0, 1])$, $j = j_0, \dots, D$, generados por las bases wavelet $\{\psi_{j:k}, k \in \Gamma_j\}$, $j = j_0, \dots, D$, respectivamente. Las figuras 5.1–5.4 muestran la variabilidad temporal representada a diferentes escalas de los datos curva generados, en algunos de los nodos ($N = 900$) de una rejilla regular espacial 30×30 .

En la estimación de los parámetros multiescalares (5.17) y (5.24), las ecuaciones (5.16)–(5.23), y sus contrapartes no diagonales son respectivamente calculadas. La función η es constante en los nodos de las D escalas consideradas en la transformación wavelet bidimensional de los operadores L_i , $i = 1, 2$. En particular, la elección $\eta(\varpi) = |\varpi_1|^2 |\varpi_2|^2$ se ha hecho, para cada $\varpi = (\varpi_1, \varpi_2) \in [0, 2\pi) \times [0, 2\pi)$. El promedio por escala de los errores cuadráticos medios empíricos, asociados con los estimadores paramétricos multiescalares de (5.17) y (5.24), para $j_0 = 6$, $D = 9$, basado en 100 generaciones de las muestras funcionales de tamaño

$$N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500,$$

se muestran en las tablas 5.2–5.3.

La Figura 5.5 muestra los errores cuadráticos medios empíricos, asociados a las estimaciones $\left\{ \hat{\lambda}_{N,p,1}, \hat{\lambda}_{N,p,2}, p = 1, \dots, k_N \right\}$ del espectro de puntos puros de L_1 y L_2 , calculados a partir de las reconstrucciones wavelet bidimensionales empíricas de L_1 y L_2 a escala $D = 10$, basadas en 100 realizaciones de los estimadores paramétricos multiescalares. Los diagramas de cajas de sus valores muestrales se pueden ver en la Figura 5.6. Finalmente, los verdaderos operadores L_1 , L_2 , y sus estimaciones funcionales, a escala $j = 7, 8, 9, 10$, se muestran en la Figura 5.7. Los gráficos de contorno de la Figura 5.9 proporcionan la descripción múltiescala (escalas 7–10) del campo de log-intensidad espacial, original y estimado, \mathbf{X} , en $t = 1/2$. Al mismo tiempo, la Figura 5.8 muestra los valores de log-intensidad suavizada, original y estimada, a diferentes escalas o niveles de resolución ($j = 7, 8, 9, 10$). Se puede observar el efecto del procedimiento de

Capítulo 5. Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo

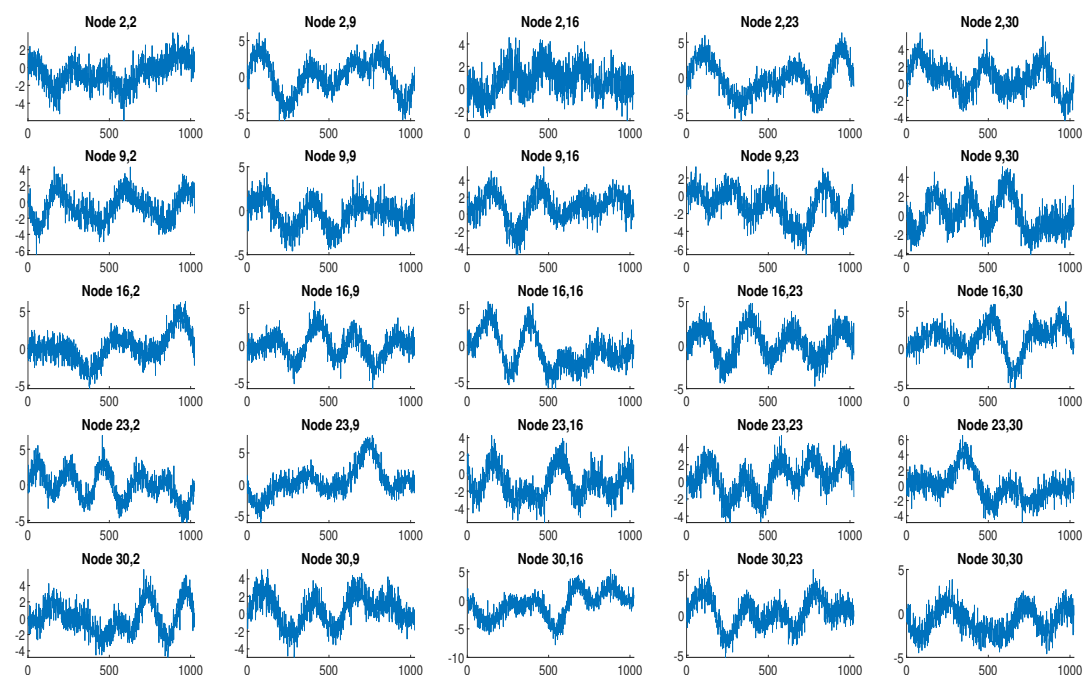


Figura 5.1: Escala 10, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30

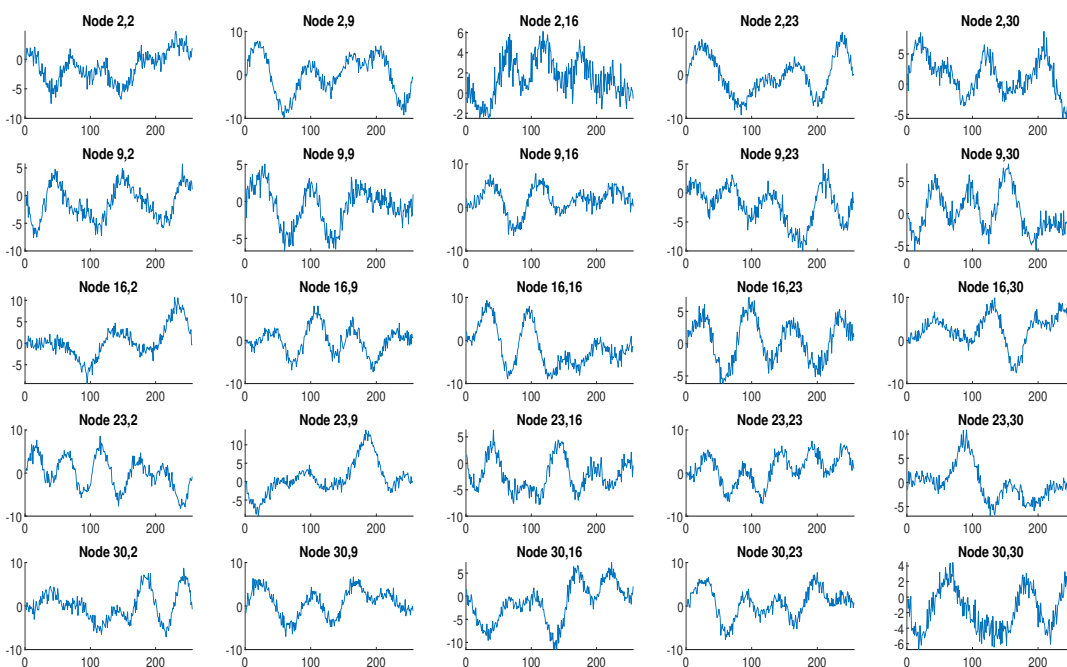


Figura 5.2: Escala 9, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30

Capítulo 5. Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo

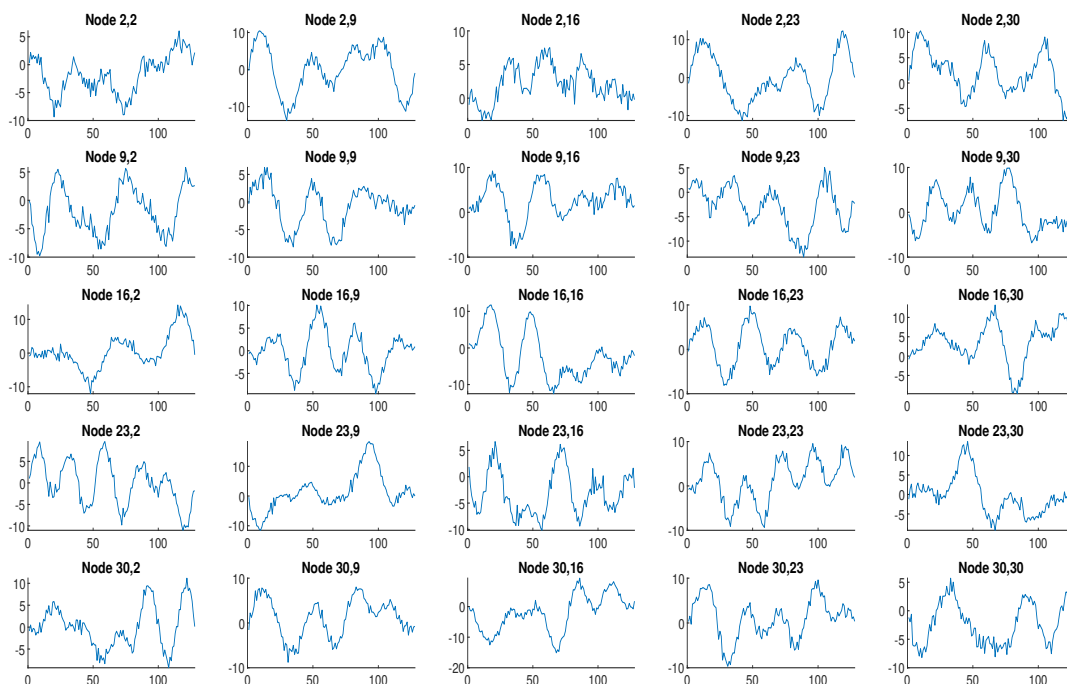


Figura 5.3: Escala 8, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30

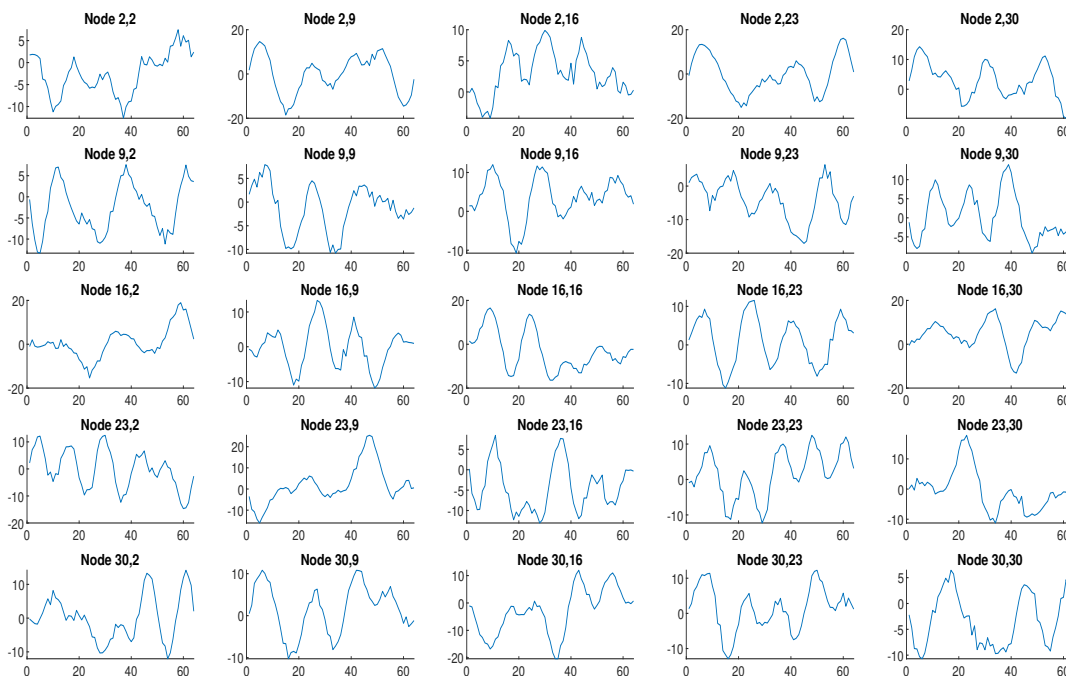


Figura 5.4: Escala 7, $N = 900$. Datos curva sobre algunos nodos de una rejilla regular espacial 30×30

Tabla 5.2: Promedio por escala de los errores cuadráticos medios empíricos asociados a los estimadores paramétricos implicados en la aproximación de L_1 , a partir de 100 generaciones de las muestras funcionales de tamaño $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$. Se consideran las escalas $j = 6, 7, 8, 9$, y el parámetro de truncamiento $k_N = 10$

N	Scale 6	Scale 7	Scale 8	Scale 9
100	6.828e-02	4.158e-03	2.606e-04	1.630e-05
900	9.171e-03	5.678e-04	3.558e-05	2.226e-06
2500	3.540e-03	2.174e-04	1.362e-05	8.520e-07
4900	1.930e-03	1.309e-04	8.205e-06	5.132e-07
8100	1.151e-03	7.640e-05	4.789e-06	2.995e-07
12100	9.036e-04	5.234e-05	3.280e-06	2.052e-07
16900	6.120e-04	3.712e-05	2.326e-06	1.455e-07
22500	4.663e-04	3.066e-05	1.921e-06	1.202e-07

Tabla 5.3: Promedio por escala de los errores cuadráticos medios empíricos asociados a los estimadores paramétricos implicados en la aproximación de L_2 , a partir de 100 generaciones de las muestras funcionales de tamaño $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$. Se consideran las escalas $j = 6, 7, 8, 9$, y el parámetro de truncamiento $k_N = 10$

N	Scale 6	Scale 7	Scale 8	Scale 9
100	7.080e-02	4.476e-03	2.806e-04	1.755e-05
900	9.360e-03	5.916e-04	3.708e-05	2.319e-06
2500	3.642e-03	2.302e-04	1.443e-05	9.025e-07
4900	1.787e-03	1.129e-04	7.078e-06	4.427e-07
8100	1.093e-03	6.909e-05	4.330e-06	2.708e-07
12100	9.795e-04	6.188e-05	3.878e-06	2.425e-07
16900	6.367e-04	4.022e-05	2.521e-06	1.577e-07
22500	4.472e-04	2.825e-05	1.771e-06	1.107e-07

preprocesamiento del Análisis Funcional de Datos (FDA), y el efecto de aumentar el número de nodos espaciales, cuando se comparan los patrones espaciales observados a diferentes escalas temporales en las figuras 5.8 y 5.9.

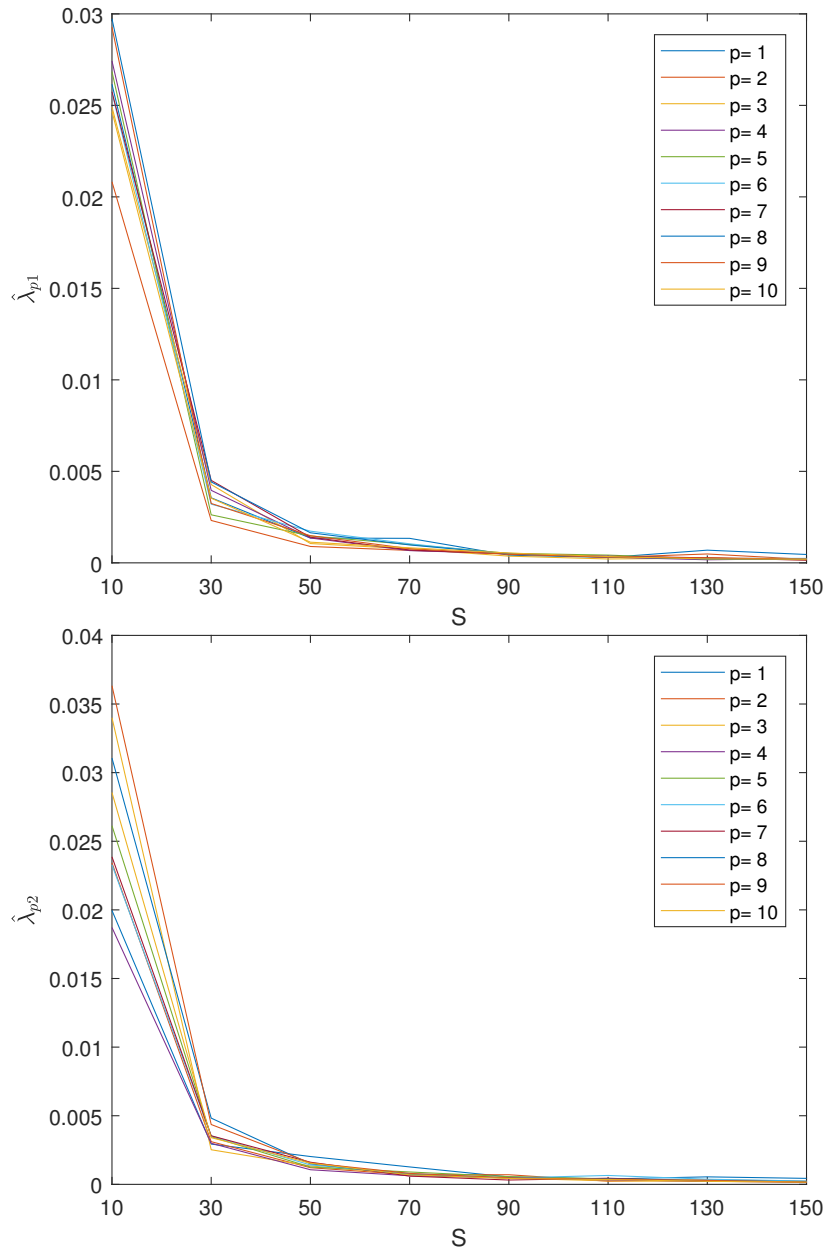


Figura 5.5: Errores cuadráticos medios empíricos (EMSE), asociados con $\{\hat{\lambda}_{N,p,1}, \hat{\lambda}_{N,p,2}, p = 1, \dots, k_N\}$, $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$

Capítulo 5. Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo

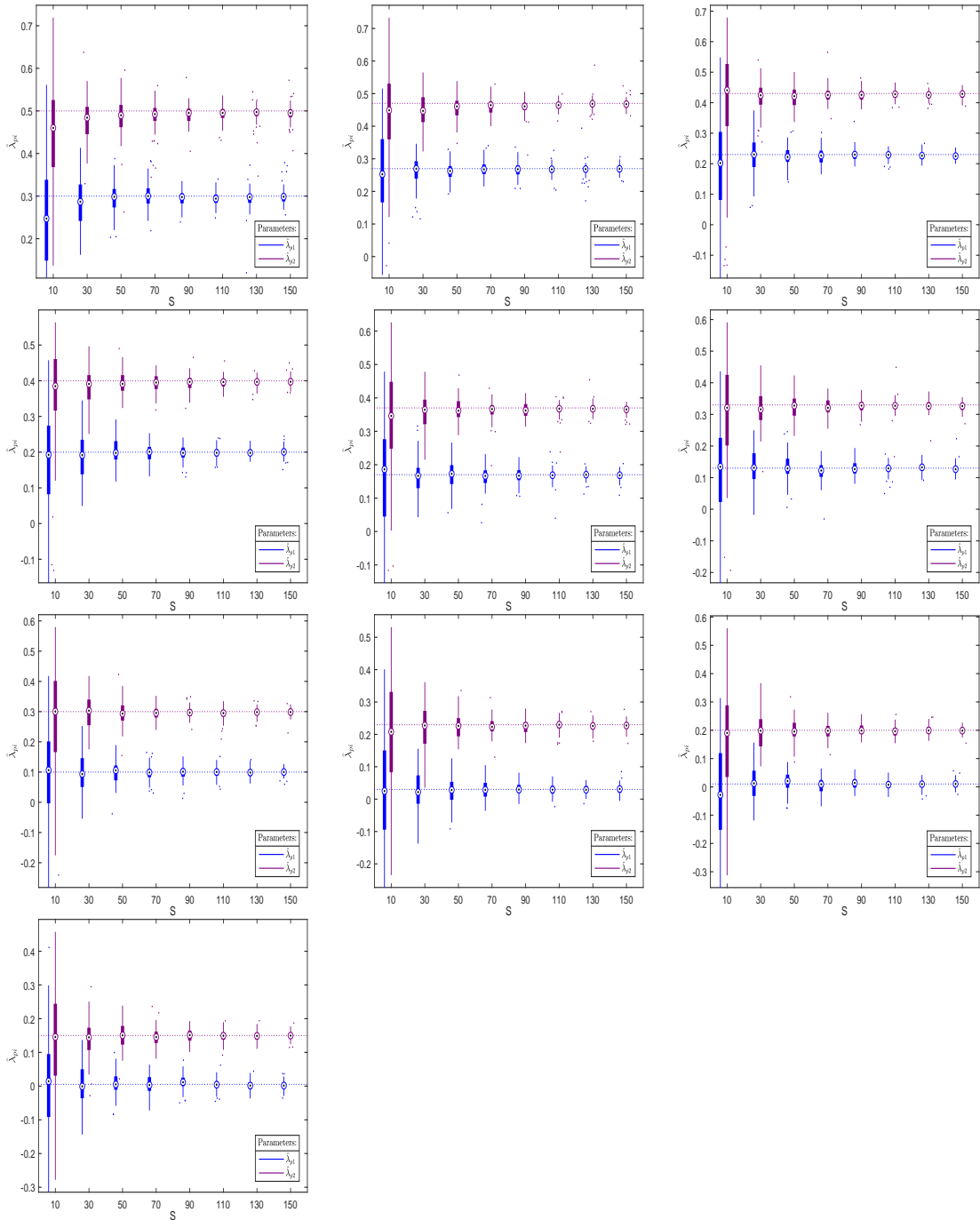


Figura 5.6: Diagrama de cajas con los valores muestrales de $\hat{\lambda}_{N,p,i}$, $p = 1, \dots, 10$, $i = 1, 2$ de izquierda a derecha, y de arriba a abajo), basados en 100 generaciones de muestras funcionales de tamaño $N = 100, 900, 2500, 4900, 8100, 12100, 16900, 22500$. El verdadero valor del parámetro se refleja en la línea punteada

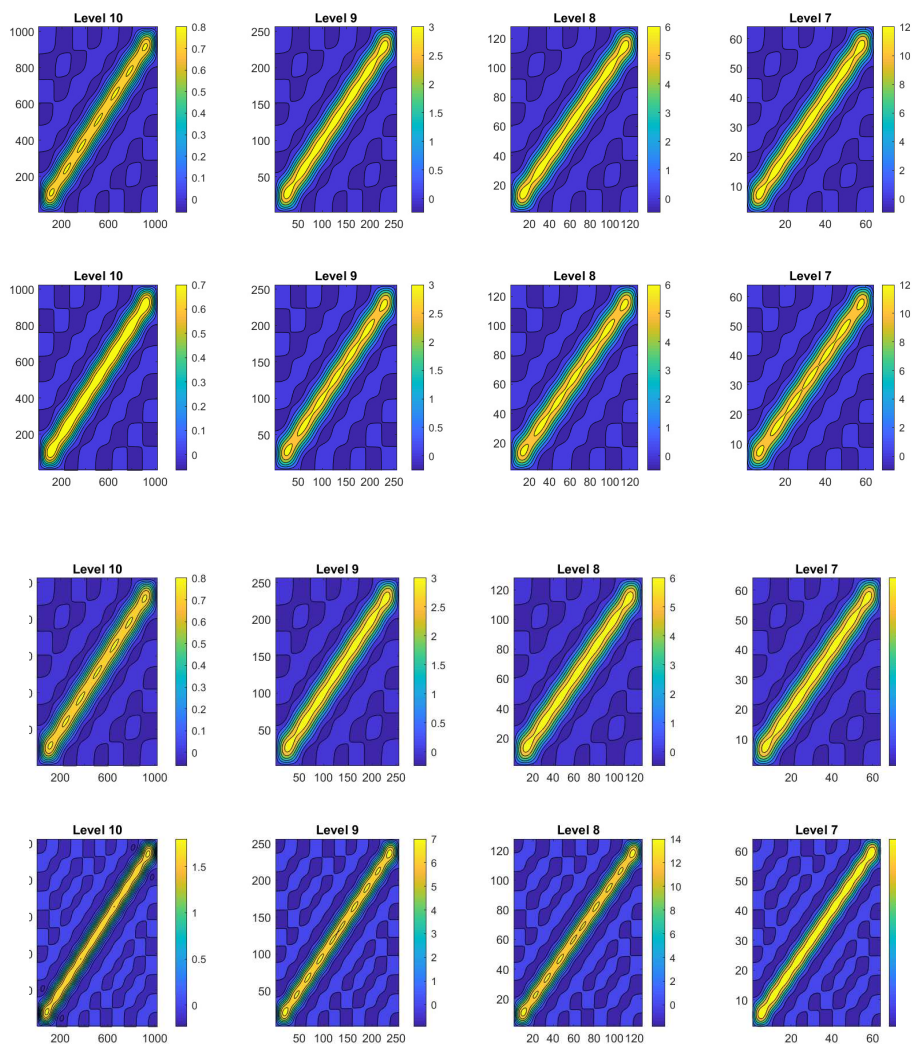


Figura 5.7: Verdadero L_1 en la fila superior, y su estimación multiscalar en la segunda fila, para las escalas $j = 7, 8, 9, 10$ (de derecha a izquierda). Verdadero L_2 en la tercera fila, y su estimación multiscalar en la fila inferior, para las escalas $j = 7, 8, 9, 10$ (de derecha a izquierda), sobre una rejilla regular espacial 30×30

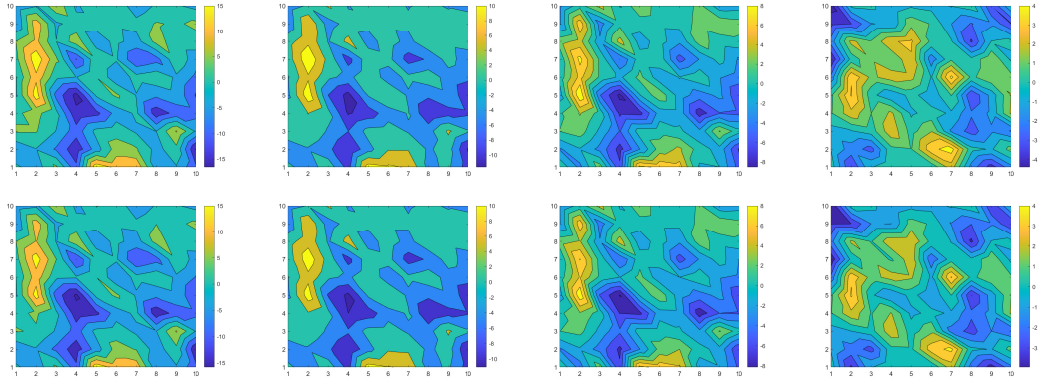


Figura 5.8: Campo X de log-intensidad espacial original (fila superior) y estimado (fila inferior), para el tiempo $t = 1/2$, a través de las escalas $j = 7, 8, 9, 10$ (de izquierda a derecha), sobre una rejilla regular espacial 10×10 , a partir de los datos curva suavizados

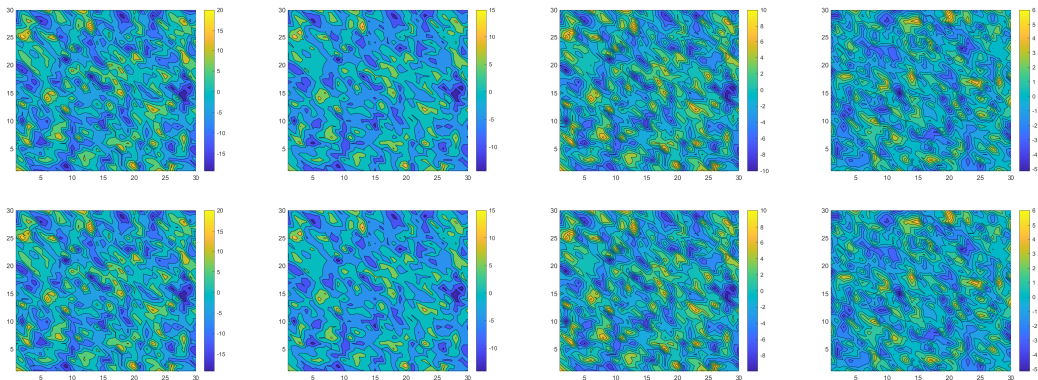


Figura 5.9: Campo X de log-intensidad espacial original (fila superior) y estimado (fila inferior), para el tiempo $t = 1/2$, a través de las escalas $j = 7, 8, 9, 10$ (de izquierda a derecha), sobre una rejilla regular espacial 30×30 , a partir de los datos curva suavizados

5.5. Ejemplo de datos reales

El Instituto Nacional de Estadística proporcionó los datos sobre los casos observados de muertes por enfermedades respiratorias, consistentes en 432 registros mensuales, en el período 1980-2015, recogidos en 48 provincias españolas. Los datos son temporales y espaciales interpolados en una rejilla regular 20×20 . En concreto, se consideran 1725 nodos temporales y 400 nodos espaciales. Se obtiene un ajuste flexible del comportamiento (o singularidad) local subyacente de los datos observados e interpolados, a partir de una elección adecuada de la escala o nivel de resolución (ver figuras 5.10–5.11). Nótese que el preprocesamiento de FDA normalmente lleva a un sobresuavizado. Ese es el caso del suavizado mediante B-splines que se aplica a menudo para construir conjuntos de datos curva (véase la Figura 5.12).

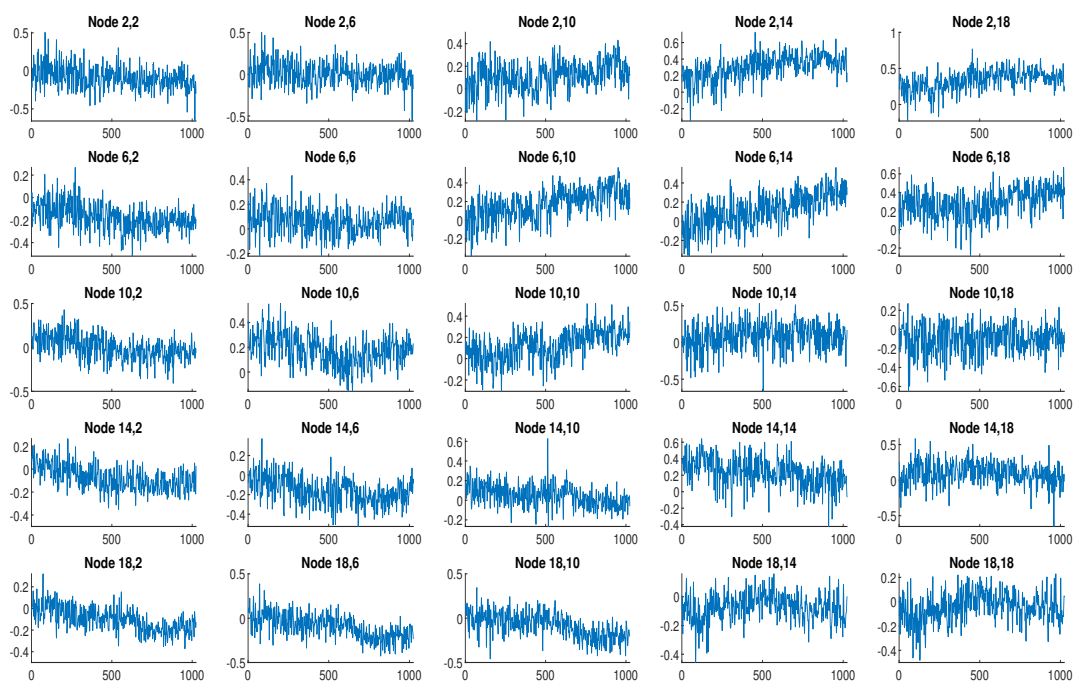


Figura 5.10: Datos temporales y espaciales interpolados en una rejilla regular espacial 20×20

Capítulo 5. Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo

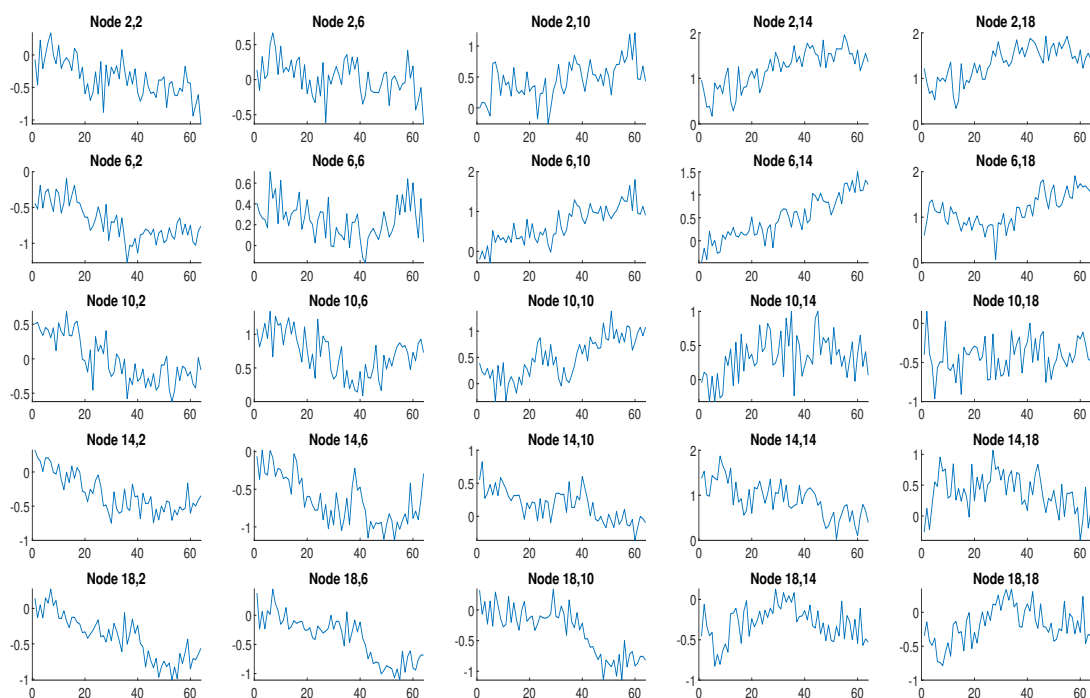


Figura 5.11: Datos temporales y espaciales interpolados en una rejilla regular espacial 20×20 a escala (nivel de resolución) 7

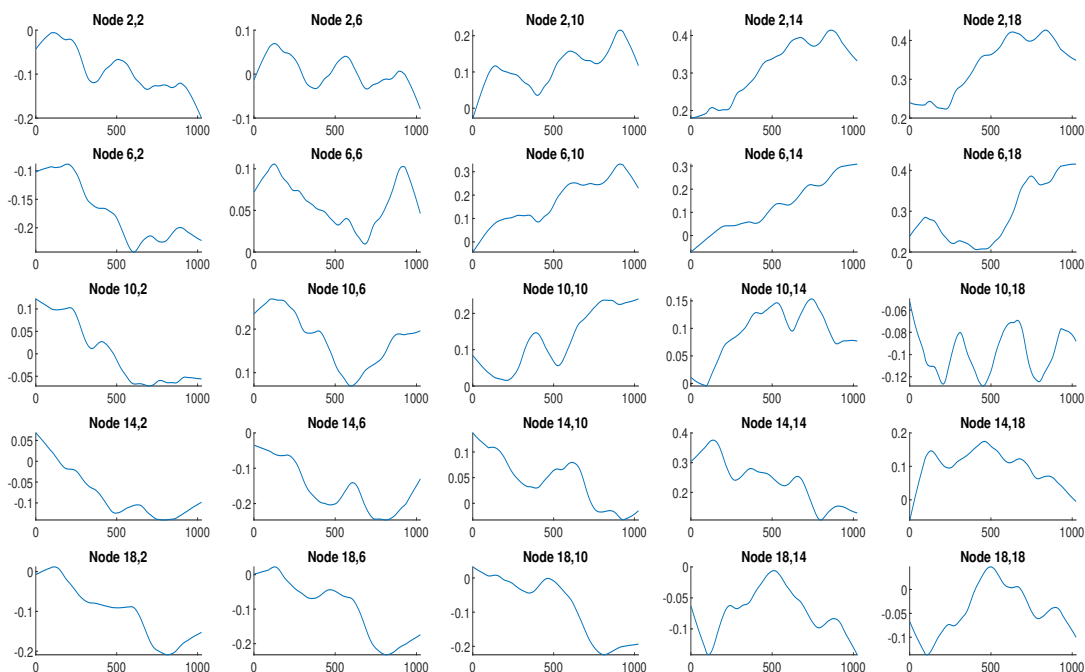


Figura 5.12: Datos curva suavizados mediante B-splines en una rejilla regular espacial 20×20

5.5.1. Estimación multiescalar

Las ecuaciones (5.17)–(5.25) se implementan en términos de los autovectores empíricos, y la base de wavelets de Haar. Las estimaciones calculadas a escalas (niveles de resolución) $j = 7, 8, 9, 10$, de los operadores de autocorrelación L_1 y L_2 se pueden ver en la Figura 5.13, para $k_N = [\ln(N)]^- = [\ln(400)]^- = 5 = k_{400}$. Los gráficos de contorno en la Figura 5.14 muestran los patrones espaciales del campo de log-intensidad observada y estimada sobre una rejilla regular espacial 20×20 , en los valores mensuales $t = 108$ y $t = 216$, a través de las escalas $j = 7, 8, 9, 10$. En este caso, el análisis multiescalar se ha llevado a cabo a partir de los datos interpolados no suavizados. La Figura 5.15 muestra los valores originales y estimados del campo de log-intensidad sobre los mismos nodos temporales y espaciales, a partir de los datos curva suavizados mediante B-splines. Se puede observar la pérdida de información en la Figura 5.15, sobre la variabilidad espacial mostrada por el campo de log-intensidad a escalas $j = 9, 10$, con respecto a la Figura 5.14. Así, se observan patrones espaciales similares, a escalas $j = 7, 8, 9, 10$, cuando se consideran los datos curva suavizados mediante B-splines, ocultando las heterogeneidades que el campo de log-intensidad presenta a través de diferentes escalas.

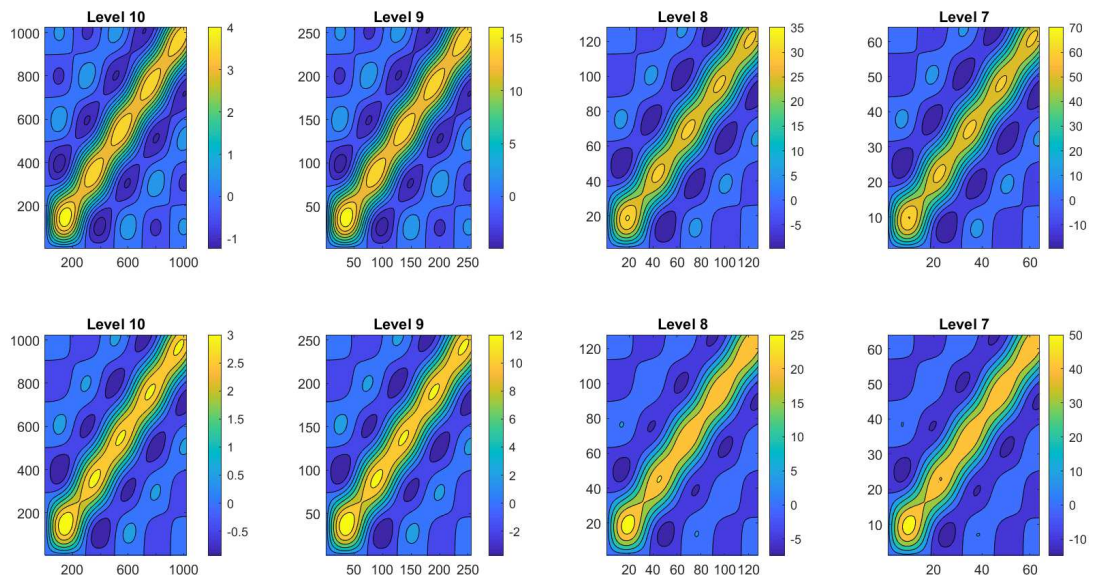


Figura 5.13: Estimación multiescalar del operador L_1 (arriba), y L_2 (abajo) a través de escalas (niveles de multiresolución) $j = 7, 8, 9, 10$ (de derecha a izquierda)

Capítulo 5. Un modelo de recuento funcional espacial para el análisis de la heterogeneidad en el tiempo

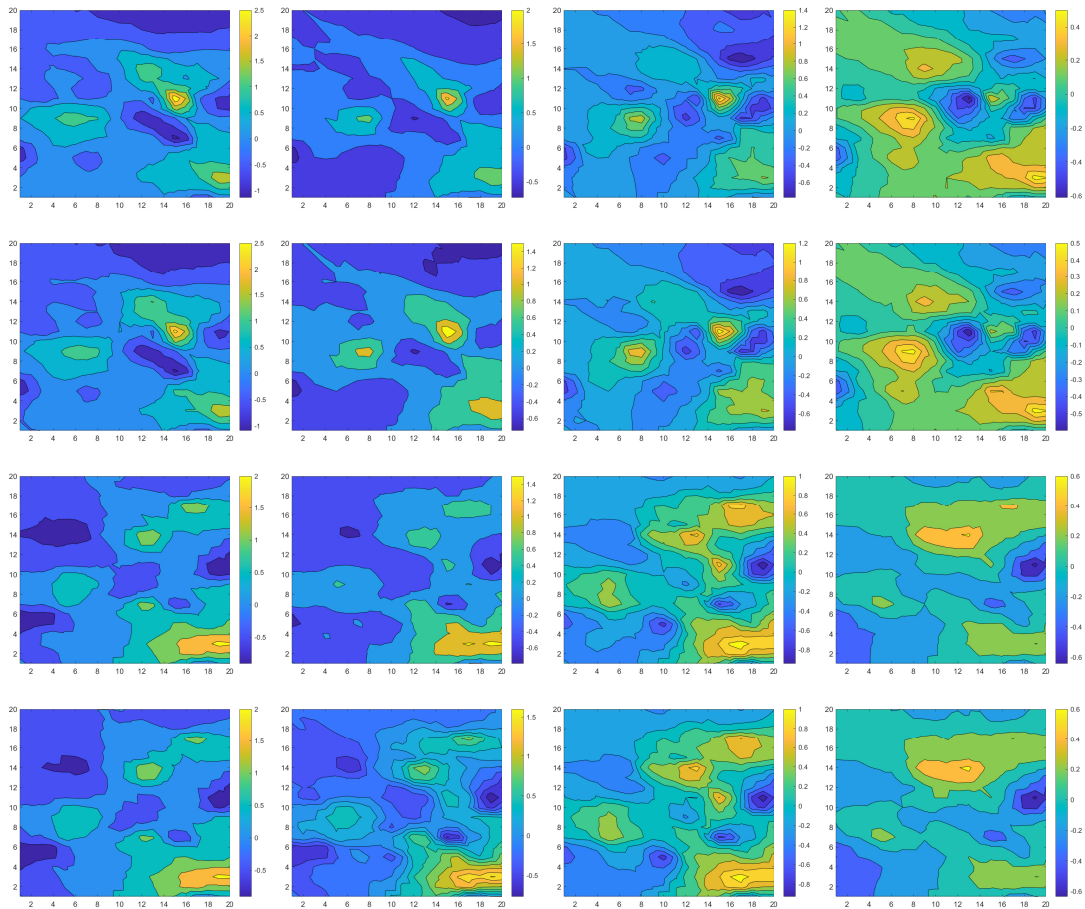


Figura 5.14: Gráficos de contorno del campo de log-intensidad observado para los valores mensuales $t = 108$ (fila superior) y $t = 216$ (tercera fila), y el campo de log-intensidad estimada para $t = 108$ (segunda fila) y $t = 216$ (fila inferior). Tanto los valores observados como los estimados para los tiempos $t = 108$ y $t = 216$ se muestran a través de las escalas $j = 7, 8, 9, 10$ (de izquierda a derecha), en el sistema wavelet de Haar, a partir de los datos interpolados temporales y espaciales sobre una rejilla regular 20×20

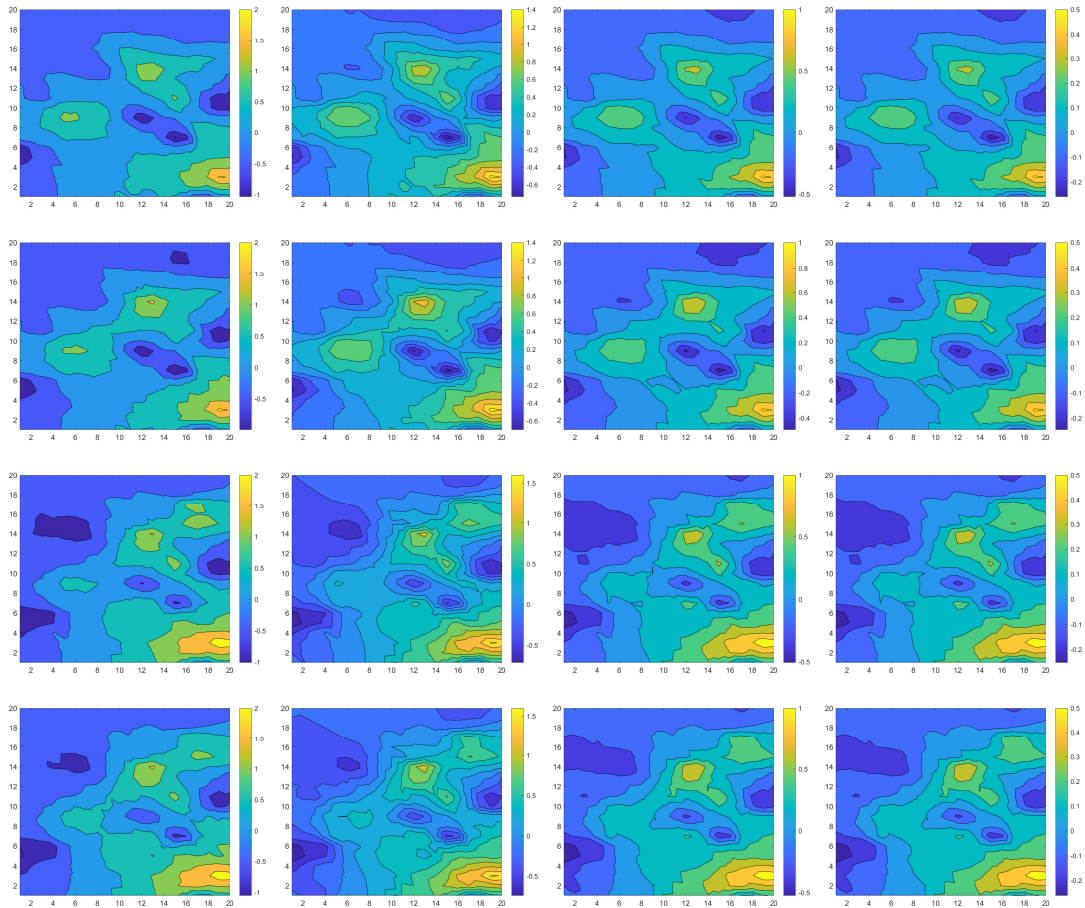


Figura 5.15: Gráficos de contorno del campo de log-intensidad observado para los valores mensuales $t = 108$ (fila superior) y $t = 216$ (tercera fila), y el campo de log-intensidad estimada para $t = 108$ (segunda fila) y $t = 216$ (fila inferior). Tanto los valores observados como los estimados para los tiempos $t = 108$ y $t = 216$ se muestran a través de las escalas $j = 7, 8, 9, 10$ (de izquierda a derecha), en el sistema wavelet de Haar, a partir de los datos temporales interpolados y suavizados sobre una rejilla regular 20×20

Tabla 5.4: ALOOCVE. Promedio puntual anual de los errores de validación cruzada

Año	ALOOCVE	Año	ALOOCVE	Año	ALOOCVE
1980	0.0247	1992	0.0118	2004	0.0132
1981	0.0144	1993	0.0130	2005	0.0117
1982	0.0112	1994	0.0163	2006	0.0135
1983	0.0125	1995	0.0159	2007	0.0140
1984	0.0144	1996	0.0111	2008	0.0118
1985	0.0122	1997	0.0099	2009	0.0113
1986	0.0126	1998	0.0108	2010	0.0143
1987	0.0155	1999	0.0141	2011	0.0131
1988	0.0161	2000	0.0167	2012	0.0122
1989	0.0144	2001	0.0161	2013	0.0115
1990	0.0125	2002	0.0143	2014	0.0145
1991	0.0118	2003	0.0140	2015	0.0221

5.5.2. Resultados de la validación

Nuestro enfoque funcional espacial multiescalar se valida, a continuación, a partir de los datos. En concreto, dejando fuera las curvas observadas en los nodos pertenecientes a una provincia que define la región de interés (el conjunto de datos funcionales de validación), se calculan las ecuaciones (5.17)–(5.25), a partir del resto de observaciones funcionales, distribuidas espacialmente en el resto de provincias españolas (el conjunto de datos funcionales de entrenamiento). A continuación, se obtienen los correspondientes estimadores y predictores paramétricos componente a componente $SAR_{\ell^2}(1)$ multiescales, a partir de la fórmula empírica de reconstrucción wavelet a nivel de resolución 10 (versión truncada de la ecuación (5.25)). Este proceso se repite 48 veces (una en cada provincia). Así, el error funcional de validación cruzada se calcula como la media de los errores funcionales absolutos calculados en cada una de las 48 iteraciones. La media puntual anual del error funcional de validación cruzada calculado puede consultarse en la Tabla 5.4. El promedio anual original y estimado del número de muertes en cada provincia, para cada uno de los años analizados, también se muestra en la Figura 5.16.

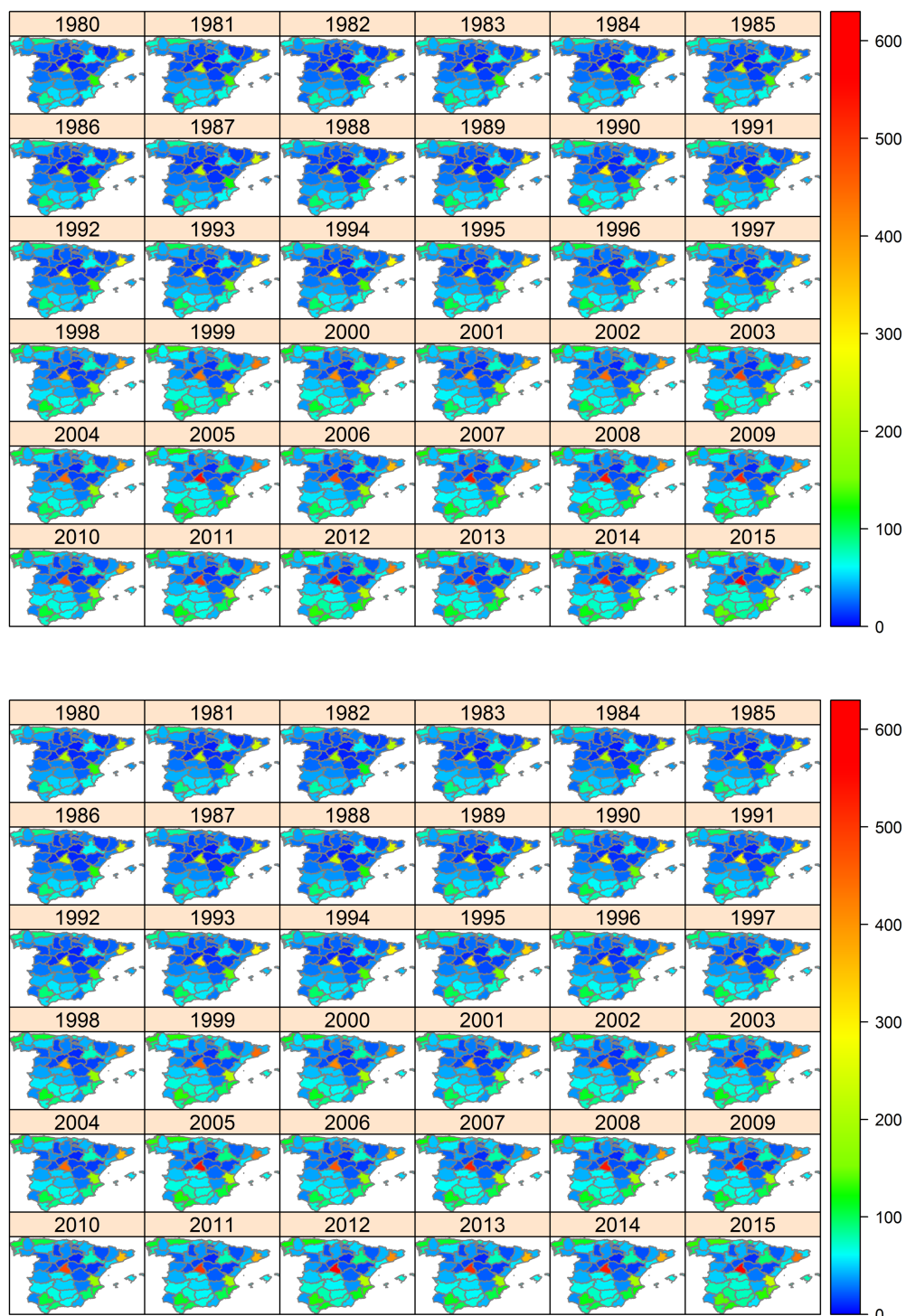


Figura 5.16: Promedio anual del número de muertes por enfermedades respiratorias observadas y estimadas (arriba y abajo, respectivamente) en cada una de las 48 provincias españolas desde enero de 1980 hasta diciembre de 2015

APLICACIONES A DATOS REALES

Esta parte de la tesis se dedica al estudio de ejemplos de aplicaciones a datos reales, mediante las técnicas y metodologías vistas en los capítulos anteriores. Así, los capítulos 6 y 7, representan aplicaciones prácticas sobre la evolución de dos enfermedades distintas en territorios distintos. En el Capítulo 6 se estudia la evolución de la incidencia de Fiebre del Dengue (DF) en países de América, mientras que en el Capítulo 7 se estudia la evolución de la mortalidad por COVID-19 sobre las comunidades autónomas de España. De esta manera, con algunas variaciones e innovaciones sobre las metodologías presentadas en la tesis, se puede validar y comparar, con otras metodologías, la utilidad de las técnicas que se proponen para el análisis de incidencia y mortalidad de enfermedades.

Los resultados obtenidos en esta parte de la tesis se recogen en el trabajo titulado “A Bayesian Functional Methodology for Dengue Risk Mapping in Latin America and the Caribbean”, publicado en “*Acta Tropica*”, Vol. 215, pp 105788, (Torres-Signes y Dip, 2021), para el Capítulo 6; y en el trabajo titulado “COVID–19 mortality analysis from soft-data multivariate curve regression and *machine learning*”, actualmente sometido (arXiv:2008.06344), cuyos autores son A. Torres-Signes, M. P. Frías, M. D. Ruiz-Medina, (Torres-Signes et al., 2021b), para el Capítulo 7.

Capítulo 6

Una metodología funcional bayesiana para el mapeo del riesgo de dengue en América Latina y el Caribe

En el presente capítulo se implementa un enfoque ARH(1) mediante un modelo bayesiano para el mapeo de la evolución del riesgo de la fiebre del dengue en 32 países de América durante el período 1998-2018.

El dengue es una grave enfermedad viral que se ha establecido en todo el mundo en ciclos de transmisión tanto endémicos como epidémicos como la enfermedad viral más importante transmitida por mosquitos. La incidencia y la prevalencia del dengue están aumentando en las zonas endémicas de las regiones tropicales y subtropicales. Según la Organización Mundial de la Salud, antes de 1970 sólo nueve países habían sufrido epidemias agravadas de dengue y hoy en día la enfermedad es endémica en más de 100 países, siendo América, Asia sudoriental y el Pacífico occidental las regiones más gravemente afectadas. Uruguay y Chile han sido los únicos países de América Latina que no se han visto afectados por la transmisión del dengue.

La investigación sobre el mapeo de enfermedades espaciales y espacio-temporales se ha llevado a cabo en un marco bayesiano. Los modelos de mapeo de enfermedades espaciales y temporales son instrumentos comunes para explicar los patrones de recuento de enfermedades y suelen formularse en modelos bayesianos jerárquicos. La utilidad de estos modelos para estos fines, como se

señala en [Aswi et al. \(2019\)](#), radica en su flexibilidad para incorporar a diversos niveles la variabilidad del modelo, lo que permite una evaluación más completa de la incertidumbre de la predicción.

En cuanto a los enfoques espaciales y espacio-temporales bayesianos para la modelización de los casos de dengue, la bibliografía se ha centrado principalmente en modelos plenamente bayesianos con un efecto aleatorio estructurado espacialmente, utilizando una estructura previa autorregresiva condicionada (CAR) para examinar la relación entre el riesgo de dengue y covariables seleccionadas.

Los modelos teóricos obtenidos a partir de estadísticos funcionales espaciales han surgido como una nueva rama de la estadística que permite también el mapeo de enfermedades. En este capítulo se propone un modelo de estimación funcional para analizar el riesgo de incidencia del dengue en México, América Central y América del Sur. En concreto, se utiliza un proceso autorregresivo de orden uno Hilbert valuado, ARH(1), para estimar la evolución funcional temporal de los patrones geográficos espaciales de riesgo de incidencia en el mapeo de enfermedades durante 1998–2018.

El operador de autocovarianza del proceso ARH(1) se estima desde una perspectiva clásica y bayesiana. El análisis exploratorio de datos efectuado en el Capítulo 6 se basa en la descomposición en valores singulares del operador de autocovarianza empírico, que permite proponer los modelos de familias conjugadas para la estimación bayesiana de un conjunto finito de valores singulares del operador de autocovarianza teórico subyacente, mediante aplicación de los resultados publicados en el trabajo de [Bosq y Ruiz-Medina \(2014\)](#). Los resultados se comparan con los obtenidos mediante la aplicación de tres modelos espacio-temporales jerárquicos alternativos: un modelo de Leroux, un modelo autorregresivo condicionado intrínseco (iCAR) y un modelo de Besag, York y Mollie (BYM), véase [Adin et al. \(2019\)](#).

Sobre la base de los estudios y la literatura anteriores, la aplicación de un modelo ARH(1) para hacer predicciones es amplia pero, hasta donde sabemos, no se ha utilizado un proceso autorregresivo hilbertiano con metodología de estimación bayesiana para modelizar, en concreto, los casos de DF. En este capítulo se presenta un análisis preliminar a nivel empírico y exploratorio donde se aplica esta metodología para el estudio de la fiebre del dengue en América Latina y el Caribe.

6.1. Métodos

6.1.1. El modelo

En esta sección se presenta el modelo utilizado para la representación de la incidencia de DF en los diferentes países considerados en el estudio. Se analiza la evolución del mapeo de la enfermedad desde la perspectiva de estadísticos espacio-temporales. Se adopta un enfoque funcional para la representación de la evolución temporal de los patrones espaciales, en términos de un modelo hilbertiano autorregresivo de primer orden, ARH(1).

El operador de autocovarianza de este modelo se obtiene a partir de la inferencia bayesiana y clásica. En consecuencia, se calcula el operador de autocorrelación, véase Bosq (2000), que puede interpretarse como un operador de interacción espacio-temporal para secuencias de superficies en el tiempo. Se considera la *razón de mortalidad estandarizada*, SMR, como índice utilizado habitualmente para la incidencia de enfermedades y mortalidad. En nuestro caso, las zonas donde observamos la enfermedad son países, es decir, grandes regiones geográficas, por lo que utilizamos la SMR como una medida fiable del riesgo relativo de DF, véase Meza (2003) y Mukhsar. et al. (2016). Así pues, el estimador de máxima verosimilitud del riesgo relativo en una zona viene dado por la siguiente fórmula

$$\text{SMR} = \text{casos observados} / \text{casos esperados}.$$

El proceso de Poisson de media $\theta_{i,t}$ se considera para modelizar el número de nuevas infecciones de DF, $C_{i,t}$, en la región i en el tiempo t . Para cada región i y tiempo t , se obtiene el número esperado de casos de DF, $e_{i,t}$, por lo que $\theta_{i,t} = r_{i,t} e_{i,t}$, donde $r_{i,t}$ es el riesgo relativo en la región i y tiempo t . Por lo tanto,

$$\begin{aligned} C_{i,t} | r_{i,t} &\sim \mathcal{P}(\theta_{i,t} = e_{i,t} r_{i,t}), \\ \log \theta_{i,t} &= \log e_{i,t} + \log r_{i,t} \end{aligned} \quad (6.1)$$

Entonces, el modelo de observación, $Z_{i,t}$, se define por la siguiente ecuación, derivada al considerar la expansión de Taylor de la función logarítmica de la estimación de $r_{i,t}$ en el valor observado $\theta_{i,t}$. Para cada $t = 1, \dots, T$, y para cada

$i = 1, \dots, N,$

$$\begin{aligned}
 Z_{i,t} &= \log \left(\frac{C_{i,t}}{e_{i,t}} \right) = \log C_{i,t} - \log e_{i,t} \\
 &= \log \theta_{i,t} + \frac{C_{i,t} - \theta_{i,t}}{\theta_{i,t}} - \log e_{i,t} \\
 &= \log r_{i,t} + \frac{C_{i,t} - \theta_{i,t}}{\theta_{i,t}} \\
 &= \log r_{i,t} + v_{i,t},
 \end{aligned} \tag{6.2}$$

donde, a partir de la aproximación gaussiana de la distribución de Poisson, para cada $t = 1, \dots, T$, y para cada $i = 1, \dots, N$, $v_{i,t}$ puede considerarse como un ruido blanco gaussiano con varianza $1/\theta_{i,t}$, véase el Capítulo 3. También asumimos las condiciones necesarias para considerar este ruido blanco espacio-temporal, $v_{i,t} = v(\mathbf{u}_i, \mathbf{t})$, $\mathbf{u}_i \in D \subseteq \mathbb{R}^2$, $\mathbf{t} \in \mathbb{Z}$, como un ruido blanco Hilbert-valorado en sentido fuerte en el tiempo, ver Bosq (2000).

En lo que sigue, asumimos que el operador de autocovarianza, R_0 , es un operador traza que admite una descomposición espectral en términos de un sistema de autovectores. Para $k \geq 1$,

$$R_0 \phi_k = \lambda_k(R_0) \phi_k.$$

A continuación, se describe la estimación bayesiana de los parámetros que definen la descomposición espectral, $\lambda_k(R_0)$,

$$R_0 = \sum_{k=1}^{\infty} \lambda_k(R_0) \phi_k \otimes \phi_k.$$

Los resultados se obtienen en un enfoque de modelo de dos etapas. en concreto, en la primera etapa, los parámetros se estiman en un enfoque bayesiano y clásico. En el caso bayesiano, se consideran familias conjugadas con respecto a la probabilidad gaussiana univariante en términos de cada proyección, suponiendo un marco gaussiano ARH(1). En la segunda etapa, se aproxima la transformación logarítmica del riesgo relativo con la ecuación de estados de un modelo ARH(1), véase la Sección 6.1.2.

Para estimar el operador de autocovarianza, \mathbf{R}_0 , para cada instante $t = 1, \dots, T$, consideramos $Y_k(t) = \sqrt{\lambda_k(\mathbf{R}_0)} Z_k(t) \sim \mathcal{N}(0, \lambda_k(\mathbf{R}_0))$; $k \geq 1$. La fami-

lia de distribuciones a-priori conjugadas para el parámetro de escala de la verosimilitud gaussiana univariante es una distribución gamma inversa, $\mathcal{IG}(\alpha, \beta)$. Consideramos que, para cada $k \geq 1$, el parámetro λ_k se distribuye como $\mathcal{IG}(\alpha_k, \beta_k)$; entonces, la distribución a-posteriori de λ_k viene dada por $\mathcal{IG}\left(\alpha_k + T/2, \beta_k + \sum_{t=1}^T Z_k^2(t)/2\right)$, ver, por ejemplo, Lehmann y Casella (1998), entre otros. El estimador bayesiano, bajo la función de pérdida al cuadrado, se obtiene bajo la siguiente expresión:

$$\hat{\lambda}_k(\mathbf{R}_0) = \frac{2\beta_k + \sum_{t=1}^T Z_k^2(t)}{T + 2\alpha_k - 2}, \quad k = 1, \dots, M, \quad (6.3)$$

donde M es un orden de truncamiento.

En consonancia con la metodología bayesiana, también adoptada en el Capítulo 7, en este capítulo se ha considerado el uso de una a-priori no informativa en la familia de distribuciones de la Gamma Inversa, la $\mathcal{IG}(\epsilon, \epsilon)$, con $\epsilon > 0$ y de un valor pequeño. Esta distribución se conoce como distribución débilmente informativa, véase Gelman (2006). Las principales ventajas de su uso, en comparación con otras distribuciones a-priori no informativas, es que son distribuciones a-priori conjugadas. Estas distribuciones se usan frecuentemente, véase Ugarte et al. (2009b), para estimar los parámetros de escala en modelos jerárquicos espaciales y espacio-temporales para el mapeo de enfermedades.

Este estimador bayesiano, para cada $k = 1, \dots, M$, se compara con el clásico, es decir, el estimador de máxima verosimilitud:

$$\tilde{\lambda}_k(\mathbf{R}_0) = \frac{\sum_{t=1}^T Z_k^2(t)}{T}. \quad (6.4)$$

Obsérvese que el caso clásico consiste en el estimador bayesiano donde, para cada $k = 1, \dots, M$, los hiperparámetros de la distribución a-priori son $\alpha_k = 1$ y $\beta_k = 0$.

6.1.2. Modelo ARH(1)

Consideremos $Y_t \in H$, $t \in \mathbb{Z}$, un proceso centrado de Hilbert definido en el espacio de probabilidad básico (Ω, \mathcal{A}, P) , como un proceso ARH(1), introducido en la Sección 1.2.5, es decir, que satisface

$$Y_t(\mathbf{x}) = A(Y_{t-1})(\mathbf{x}) + v_t(\mathbf{x}), \quad \forall \mathbf{x} \in D \subseteq \mathbb{R}^n, \quad t \in \mathbb{Z}, \quad (6.5)$$

donde v es un ruido blanco hilbertiano fuerte (Bosq, 2000), cuyo operador de autocovarianza viene dado por $R_v = E(Y_t \otimes Y_t)$, con $t \in \mathbb{Z}$. Aquí, \otimes representa el producto tensorial de funciones en H . El operador de interacción espacio-temporal A admite la descomposición espectral en términos de sus autovalores y autovectores. Por lo tanto, $A = \Psi \Lambda \Phi^*$, en términos del sistema de autovectores izquierdo $\{\psi_i, i \in \mathbb{N}\}$, y el derecho $\{\phi_i, i \in \mathbb{N}\}$, donde Λ es el operador diagonal definido por la secuencia de autovalores $\{\lambda_i, i \in \mathbb{N}\}$. Entonces, de (6.5),

$$\Phi^* Y_t = \Lambda \Phi^* Y_{t-1} + \Phi^* v_t. \quad (6.6)$$

La estructura de proceso de Hilbert de segundo orden, Y_t , se caracteriza en términos de los operadores de autocovarianza y de covarianza cruzada. Se considera una formulación vectorial truncada para obtener la renombrada ecuación autorregresiva diagonal (6.6). Para un orden de truncamiento dado $M \leq N$, y teniendo en cuenta (6.5), la aproximación diagonal del proceso ARH(1) es, para $t = 1, \dots, T$,

$$\Phi^* Y_t = \mathbf{Z}(t) = \Lambda \mathbf{Z}(t-1) + \mathbf{u}(t), \quad (6.7)$$

donde $\mathbf{Z}(t)$ es un vector $M \times 1$ con entradas $Z_m(t) = \langle \mathbf{Y}_t, \phi_m \rangle_H$, $m = 1, \dots, M$ ($\langle \cdot, \cdot \rangle_H$ denota el producto escalar en el espacio de funciones de cuadrado integrable Hilbert separable, H), Λ es una matriz diagonal $M \times M$ con entradas λ_m , $m = 1, \dots, M$, y $\mathbf{u}(t)$ es un vector $M \times 1$ con entradas $u_m(t) = \langle \mathbf{v}_t, \phi_m \rangle_H$, $m = 1, \dots, M$.

6.1.3. Datos

La base de datos ha sido proporcionada por *Pan American Health Organization* (PAHO, 2019). Incluye 32 países con datos anuales, desde 1998 hasta 2018, y se han considerado todos los casos de dengue notificados. Según la PAHO, una enorme cantidad de casos son asintomáticos y, por lo tanto, las cifras reales de casos de dengue no se conocen suficientemente con muchos casos clasificados erróneamente. Aunque la base de datos tiene observaciones para los países desde 1980, el número de casos de dengue igual a cero ha sido menor desde 1998. Además, es difícil distinguir entre un valor no reportado (faltante) y casos nulos de dengue, por lo que hemos considerado todos los ceros como verdaderos valores observados. Así, hemos seleccionado para el estudio todos

los países de América, que presentan hasta cinco ceros, desde 1998 hasta 2018. Estos datos, que incluyen 672 valores con 24 ceros, están representados en la Figura 6.1.

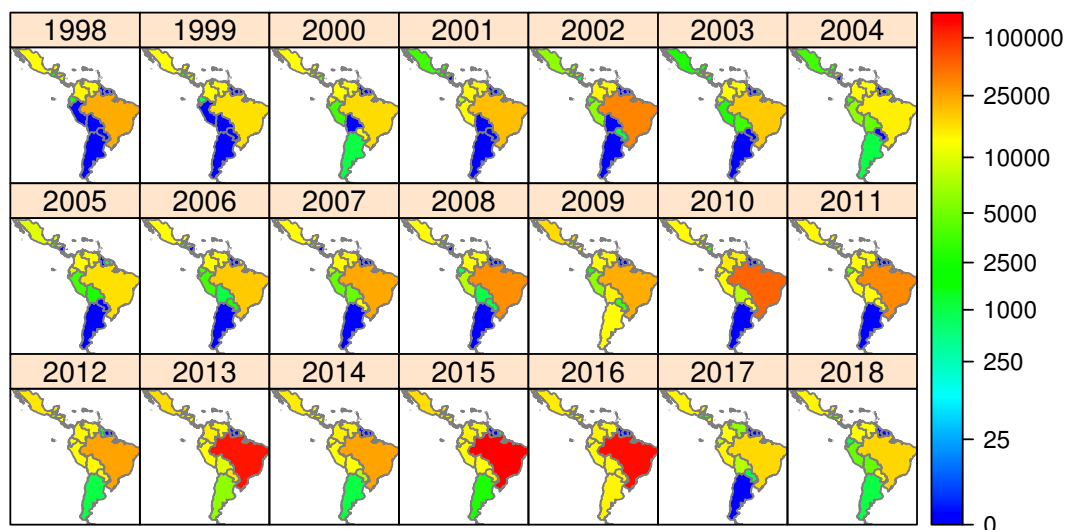


Figura 6.1: Evolución del número de DF observado desde 1998 hasta 2018 en los países bajo estudio

La heterogeneidad de los datos dentro de cada país al transcurrir de los años es notable. Por ejemplo, Guadalupe suele tener pocos casos de DF, que incluye ceros, pero en 2010 alcanza 40.737 casos de DF. Por otro lado, Brasil siempre muestra un gran número de casos de DF aunque también tiene una gran población. Es un país endémico del dengue y a lo largo de los años se ha enfrentado a varios brotes (2013, 2015 y 2016) siendo golpeados también los países vecinos. El elevado número de casos de dengue observados en algunos países sugiere que los casos importados de países vecinos contribuyen a la propagación de la enfermedad.

La heterogeneidad de los países para cada año es clara en la Figura 6.1. En ese sentido, 2015 es un año notable, en el que hay países con un bajo número de casos de DF, como San Cristóbal y Nieves (5 casos, mientras que su media para los años del estudio es de alrededor de 27), y otros con un gran número de casos de DF, como México (219.593 casos mientras que su media para los años del estudio es de alrededor de 75.184). A pesar de ello, como ya han señalado

otros autores en la literatura previa, (Carbajo et al., 2001; Martínez-Bello et al., 2018; Restrepo et al., 2014), en general, se observa una tendencia creciente en el número de casos de DF, lo que es evidente ya que a partir de 2012 los colores en los mapas son más intensos, aunque los últimos dos años están fuera de esta tendencia.

6.2. Resultados

Para estimar los autovalores del operador de covarianza hemos considerado tanto el enfoque clásico como el bayesiano. Así, $\hat{\lambda}_k(\mathbf{R}_0)$ y $\tilde{\lambda}_k(\mathbf{R}_0)$ se calculan para $k = 1, \dots, M$. Para obtener el estimador bayesiano, hemos utilizado la clase de distribuciones a-priori gamma inversas, introducida en la Sección 6.1, dando el valor $\epsilon = 0.05$, es decir, una distribución $\mathcal{IG}(0.05, 0.05)$.

En la estimación de parámetros bayesianos implementada en la Sección 6.1, se han considerado los valores empíricos de α_k y β_k , para $k = 1, \dots, M = 19$, para la estimación bayesiana y clásica del operador de autocovarianza.

La elección del orden de truncamiento, $M = 19$, se ha hecho teniendo en cuenta que los autovalores de un orden superior a T , cuando $T < N$, teóricamente son nulos, véase Bosq y Blanke (2007). Por otra parte, la determinación del orden de truncamiento es compleja porque un valor demasiado pequeño produce malas estimaciones pero, con un orden demasiado grande, el error de predicción podría explotar, ver Bosq (2000). Hemos comprobado que ninguna de estas dos cosas es el caso de nuestra elección. Además, como criterio de truncamiento, este orden de truncamiento lleva a un porcentaje muy alto de variabilidad empírica explicada en términos de la traza aproximada del operador de autocovarianza del proceso ARH(1), ver Ruiz-Medina et al. (2014).

Una medida de la variabilidad entre los valores observados y los estimados se muestra en la Tabla 6.1. Se trata del Error Absoluto Empírico del log-riesgo estimado en ambos casos, estimación bayesiana y clásica, $|\log(r_{i,t}) - \log(\hat{r}_{i,t})|$, respectivamente, para cada año, Promediado para todos los países ($i = 1, \dots, 32$) bajo estudio (EAEA).

Para captar la heterogeneidad de las medidas de error desde otra perspectiva, la Tabla 6.2 muestra el Error Cuadrático Empírico para el log-riesgo tanto en la estimación bayesiana como en la clásica, $(\log(r_{i,t}) - \log(\hat{r}_{i,t}))^2$ y $(\log(r_{i,t}) - N - \tilde{r}_{i,t})^2$, respectivamente, para los 32 países considerados Prome-

Tabla 6.1: Errores absolutos empíricos Promediados para el riesgo relativo de DF según el promedio de los países bajo estudio, para la estimación bayesiana y clásica, (BEAEA) y (CEAEA), respectivamente

Año	CEAEA	BEAEA	Año	CEAEA	BEAEA	Año	CEAEA	BEAEA
1998	0	0	2005	0.1038	0.0636	2012	0.1794	0.0983
1999	0.0940	0.6254	2006	0.3079	0.2324	2013	0.1823	0.1431
2000	0.1571	0.1258	2007	0.4676	0.1554	2014	0.1692	0.0989
2001	0.0910	0.0609	2008	0.3570	0.0789	2015	0.2865	0.1988
2002	0.1254	0.0891	2009	0.2442	0.0843	2016	0.3734	0.1198
2003	0.0973	0.1137	2010	0.3078	0.1492	2017	0.2011	0.0852
2004	0.0698	0.1101	2011	0.3283	0.1483	2018	0.3999	0.3405

diado para todos los años bajo estudio (ESEA).

Tabla 6.2: Errores Cuadráticos Empíricos Promediados para el riesgo relativo de DF según el promedio de los años bajo estudio, para la estimación bayesiana y clásica, (BESEA) y (CESEA), respectivamente

País	CESEA	BESEA	País	CESEA	BESEA
Puerto Rico	0.3061	0.1619	Argentina	0.0910	0.0444
República Dominicana	0.0068	0.0058	Brasil	0.0015	0.0007
Antigua y Barbuda	0.0578	0.0339	Paraguay	0.1669	0.0763
Barbados	0.1165	0.0757	Belice	0.0455	0.0298
Dominica	0.2094	0.1206	Costa Rica	0.0526	0.0164
Grenada	0.6863	0.3907	El Salvador	0.0156	0.0232
Guadalupe	0.0344	0.0239	Guatemala	0.0064	0.0050
Guyana Francesa	0.1225	0.0463	Honduras	0.0165	0.0058
Guyana	0.3861	0.1653	México	0.0054	0.0034
Jamaica	0.4837	0.2895	Nicaragua	0.0097	0.0100
Martinica	0.0387	0.0185	Panamá	0.0740	0.0585
San Cristóbal y Nieves	0.0103	0.0143	Bolivia	0.0245	0.0128
S. Vicente y las <i>Gr^{dinas}</i>	0.0563	0.0561	Colombia	0.0202	0.0123
Santa Lucía	0.0194	0.0271	Ecuador	0.0398	0.0275
Suriname	0.4153	0.5046	Perú	0.0226	0.0199
Trinidad y Tobago	0.0557	0.0468	Venezuela	0.0109	0.0115

La Figura 6.2 muestra el log-riesgo estimado de los 32 países americanos considerados de 1998 a 2018 para la estimación bayesiana, ya que tiene el EM-SE más más pequeño, $(\sum_{i=1}^{32} ESEA/32)$. De esta figura se ha extraído la Figura

6.3 para mostrar el comportamiento particular de los dos tipos de estimación junto con el log-riesgo observado en seis países en particular. Esta medida ayuda a entender el comportamiento de la incidencia del riesgo de enfermedad por país y año. Cuando la medida está alrededor de cero, el número de incidencias estimadas es similar a las esperadas, según los casos observados en todos los países. Los valores negativos indican que el número de incidencia estimada es inferior a lo que se esperaría, y de forma inversa cuando la medición es positiva.

El hecho de tener ceros en el conjunto de datos es la razón de que los valores de log-riesgos sean tan bajos (en este caso, se ha considerado 10^{-4} para los riesgos nulos). En el lado opuesto, hay episodios de alto riesgo en países que provocan valores grandes y atípicos. Por ejemplo, la Guyana Francesa, en 2006, tiene 15.904 casos de DF, uno de sus picos más altos mientras que este año no fue especialmente agresivo en el resto de los países (véase la Figura 6.1). Esto provoca un valor extremo para la observación del riesgo (véase el año 9 en la Figura 6.3). Por otro lado, al ser un país pequeño, no se puede ver en la Figura 6.1 y 6.2. Hay muchos países con importantes diferencias en sus extensiones, que dificultan la observación de toda la gama de colores utilizados en los mapas. Esta es una de las razones por las que los países pequeños no se incluyen generalmente en los estudios de cartografía de riesgo de enfermedades. Aún así, hemos considerado apropiado el uso de todos los países en las condiciones comentadas en la Sección 6.1.3, ya que la metodología de estimación propuesta y la medición utilizada permiten hacerlo.

6.2.1. Una comparativa con enfoques previos

En esta sección, comparamos nuestro modelo con tres modelos espacio-temporales alternativos y tradicionales: un modelo de Leroux, un modelo iCAR y un modelo BYM, utilizando el software SSTCDapp, ver Adin et al. (2019).

En la Tabla 6.3 se detalla, para cada metodología aplicada, el Error Absoluto Medio Empírico (EMAE) y el Error Cuadrático Medio Empírico (EMSE) del log-riesgo estimado de DF para todos los países y años bajo estudio, es decir,

$$\begin{aligned} \text{EMAE} &= \frac{1}{21} \frac{1}{32} \sum_{t=1}^{21} \sum_{i=1}^{32} |\log(r_{i,t}) - \log(\hat{r}_{i,t})|, \\ \text{EMSE} &= \frac{1}{21} \frac{1}{32} \sum_{t=1}^{21} \sum_{i=1}^{32} (\log(r_{i,t}) - \log(\hat{r}_{i,t}))^2, \end{aligned}$$

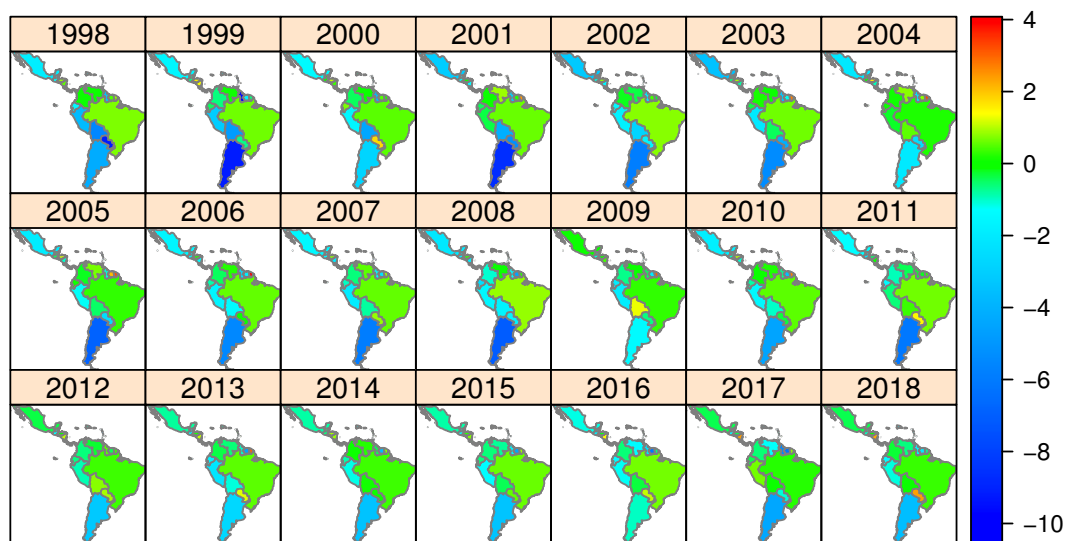


Figura 6.2: Log-riesgo estimado a partir de la estimación bayesiana en el modelo propuesto desde 1998 hasta 2018 en los países bajo estudio

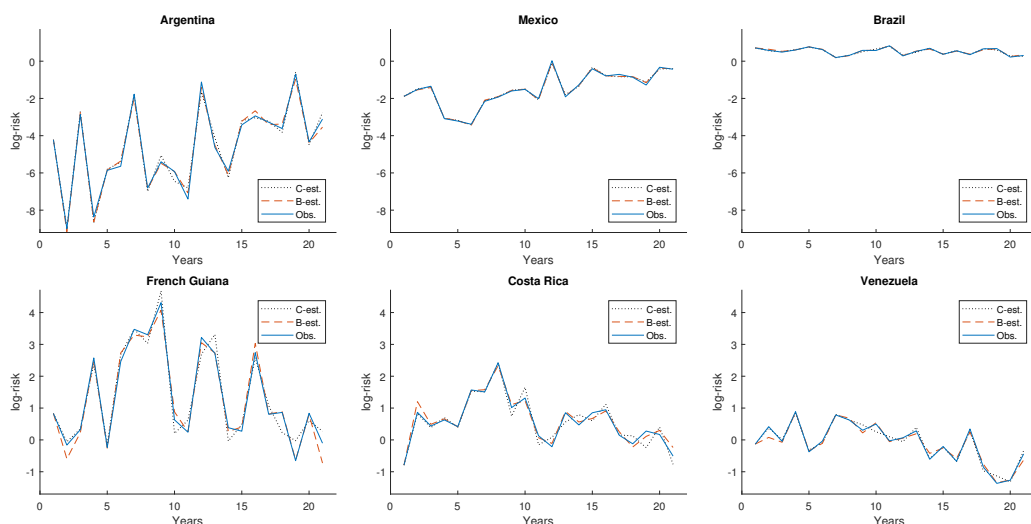


Figura 6.3: Log-riesgo estimado a partir de la estimación clásica (C-est., en línea a puntos) y la estimación bayesiana (B-est., en línea discontinua), y el log-riesgo observado (Obs., en línea continua) desde 1998 hasta 2018 para los países: Argentina, México, Brasil, Guyana Francesa, Costa Rica y Venezuela

donde $r_{i,t}$ y $\hat{r}_{i,t}$ son el riesgo relativo observado y estimado, respectivamente. El mejor resultado se obtiene para la estimación bayesiana, en ambas medi-

ciones, EMAE y EMSE. La Figura 6.4 muestra las estimaciones de log-riesgo a partir de los modelos espacio-temporales tradicionales como por el modelo que proponemos nosotros, junto con las observaciones reales en dos países.

Tabla 6.3: Comparación del Error Absoluto Medio Empírico (EMAE) y el Error Cuadrático Medio Empírico (EMSE) desde 1998 hasta 2018 para todos los países, según las cinco metodologías aplicadas

Metodología	EMAE	EMSE
LEROUX	0.1999	0.7854
iCAR	0.1987	0.7790
BYM	0.1974	0.7723
C-est.	0.2163	0.1128
B-est.	0.1486	0.0731

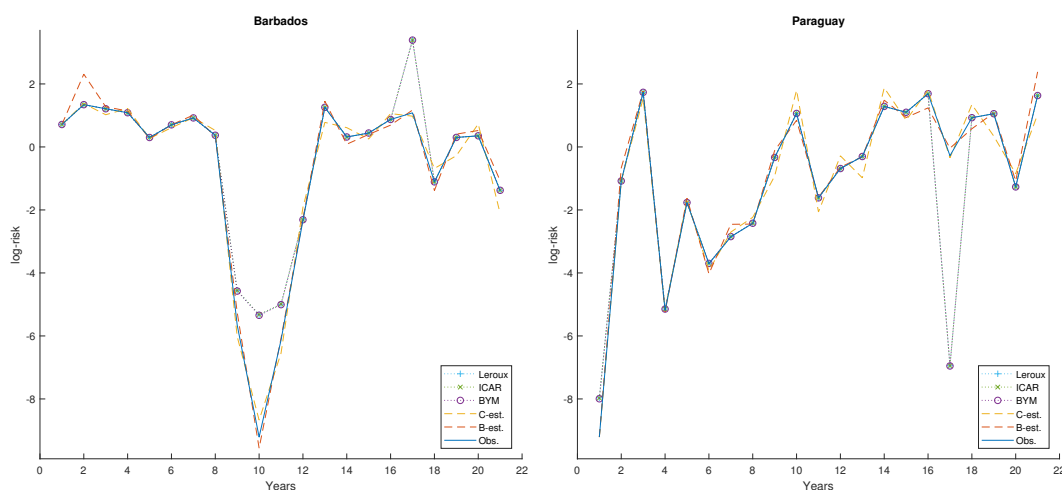


Figura 6.4: Log-riesgo estimado a partir de los modelos espacio-temporales tradicionales (con signo más, cruz y círculo, en líneas punteadas), la metodología de estimación propuesta (en líneas discontinuas) y el log-riesgo observado (en línea continua), desde 1998 hasta 2018, para los países: Barbados y Paraguay

6.3. Discusión

Primeramente, se comparan los estimadores bayesianos y clásicos propuestos en el modelo. En las tablas 6.1 y 6.2 se muestran los pequeños valores de estas medidas de resumen (errores absolutos y cuadráticos empíricos para las estimaciones bayesiana y clásica), teniendo en cuenta la gran variabilidad de los valores de riesgo debido a la heterogeneidad tanto de la población como del número de casos de DF. El año 1998 (Tabla 6.1) tiene valor nulo debido al uso de un modelo autorregresivo de orden uno. El resto de los años, ambos estimadores presentan un comportamiento irregular donde a veces el estimador clásico tiene mejores resultados y viceversa. En ese sentido, podemos destacar, por ejemplo, los años 2003 y 2007, respectivamente. Pero, el comportamiento general es un aumento considerable del error empírico en el caso clásico, ver por ejemplo los años 2000, 2001, 2002, 2005, etc.

Un comportamiento similar se puede encontrar en la Tabla 6.2, donde países como Venezuela muestran un mayor nivel de error para el caso bayesiano, pero el comportamiento general es un considerable incremento en el clásico, ver, por ejemplo, Argentina, Brasil, Paraguay, Belice, Costa Rica, etc. Además, en la Tabla 6.2 se puede apreciar la diferencia entre las medidas de error presentadas según el tamaño de los países en estudio. Así, en general, la medición del error empírico es menor para los países grandes (Brasil, México y otros) y mayor para los pequeños (Granada, Jamaica, Suriname y otros). En concreto, esto se acentúa para el caso bayesiano, donde el BESEA llega a la mitad del CESEA, en algunos países grandes como el Brasil.

En la Figura 6.2 se muestra el log-riesgo estimado, para todos los países y años, según el modelo bayesiano explicado en la Sección 6.1. La heterogeneidad de los patrones se mantiene en la metodología de estimación propuesta sin un exceso de suavizado, lo que permite detectar zonas de alto riesgo y evitar posibles falsos negativos. La evolución de esta medida es muy similar a la de los datos reales (ver las tablas de error), manteniendo la heterogeneidad observada de los patrones de riesgo en los países y en el tiempo, es decir, la heterogeneidad y la variabilidad siguen presentes. Ver también la Figura 6.3, donde se puede comparar el riesgo estimado y el observado para algunos países. Este comportamiento sugiere que los modelos propuestos se han ajustado bien a los datos.

Obsérvese que la Figura 6.2 informa sobre el riesgo en un país con respecto a los demás en un año concreto. Podemos observar la evolución del riesgo en cada país (comparado con el resto de los países) por año. Sin embargo, en general, no podemos establecer una conexión directa entre esta figura y la Figura 6.1, aunque pueda aparecer en algunos países. Por ejemplo, podemos observar que el número de casos de DF en Argentina, a pesar de ser generalmente bajo, exhibe una tendencia creciente a nivel global, además, el riesgo a nivel global aumenta en comparación con el resto de los países del estudio (ver también la Figura 6.3). Cabe destacar que en el caso argentino, el riesgo se ve mitigado por la inclusión de toda la población en el estudio, cuando en realidad la parte sur no es una población expuesta a la enfermedad. Por otro lado, Brasil presenta tanto un alto número de casos de DF como un alto nivel de riesgo en todos los años del estudio. Sin embargo, esta estabilidad del riesgo a lo largo del tiempo (véanse las figuras 6.2 y 6.3) no refleja la gran variabilidad del número de casos (véase la Figura 6.1).

Nuestros resultados son similares a los encontrados recientemente en [Messina et al. \(2018\)](#). Estos autores muestran altos niveles de idoneidad medioambiental para el dengue utilizando datos de 2015. Contemplan que cualquier pico con un valor predicho de idoneidad para el dengue por encima de 0,467 se considera en riesgo. En el caso de América del Sur, cuando ajustan un modelo de árbol de regresión “*boosted*” para predecir, el riesgo de dengue aumenta en 2020 y 2050.

De manera similar, [Bhatt et al. \(2013\)](#) predice que la transmisión del dengue es omnipresente en todos los trópicos, siendo los de América y Asia, los de mayor riesgo. También muestran un mapa con niveles de riesgo durante 2010, comparables a los calculados en nuestro trabajo. Sin embargo, [Cattarino et al. \(2020\)](#) explica que los actuales mapas de riesgo del dengue proporcionan estimaciones de la carga de la enfermedad o de los límites de endemidad en lugar de la intensidad de la transmisión. Además, predicen una alta intensidad de transmisión en todos los continentes, a ambos lados de los trópicos, con puntos calientes en Sudamérica (Colombia, Venezuela y Brasil). Es pertinente aclarar que los resultados de los trabajos citados no pueden compararse directamente. Emplean diferentes metodologías estadísticas y se centran en zonas y años concretos. Sin embargo, en todos ellos, los países presentan un riesgo de dengue similar a los encontrados con nuestro método.

En la comparación con los enfoques previos (ver Tabla 6.3), se observa que los errores de la metodología de estimación basada en un proceso ARH(1) tienen un comportamiento más estable en las dos medidas propuestas, EMAE y EMSE. Al comparar con los modelos espacio-temporales previos, los propuestos aquí alcanzan un mejor ajuste con menores niveles de error. En particular, la estimación bayesiana que proponemos reduce los errores encontrados con los otros modelos y capta mejor la heterogeneidad presente en los datos.

Es importante destacar la información que dan las mediciones de la Tabla 6.3. El hecho de que, en los modelos espacio-temporales tradicionales considerados, el EMSE sea grande comparado con el EMAE es indicativo de que estas técnicas pueden presentar falsos negativos y positivos. El EMAE refleja que, en la mayoría de los casos, estos modelos se ajustan correctamente a los datos espacio-temporales, pero la gran diferencia con la EMSE refleja que, en momentos específicos, hay grandes diferencias entre el riesgo estimado y el observado, ver Figura 6.4 y también 6.3. Al comparar EMAE y EMSE (ver Tabla 6.3), buscamos valores pequeños y resultados similares, para que haya robustez en las estimaciones.

En nuestra propuesta, EMAE y EMSE son más pequeños que en los modelos espacio-temporales tradicionales analizados, por lo que la metodología de estimación presentada muestra un mejor comportamiento y mayor estabilidad, ya que las medidas de error son pequeñas y también cercanas. El EMSE es más pequeño en nuestros enfoques que en los otros modelos y en particular, en nuestra propuesta, el EMSE es más pequeño que el EMAE, lo que corrobora la bondad del ajuste ya que, al ser una medida cuadrática, penaliza grandes errores y sólo puede ser más pequeña si los errores están por debajo de uno.

En el proceso de elaboración del presente trabajo han surgido varios problemas que limitan el estudio. El número de casos de dengue notificados contiene incoherencias debido a los casos asintomáticos y mal clasificados en la base de datos primaria. Los problemas de la base de datos suelen ser las discrepancias en las tasas de confirmación de los laboratorios (sensibilidad y especificidad) y, a veces, los países revisan las definiciones y clasificaciones de los casos conforme pasa el tiempo. Estas diferencias en la calidad de los datos de los distintos países causan dificultades para el análisis de los datos. La modelización de zonas con grandes diferencias en sus extensiones da lugar a estimaciones un poco peores, especialmente en países pequeños como las islas del Caribe, donde la

estimación del riesgo relativo de enfermedades es más inestable.

Este fenómeno se puede apreciar en la Figura 6.3, en la que el modelo se ajusta mejor en países grandes (México, Brasil) y con un considerable número de casos que en países pequeños (Guyana Francesa, Costa Rica). Esta figura también muestra que a pesar de que Argentina y Venezuela son países grandes, exhiben un peor ajuste, probablemente debido a la calidad de los datos, además, en el caso argentino, no toda la extensión del país está claramente expuesta a la enfermedad (ver también la Tabla 6.2).

6.4. Comentarios finales al capítulo

Los casos de fiebre del dengue, en México, América Central y América del Sur, se han estudiado mediante un modelo bayesiano basado en procesos autorregresivos Hilbert-valorados de orden uno, para explicar la evolución temporal de la incidencia de la enfermedad. Los valores funcionales espaciales permiten preservar la heterogeneidad de las áreas consideradas con la estimación temporal. Esto representa una importante contribución al mapeo de enfermedades en el enfoque espacial y temporal, especialmente para los casos de dengue. Una limitación de la investigación de la enfermedad del dengue es que algunos casos de DF no se reportan, o se clasifican erróneamente. Utilizando la metodología de estimación que proponemos, se evita un exceso de suavizamiento y se conserva la heterogeneidad a lo largo de los años estudiados, de modo que el modelo detecta tanto altos como bajos niveles de riesgo.

Los resultados empíricos de la metodología propuesta apoyan nuestras conclusiones, sin subestimar el posible problema derivado de la fuente de los datos, como ya se ha señalado. Los errores empíricos del riesgo estimado, para cada año y país, son casi insignificantes como se muestra en las mediciones resumidas. Además, una comparación empírica con enfoques previos respalda nuestro modelo, produciendo mediciones de error mucho más bajas para la estimación bayesiana dentro del marco propuesto de FDA. Podemos concluir que, a la vista de los resultados empíricos, ha sido posible encontrar distribuciones a-priori en un marco bayesiano de FDA con las que lograr una estimación basada en un modelo ARH(1) que funcione de forma adecuada.

De esta manera, la naturaleza direccional en el tiempo y la variabilidad regional en el espacio de los datos de DF se han modelizado con la metodología

de estimación presentada. Este enfoque proporciona instrumentos para poder abordar medidas de prevención y control, así como toma de decisiones, ya que, para los datos observados, el modelo presentado parece explicar adecuadamente el comportamiento de la evolución del riesgo de la fiebre del dengue.

Esta característica es un punto de partida necesario para la predicción de valores futuros. No sólo eso, la metodología introducida podría ser ampliada, como en el capítulo anterior, con la estimación de intervalos de confianza y aproximación de las distribuciones a-priori de los estadísticos implicados en las respectivas metodologías de estimación, utilizando *bootstrap*. También se puede realizar un estudio de simulación para analizar interesantes propiedades de los estimadores bayesianos considerados según la elección de los hiperparámetros α_k y β_k que definen la distribución a-priori. Esto permitiría la comparación con otros estimadores, además del clásico utilizado aquí, para analizar el comportamiento de la correlación entre las tendencias temporales de las regiones vecinas, y estudiar más profundamente el problema de selección asociado al orden de truncamiento, M .

Capítulo 7

Análisis de la mortalidad por COVID-19 mediante curvas multivariantes de regresión a partir de datos blandos y mediante aprendizaje automático

En este capítulo, se presenta un enfoque de predicción espacio-temporal con múltiples objetivos, en el que se combina regresión logarítmica trigonométrica de curvas con el análisis residual espacial de series temporales multivariantes. En nuestro ajuste del modelo de curvas de regresión, nos interesa reflejar el comportamiento cíclico de la mortalidad por COVID-19 inducido por el endurecimiento o la relajación de las medidas de contención, adoptadas para mitigar el aumento de las infecciones y la mortalidad. La base trigonométrica (senos y cosenos) es entonces seleccionada en nuestro ajuste del modelo de regresión logarítmica de curva heterogénea espacial.

Para la selección del modelo, se considera la relación entre el riesgo mínimo empírico esperado y el correspondiente valor esperado de la función de pérdida cuadrática en dicho minimizador, véase, por ejemplo, [Chapelle et al. \(2002\)](#). Obsérvese que este procedimiento de selección proporciona una concordancia entre el riesgo empírico mínimo esperado, y el correspondiente valor esperado de la función de pérdida teórica.

El factor penalizado propuesto en [Chapelle et al. \(2002\)](#), aplicado a nuestra

elección del parámetro de truncamiento, lleva a la dimensión del subespacio, donde se aproxima nuestro estimador curva de regresión, en cualquier ubicación espacial. Este procedimiento de selección de modelo es asintóticamente equivalente al factor de corrección de Akaike. Una modificación robusta del criterio de información de Akaike puede encontrarse, por ejemplo, en Agostinelli (2001). Como alternativa, se puede considerar el criterio de validación cruzada para seleccionar el mejor subconjunto de variables explicativas (véase Takano y Miyashiro (2020), donde se propone un enfoque de optimización mixto-integro en este contexto).

Más allá del análisis asintótico, la selección del número de variables explicativas en el modelo de regresión no lineal paramétrica implementado, a partir de tamaños de muestra finitos, constituye uno de los principales desafíos a abordar en este capítulo de la tesis. Específicamente, se calcula el estimador *bootstrap* del error cuadrático medio del modelo teórico, a partir del cálculo del error empírico, obtenido de una muestra de entrenamiento. Se proporcionan intervalos de confianza *bootstrap*, para la media espacial del predictor curva de regresión, y para la media del error empírico, en el cálculo del predictor funcional de regresión, así como para la media de los errores empíricos estándar de los predictores residuales, obtenidos mediante el análisis de series temporales multivariantes implementado. También se calcula la distribución de probabilidad *bootstrap* de estos estadísticos.

En nuestro análisis de series temporales multivariantes de los residuos de regresión, se obtiene una aproximación empírica, mediante *bootstrap*, de la distribución a-priori de las proyecciones empíricas del modelo finito-dimensional de correlación lineal espacial considerado. Posteriormente, se selecciona el modelo teórico, dentro de la familia de distribuciones de probabilidad Beta, que mejor se ajusta al modelo empírico calculado mediante *bootstrap*. Finalmente, la estimación de los valores singulares del modelo de correlación lineal espacial se obtiene, mediante maximización de una función proporcional a la distribución a-posteriori, calculada numéricamente para la a-priori multivariante separable propuesta, basada en la familia Beta.

El enfoque de predicción con múltiples objetivos presentado, se aplica al análisis espacio-temporal de la mortalidad por COVID-19 en la primera ola que afecta a las comunidades españolas, desde el 8 de marzo de 2020 hasta el 13 de mayo de 2020. Nótese que el enfoque, presentado en este capítulo, tiene en

cuenta el error de observación en los recuentos, y la incertidumbre en el espacio de parámetros, así como la incertidumbre presente, en el proceso espacio-temporal subyacente. Nuestros resultados muestran una notable concordancia cualitativa con los datos epidemiológicos reportados.

En el estudio empírico se distingue entre dos categorías de datos, definidas en sentido fuerte y débil, para referirnos a datos duros y débiles, respectivamente. Es bien sabido que la limitada disponibilidad de datos duros afecta al análisis espacio-temporal y, por lo tanto, la incorporación de datos blandos en los modelos de regresión ML puede ayudar a este análisis, proporcionando una visión global de la información de la muestra disponible, véase, por ejemplo, Christakos et al. (2001). En el análisis empírico comparativo realizado, con modelos de regresión ML, y el enfoque presentado, se consideran los datos duros y blandos de entrada.

Este análisis se basa en la validación cruzada y en la estimación de la densidad de probabilidad y de los intervalos de confianza mediante *bootstrap*. En concreto, la validación cruzada aleatoria k -fold ($k = 5, 10$), en términos de los Errores Porcentuales Absolutos Medios Simétricos (SMAPE), evalúa el rendimiento de los modelos de regresión comparados a partir de datos duros y blandos. Los intervalos de confianza *bootstrap* y las estimaciones de la densidad de probabilidad de los SMAPE promediados espacialmente proporcionan otra visión de este análisis empírico comparativo, apoyando la anterior clasificación de modelos basados en la validación cruzada aleatoria k -fold. En la categoría de datos blandos, donde nuestro enfoque se compara con la regresión ML, también se incorporan las correlaciones espaciales a corto plazo.

Podemos concluir que casi el mejor rendimiento, tanto en la categoría de datos duros como en la de datos blandos, lo muestran las Redes Neuronales definidas a partir de Bases de Funciones Radiales (RBF) y los Procesos Gaussianos (GP). Ambos enfoques mejoran cuando los datos blandos se incorporan al análisis de regresión. Se observan ligeras diferencias en el rendimiento de la Regresión de Soporte Vectorial (SVR) y las Redes Neuronales Bayesianas (BNN). El Perceptrón Multicapa (MLP) supera a las Redes Neuronales de Regresión Generalizada (GRNN), presentando mejores resultados de estimación cuando se analizan los datos duros.

Nuestras curvas de regresión trigonométricas, y el análisis clásico, basado en momentos empíricos de series temporales residuales multivariantes, presen-

ta los SMAPE similares a GRNN, en relación a las características de distribución aleatoria k -fold y bootstrapping, mostrando el peor rendimiento en la categoría de datos blandos. Obsérvese que GRNN también se ve favorecido por la categoría de datos blandos. En esta categoría, BNN y nuestro enfoque muestran un rendimiento muy similar, cuando la regresión trigonométrica se combina con la predicción bayesiana multivariante de series temporales residuales. De hecho, se observan algunas características de distribución bootstrapping ligeramente mejores de nuestro enfoque respecto a BNN en la categoría de datos blandos.

El esquema del capítulo es el siguiente. El enfoque de modelización se presenta en la Sección 7.1. En la Sección 7.2, se describe la metodología de predicción con múltiples objetivos. Esta metodología se aplica al análisis estadístico espacio-temporal de la mortalidad por COVID-19 en España en la Sección 7.3. El estudio empírico comparativo con los modelos de regresión ML se presenta en la Sección 7.4. Las conclusiones sobre nuestro ranking de modelos a partir de los datos se encuentran en la Sección 7.6. En la Sección 7.5, se ofrece una breve introducción a nuestra implementación de modelos ML a partir de datos duros y blandos. También se muestra información adicional sobre nuestro estudio empírico comparativo. En particular, se muestran los casos acumulados de mortalidad observados y predichos, y las curvas de log-riesgo.

7.1. Modelización de los datos

Sea $(\Omega, \mathcal{A}, \mathcal{P})$ el espacio de probabilidad básico. Consideramos que $H = L^2(\mathbb{R}^d)$, $d \geq 2$, el espacio de funciones de cuadrado integrable sobre \mathbb{R}^d , es el espacio de Hilbert real separable subyacente. En lo sucesivo, denotaremos por \mathcal{B}^d la σ -álgebra de Borel en \mathbb{R}^d , $d \geq 1$. Sea $X = \{X_t(\mathbf{z}), \mathbf{z} \in \mathbb{R}^d, t \in \mathbb{R}_+\}$ nuestro proceso espacio-temporal de datos duros de entrada sobre $(\Omega, \mathcal{A}, \mathcal{P})$, satisfaciendo $E[\|X_t(\cdot)\|_H^2] < \infty$, para cualquier tiempo $t \in \mathbb{R}_+$. El proceso de datos blandos de entrada sobre cualquier conjunto espacial delimitado $D \in \mathcal{B}^d$ se define entonces como

$$\left\{ X_t(h) = \int_D X_t(\mathbf{z})h(\mathbf{z})d\mathbf{z}, h \in \mathcal{C}_0^\infty(D), t \in \mathbb{R}_+ \right\}, \quad (7.1)$$

donde $\mathcal{C}_0^\infty(D)$ denota el espacio de funciones infinitas diferenciables, con soporte compacto contenido en D . Para cada conjunto acotado $D \in \mathcal{B}^d$, se define

$$\Lambda = \{\Lambda_t(h) = \exp(X_t(h)), h \in \mathcal{C}_0^\infty(D), t \in \mathbb{R}_+\}.$$

Supongamos que, para cualquier intervalo positivo finito $\mathcal{T} \in \mathcal{B}$, y conjunto acotado $D \in \mathcal{B}^d$,

$$\mathcal{I}_{\mathcal{T}}(h) = \int_{\mathcal{T}} \exp(X_t(h)) dt < \infty, \quad \forall h \in \mathcal{C}_0^\infty(D), \quad (7.2)$$

casi seguramente (c. s.). Sea $\{N_h : (\Omega, \mathcal{A}, \mathcal{P}) \times \mathcal{B} \rightarrow \mathbb{N}, h \in H\}$ una familia de medidas de recuento aleatorias. Supongamos que, para $h \in \mathcal{C}_0^\infty(D)$, siendo $D \in \mathcal{B}^d$, un conjunto acotado, dada la observación $\{x_t(h), t \in \mathcal{T}\}$, en el intervalo temporal finito $\mathcal{T} \in \mathcal{B}$, del proceso de datos blandos de entrada sobre la ventana espacial h en D , la distribución de probabilidad condicionada del número de sucesos aleatorios $N_h(\mathcal{T})$, que ocurren en $\mathcal{T} \in \mathcal{B}$, es una distribución de probabilidad de Poisson de media $\int_{\mathcal{T}} \exp(x_t(h)) dt$. Nos referimos a $\mathcal{I}_{\mathcal{T}}(h)$ como el proceso aleatorio generalizado de riesgo de mortalidad acumulada sobre el intervalo \mathcal{T} . Por lo tanto, el proceso de datos duros de entrada $X = \{X_t(\mathbf{z}), \mathbf{z} \in \mathbb{R}^d, t \in \mathbb{R}_+\}$ define el proceso de log-riesgo de mortalidad espacio-temporal.

El siguiente modelo de curvas de regresión se ajusta a partir del proceso de datos blandos de entrada, para $p = 1, \dots, P$,

$$\begin{aligned} \ln(\Lambda_t)(\psi_{p, \varpi_p}) &= g_t(\psi_{p, \varpi_p}, \boldsymbol{\theta}(p)) + \varepsilon_t(\psi_{p, \varpi_p}) \\ &= \langle g_t(\cdot, \boldsymbol{\theta}(p)), \psi_{p, \varpi_p}(\cdot) \rangle_H + \langle \varepsilon_t(\cdot), \psi_{p, \varpi_p}(\cdot) \rangle_H, \quad t \in \mathbb{R}_+, \end{aligned} \quad (7.3)$$

donde

$$g_t(\psi_{p, \varpi_p}, \boldsymbol{\theta}(p)) = \int_{\mathcal{D}_p} g_t(\mathbf{z}, \boldsymbol{\theta}(p)) \psi_{p, \varpi_p}(\mathbf{z}) d\mathbf{z}$$

$$\langle f, g \rangle_H = \int_{\mathbb{R}^d} f(\mathbf{z})g(\mathbf{z})d\mathbf{z}, \quad (7.4)$$

con $\{\psi_{p, \varpi_p}, p = 1, \dots, P\} \subset H$ denotando una familia de funciones en H , cuyos elementos tienen respectivos soportes compactos $\mathcal{D}_p, p = 1, \dots, P$, definiendo las áreas pequeñas p donde se agregan los recuentos, satisfaciendo condiciones de regularidad adecuadas. Para cada $p = 1, \dots, P$, el vector ϖ_p contiene los parámetros de centro y ancho de banda, definiendo la ventana seleccionada en el análisis de la zona pequeña p . Para cada $p \in \{1, \dots, P\}$, $\boldsymbol{\theta}(p) = (\theta^1(p), \dots, \theta^q(p)) \in \Theta$ representa el vector de parámetros desconocidos a estimar en la región p , y Θ es el conjunto abierto que define el espacio de parámetros, cuyo cierre Θ^c es un conjunto compacto en \mathbb{R}^q .

Asumimos que g_t es de la forma, véase, por ejemplo, [Ivanov et al. \(2015\)](#),

$$g_t(\boldsymbol{\theta}(p)) = \sum_{k=1}^N (A_k(p) \cos(\varphi_k(p)t) + B_k(p) \text{sen}(\varphi_k(p)t)), \quad p = 1, \dots, P, \quad t \in \mathbb{R}_+, \quad (7.5)$$

cuyos parámetros dependientes del espacio están dados por las escalas temporales $(\varphi_1(\cdot), \dots, \varphi_N(\cdot))$, y los coeficientes de Fourier $(A_1(\cdot), B_1(\cdot), \dots, A_N(\cdot), B_N(\cdot))$. Para simplificar, consideraremos que los parámetros de escala $\varphi_k, k = 1, \dots, N$, son conocidos, y fijos sobre las regiones espaciales P . Además, $C_k^2(\cdot) = A_k^2(\cdot) + B_k^2(\cdot) > 0$, para $k = 1, \dots, N$, donde N denota el parámetro de truncamiento, que se seleccionará según el factor penalizado propuesto en [Chapelle et al. \(2002\)](#), como explicamos con más detalle en la Sección 7.2. Así,

$$\boldsymbol{\theta}(p) = (A_1(p), B_1(p), \dots, A_N(p), B_N(p)), \quad p = 1, \dots, P.$$

Para analizar la correlación espacial entre las regiones, se considera un modelo autorregresivo multivariante para la predicción del término residual de regresión en cada región $p \in \{1, \dots, P\}$. En particular, para cualquier $T \geq 2$, se supone que ε_t en la ecuación (7.3) satisface, para $p = 1, \dots, P$, la siguiente ecuación de estados,

$$\varepsilon_t(\psi_{p,\varpi_p}) = \sum_{q=1}^P \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p})\varepsilon_{t-1}(\psi_{q,\varpi_q}) + \nu_t(\psi_{p,\varpi_p}), \quad (7.6)$$

donde, para cualquier $t \in \mathbb{R}_+$, y $p, q = 1, \dots, P$,

$$\begin{aligned} \varepsilon_t(\psi_{p,\varpi_p}) &= \int_{\mathcal{D}_p} \varepsilon_t(\mathbf{z})\psi_{p,\varpi_p}(\mathbf{z})d\mathbf{z} \\ \nu_t(\psi_{p,\varpi_p}) &= \int_{\mathcal{D}_p} \nu_t(\mathbf{z})\psi_{p,\varpi_p}(\mathbf{z})d\mathbf{z} \\ \rho(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) &= \int_{\mathcal{D}_p \times \mathcal{D}_q} \rho(\mathbf{z}, \mathbf{y})\psi_{p,\varpi_p}(\mathbf{z})\psi_{q,\varpi_q}(\mathbf{y})d\mathbf{y}d\mathbf{z}. \end{aligned}$$

Aquí, $(\nu_t(\psi_{p,\varpi_p}), p = 1, \dots, P), t \in \mathbb{R}_+$, se supone que son vectores gaussianos independientes de media cero P -dimensional. Para $p, q \in \{1, \dots, P\}$, la proyección $\rho(\psi_{p,\varpi_p})(\psi_{q,\varpi_q})$ mantiene la autocorrelación lineal temporal en cada región espacial para $p = q$, y la correlación cruzada lineal temporal entre regiones para $p \neq q$ del error de regresión $\{\varepsilon_t(\cdot), t \in \mathbb{R}_+\}$, véase Bosq (2000).

7.2. Aplicación del modelo de curvas de regresión y del análisis residual espacial

Sean $\mathcal{D}_1, \dots, \mathcal{D}_P$ las áreas pequeñas, en las que se agregan los recuentos, y $\{\psi_{p,\varpi_p}, \varpi_p = (c_p, \rho_p), p = 1, \dots, P\} \subset H$ las funciones con respectivos soportes compactos $\mathcal{D}_1, \dots, \mathcal{D}_P$. En particular, denotamos por $c_p, p = 1, \dots, P$, los centros respectivamente asignados a las regiones $\mathcal{D}_1, \dots, \mathcal{D}_P$, y por ρ_1, \dots, ρ_P , los parámetros de ancho de banda que proporcionan los tamaños de ventana asociados.

En la práctica, a partir del modelo de observación (7.3), para encontrar g_t en (7.5) que minimice la función de pérdida cuadrática esperada, o el riesgo esperado, buscamos el minimizador $\hat{\theta}_T(p)$ del riesgo de regresión empírico

$$L_T(\hat{\theta}_T(p)) = \inf_{\theta(p) \in \Theta^c} L_T(\theta(p)) = \inf_{\theta(p) \in \Theta^c} \frac{1}{T} \sum_{t=1}^T |\ln(\Lambda_t)(\psi_{p,\varpi_p}) - g_t(\theta(p))|^2. \quad (7.7)$$

El parámetro de truncamiento N se selecciona entonces para controlar la rela-

ción entre la función de pérdida cuadrática esperada en $\widehat{\boldsymbol{\theta}}_T(p)$, y el valor esperado del riesgo mínimo empírico a partir de la identidad

$$\begin{aligned} E \left[\ln(\Lambda_t)(\psi_{p,\varpi_p}) - g_t(\psi_{p,\varpi_p}, \widehat{\boldsymbol{\theta}}_T(p)) \right]^2 \\ = E \left[L_T(\widehat{\boldsymbol{\theta}}_T(p)) \right] \left(1 - \frac{N}{T} \right)^{-1} \left(1 + \frac{\sum_{i=1}^N 1/\lambda_i}{T} \right), \end{aligned} \quad (7.8)$$

donde, para $i = 1, \dots, N$, $1/\lambda_i$ denota la inversa del i -ésimo autovalor de la matriz $\Phi^T \Phi$, siendo Φ una matriz $T \times N$, cuyos elementos son los valores de las N funciones de base trigonométrica seleccionadas en los puntos temporales $t = 1, \dots, T$. El parámetro N debe ser tal que $N < T$. Obsérvese que, asintóticamente, cuando $N \rightarrow \infty$, $\Phi^T \Phi$ va a la matriz identidad, y para $i = 1, \dots, N$, $1/\lambda_i \sim 1$. Para $p = 1, \dots, P$, hemos considerado el riesgo mínimo empírico

$$L_T(\widehat{\boldsymbol{\theta}}_T(p)) = \frac{1}{T} \widetilde{\mathcal{R}}^T(p) \left(I_{T \times T} - \Phi (\Phi^T \Phi)^{-1} \Phi^T \right) \widetilde{\mathcal{R}}(p), \quad (7.9)$$

con, para $p = 1, \dots, P$,

$$\widetilde{\mathcal{R}}(p) = \left(\sum_{k=N+1}^{\infty} (A_k(p) \cos(\varphi_k t) + B_k(p) \text{sen}(\varphi_k t)) + \varepsilon_t(\psi_{p,\varpi_p}), t = 1, \dots, T \right).$$

Así, nuestro predictor de regresión viene dado por

$$\widehat{\ln(\Lambda_t)}(\psi_{p,\varpi_p}) = g_t(\widehat{\boldsymbol{\theta}}_T(p)), \quad p = 1, \dots, P, \quad (7.10)$$

véase el Teorema 1 en Ivanov et al. (2015) sobre las condiciones para la consistencia débil de (7.10).

Se consideran los residuos de regresión,

$$\mathbf{Y} = \left\{ Y_t(\psi_{p,\varpi_p}) = \ln(\Lambda_t)(\psi_{p,\varpi_p}) - g_t(\widehat{\boldsymbol{\theta}}_T(p)), t = 1, \dots, T, p = 1, \dots, P \right\},$$

y los operadores nucleares empíricos de autocovarianza y covarianza cruzada,

$$\begin{aligned}\widehat{R}_{0,T}^{\mathbf{Y}}(\psi_{p,\varpi_p})(\psi_{q,\varpi_q}) &= \frac{1}{T} \sum_{t=1}^T Y_t(\psi_{p,\varpi_p}) Y_t(\psi_{q,\varpi_q}), \\ \widehat{R}_{1,T}^{\mathbf{Y}}(\psi_{p,\varpi_p})(\psi_{q,\varpi_q}) &= \frac{1}{T-1} \sum_{t=1}^{T-1} Y_t(\psi_{q,\varpi_q}) Y_{t+1}(\psi_{p,\varpi_p}), \quad p, q = 1, \dots, P,\end{aligned}\tag{7.11}$$

respectivamente, en la estimación de la correlación residual lineal espacial (Bosq, 2000). También se considera aquí un parámetro de truncamiento $k(T)$ para eliminar la inestabilidad de este problema de estimación por la discontinuidad respecto a los datos de entrada. En particular, $k(T)$ debe satisfacer $k(T) \rightarrow \infty$, $k(T)/T \rightarrow 0$, $T \rightarrow \infty$. Una elección adecuada de $k(T)$ también garantiza una fuerte consistencia del estimador

$$\widehat{\rho}_{k(T)}(\psi_{p,\varpi_p})(\psi_{q,\varpi_q}) = \sum_{k,l=1}^{k(T)} \frac{\langle \psi_{p,\varpi_p}, \phi_{k,T} \rangle_H \langle \psi_{q,\varpi_q}, \phi_{l,T} \rangle_H}{\lambda_{k,T}(\widehat{R}_{0,T}^{\mathbf{Y}})} \widehat{R}_{1,T}^{\mathbf{Y}}(\phi_{k,T})(\phi_{l,T}),\tag{7.12}$$

para $p, q = 1, \dots, P$, (Bosq, 2000). Aquí,

$$\widehat{R}_{0,T}^{\mathbf{Y}} = \sum_{k=1}^T \lambda_{k,T}(\widehat{R}_{0,T}^{\mathbf{Y}}) [\phi_{k,T} \otimes \phi_{k,T}],\tag{7.13}$$

donde $\{\lambda_{k,T}(\widehat{R}_{0,T}^{\mathbf{Y}}), k = 1, \dots, T\}$ y $\{\phi_{k,T}, k \geq 1\}$ denotan los autovalores empíricos y los autovectores de $\widehat{R}_{0,T}^{\mathbf{Y}}$, respectivamente. En particular, consideramos $k(T) = \ln(T)$, ver Bosq (2000). El predictor *plug-in* clásico se calcula entonces, para cada $p = 1, \dots, P$, como

$$\widehat{Y}_t^{k(T)}(\psi_{p,\varpi_p}) = \sum_{q=1}^P \widehat{\rho}_{k(T)}(\psi_{q,\varpi_q})(\psi_{p,\varpi_p}) Y_{t-1}(\psi_{q,\varpi_q}), \quad t \geq 1.\tag{7.14}$$

Bajo la distribución gaussiana de ν_t , en la estimación bayesiana de ρ , de (7.6), la función de verosimilitud, que define la función objetivo, viene dada por, para cada $p = 1, \dots, P$,

$$\begin{aligned}
& \tilde{L}_p(\varepsilon_{1p}, \dots, \varepsilon_{Tp}, \varepsilon_{0q}, \dots, \varepsilon_{(T-1)q} \rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}), q = 1, \dots, P) \\
&= \frac{\exp\left(-\frac{1}{2\sigma_p^2} \sum_{t=1}^T \left(\varepsilon_t(\psi_{p, \varpi_p}) - \sum_{q=1}^P \varepsilon_{t-1}(\psi_{q, \varpi_q}) \rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p})\right)^2\right)}{(\sigma_p \sqrt{2\pi})^T} \\
&\quad \times \prod_{q=1}^P [\rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p})]^{a_{pq}-1} (1 - \rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}))^{b_{pq}-1} \\
&\quad \times \frac{\mathbb{I}_{\{0 < \rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}) < 1\}}}{\mathbb{B}(a_{pq}, b_{pq})} \\
&= \frac{1}{(\sigma_p \sqrt{2\pi})^T} \exp\left(-\frac{1}{2\sigma_p^2} \sum_{t=1}^T [\nu_t(\psi_{p, \varpi_p})]^2\right) \\
&\quad \times \prod_{q=1}^P [\rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p})]^{a_{pq}-1} (1 - \rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}))^{b_{pq}-1} \\
&\quad \times \frac{\mathbb{I}_{\{0 < \rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}) < 1\}}}{\mathbb{B}(a_{pq}, b_{pq})},
\end{aligned} \tag{7.15}$$

donde, para cada $p = 1, \dots, P$, las distribuciones de probabilidad Beta con parámetros de forma a_{pq} y b_{pq} , $q = 1, \dots, P$, definen respectivamente las distribuciones de probabilidad a-priori de las variables aleatorias independientes $\{\rho(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}), q = 1, \dots, P\}$. Aquí, para cada $p = 1, \dots, P$, $\varepsilon_{tp} = \varepsilon_t(\psi_{p, \varpi_p}) = \langle \varepsilon_t, \psi_{p, \varpi_p} \rangle_H$, y $\sigma_p = \sqrt{E[\varepsilon_t(\psi_{p, \varpi_p})]^2}$, para $t = 0, \dots, T$. Como antes, ψ_{p, ϖ_p} pondera la información de la muestra espacial sobre el área pequeña p , para $p = 1, \dots, P$. Como es habitual, $\mathbb{I}_{0 < \cdot < 1}$ denota la función indicadora en el intervalo $(0, 1)$, y $\mathbb{B}(a_{pq}, b_{pq})$ es la función Beta,

$$\mathbb{B}(a_{pq}, b_{pq}) = \frac{\Gamma(a_{pq})\Gamma(b_{pq})}{\Gamma(a_{pq} + b_{pq})}.$$

A partir de (7.15), el predictor bayesiano se obtiene, para $p = 1, \dots, P$, como

$$\tilde{\varepsilon}_t(\psi_{p, \varpi_p}) = \sum_{q=1}^P \tilde{\rho}(\psi_{q, \varpi_q})(\psi_{p, \varpi_p}) \varepsilon_{t-1}(\psi_{q, \varpi_q}), \quad t \geq 1, \tag{7.16}$$

con $(\tilde{\rho}(\psi_{1, \varpi_1})(\psi_{p, \varpi_p}), \dots, \tilde{\rho}(\psi_{P, \varpi_P})(\psi_{p, \varpi_p}))$ que se calcula maximizando (7.15), para encontrar la moda de la distribución a-posteriori (véase Bosq y Ruiz-Medina (2014), donde se introduce la estimación bayesiana en un marco

infinito-dimensional). Nos referimos a (7.16) como el predictor bayesiano *plug-in* del proceso de log-riesgo de mortalidad residual en el área pequeña p , para $p = 1, \dots, P$. En la práctica, la ecuación (7.15) se aproxima a partir de los valores calculados del proceso residual de regresión.

7.3. Análisis estadístico de la mortalidad por COVID-19

Nuestro análisis se basa en los registros diarios de mortalidad por COVID-19 comunicados por el Instituto de Salud Carlos III (ISCHII, 2020), desde el 8 de marzo hasta el 13 de mayo de 2020, en las 17 comunidades españolas. Se ha considerado la interpolación en 265 nodos temporales, y el suavizado con B-spline cúbico en la implementación de (7.3)–(7.5). El predictor de regresión se calcula a partir de la ecuación (7.10), aplicando (7.7)–(7.8). En particular,

$$T(265, 12) = \left(1 - \frac{12}{265}\right)^{-1} \left(1 + \frac{\sum_{i=1}^N 1/\lambda_i}{T}\right) = 1.1304, \quad (7.17)$$

corresponde a nuestra elección $N = 6$ en (7.5). Las tablas 7.1–7.2 siguientes muestran las estimaciones de los parámetros $\hat{A}_k(\cdot)$ y $\hat{B}_k(\cdot)$, $k = 1, \dots, 6$, donde se ha considerado $\varphi_k = \frac{2\pi}{265}$, para $k = 1, \dots, N = 6$.

En dichas tablas y siguientes, aparecen los siguientes códigos de la Comunidad Española (SC): C1 para Andalucía; C2 para Aragón; C3 para Asturias; C4 para Islas Baleares; C5 para Canarias; C6 para Cantabria; C7 para Castilla La Mancha; C8 para Castilla y León; C9 para Cataluña; C10 para Comunidad Valenciana; C11 para Extremadura; C12 para Galicia; C13 para Comunidad de Madrid; C14 para Murcia; C15 para Navarra; C16 para País Vasco, y C17 para La Rioja.

Se calculan los intervalos de confianza de la curva *bootstrap* con un nivel de confianza de $1 - \alpha = 0.95$, basados en muestras *bootstrap* de 1000 para la media espacial, sobre las 17 comunidades españolas, de los predictores curva de regresión. Su construcción se basa en el método del percentil corregido por el sesgo y acelerado (\mathcal{I}_1); intervalo normal aproximado con sesgo y error estándar *bootstrap* (\mathcal{I}_2); método del percentil básico (\mathcal{I}_3), y método del percentil corregido por el sesgo (\mathcal{I}_4), véase la Figura 7.1 a continuación. Los valores de

Tabla 7.1: Estimaciones de los parámetros de regresión $\hat{A}_k(\cdot)$, $k = 1, \dots, 6$, en las 17 comunidades españolas

SC/PE	$\hat{A}_1(\cdot)$	$\hat{A}_2(\cdot)$	$\hat{A}_3(\cdot)$	$\hat{A}_4(\cdot)$	$\hat{A}_5(\cdot)$	$\hat{A}_6(\cdot)$
C1	3.6343	-0.4814	-0.0075	-0.0258	0.0189	0.0193
C2	3.4345	-0.3923	0.0416	0.0265	-0.0709	-0.0572
C3	3.2031	-0.1364	-0.0088	0.0221	0.0430	0.0289
C4	3.1445	-0.1118	0.0041	0.0337	0.0062	0.0072
C5	3.1015	-0.0693	-0.0345	0.0352	-0.0112	0.0003
C6	3.1347	-0.1397	0.0020	0.0300	-0.0061	-0.0002
C7	4.0591	-0.5487	-0.0907	0.0951	0.0992	0.0842
C8	3.8032	-0.5500	-0.1007	0.0633	0.0139	0.0277
C9	4.5095	-0.7435	-0.1134	0.1809	0.2231	0.2026
C10	3.6321	-0.4685	-0.0540	0.0384	-0.0152	0.0011
C11	3.2967	-0.2274	-0.0083	0.0553	0.0250	0.0240
C12	3.3454	-0.2122	-0.0927	-0.0330	0.0724	0.0679
C13	4.8419	-0.6790	-0.2455	0.0311	0.0554	0.0667
C14	3.0941	-0.1037	0.0210	0.0141	-0.0016	0.0041
C15	3.2877	-0.2598	-0.0524	0.0842	-0.0423	-0.0348
C16	3.6870	-0.4302	-0.0086	0.0078	-0.0027	-0.0017
C17	3.2197	-0.2071	0.0162	0.0079	0.0206	0.0110

Tabla 7.2: Estimaciones de los parámetros de regresión $\hat{B}_k(\cdot)$, $k = 1, \dots, 6$, en las 17 comunidades españolas

SC/PE	$\hat{B}_1(\cdot)$	$\hat{B}_2(\cdot)$	$\hat{B}_3(\cdot)$	$\hat{B}_4(\cdot)$	$\hat{B}_5(\cdot)$	$\hat{B}_6(\cdot)$
C1	0	-0.0052	-0.1330	-0.0123	0.0064	-0.0195
C2	0	-0.0367	-0.0998	-0.0462	-0.0343	-0.0107
C3	0	-0.0531	-0.0074	-0.0142	-0.0003	0.0020
C4	0	-0.0074	-0.0284	-0.0151	-0.0092	0.0012
C5	0	0.0433	-0.0438	-0.0116	-0.0118	0.0046
C6	0	0.0018	-0.0174	-0.0068	-0.0089	0.0000
C7	0	-0.0365	-0.2451	-0.1791	-0.0820	0.0026
C8	0	0.0953	-0.2389	-0.0431	-0.0313	-0.0045
C9	0	-0.1587	-0.4054	-0.2269	-0.1010	0.0047
C10	0	0.1118	-0.1579	-0.0458	-0.0418	-0.0220
C11	0	0.0754	-0.1138	-0.0166	-0.0048	0.0072
C12	0	-0.1104	-0.1338	0.1330	0.0761	-0.0017
C13	0	0.4654	-0.1302	-0.1602	-0.1061	-0.0038
C14	0	0.0355	-0.0560	0.0119	0.0025	-0.0044
C15	0	-0.0187	-0.0021	-0.0897	-0.0562	0.0134
C16	0	0.0025	-0.0707	-0.0638	-0.0439	-0.0267
C17	0	0.0389	-0.0270	-0.0174	-0.0006	0.0019

riesgo de regresión mínimo empírico $L_{265}(\hat{\theta}_{265}(p))$, $p = 1, \dots, 17$, se muestran en la Tabla 7.3.

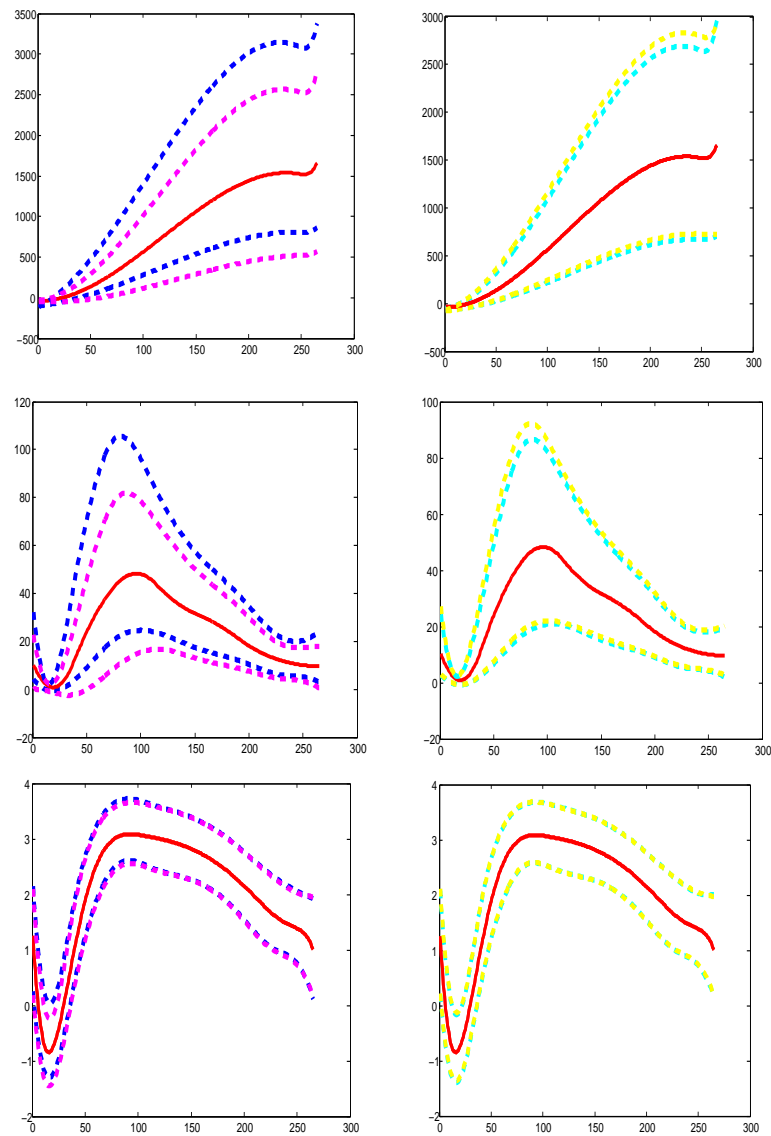


Figura 7.1: En la parte izquierda, curva acumulada de la media de mortalidad por COVID-19 en España, desde el 8 de marzo de 2020 hasta el 13 de mayo de 2020 (línea roja continua, 265 nodos temporales), e intervalos de confianza de la curva *bootstrap*, arriba, (líneas azules discontinuas) y \mathcal{I}_1 (líneas magenta discontinuas), y abajo, \mathcal{I}_2 (líneas verdes discontinuas) y \mathcal{I}_3 (líneas verdes discontinuas) y \mathcal{I}_4 (líneas amarillas discontinuas). Los gráficos del centro y de la parte derecha reflejan la misma información, respectivamente, referida a las curvas de intensidad media (curva de riesgo de mortalidad COVID-19 promediada espacialmente), y log-intensidad (curva de log-riesgo de mortalidad por COVID-19 promediada espacialmente) en España. Todos los intervalos *bootstrap* de confianza se calculan con un nivel de confianza de $1 - \alpha = 0.95$, a partir de 1000 muestras *bootstrap*

Tabla 7.3: Valores observados $L_{265}(\hat{\theta}_{265}(p))$, $p = 1, \dots, 17$

$L_{265}(\hat{\theta}_{265}(p))$	$p =$	1 ... 17		
0.0155	0.0259	0.0668	0.0408	0.0927
0.0623	0.1642	0.0883	0.2174	0.0313
0.0559	0.1904	0.0054	0.1602	0.1640
0.0003	0.1238			

La Figura 7.2, en la parte superior muestra 1000 valores muestrales *bootstrap*

$$\overline{L}_{265}(\omega_i) = \frac{1}{P} \sum_{p=1}^P L_{265}(\omega_i, \hat{\theta}_{265}(p)), \quad \omega_i \in \Omega, \quad i = 1, \dots, 1000,$$

del riesgo mínimo cuadrático empírico promediado espacialmente en la regresión trigonométrica. Obsérvese que la media muestral de estos valores es $\overline{L} = 0.0262$, lo que muestra un buen rendimiento del predictor de regresión por mínimos cuadrados, según el valor $T(265, 12) = 1.1304$ obtenido. El histograma *bootstrap* y la correspondiente aproximación de la función de densidad de probabilidad, calculada a partir de $\overline{L}_{265}(\omega_i)$, $i = 1, \dots, 1000$, también se representan en la parte inferior de la Figura 7.2.

También se han calculado los intervalos de confianza *bootstrap* para \overline{L}_{265} al nivel $1 - \alpha = 0.95$, a partir de 1000 y 10000 muestras *bootstrap*. La Tabla 7.4 muestra estos intervalos respectivamente basados en el método del percentil corregido y acelerado (\mathcal{I}_1); intervalo aproximado Normal con sesgo *bootstrap* y error estándar (\mathcal{I}_2); método del percentil básico (\mathcal{I}_3); método del percentil corregido por el sesgo (\mathcal{I}_4), e intervalo de confianza basado en el estudiante (\mathcal{I}_5).

Los predictores clásico y bayesiano del proceso de log-riesgo de mortalidad por COVID-19 residual en cada una de las comunidades españolas se calculan respectivamente a partir de las ecuaciones (7.14) y (7.16), para $P = 17$.

Dadas las características espectrales empíricas observadas en la aproximación regularizada $\hat{\rho}_{k(T)}$ de ρ en (7.12), a partir de la descomposición del valor singular de los operadores empíricos en (7.11) nuestra elección de la distribución a-priori para las proyecciones de ρ ha sido una probabilidad a-priori Beta escalada, por un factor de $1/3$, con hiperparámetros $a_{pq} = 14$, y $b_{pq} = 13$, para $p, q = 1, \dots, 17$. La idoneidad de esta elección basada en los datos, en lo que

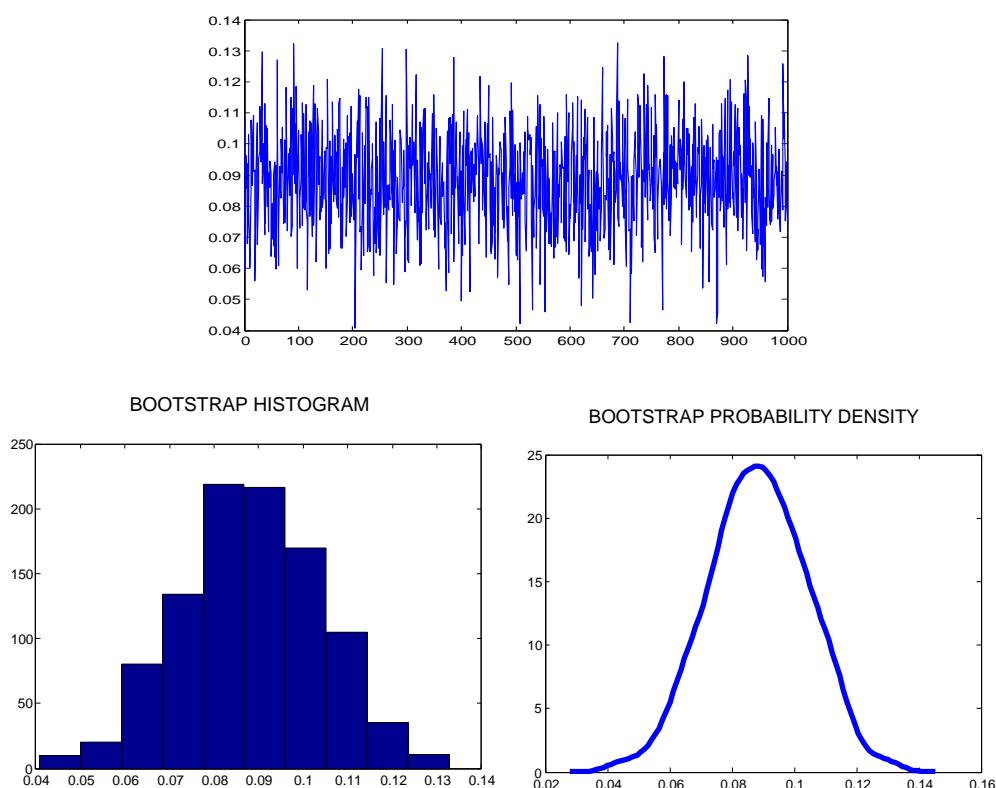


Figura 7.2: Valores muestrales de la media espacial del riesgo mínimo empírico en la regresión trigonométrica, en la parte superior. Histograma *bootstrap* (parte inferior izquierda) y densidad de probabilidad *bootstrap* (parte inferior derecha) de la media espacial del riesgo mínimo empírico de regresión, basado en 1000 muestras *bootstrap*

Tabla 7.4: Intervalos de confianza *bootstrap* para $\overline{L_{265}}$ (nivel de confianza $1 - \alpha = 0.95$)

CI/S	1000	10000
\mathcal{I}_1	[0.0593, 0.1222]	[0.0594, 0.1236]
\mathcal{I}_2	[0.0564, 0.1196]	[0.0567, 0.1207]
\mathcal{I}_3	[0.0584, 0.1215]	[0.0579, 0.1217]
\mathcal{I}_4	[0.0592, 0.1233]	[0.0581, 0.1208]
\mathcal{I}_5	[0.0484, 0.1281]	[0.0494, 0.1215]

respecta a la localización de la moda, y el grosor de las colas, se ilustra en la Figura 7.3. En concreto, en el gráfico de la derecha de la Figura 7.3, se muestran tanto la densidad de probabilidad Beta escalada, con los parámetros de forma con valores 14 y 13 (línea cuadrada roja), como la densidad de probabilidad

ajustada (línea cuadrada azul), a partir de las muestras *bootstrap* generadas, basadas en las proyecciones empíricas de ρ . Obsérvese que el rango observado de las proyecciones empíricas de ρ está bien ajustado, como puede verse en el gráfico de la izquierda de la Figura 7.3.

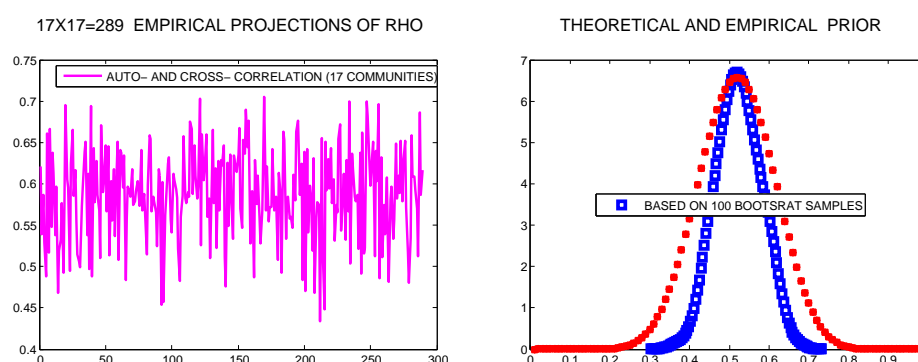


Figura 7.3: Proyecciones empíricas del operador de autocorrelación ρ (lado izquierdo), que reflejan la autocorrelación temporal y la correlación cruzada entre las 17 comunidades españolas analizadas. En el lado derecho, la densidad de probabilidad a-priori considerada (cuadrados rojos) de una variable aleatoria con distribución Beta escalada por el factor $1/3$, con parámetros de forma con valores 14 y 13, se compara con el ajuste *bootstrap* a priori de una variable empírica (cuadrados azules)

Los intervalos de confianza *bootstrap* $\mathcal{I}_1, \dots, \mathcal{I}_5$ al nivel $1 - \alpha = 0.95$, para el error estándar de entrenamiento esperado de los predictores del log-riesgo de mortalidad por COVID-19 de series temporales multivariadas clásicas y bayesianas residuales, basados en 1000 muestras *bootstrap*, se muestran en la Tabla 7.5.

Tabla 7.5: Intervalos de confianza *bootstrap* para el error estándar de entrenamiento esperado de los predictores de log-riesgo de mortalidad por COVID-19 clásico y bayesiano ($1 - \alpha = 0.95$)

CI/S	Clásico	Bayesiano
\mathcal{I}_1	[0.0474, 0.0597]	[0.0173, 0.0228]
\mathcal{I}_2	[0.0455, 0.0578]	[0.0167, 0.0220]
\mathcal{I}_3	[0.0463, 0.0588]	[0.0169, 0.0225]
\mathcal{I}_4	[0.0460, 0.0586]	[0.0172, 0.0226]
\mathcal{I}_5	[0.0421, 0.0563]	[0.0158, 0.0215]

Los mapas representados en la Figura 7.4 muestran la evolución espacio-temporal observada del riesgo de mortalidad por COVID-19, y su predicción, a partir del modelo de regresión trigonométrica de la curva ajustada, y el posterior análisis de series temporales clásico y bayesiano.

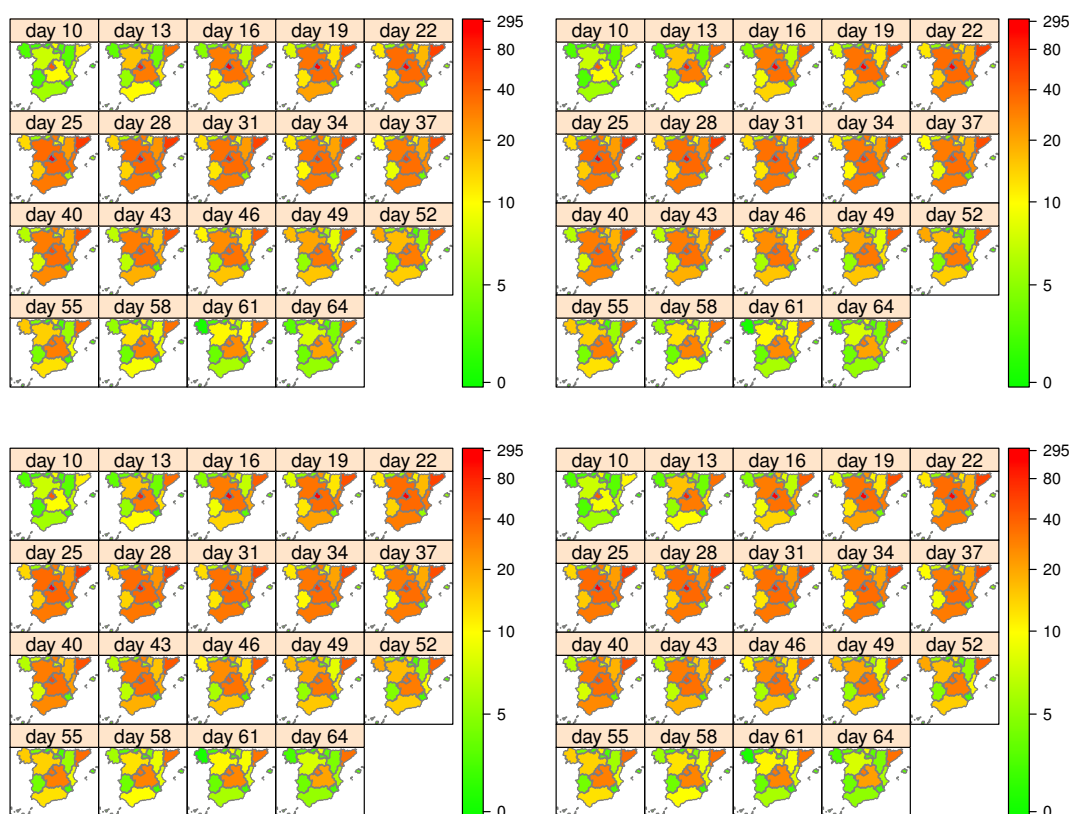


Figura 7.4: Mapas de riesgo de mortalidad por COVID-19, desde el 8 de marzo hasta el 13 de mayo de 2020. Mapas observados (arriba) y estimados (abajo), calculados a partir de la regresión trigonométrica, combinados con predictores residuales clásicos (abajo izquierda) y bayesianos (abajo derecha)

7.4. Un estudio empírico comparativo

Los modelos de regresión ML introducidos en la Sección 7.5 se aplican al análisis de mortalidad por COVID-19 y se comparan, a través de la validación cruzada aleatoria k -fold y los estimadores *bootstrap*, con el enfoque de predicción espacio-temporal con múltiples objetivos presentado. Distinguimos dos categorías, respectivamente, referidas a la definición de sentido fuerte (da-

tos duros) y sentido débil (datos blandos) de nuestro conjunto de datos. La validación cruzada aleatoria k -fold ($k = 5, 10$), en términos de los Errores Porcentuales Absolutos Medios Simétricos (SMAPE), evalúa el rendimiento de los modelos de regresión comparados, a partir de datos duros y blandos. También se calculan los intervalos de confianza *bootstrap* y las estimaciones de densidad de probabilidad de los SMAPE promediados espacialmente. La Sección 7.6 proporciona una clasificación de los modelos basada en las dos categorías analizadas de datos, a partir de los SMAPE obtenidos mediante la validación cruzada aleatoria k -fold, y los procedimientos de estimación *bootstrap* aplicados.

7.4.1. Resultados de la validación cruzada aleatoria k -fold

Tras la interpolación y el suavizado B -spline cúbico de nuestro conjunto de datos original, se aplica la transformación logarítmica y el escalado lineal. Hemos eliminado los diez primeros puntos y los tres últimos, para cada curva de log-riesgo de mortalidad por COVID-19, como un conjunto fuera de la muestra. Nuestro enfoque se implementa en la segunda categoría a partir de los datos blandos. En esta implementación, consideramos $N = 6$, adoptando el criterio de selección del modelo dado en la Sección 7.3, véase la ecuación (7.8) y la referencia Chapelle et al. (2002).

En la predicción clásica y bayesiana de series temporales multivariantes, nuestra elección de $k(T) = k(265) = 8$ proporciona un equilibrio entre $k(T) = [\ln(T)]^- = [\ln(265)]^- = 5$, estableciendo un acuerdo con la separación y velocidad de decaimiento de los autovalores empíricos del operador de autocovarianza, y el valor del parámetro $k(T) = 9$, controlando la complejidad del modelo según el tamaño muestral $T = 265$. Las fluctuaciones aleatorias, observadas en las proyecciones empíricas del operador de autocorrelación espacial ρ para este valor de parámetro $k(T)$, también están bien ajustadas, teniendo en cuenta la elección comentada de los hiperparámetros de forma, en la densidad de probabilidad a-priori Beta.

El ajuste del modelo se evalúa en términos de los Errores Porcentuales Absolutos Medios Simétricos (SMAPE), dados por, para $P = 17$, y $T = 265$,

$$\frac{1}{T} \sum_{t=1}^T \frac{|\widehat{\ln(\Lambda_t)}(\psi_{p,\rho_p}) - \ln(\Lambda_t)(\psi_{p,\rho_p})|}{\left(|\ln(\Lambda_t)(\psi_{p,\rho_p})| + |\widehat{\ln(\Lambda_t)}(\psi_{p,\rho_p})| \right) / 2}, \quad p = 1, \dots, P. \quad (7.18)$$

Hemos calculado la media de los SMAPE obtenidos en cada una de las k iteraciones del procedimiento de validación cruzada aleatoria k -fold. Esta técnica de validación consiste en dividir aleatoriamente la muestra funcional en una muestra de entrenamiento y otra de validación en cada una de las iteraciones k . El ajuste del modelo se realiza a partir de la muestra de entrenamiento, y las salidas objetivo se definen a partir de la muestra de validación o de prueba. Al ejecutar cada modelo diez veces y promediar los SMAPE, eliminamos las fluctuaciones debidas a los pesos iniciales aleatorios (para los modelos MLP y BNN), y las diferencias en la estimación de los parámetros en todos los métodos, debido a la especificación aleatoria de la división de la muestra en el procedimiento de validación cruzada aleatoria k -fold.

Los SMAPE de validación cruzada aleatoria 10-fold basados en diez ejecuciones se muestran en la Tabla 7.6, para las seis técnicas ML probadas, GRNN, MLP, SVR, BNN, RBF y GP, cuando se consideran datos duros (véase también la Tabla 7.13 de la Sección 7.5 sobre los resultados de validación cruzada aleatoria 5-fold). La Tabla 7.7 proporciona los resultados de validación cruzada aleatoria 10-fold, de la categoría de datos blandos (véase también la Tabla 7.14 de la Sección 7.5 sobre los resultados de validación cruzada aleatoria 5-fold). Los correspondientes resultados de validación cruzada del enfoque presentado a partir de datos blandos se muestran en la Tabla 7.8.

La selección de hiperparámetros del modelo ML se ha realizado aplicando una validación cruzada aleatoria k -fold ($k = 5, 10$). Nuestra selección se ha realizado a partir de un conjunto adecuado de candidatos. En concreto, los números óptimos de nodos ocultos (NH) en la implementación de MLP y BNN se han seleccionado de los conjuntos de candidatos $[0, 1, 3, 5, 7, 9]$ y $[1, 3, 5, 7, 9]$, respectivamente. Los resultados de la validación cruzada aleatoria en ambos casos, $k = 5, 10$, llevan a la misma elección del valor óptimo de NH, a saber, $NH = 1$ para MLP, y $NH = 5$ para BNN. Este último muestra ligeras diferencias con respecto a los valores $NH = 3, 7$, en la implementación de validación cruzada aleatoria 10-fold. Del mismo modo, hemos seleccionado los respectivos parámetros de dispersión β y ancho de banda h en los procedimientos RBF y GRNN. Así, después de aplicar una validación cruzada aleatoria k -fold, con $k = 5, 10$, se obtienen los valores óptimos $\beta = 2.5$, y $h = 0.05$, a partir de los conjuntos de candidatos $[2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20]$ y $[0.05, 0.1, 0.2, 0.3, 0.5, 0.6, 0.7]$, respectivamente (véase la Sección 7.5).

Tabla 7.6: **Categoría de datos duros**. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 10-fold. Como se indica, los valores mostrados deben ser multiplicados por 10^{-2}

SC($\times 10^{-2}$)	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1957	0.0777	0.0700	0.0594	0.0543	0.0554
C2	0.6132	0.1490	0.0663	0.0738	0.0680	0.0654
C3	0.1556	0.0473	0.0350	0.0303	0.0331	0.0304
C4	0.0971	0.0342	0.0135	0.0200	0.0182	0.0211
C5	0.2049	0.0457	0.0318	0.0370	0.0369	0.0372
C6	0.1572	0.0368	0.0177	0.0234	0.0233	0.0247
C7	0.4898	0.0698	0.0644	0.0590	0.0616	0.0588
C8	0.0804	0.0340	0.0171	0.0191	0.0211	0.0177
C9	0.7258	0.1976	0.0979	0.0812	0.0326	0.0437
C10	0.2191	0.0704	0.0556	0.0482	0.0471	0.0463
C11	0.1262	0.0530	0.0310	0.0395	0.0375	0.0355
C12	0.5228	0.1578	0.1341	0.1282	0.0940	0.0993
C13	0.3594	0.0647	0.0576	0.0579	0.0533	0.0458
C14	0.1345	0.0366	0.0209	0.0204	0.0194	0.0207
C15	0.6080	0.1523	0.1411	0.1141	0.0982	0.1039
C16	0.2464	0.0889	0.0709	0.0622	0.0568	0.0594
C17	0.0660	0.0370	0.0148	0.0222	0.0203	0.0227
M.	0.2942	0.0796	0.0553	0.0527	0.0456	0.0463
T.	5.0022	1.3528	0.9397	0.8959	0.7757	0.7879

Tabla 7.7: **Categoría de datos blandos**. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 10-fold. Como se indica, los valores mostrados deben ser multiplicados por 10^{-2}

SC($\times 10^{-2}$)	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1545	0.0983	0.0666	0.0573	0.0234	0.0312
C2	0.1844	0.1730	0.0660	0.0749	0.0277	0.0301
C3	0.1029	0.1192	0.0481	0.0452	0.0273	0.0274
C4	0.0432	0.0286	0.0165	0.0158	0.0124	0.0123
C5	0.0610	0.0476	0.0258	0.0248	0.0144	0.0149
C6	0.0260	0.0217	0.0133	0.0140	0.0124	0.0125
C7	0.3750	0.2026	0.1095	0.0924	0.0307	0.0399
C8	0.0764	0.0482	0.0305	0.0300	0.0262	0.0187
C9	0.4894	0.3198	0.1753	0.1212	0.0229	0.0372
C10	0.1680	0.0815	0.0521	0.0462	0.0252	0.0290
C11	0.1537	0.0839	0.0436	0.0397	0.0199	0.0219
C12	0.3689	0.2558	0.1505	0.1249	0.0401	0.0490
C13	0.2848	0.1582	0.0968	0.0792	0.0240	0.0320
C14	0.0367	0.0226	0.0120	0.0143	0.0106	0.0104
C15	0.3618	0.2264	0.1201	0.1227	0.0317	0.0522
C16	0.1773	0.0835	0.0651	0.0545	0.0264	0.0318
C17	0.0884	0.0623	0.0210	0.0231	0.0125	0.0136
M.	0.1854	0.1196	0.0655	0.0577	0.0228	0.0273
T.	3.1524	2.0333	1.1129	0.9801	0.3877	0.4642

Se observa un mejor rendimiento de los datos duros en el SVR lineal. En su implementación, se aplica la optimización de hiperparámetros automática de la función de Matlab *fitrsvm*. Mientras que, desde la categoría de datos blandos, la mejor opción corresponde al ajuste del modelo SVR no lineal basado en el *kernel* gaussiano (aplicando la misma opción de optimización de hiperparámetros automática, en el argumento de la función de matlab *fitrsvm*). En la implementación de GP, seguimos el mismo procedimiento de ajuste para la selección del modelo. En este caso, para ambas categorías, hemos seleccionado la optimización bayesiana de validación cruzada (en el argumento de optimización de hiperparámetros de la función de Matlab *fitrgp*).

En todos los resultados mostrados, se ha calculado la media de los SMAPE (M.) y el total de los SMAPE (T.) como medidas de rendimiento, para comparar los modelos ML probados, y nuestro enfoque.

Tabla 7.8: **Nuestro enfoque**. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 10-fold de los análisis residuales clásico y bayesiano probados

SC	C. k10	B. k10
C1	0.0024	0.7106(10) ⁻³
C2	0.0019	0.4003(10) ⁻³
C3	0.0016	0.6797(10) ⁻³
C4	0.0017	0.4367(10) ⁻³
C5	0.0023	0.6530(10) ⁻³
C6	0.0018	0.5854(10) ⁻³
C7	0.0017	0.6341(10) ⁻³
C8	0.0016	0.6593(10) ⁻³
C9	0.0013	0.5979(10) ⁻³
C10	0.0019	0.6954(10) ⁻³
C11	0.0017	0.5444(10) ⁻³
C12	0.0016	0.5016(10) ⁻³
C13	0.0020	0.4832(10) ⁻³
C14	0.0026	0.6544(10) ⁻³
C15	0.0023	0.6616(10) ⁻³
C16	0.0015	0.7134(10) ⁻³
C17	0.0022	0.6781(10) ⁻³
M.	0.0019	0.60524(10) ⁻³
T.	0.0321	0.0103

7.4.2. Resultados de clasificación basados en *bootstrap*

Para los modelos de regresión ML probados, en las categorías de datos duros y blandos, se construyen intervalos de confianza *bootstrap* (nivel de confianza $1 - \alpha = 0.95$) para los SMAPE promediados espacialmente, basados en 1000 muestras *bootstrap*.

Nuestro enfoque requiere que se incorpore la información de los datos blandos. Al igual que antes, los intervalos de confianza *bootstrap* calculados \mathcal{I}_i , $i = 1, \dots, 5$, se basan respectivamente en el sesgo corregido y método de percentil acelerado (\mathcal{I}_1); intervalo normal aproximado con sesgo *bootstrap* y error estándar (\mathcal{I}_2); método del percentil básico (\mathcal{I}_3); método del percentil de sesgo corregido (\mathcal{I}_4); e intervalo de confianza basado en Student (\mathcal{I}_5); véanse las tablas 7.9 y 7.10. El histograma *bootstrap* y la densidad de probabilidad de los SMAPE promediados espacialmente se muestran en las Figuras 7.5 y 7.6, para la categoría de datos duros, y en las Figuras 7.7, 7.8 y 7.9, para la categoría de datos blandos. Los resultados de la clasificación de los modelos basados en datos obtenidos se analizan en la Ssección 7.6.

Tabla 7.9: *Categoría de datos duros*. Intervalos de confianza *bootstrap* ($1 - \alpha = 0.95$) para los SMAPE promediados espacialmente a partir de 1000 muestras *bootstrap* ($T = 265$, $P = 17$)

IC\ML	GRNN	MLP
\mathcal{I}_1	$[2.1(10)^{-3}, 4.1(10)^{-3}]$	$[0.5(10)^{-3}, 1(10)^{-3}]$
\mathcal{I}_2	$[2(10)^{-3}, 3.9(10)^{-3}]$	$[0.4776(10)^{-3}, 0.9483(10)^{-3}]$
\mathcal{I}_3	$[2(10)^{-3}, 4(10)^{-3}]$	$[0.4746(10)^{-3}, 0.9713(10)^{-3}]$
\mathcal{I}_4	$[2(10)^{-3}, 4(10)^{-3}]$	$[0.5118(10)^{-3}, 0.9878(10)^{-3}]$
\mathcal{I}_5	$[1.7(10)^{-3}, 3.9(10)^{-3}]$	$[0.2780(10)^{-3}, 0.9244(10)^{-3}]$
IC\ML	SVR	BNN
\mathcal{I}_1	$[0.3682(10)^{-3}, 0.7219(10)^{-3}]$	$[0.3720(10)^{-3}, 0.6659(10)^{-3}]$
\mathcal{I}_2	$[0.3516(10)^{-3}, 0.6763(10)^{-3}]$	$[0.3587(10)^{-3}, 0.6379(10)^{-3}]$
\mathcal{I}_3	$[0.3493(10)^{-3}, 0.6770(10)^{-3}]$	$[0.3668(10)^{-3}, 0.6509(10)^{-3}]$
\mathcal{I}_4	$[0.3508(10)^{-3}, 0.6865(10)^{-3}]$	$[0.3654(10)^{-3}, 0.6379(10)^{-3}]$
\mathcal{I}_5	$[0.3050(10)^{-3}, 0.6661(10)^{-3}]$	$[0.3099(10)^{-3}, 0.6335(10)^{-3}]$
IC\ML	RBF	GP
\mathcal{I}_1	$[0.3260(10)^{-3}, 0.5310(10)^{-3}]$	$[0.3243(10)^{-3}, 0.5350(10)^{-3}]$
\mathcal{I}_2	$[0.3155(10)^{-3}, 0.5159(10)^{-3}]$	$[0.3065(10)^{-3}, 0.5126(10)^{-3}]$
\mathcal{I}_3	$[0.3140(10)^{-3}, 0.5270(10)^{-3}]$	$[0.3095(10)^{-3}, 0.5188(10)^{-3}]$
\mathcal{I}_4	$[0.3247(10)^{-3}, 0.5338(10)^{-3}]$	$[0.3152(10)^{-3}, 0.5222(10)^{-3}]$
\mathcal{I}_5	$[0.2677(10)^{-3}, 0.5141(10)^{-3}]$	$[0.2505(10)^{-3}, 0.5046(10)^{-3}]$

Tabla 7.10: **Categoría de datos blandos.** Intervalos de confianza *bootstrap* ($1 - \alpha = 0.95$) para los SMAPE promediados espacialmente a partir de 1000 muestras *bootstrap* ($T = 265, P = 17$)

IC\ML	GRNN	MLP
\mathcal{I}_1	$[1.3(10)^{-3}, 2.6(10)^{-3}]$	$[0.6(10)^{-3}, 1.3(10)^{-3}]$
\mathcal{I}_2	$[1.3(10)^{-3}, 2.6(10)^{-3}]$	$[0.6(10)^{-3}, 1.3(10)^{-3}]$
\mathcal{I}_3	$[1.3(10)^{-3}, 2.7(10)^{-3}]$	$[0.6(10)^{-3}, 1.3(10)^{-3}]$
\mathcal{I}_4	$[1.3(10)^{-3}, 2.6(10)^{-3}]$	$[0.7(10)^{-3}, 1.3(10)^{-3}]$
\mathcal{I}_5	$[1(10)^{-3}, 2.7(10)^{-3}]$	$[0.5(10)^{-3}, 1.3(10)^{-3}]$
IC\ML	SVR	BNN
\mathcal{I}_1	$[0.4096(10)^{-3}, 0.8221(10)^{-3}]$	$[0.3588(10)^{-3}, 0.6177(10)^{-3}]$
\mathcal{I}_2	$[0.3764(10)^{-3}, 0.7763(10)^{-3}]$	$[0.3433(10)^{-3}, 0.6053(10)^{-3}]$
\mathcal{I}_3	$[0.3900(10)^{-3}, 0.7889(10)^{-3}]$	$[0.3454(10)^{-3}, 0.6037(10)^{-3}]$
\mathcal{I}_4	$[0.4108(10)^{-3}, 0.7805(10)^{-3}]$	$[0.3559(10)^{-3}, 0.5988(10)^{-3}]$
\mathcal{I}_5	$[0.3105(10)^{-3}, 0.7818(10)^{-3}]$	$[0.3003(10)^{-3}, 0.6129(10)^{-3}]$
IC\ML	RBF	GP
\mathcal{I}_1	$[0.1794(10)^{-3}, 0.2478(10)^{-3}]$	$[0.2095(10)^{-3}, 0.3248(10)^{-3}]$
\mathcal{I}_2	$[0.1754(10)^{-3}, 0.2474(10)^{-3}]$	$[0.2065(10)^{-3}, 0.3215(10)^{-3}]$
\mathcal{I}_3	$[0.1785(10)^{-3}, 0.2485(10)^{-3}]$	$[0.2079(10)^{-3}, 0.3262(10)^{-3}]$
\mathcal{I}_4	$[0.1743(10)^{-3}, 0.2494(10)^{-3}]$	$[0.2091(10)^{-3}, 0.3258(10)^{-3}]$
\mathcal{I}_5	$[0.1616(10)^{-3}, 0.2542(10)^{-3}]$	$[0.1941(10)^{-3}, 0.3232(10)^{-3}]$
IC\NE	Clásico	Bayesiano
\mathcal{I}_1	$[2.2(10)^{-3}, 3.7(10)^{-3}]$	$[0.2943(10)^{-3}, 0.5177(10)^{-3}]$
\mathcal{I}_2	$[2.1(10)^{-3}, 3.3(10)^{-3}]$	$[0.2802(10)^{-3}, 0.4854(10)^{-3}]$
\mathcal{I}_3	$[2.1(10)^{-3}, 3.3(10)^{-3}]$	$[0.2833(10)^{-3}, 0.4884(10)^{-3}]$
\mathcal{I}_4	$[2.2(10)^{-3}, 3.4(10)^{-3}]$	$[0.2900(10)^{-3}, 0.5124(10)^{-3}]$
\mathcal{I}_5	$[1.8(10)^{-3}, 3.2(10)^{-3}]$	$[0.2418(10)^{-3}, 0.4664(10)^{-3}]$

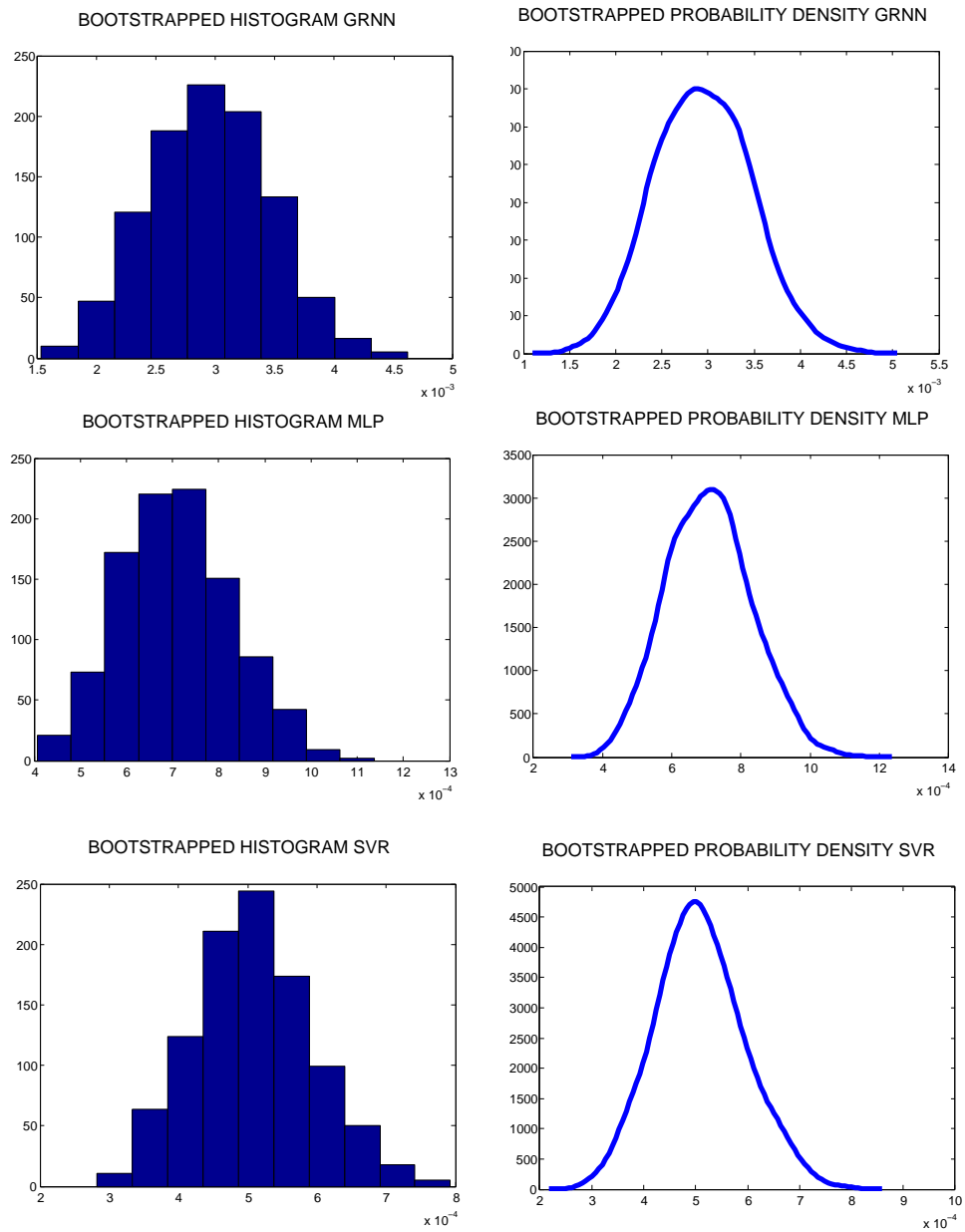


Figura 7.5: *Categoría de datos duros*. A partir de 1000 muestras *bootstrap*, se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para GRNN (arriba), MLP (centro) y SVR lineal (abajo)

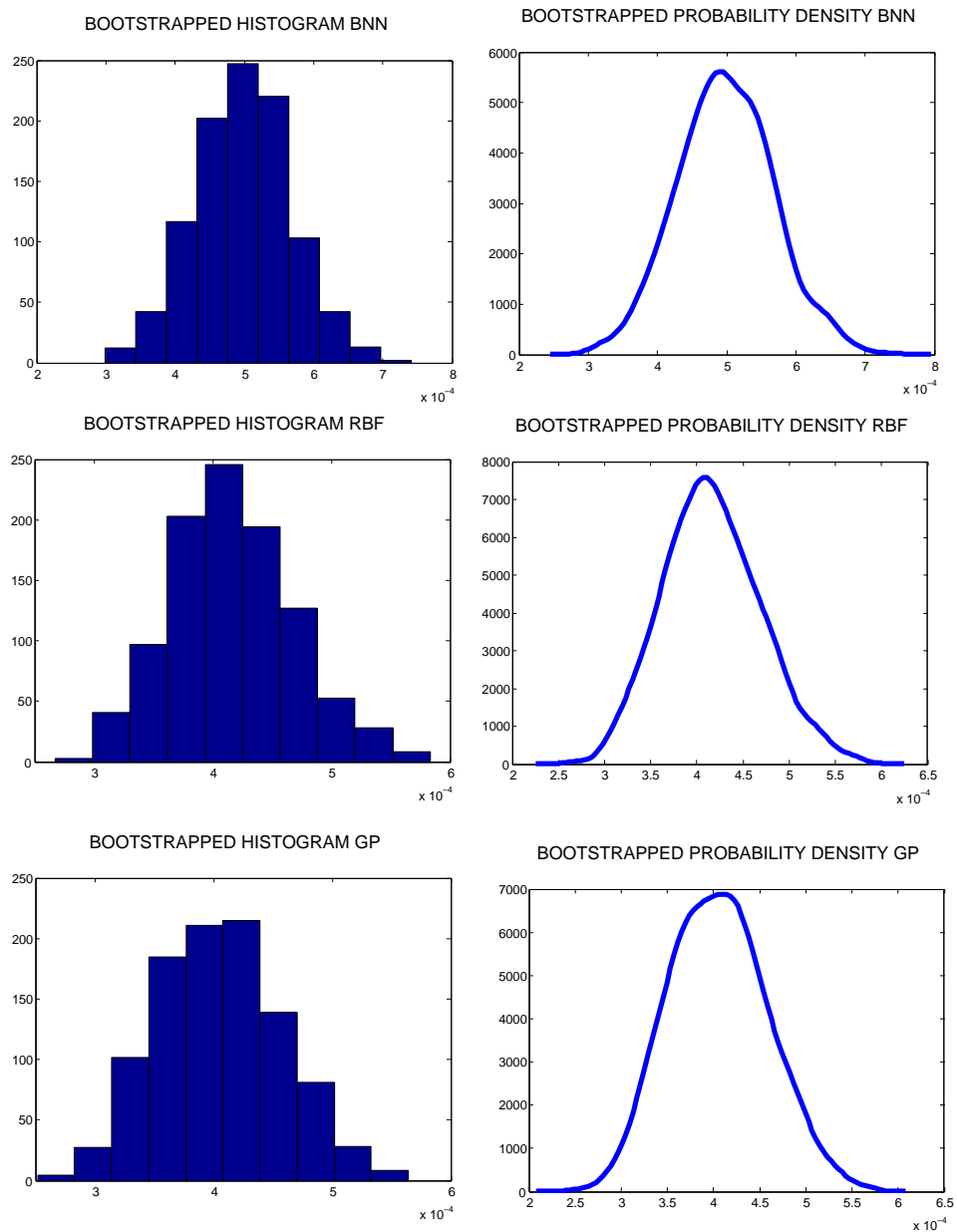


Figura 7.6: *Categoría de datos duros*. A partir de 1000 muestras *bootstrap*, se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para BNN (arriba), RBF (centro) y GP (abajo)

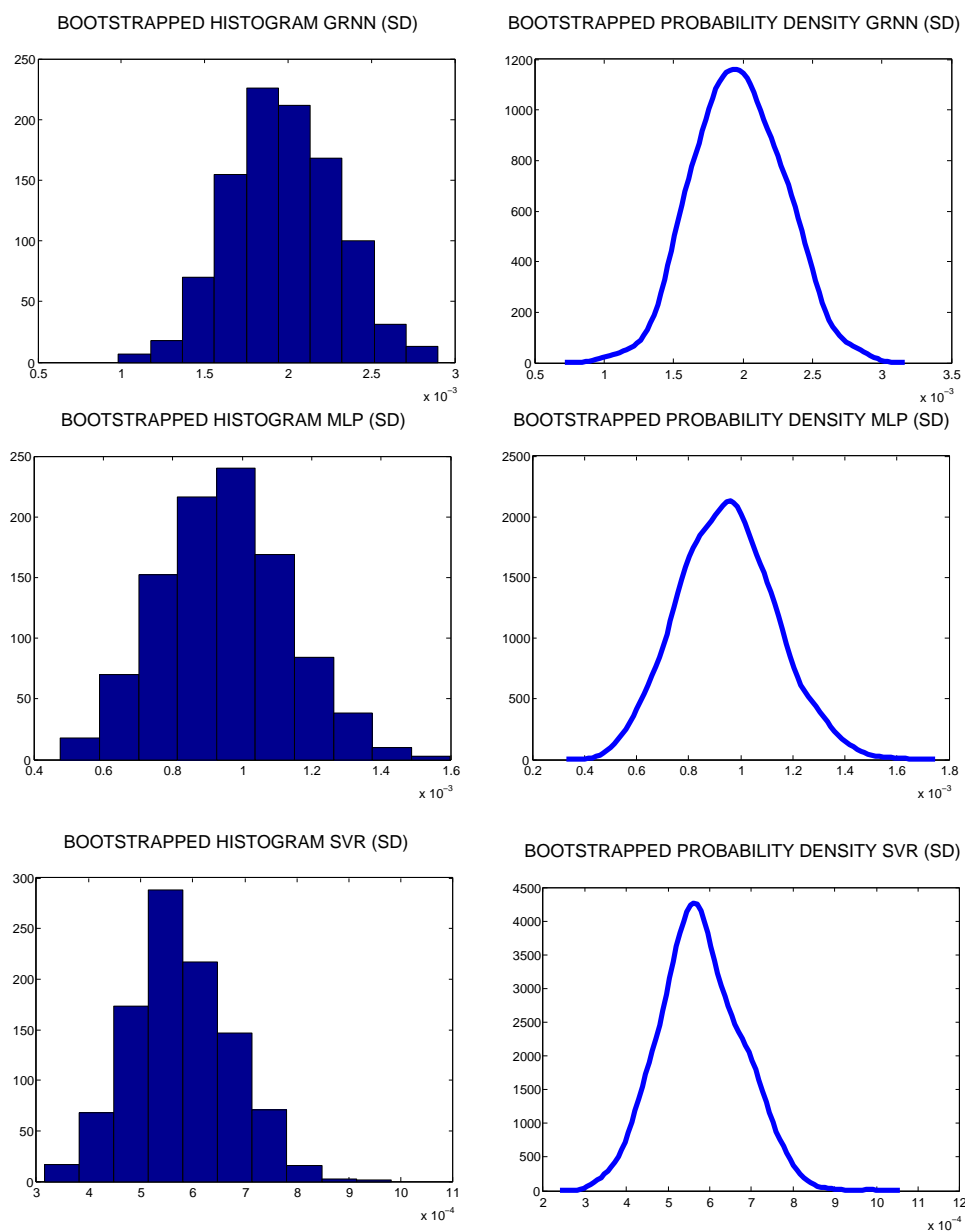


Figura 7.7: *Categoría de datos blandos*. A partir de 1000 muestras *bootstrap*, se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para GRNN (arriba), MLP (centro) y SVR no lineal (abajo)

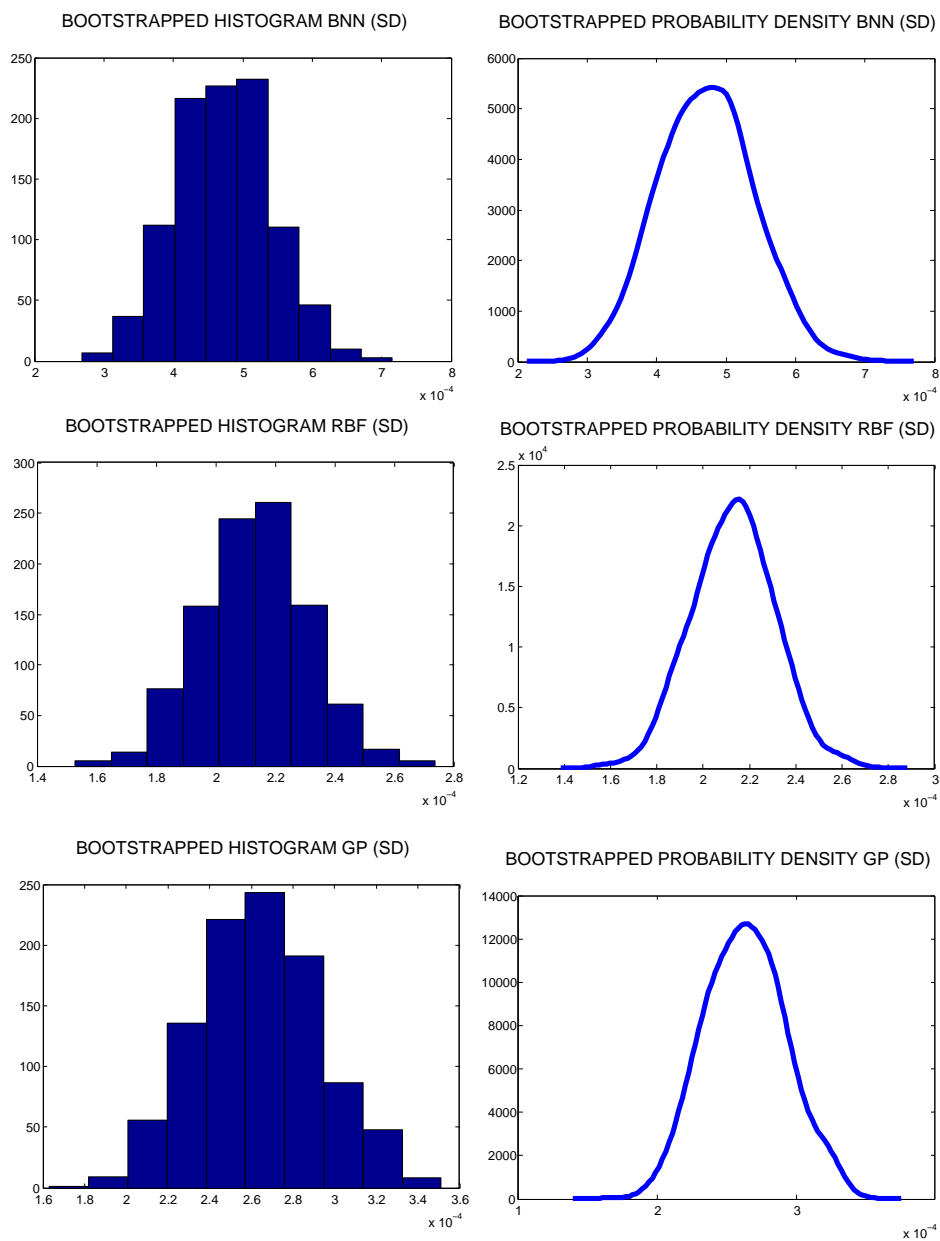


Figura 7.8: *Categoría de datos blandos*. A partir de 1000 muestras *bootstrap*, se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para BNN (arriba), RBF (centro) y GP (abajo)

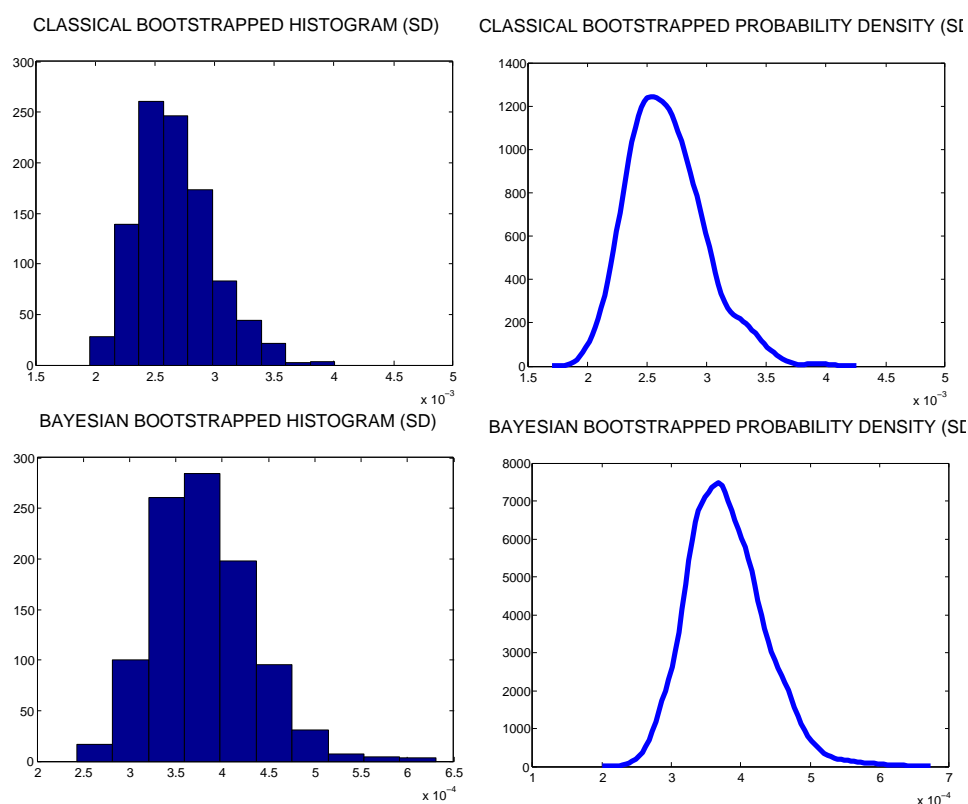


Figura 7.9: *Categoría de datos blandos*. A partir de 1000 muestras *bootstrap*, se representan las densidades de probabilidad y los histogramas de los SMAPE promediados espacialmente, para la regresión trigonométrica combinada con la predicción residual, basada en los momento empíricos, clásica (arriba) y bayesiana (abajo)

7.5. Implementación de datos duros y blandos en modelos ML

Describimos brevemente la implementación de los modelos ML en las categorías de datos duros y blandos.

Perceptrón multicapa (MLP)

MLP comparte la filosofía de la regresión no lineal, en términos de una función de enlace g , la salida del nodo oculto, que define la siguiente aproximación

de la respuesta:

$$\hat{y} = \eta_0 + \sum_{k=1}^{NH} \eta_k g(\beta_k^T \mathbf{x}), \quad (7.19)$$

del vector de entrada $\mathbf{x}^T = (1, \mathbf{x})$ aumentado con 1, y el vector de pesos β_k asociado al k -ésimo nodo oculto, $k = 1, \dots, NH$, definiendo $\beta = (\beta_1^T, \dots, \beta_{NH}^T)^T$. Normalmente, se considera la función logística $g(u) = \frac{1}{1 + \exp(-u)}$. Las salidas de los nodos ocultos también se ponderan por las componentes del vector $(\eta_0, \dots, \eta_{NH})$. En este contexto, \hat{y} se suele denominar la salida de red. MLP permite la aproximación de cualquier función continua dada sobre un conjunto compacto, a partir de una red dada con un número finito de nodos ocultos.

Se aplican algoritmos de optimización para obtener los pesos de la función de pérdida de mínimos cuadrados. Teniendo en cuenta (7.19), nuestra implementación de MLP a partir de datos blandos viene dada por

$$\hat{y}_m = \hat{y}(\mathbf{x}_m) = \ln(\widehat{\lambda_{t+1}}(\phi_m)) = \eta_0 + \sum_{k=1}^{NH} \eta_k g(\beta_k^T \mathbf{x}_m), \quad (7.20)$$

donde, para $m = 1, \dots, M(T)$,

$$\mathbf{x}_m = (\ln(\lambda_{t+1-j_0}) (\phi_m), \dots, \ln(\lambda_{t+1-1}) (\phi_m))^T. \quad (7.21)$$

El parámetro j_0 se refiere al número de retardos temporales incorporados en la predicción. El parámetro de truncamiento $M(T)$ desempeña un papel similar al parámetro $k(T)$ (comentado en secciones previas). Aquí, T denota el número de nodos temporales de entrenamiento.

Redes Neuronales a partir de Bases de Funciones Radiales (RBF)

RBF trabaja con funciones nodo, que dependen de unos parámetros de centro y escala, ajustándose a la suavidad local de la respuesta. En concreto, a partir de una red inicial en blanco, los nodos se añaden secuencialmente, en torno al patrón de entrenamiento, hasta alcanzar un error aceptable. A continuación, se vuelven a calcular todos los pesos de la capa de salida mediante la fórmula de mínimos cuadrados. Las funciones gaussianas han sido ampliamente seleccionadas como funciones nodo.

En particular, en este caso, nuestra implementación a partir de datos blandos se ha logrado a partir de la fórmula: Para $m = 1, \dots, M(T)$,

$$\widehat{y}_{t+1}(\mathbf{x}_m) = \ln(\widehat{\lambda_{t+1}}(\phi_m)) = \sum_{j=1}^{NH} \eta_j \exp\left(\frac{\|\mathbf{x}_m - c_j\|^2}{\beta^2}\right) \quad (7.22)$$

donde, como es habitual, los parámetros de peso η_j , $j = 1, \dots, NH$, definen la combinación lineal de las funciones de base radial. Aquí, el parámetro escalar de dispersión β , y los parámetros vectoriales c_j , $j = 1, \dots, NH$, proporcionan respectivamente el ancho y los centros de las funciones nodo. Los vectores de entrada se definen como en (7.21).

Regresión de Soporte Vectorial (SVR)

La implementación de SVR implica una función de pérdida que lleva a un equilibrio entre la complejidad del modelo y la precisión (predicción exacta). En su formulación también se considera un parámetro de sesgo, b , que se refleja en la siguiente ecuación

$$y = f(\mathbf{x}) = \beta^T \mathbf{x} + b \quad (7.23)$$

donde se considera la función de pérdida

$$\mathcal{L} = \frac{1}{2} \|\beta\|^2 + C \sum_{m=1}^M |y_m - f(x_m)|_\epsilon. \quad (7.24)$$

Aquí, x_m e y_m denotan, respectivamente, el m -ésimo vector de entrada de entrenamiento y la salida objetivo, para $m = 1, \dots, M$, y

$$|y_m - f(x_m)|_\epsilon = \max\{0, |y_m - f(x_m)| - \epsilon\}.$$

Así, los errores por debajo de ϵ no se penalizan. La solución del problema de optimización asociado a la función de pérdida (7.24) se obtiene a partir del gradiente correspondiente de la función lagrangiana, que involucra a multiplicadores de Lagrange, que determinan los pesos óptimos a partir de los puntos de datos de entrenamiento.

A partir de la ecuación (7.23), en la categoría de datos suaves, resolvemos

el problema de optimización restringido:

$$\begin{aligned}
 y_{t+1}(\mathbf{x}_m) &= f(\mathbf{x}_m) = \boldsymbol{\beta}^T \mathbf{x}_m + b, \\
 \mathcal{L} &= \frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{\beta} + C \sum_{m=1}^{M(T)} \left| \ln(\lambda_{t+1})(\phi_m) - \sum_{j=1}^{j_0} \ln(\lambda_{t+1-j})(\phi_m) \beta_j - b \right|_{\epsilon} \\
 &= \frac{1}{2} \boldsymbol{\beta}^T \boldsymbol{\beta} + C \sum_{m=1}^{M(T)} |y_m - f(\mathbf{x}_m)|_{\epsilon}, \tag{7.25}
 \end{aligned}$$

donde, para $m = 1, \dots, M(T)$, el vector de entrada \mathbf{x}_m se define como en (7.21).

Redes Neuronales Bayesianas (BNN)

El diseño de BNN implica la estimación bayesiana y el concepto de regularización. Los parámetros o pesos de red se consideran variables aleatorias, siguiendo una distribución de probabilidad a-priori. En la selección de la distribución de probabilidad a-priori de los pesos se suelen favorecer ajustes suaves para reducir la complejidad del modelo. La distribución de probabilidad a-posteriori de los pesos se obtiene después de observar los datos. A continuación, se calcula la predicción de la red.

En concreto, la predicción óptima se obtiene minimizando la siguiente expresión:

$$J = \nu E_{\mathbf{O}} + (1 - \nu) E_{\mathbf{W}}, \tag{7.26}$$

donde $E_{\mathbf{O}}$ es la suma de los errores cuadráticos en la salida de red, basada en la distribución a-posteriori de los parámetros, $E_{\mathbf{W}}$ representa la suma de los cuadrados de los pesos o los parámetros de la red, y $\nu \in (0, 1)$ denota el parámetro de regularización. Sea L el número de parámetros, habitualmente se asume una distribución de probabilidad gaussiana a-priori para los parámetros de red con media cero y matriz de varianzas-covarianzas $\frac{1}{2(1-\nu)} I_{L \times L}$, con $I_{L \times L}$ denotando la matriz identidad $L \times L$. Por lo tanto,

$$p(\mathbf{w}) = \left[\frac{1 - \nu}{\pi} \right]^{L/2} \exp(-(1 - \nu) E_{\mathbf{W}}(\mathbf{w})). \tag{7.27}$$

Esta distribución a-priori pone más peso en los valores pequeños de los parámetros de red cercanos a cero. La densidad de probabilidad a-posteriori, dados los datos observados $\mathbf{O} = \mathbf{o}$, y el valor ν del parámetro de regularización, se

define como

$$p(\mathbf{w}/\mathbf{o}, \nu) = \frac{p(\mathbf{o}/\mathbf{w}, \nu)p(\mathbf{w}/\nu)}{p(\mathbf{o}/\nu)}. \quad (7.28)$$

Considerando que los errores también se distribuyen de forma gaussiana, la probabilidad condicionada de los datos \mathbf{O} dados los parámetros ν y \mathbf{w} , se obtiene como

$$p(\mathbf{o}/\mathbf{w}, \nu) = \left(\frac{\nu}{\pi}\right)^{M/2} \exp(-\nu E_{\mathbf{O}}(\mathbf{o})), \quad (7.29)$$

donde M denota el número de puntos de datos de entrenamiento. A partir de las ecuaciones (7.27)–(7.29),

$$p(\mathbf{w}/\mathbf{o}, \nu) = c \exp(-J(\mathbf{w}, \mathbf{o}, \nu)), \quad (7.30)$$

donde c denota la constante de normalización. La probabilidad condicionada del parámetro ν dado los datos observados $\mathbf{O} = \mathbf{o}$ también se calcula bajo un marco bayesiano como

$$p(\nu/\mathbf{o}) = \frac{P(\mathbf{o}/\nu)p(\nu)}{p(\mathbf{o})}. \quad (7.31)$$

Las ecuaciones (7.30) y (7.31) se maximizan para obtener los pesos óptimos y el parámetro de regularización ν , respectivamente.

En nuestra implementación de datos blandos a partir de (7.21), el problema de optimización correspondiente se formula condicionando a las proyecciones espaciales empíricas. Obsérvese que, en el marco de la probabilidad a-priori gaussiana seleccionada, las proyecciones de error son también gaussianas, y nuestra elección de la función base, diagonalizando el operador de autocovarianza empírica de los errores, lleva a un vector de errores proyectados con componentes gaussianas independientes, normalizadas adecuadamente por los autovalores empíricos. Por lo tanto, la optimización a partir de las ecuaciones (7.30) y (7.31) puede implementarse de forma similar al caso BNN de datos duros.

Redes Neuronales de Regresión Generalizada (GRNN)

GRNN se basa en la regresión *kernel*. El estimador *kernel* se calcula a partir de la suma ponderada de las respuestas observadas, o de los resultados objetivos asociados con los puntos de datos de entrenamiento en un entorno del punto de datos objetivos x , donde se debe calcular la predicción. Así, los pun-

tos de datos de entrenamiento se seleccionan en la vecindad del punto objetivo dado x .

En concreto, se aplica la siguiente fórmula en la aproximación del valor de la respuesta en el punto x :

$$\widehat{y}(x) = \sum_{j=1}^T \frac{\mathcal{K}\left(\frac{\|x-x_j\|}{h}\right)}{\sum_{l=1}^T \mathcal{K}\left(\frac{\|x-x_l\|}{h}\right)} y_j, \quad (7.32)$$

donde y_j es la salida del objetivo para el punto de datos de entrenamiento x_j , para $j = 1, \dots, M$, y \mathcal{K} es la función *kernel*. Normalmente se considera una función *kernel* isotrópica de rápido decrecimiento, por ejemplo, el *kernel* gaussiano $\mathcal{K}(u) = \exp(-u^2/2) / \sqrt{2\pi}$ constituye una elección común. El parámetro de ancho de banda h define la suavidad del ajuste. Así, h controla el tamaño de la región de suavizado. Por lo tanto, valores grandes de h corresponden a un suavizamiento más fuerte que valores más pequeños, permitiendo un mayor grado de variación local.

En nuestra implementación de datos blandos de (7.32), denotamos por $\Phi_{M(T)}(H)$, el subespacio de H obtenido por la proyección de las funciones en H sobre $\{\phi_1, \dots, \phi_{M(T)}\}$, y $\mathbf{y}_t = (\ln(\lambda_t)(\phi_1), \dots, \ln(\lambda_t)(\phi_{M(T)}))^T$, $t \geq 1$, entonces,

$$\widehat{\mathbf{y}}_{t+1} = \sum_{j=1}^{j_0-1} \frac{\mathcal{K}\left(\frac{\|\mathbf{y}_t - \mathbf{y}_{t-j}\|_{\Phi_{M(T)}(H)}}{h}\right)}{\sum_{l=1}^{j_0-1} \mathcal{K}\left(\frac{\|\mathbf{y}_t - \mathbf{y}_{t-l}\|_{\Phi_{M(T)}(H)}}{h}\right)} \mathbf{y}_{t-j+1}, \quad (7.33)$$

donde, como antes, el parámetro j_0 se refiere a el número de retardos temporales incorporados en la predicción.

Procesos Gaussianos (GP)

Se suele observar un buen rendimiento en la aplicación de la regresión de GP, basada en el supuesto de distribución de probabilidad normal multivariante, que caracteriza las respuestas observadas en los diferentes puntos de datos de entrenamiento.

En concreto, consideramos que el modelo de observación

$$\mathbf{Y} = \mathbf{Z} + \boldsymbol{\epsilon}, \quad (7.34)$$

donde el vector de ruido aditivo ϵ es independiente de \mathbf{Z} , y tiene componentes gaussianas idénticamente distribuidas e independientes, con media cero y varianza σ_ϵ^2 . Se asume una distribución normal multivariante del vector aleatorio $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, con matriz de covarianzas $\Sigma_{\mathbf{Z}}$. Esta matriz proporciona las varianzas y covarianzas entre los valores de la función $Z(x_i)$, y $Z(x_j)$, $i, j = 1, \dots, N$, en los puntos de datos de entrenamiento $X = (x_1, \dots, x_N)$. La distribución gaussiana condicionada de \mathbf{Z} , dada \mathbf{Y} , lleva a la solución del problema de estimación inversa (7.34). Por lo tanto, la estimación de \mathbf{Z} , para un vector de entrada dado \mathbf{x}_* , se obtiene como

$$\hat{\mathbf{Z}}_{\mathbf{x}_*} = E[\mathbf{Z}/\mathbf{y}, \mathbf{x}_*, X] = \Sigma_{\mathbf{Z}}(\mathbf{x}_*, X) [\Sigma_{\mathbf{Z}}(X, X) + \sigma_\epsilon^2 \mathbf{I}]^{-1} \mathbf{y}. \quad (7.35)$$

En la categoría de datos blandos, se consigue una implementación multivariante alternativa en términos de proyección $\Phi_{M(T)}$ sobre $\{\phi_1, \dots, \phi_{M(T)}\}$. Por lo tanto, $\Sigma_{\mathbf{Z}}$ se sustituye por la matriz del operador de covarianzas

$$R_{\mathbf{Z}}^{X,X} = \begin{pmatrix} \Phi_{M(T)}^* \mathcal{R}_{1,1} \Phi_{M(T)} & \dots & \Phi_{M(T)}^* \mathcal{R}_{1,T} \Phi_{M(T)} \\ \Phi_{M(T)}^* \mathcal{R}_{2,1} \Phi_{M(T)} & \dots & \Phi_{M(T)}^* \mathcal{R}_{2,T} \Phi_{M(T)} \\ \vdots & \dots & \vdots \\ \Phi_{M(T)}^* \mathcal{R}_{T,1} \Phi_{M(T)} & \dots & \Phi_{M(T)}^* \mathcal{R}_{T,T} \Phi_{M(T)} \end{pmatrix},$$

asociada a los nodos de entrenamiento temporal T considerados, y el operador de proyección $\Phi_{M(T)}$ sobre $\{\phi_1, \dots, \phi_{M(T)}\}$, implicado en la definición de nuestros puntos de datos blandos de entrenamiento X . Aquí, $\mathcal{R}_{i,j} = E[[\ln(\Lambda_i) - E[\ln(\Lambda_i)]] \otimes [\ln(\Lambda_j) - E[\ln(\Lambda_j)]]]$, $i, j = 1, \dots, T$, definen los operadores de autocovarianza y covarianza cruzada de los valores funcionales en los puntos de datos de entrenamiento. En este caso, $\sigma_\epsilon^2 \mathbf{I}$ se convierte en el operador de autocovarianza de la matriz diagonal $diag(R_{0,\epsilon})$ del proceso de innovación H^T -valuado $\epsilon = (\epsilon_1, \dots, \epsilon_T)$. Es decir,

$$\begin{aligned} \hat{\mathbf{Z}}_{\mathbf{x}_*} &= E[\mathbf{Z}/\mathbf{y}, \mathbf{x}_*, X, \Phi_{M(T)}] \\ &= R_{\mathbf{Z}}^{\mathbf{x}_*, X} \left[R_{\mathbf{Z}}^{X,X} + \Phi_{M(T)}^* diag(R_{0,\epsilon}) \Phi_{M(T)} \right]^{-1} \Phi_{M(T)}(\mathbf{y}). \end{aligned}$$

Valores muestrales SMAPE mediante ML a partir de los datos observados duros y blandos

Los Errores Porcentuales Absolutos Medios Simétricos (SMAPE) asociados a los resultados de la estimación ML, basados en la muestra funcional global, a partir de datos duros y blandos, se muestran respectivamente en las tablas 7.11 y 7.12. Véanse también las Figuras 7.10–7.13 más abajo, donde se muestran respectivamente las curvas observadas y estimadas de log-riesgo y casos acumulados de mortalidad por COVID-19.

Validación cruzada aleatoria 5-fold

Los SMAPE de la validación cruzada aleatoria 5-fold, obtenidos a partir de la implementación de los seis modelos de regresión ML probados, se muestran en la Tabla 7.13, para la categoría de datos duros, y en la Tabla 7.14, para la categoría de datos blandos.

Tabla 7.11: *Categoría de datos duros*. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold. Como se indica, los valores mostrados deben ser multiplicados por 10^{-2}

SC ($\times 10^{-2}$)	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1964	0.0611	0.0665	0.0535	0.0483	0.0490
C2	0.6117	0.1172	0.0588	0.0678	0.0609	0.0567
C3	0.1565	0.0375	0.0328	0.0284	0.0300	0.0273
C4	0.0969	0.0314	0.0129	0.0191	0.0167	0.0185
C5	0.2044	0.0427	0.0290	0.0356	0.0334	0.0328
C6	0.1571	0.0319	0.0161	0.0230	0.0217	0.0218
C7	0.4889	0.0545	0.0619	0.0583	0.0569	0.0503
C8	0.0808	0.0273	0.0161	0.0180	0.0189	0.0153
C9	0.7231	0.1976	0.0850	0.1102	0.0318	0.0352
C10	0.2185	0.0526	0.0532	0.0446	0.0428	0.0415
C11	0.1255	0.0487	0.0298	0.0350	0.0338	0.0316
C12	0.5216	0.1666	0.1210	0.1022	0.0870	0.0887
C13	0.3584	0.0592	0.0548	0.0613	0.0490	0.0412
C14	0.1342	0.0286	0.0201	0.0202	0.0182	0.0180
C15	0.6064	0.1470	0.1307	0.0931	0.0864	0.0927
C16	0.2456	0.0681	0.0674	0.0573	0.0514	0.0532
C17	0.0655	0.0356	0.0147	0.0207	0.0181	0.0200
M.	0.2936	0.0710	0.0512	0.0499	0.0415	0.0408
T.	4.9916	1.2078	0.8707	0.8480	0.7053	0.6936

Tabla 7.12: **Categoría de datos blandos.** Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold. Como se indica, los valores mostrados deben ser multiplicados por 10^{-2}

SC ($\times 10^{-2}$)	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1526	0.0857	0.0592	0.0501	0.0223	0.0305
C2	0.1831	0.1393	0.0619	0.0606	0.0248	0.0285
C3	0.1024	0.0804	0.0457	0.0374	0.0268	0.0278
C4	0.0431	0.0212	0.0157	0.0154	0.0117	0.0120
C5	0.0610	0.0362	0.0242	0.0240	0.0133	0.0138
C6	0.0260	0.0167	0.0125	0.0135	0.0118	0.0121
C7	0.3734	0.1301	0.0914	0.0737	0.0298	0.0398
C8	0.0762	0.0435	0.0290	0.0238	0.0244	0.0180
C9	0.4850	0.2467	0.1470	0.0818	0.0199	0.0360
C10	0.1663	0.0607	0.0460	0.0421	0.0236	0.0276
C11	0.1533	0.0540	0.0394	0.0383	0.0180	0.0209
C12	0.3641	0.2325	0.1320	0.0887	0.0369	0.0480
C13	0.2827	0.1350	0.0707	0.0699	0.0215	0.0304
C14	0.0361	0.0197	0.0117	0.0121	0.0102	0.0104
C15	0.3590	0.1843	0.1226	0.1049	0.0290	0.0506
C16	0.1759	0.0754	0.0566	0.0489	0.0243	0.0302
C17	0.0878	0.0352	0.0190	0.0219	0.0122	0.0134
N.	0.1840	0.0939	0.0579	0.0475	0.0212	0.0265
T.	3.1282	1.5966	0.9845	0.8070	0.3605	0.4497

7.6. Comentarios finales al capítulo

Se puede observar la concordancia entre los respectivos resultados de clasificación de modelos basados en rendimiento, obtenidos a partir de la validación cruzada aleatoria k -fold, y la estimación *bootstrap* en las secciones 7.4.1 y 7.4.2.

En la categoría de datos duros, el mejor rendimiento lo muestran RBF y GP. Se observan características *bootstrapp* similares para BNN y SVR, con valores de los SMAPE promediados espacialmente ligeramente superiores, reflejados en la ubicación de la moda, en los histogramas y densidades de probabilidad mostrados en las figuras 7.5 y 7.6. Estas cuatro metodologías de regresión muestran un grado similar de variabilidad, con respecto a los valores muestrales de los SMAPE promediados espacialmente. Una mayor variabilidad que RBF, GP, BNN y SVR se muestra a partir de los valores muestrales de los SMAPE promediados espacialmente de MLP. La moda *bootstrap* de MLP también está ligeramente desplazada hacia la derecha. El peor rendimiento corresponde a GRNN (véase también la Tabla 7.6).

En la categoría de datos blandos, donde nuestro enfoque se incorpora al es-

Capítulo 7. Análisis de la mortalidad por COVID-19 mediante curvas multivariantes de regresión a partir de datos blandos y mediante aprendizaje automático

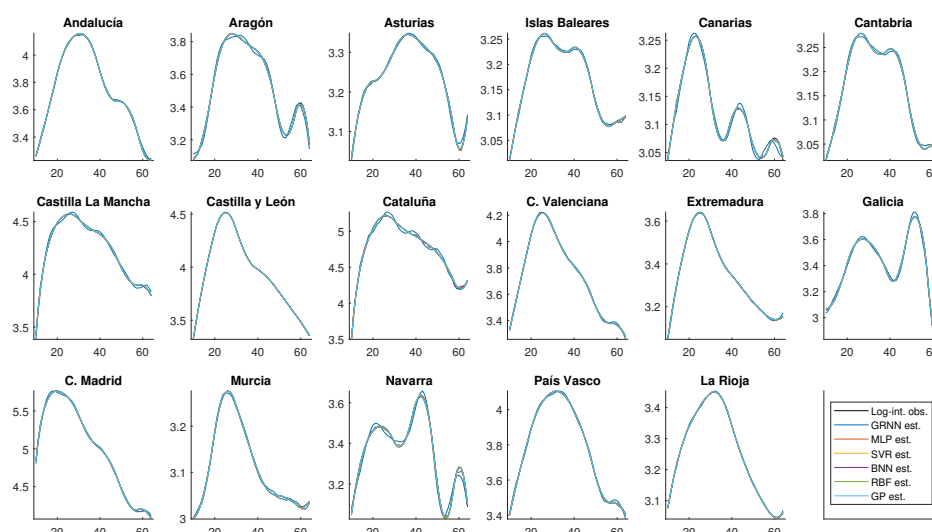


Figura 7.10: *Categoría de datos duros*. Curvas de log-riesgo de mortalidad por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), y Procesos Gaussianos (GP)

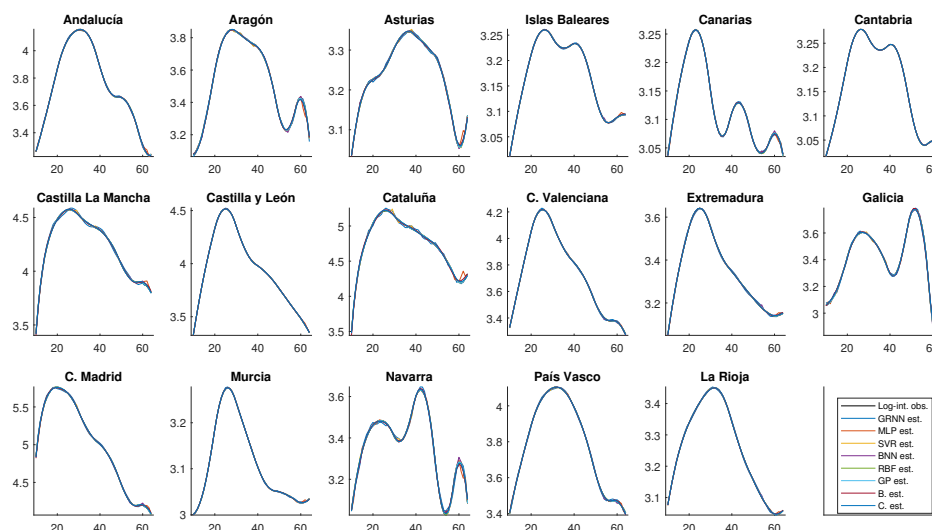


Figura 7.11: *Categoría de datos blandos*. Curvas de log-riesgo de mortalidad por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), Procesos Gaussianos (GP), y regresión trigonométrica combinada con predicción residual clásica y bayesiana

Capítulo 7. Análisis de la mortalidad por COVID-19 mediante curvas multivariantes de regresión a partir de datos blandos y mediante aprendizaje automático

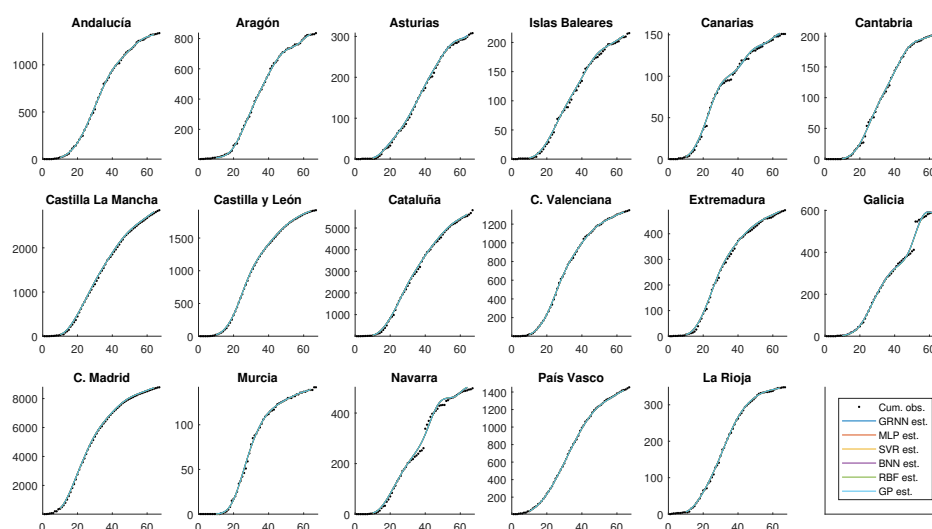


Figura 7.12: **Categoría de datos duros.** Curvas de mortalidad acumulada por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), y Procesos Gaussianos (GP)

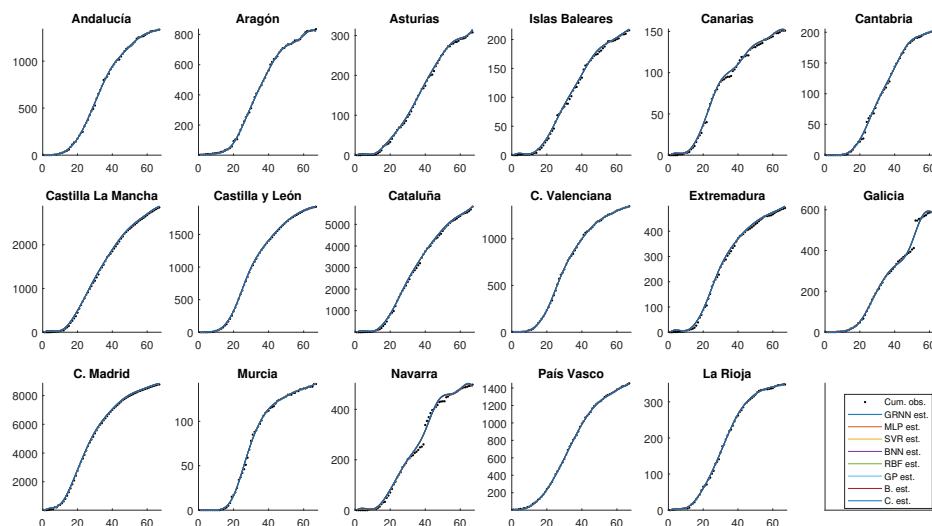


Figura 7.13: **Categoría de datos blandos.** Curvas de mortalidad acumulada por COVID-19 observadas y estimadas, a partir de la implementación de Redes Neuronales de Regresión Generalizada (GRNN), Perceptrón Multicapa (MLP), Regresión de Soporte Vectorial (SVR), Redes Neuronales Bayesianas (BNN), Redes Neuronales a partir de Bases de Funciones Radiales (RBF), Procesos Gaussianos (GP), y regresión trigonométrica combinada con predicción residual clásica y bayesiana

Tabla 7.13: **Categoría de datos duros**. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold. (Como se indica, los valores mostrados deben multiplicarse por 10^{-2})

SC($\times 10^{-2}$)	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1962	0.0845	0.0890	0.0709	0.0635	0.0592
C2	0.6150	0.1531	0.0711	0.0805	0.0782	0.0710
C3	0.1541	0.0479	0.0438	0.0338	0.0371	0.0325
C4	0.0984	0.0388	0.0190	0.0226	0.0214	0.0226
C5	0.2065	0.0555	0.0378	0.0414	0.0429	0.0397
C6	0.1585	0.0423	0.0227	0.0257	0.0266	0.0263
C7	0.4957	0.0786	0.0771	0.0707	0.0712	0.0587
C8	0.0804	0.0352	0.0226	0.0215	0.0247	0.0189
C9	0.7280	0.2170	0.1061	0.0866	0.0421	0.0475
C10	0.2208	0.0724	0.0751	0.0577	0.0541	0.0494
C11	0.1273	0.0618	0.0373	0.0460	0.0451	0.0385
C12	0.5237	0.1706	0.1441	0.1449	0.1126	0.1065
C13	0.3637	0.0717	0.0701	0.0671	0.0605	0.0480
C14	0.1359	0.0388	0.0307	0.0235	0.0220	0.0213
C15	0.6105	0.1705	0.1592	0.1228	0.1101	0.1121
C16	0.2479	0.0931	0.0926	0.0753	0.0665	0.0632
C17	0.0667	0.0414	0.0195	0.0255	0.0241	0.0246
M.	0.2958	0.0867	0.0657	0.0598	0.0531	0.0494
T.	5.0293	1.4731	1.1177	1.0166	0.9026	0.8400

tudio comparativo empírico, se mantiene casi la misma clasificación empírica de modelos ML. Se encuentran algunas diferencias en los intervalos de confianza *bootstrap* y en los histogramas y las densidades de probabilidad calculadas. Por ejemplo, GRNN parece estar favorecido por la categoría de datos blandos, mientras que MLP muestra un peor rendimiento en esta categoría. Por lo tanto, las diferencias entre GRNN y MLP son menores en la categoría de datos blandos. También en esta categoría, se observa una ligera mejora de BNN con respecto a SVR, conservando casi el mismo rendimiento. RBF y GP muestran un mejor rendimiento en la categoría de datos blandos, siendo RBF un poco superior a GP en esta categoría (ver Tabla 7.10, y Figura 7.8).

El enfoque de regresión trigonométrica y de predicción residual de series temporales multivariantes clásica, basado en los momentos empíricos, muestra resultados similares a los de GRNN, con un rendimiento ligeramente mejor que GRNN, observado en los intervalos *bootstrap* y en el histograma y densidad de probabilidad (véanse las figuras 7.7 y 7.9). Sin embargo, como se muestra en las figuras 7.8 y 7.9, la regresión trigonométrica y la predicción residual bayesiana presentan casi el mismo rendimiento que BNN, con algunas características de

Tabla 7.14: **Categoría de datos blandos**. Promedio de los SMAPE, basado en 10 ejecuciones de validación cruzada aleatoria 5-fold. (Como se indica, los valores mostrados deben multiplicarse por 10^{-2})

SC($\times 10^{-2}$)	GRNN	MLP	SVR	BNN	RBF	GP
C1	0.1569	0.0958	0.0695	0.0604	0.0240	0.0337
C2	0.1836	0.1835	0.0697	0.0880	0.0286	0.0329
C3	0.1029	0.1309	0.0474	0.0491	0.0273	0.0284
C4	0.0433	0.0299	0.0171	0.0181	0.0133	0.0130
C5	0.0609	0.0524	0.0282	0.0264	0.0153	0.0159
C6	0.0259	0.0239	0.0133	0.0148	0.0130	0.0129
C7	0.3783	0.2136	0.1018	0.0963	0.0309	0.0439
C8	0.0774	0.0467	0.0315	0.0320	0.0292	0.0207
C9	0.4968	0.2926	0.1458	0.1329	0.0254	0.0417
C10	0.1710	0.0935	0.0541	0.0489	0.0277	0.0316
C11	0.1556	0.0914	0.0456	0.0441	0.0221	0.0247
C12	0.3759	0.2235	0.1509	0.1396	0.0402	0.0560
C13	0.2894	0.1599	0.0903	0.0775	0.0259	0.0368
C14	0.0375	0.0250	0.0124	0.0153	0.0114	0.0109
C15	0.3646	0.2410	0.1378	0.1297	0.0334	0.0573
C16	0.1792	0.0828	0.0677	0.0578	0.0292	0.0344
C17	0.0900	0.0724	0.0216	0.0256	0.0129	0.0144
M.	0.1876	0.1211	0.0650	0.0622	0.0241	0.0299
T.	3.1893	2.0589	1.1047	1.0567	0.4100	0.5090

distribución de probabilidad ligeramente mejores en nuestro enfoque respecto a BNN (véase también los intervalos *bootstrap*).

Nuestro enfoque se ve menos afectado por la división aleatoria de la muestra, en la implementación del procedimiento de validación cruzada aleatoria k -fold, ya que se ajusta un modelo residual espacial dinámico en un segundo paso (objetivo). Así, el modelo residual de regresión clásica y bayesiana de series temporales multivariante propuesto, se ajusta a las correlaciones lineales espaciales a corto plazo mostradas por la categoría de datos blandos. Sin embargo, el coste que pagamos por el aumento de la complejidad del modelo se refleja en los resultados obtenidos de clasificación de modelos aleatorios basados en validación cruzada aleatoria k -fold y *bootstrap*.

El efecto de componente espacial se refleja en las tablas 7.6 (datos duros), donde se observan las heterogeneidades espaciales mostradas por los errores SMAPE de validación cruzada 10-fold (véase también la Tabla 7.13 en la Sección 7.5). Mientras que la Tabla 7.7 (véase también la Tabla 7.14 en la Sección 7.5) revela los beneficios obtenidos en algunos de los modelos de regresión ML probados a partir de información de datos blandos. En particular, en esta cate-

goría, las posibles correlaciones lineales espaciales se incorporan al análisis, en términos de datos blandos, definidos por las proyecciones de datos curva sobre los autovectores empíricos, diagonalizando la estructura de correlación lineal espacial.

CONCLUSIONES Y LÍNEAS ABIERTAS

A continuación, se lleva a cabo una revisión de las principales conclusiones obtenidas de las contribuciones teóricas, estudios de simulación y aplicaciones prácticas desarrolladas a lo largo de los capítulos 2 al 5, así como de las futuras líneas de investigación que se derivan de todos ellos. En lo que respecta a los capítulos 6 y 7, dedicados en su totalidad a dos aplicaciones con datos reales, se hace un repaso más general, ya que al final de los mismos se han dedicado las secciones de comentarios finales (6.4 y 7.6, respectivamente) para realizar un estudio más detallado tanto de las conclusiones obtenidas como de posibles ampliaciones de dichos capítulos.

La relevancia de los resultados del Capítulo 2 sobre consistencia y normalidad asintótica de estimadores de mínimo contraste para campos aleatorios estacionarios espaciales real-valorados, en un contexto general, reside en las aplicaciones que potencialmente ofrece esta teoría, según se ilustra posteriormente en el estudio de simulación. Adicionalmente, este capítulo ofrece una metodología de estimación que será extendida al contexto infinito-dimensional en los capítulos 3–7.

Respecto al desarrollo de estudios de simulación, como ilustración de los resultados derivados del mencionado Capítulo 2, de forma general, se analizan las propiedades de los estimadores de mínimo contraste obtenidos, para tamaños de muestra finitos. Contemplando en dicho análisis, el comportamiento límite anunciado en los resultados teóricos, para tamaños de muestra suficientemente grandes. En particular, se ilustra la consistencia y normalidad asintótica del estimador *plug-in* de mínimo contraste considerado, basado en el periodograma tapered, en el contexto de los procesos autorregresivos fraccionarios gaussianos espaciales y los campos aleatorios de Gegenbauer espaciales. Así, podemos ver que los resultados numéricos apoyan totalmente los hallazgos teóricos.

El Capítulo 3 abre una nueva área de investigación, en el contexto de los procesos de Cox log-gaussianos en el tiempo, que permite, en particular, el análisis de patrones estacionarios en el tiempo y heterogéneos en el espacio. El caso analizado donde la log-intensidad aleatoria es un proceso de O-U Hilbert-valorado, admite una extensión inmediata en el contexto de modelos de difusión infinito-dimensionales, o evaluados en espacios de funciones.

En dicho Capítulo 3, también se incluye una aplicación con datos reales, sobre el análisis estadístico y modelización de la evolución de enfermedades en el espacio y tiempo, las conclusiones extraídas, tras dicha aplicación, ponen

de manifiesto la influencia de la naturaleza y características de la enfermedad estudiada, sobre el ajuste del modelo de dependencia temporal, para la log-intensidad aleatoria infinito-dimensional. En particular, en relación con el ajuste de modelos de dependencia de largo rango, se observa un exceso de suavizado mientras que, para la estimación log-gaussiana ARH(1), el suavizamiento es más débil, lo que permite mayor heterogeneidad en los patrones espaciales.

En relación con las contribuciones del Capítulo 4, se destaca la extensión de la familia de procesos de Cox introducida en el Capítulo 3 al contexto espacial. En particular, esta extensión hace posible la aplicación de técnicas de estimación, basadas en estadísticos de segundo orden, bajo homogeneidad en el espacio y heterogeneidad en el tiempo, contribuyendo al análisis de patrones puntuales y procesos de recuento, mediante modelos más flexibles. Por otra parte, los recientes avances en las ciencias de la computación han hecho posible la aplicación de técnicas de predicción probabilística factibles, para el análisis estadístico de datos complejos en los espacios de Hilbert, Banach, o métricos/semi-métricos, incluyendo el caso de datos que no se encuentran en un espacio vectorial.

Las técnicas de FDA permiten el análisis de propiedades locales de regularidad de datos de elevada dimensión de naturaleza continua, o bien, que proceden de un fenómeno o proceso subyacente continuo en el espacio y/o tiempo, evitando un mal planteamiento de los problemas clásicos de estimación multivariante. En particular, referimos al lector a Dai y Müller (2018); Goia y Vieu (2016); Horváth y Kokoszka (2012); Ruiz-Medina (2012), entre otros. Este Capítulo 4 pone en funcionamiento estructuras matemáticas más complejas, al tratar con elementos de espacios de Hilbert en el marco del análisis de patrones puntuales.

Resumiendo, el enfoque presentado en dicho capítulo, supone una contribución relevante en el ámbito del análisis espectral de procesos lineales espaciales infinito-dimensionales, bajo homogeneidad espacial, así como contribuye, de forma notable, al ámbito del análisis estadístico de procesos de recuento, en el contexto de la Estadística Espacio-Temporal.

Dado que el enfoque de estimación funcional espacial que se presenta aquí se aplica en el dominio espectral, su versión de parámetro espacial continuo puede formularse directamente a partir de los resultados desarrollados en el Capítulo 4, con una reformulación adecuada de las condiciones que se dan.

Uno de los principales problemas que habrá que afrontar en un futuro próximo es la definición de una versión multivariante de la clase de procesos puntuales introducidos, así como su estimación funcional espectral en un marco multivariante infinito-dimensional.

En el Capítulo 5, la metodología de predicción funcional espacial multiescalar presentada permite el análisis de la heterogeneidad en diferentes escalas temporales del campo de log-intensidad. Es bien sabido que las técnicas de preprocesamiento de FDA (por ejemplo, suavizado mediante B-splines) suelen ocultar o eliminar la variación local a niveles de alta resolución (véase, por ejemplo, Goia y Vieu (2016); Horváth y Kokoszka (2012); Müller y Stadtmüller (2005), entre otros). El enfoque de estimación adoptado en este capítulo supera esta limitación, proporcionando un marco más flexible. Así, se puede realizar una elección adecuada de la escala en la que se debe analizar el campo de log-intensidad, según los objetivos del estudio, y las incertidumbres en los recuentos asociadas a la falta de información de la muestra.

El enfoque basado en la entropía relativa a la frecuencia funcional espacial, que se ha presentado en el Capítulo 5, no requiere información adicional sobre la distribución a-priori de probabilidad de los parámetros, como en el marco bayesiano. En particular, en Vicente et al. (2020), los autores presentan un novedoso enfoque sobre la modelización de recuento multivariante bayesiano espacio-temporal a partir de datos de zonas (véanse también las referencias en el mismo para una visión general del estado de la técnica).

La técnica de reducción de dimensiones basada en FDA que aquí se considera evita los graves problemas de cálculo que suelen surgir cuando las matrices de covarianza de alta dimensión, asociadas a variables gaussianas latentes e hiperparámetros, están involucradas en la metodología de estimación bayesiana. De hecho, los modelos autorregresivos condicionados multivariantes (MCARs) se han desarrollado ampliamente en un marco plenamente bayesiano. Los enfoques basados en Cadenas de Markov Monte Carlo (MCMC) y la Aproximación Anidada Integrada de Laplace (INLA), véase, por ejemplo, Lindgren y Rue (2015), y Rue et al. (2009), suelen aplicarse para el ajuste de los modelos en este contexto. Sin embargo, siguen abiertos varios problemas en relación con la separabilidad y la modelización de covarianza espacio-temporal, cuando se adopta un marco de FDA.

El enfoque de estimación presentado en dicho Capítulo 5 proporciona una

alternativa a INLA en el dominio de la frecuencia funcional, considerando la expansión funcional truncada espacial de Cramér-wavelet, véase Panaretos y Tavakoli (2013a), donde se introduce la expansión funcional de Cramér-Karhunen-Loève. Es bien sabido que la expansión funcional de Cramér-Karhunen-Loève proporciona una representación óptima de un proceso funcional estacionario en el sentido de momento de segundo orden. Este criterio de optimización no siempre se satisface con nuestra expansión funcional de Cramér-wavelet, como contrapartida a una descripción más flexible y precisa de la variación local a altos niveles de resolución. Esto último constituye el objetivo de este capítulo, en relación con el análisis a múltiples escalas en un marco de frecuencias de FDA.

Este enfoque puede extenderse al caso en que el análisis multiresolución se realiza en el espacio, para la aproximación del proceso de log-intensidad funcional continuo espacial oculto que dirige los recuentos, como alternativa a las técnicas espaciales habituales de suavizado mediante B-splines. El enfoque resultante permite el análisis de la heterogeneidad a través de escalas temporales y espaciales, proporcionando una aproximación multiresolución de la interacción espacio-tiempo que afecta a la evolución del proceso de log-intensidad. Esta idea constituye el tema de un trabajo de investigación posterior.

Los capítulos 6 y 7 cierran el ciclo de aportaciones de la presente tesis. La relevancia de los contenidos de estos capítulos reside no sólo en el interés que despiertan los nuevos resultados, sobre estimación clásica y bayesiana semilineal, presentados en el contexto de procesos de Cox dirigidos por una intensidad infinito-dimensional, sino que también es notable su interés práctico, especialmente, en la situación actual que estamos viviendo, donde se requiere con urgencia de avances y modelos matemáticos para el análisis y predicción de epidemias y pandemias. De ahí, que estos capítulos cierren la presente tesis, abriendo una nueva línea de investigación estadística, de especial interés, en el ámbito de la Epidemiología y, en particular, de la modelización y análisis de enfermedades infecciosas, que se fusiona, en un análisis comparativo, con otros modelos de regresión en el contexto ML y de modelos espacio-temporales tradicionales, bajo una perspectiva FDA.

Nuestro objetivo futuro más inmediato será la extrapolación de los resultados y ranking empírico de modelos derivados, sobre la mortalidad por COVID-19, para las comunidades españolas, a otros países. Se ampliará asimismo dicho

estudio al análisis y modelización de la propagación de la COVID-19, con el objeto de realizar predicciones a corto y largo plazo sobre los posibles rebrotes, así como sobre las necesidades sanitarias que se deben cubrir para atender la población de infectados asociada a dichos brotes.

CONCLUSIONS AND OPEN LINES OF RESEARCH

A review of the main conclusions drawn from the theoretical contributions, simulation studies and practical applications developed throughout Chapters 2 to 5, as well as the future lines of research derived from all of them follow. Chapters 6 and 7, which are entirely devoted to two applications with real data, are reviewed in a more general way, since at the end of these chapters the final comments sections (6.4 and 7.6, respectively) are devoted to a more detailed study of both the conclusions obtained and possible extensions of these chapters.

The relevance of the results in Chapter 2 on consistency and asymptotic normality of minimum-contrast estimators for spatially stationary real-valued spatial random fields, in a general context, lies in the applications potentially offered by this theory, as subsequently illustrated in the simulation study. In addition, this chapter offers an estimation methodology that will be extended to the infinite-dimensional context in Chapters 3–7.

Regarding the development of simulation studies, as an illustration of the results derived from the aforementioned Chapter 2, in general, the properties of the minimum contrast estimators obtained for finite sample sizes are analyzed. In this analysis, the limiting behavior announced in the theoretical results for sufficiently large sample sizes is considered. In particular, we illustrate the consistency and asymptotic normality of the considered minimum-contrast *plug-in* estimator, based on the tapered periodogram, in the context of spatial Gaussian fractional autoregressive processes and spatial Gegenbauer random fields. Thus, we can see that the numerical results fully support the theoretical findings. Likewise, it can be applied to more general models.

Chapter 3 opens up a new area of research, in the context of log-Gaussian Cox processes in time, which allows, in particular, the analysis of time-stationary and spatially heterogeneous patterns. The analyzed case where the random log-intensity is an O-U Hilbert-valued process, admits an immediate extension in the context of infinite-dimensional diffusion models, or evaluated in function spaces.

An application with real data is also included in this Chapter 3, on the statistical analysis and modeling of the evolution of diseases in space and time, the conclusions drawn, after such application, show the influence of the nature and characteristics of the disease studied, on the fit of the time dependence model, for the infinite-dimensional random log-intensity. In particular, in relation to

the fit of dependence models with different rank, an excess of smoothing is observed while, for the log-Gaussian ARH(1) estimation, the smoothing is weaker, allowing for more heterogeneity in the spatial patterns.

In relation to the contributions of Chapter 4, the extension of the Cox family of processes introduced in Chapter 3 to the spatial context stands out. In particular, this extension makes possible the application of estimation techniques, based on second-order statistics, under homogeneity in space and heterogeneity in time, contributing to the analysis of point patterns and counting processes, by means of more flexible models.

On the other hand, recent advances in computer science have made possible the application of feasible probabilistic prediction techniques for the statistical analysis of complex data in Hilbert, Banach, or metric/semi-metric spaces, including the case of data that are not in a vector space.

FDA techniques allow the analysis of local regularity properties of high-dimensional data of a continuous nature, or else, coming from an underlying phenomenon or process continuous in space and/or time, avoiding a bad approach to classical multivariate estimation problems. In particular, we refer the reader to Dai y Müller (2018); Goia y Vieu (2016); Horváth y Kokoszka (2012); Ruiz-Medina (2012), among others. This Chapter 4 puts more complex mathematical structures to work, dealing with elements of Hilbert spaces in the framework of point pattern analysis.

In summary, the approach presented in this chapter is a relevant contribution in the field of spectral analysis of infinite-dimensional spatial linear processes under spatial homogeneity, as well as a remarkable contribution to the field of statistical analysis of counting processes, in the context of Spatio-Temporal Statistics.

Since the spatial functional estimation approach presented here is applied in the spectral domain, its continuous spatial parameter version can be formulated directly from the results developed in Chapter 4, with an appropriate reformulation of the conditions given. One of the main problems to be faced in the near future is the definition of a multivariate version of the introduced class of point processes, as well as their spectral functional estimation in an infinite-dimensional multivariate framework.

In Chapter 5, the presented multiscale spatial functional prediction methodology allows the analysis of heterogeneity at different time scales of the log-

intensity field. It is well known that FDA preprocessing techniques (e.g., B-spline smoothing) often hide or remove local variation at high resolution levels (see, e.g., [Goia y Vieu \(2016\)](#); [Horváth y Kokoszka \(2012\)](#); [Müller y Stadtmüller \(2005\)](#), among others). The estimation approach adopted in this chapter overcomes this limitation by providing a more flexible framework. Thus, an appropriate choice of the scale at which the log-intensity field should be analyzed can be made, depending on the objectives of the study, and the uncertainties in the counts associated with the lack of sample information.

The entropy-based approach related to the spatial functional frequency, which has been presented in Chapter 5, does not require additional information on the prior probability distribution of the parameters, as in the Bayesian framework. In particular, in [Vicente et al. \(2020\)](#), the authors present a novel approach to Bayesian spatio-temporal multivariate count modeling from area data (see also the references therein for an overview of the state of the art).

The FDA-based dimension reduction technique considered here avoids the severe computational problems that often arise when high-dimensional covariance matrices, associated with latent Gaussian variables and hyperparameters, are involved in Bayesian estimation methodology. Indeed, multivariate conditional autoregressive models (MCARs) have been extensively developed in a fully Bayesian framework. Approaches based on Markov Chain Monte Carlo (MCMC) and the Integrated Nested Laplace Approximation (INLA), see, for example, [Lindgren y Rue \(2015\)](#), and [Rue et al. \(2009\)](#), are often applied for model fitting in this context. However, several problems remain open in relation to separability and spatio-temporal covariance modeling, when adopting an FDA framework.

The estimation approach presented in that Chapter 5 provides an alternative to INLA in the functional frequency domain by considering the spatially truncated Cramér-wavelet functional expansion, see [Panaretos y Tavakoli \(2013a\)](#), where the Cramér-Karhunen-Loève functional expansion is introduced. It is well known that the Cramér-Karhunen-Loève functional expansion provides an optimal representation of a stationary functional process in the sense of second-order moment. This optimization criterion is not always satisfied by our Cramér-wavelet functional expansion, but this is the price to pay for a more flexible and accurate description of local variation at high levels of resolution. That, on the other hand, is the aim of this chapter, in relation to the analysis at multiple sca-

les in an FDA frequency framework. This optimization criterion is not always satisfied by our Cramér-wavelet functional expansion, as a trade-off for a more flexible and accurate description of local variation at high levels of resolution. The latter is the focus of this chapter, in relation to multiscale analysis in an FDA frequency framework.

This approach can be extended to the case where the multiresolution analysis is performed in space, for the approximation of the hidden spatial continuous functional log-intensity process driving the counts, as an alternative to the usual spatial B-spline smoothing techniques. The resulting approach allows the analysis of heterogeneity across temporal and spatial scales, providing a multiresolution approximation of the space-time interaction affecting the evolution of the log-intensity process. This topic is the subject of further research work.

Chapters 6 and 7 close the cycle of contributions of this thesis. The relevance of the contents of these chapters lies not only in the interest aroused by the new results on classical and Bayesian semilinear estimation, presented in the context of Cox processes driven by an infinite-dimensional intensity, but also in their practical interest, especially in the current situation, where mathematical advances and models are urgently required for the analysis and prediction of epidemics and pandemics. Hence, these chapters close this thesis, opening a new line of statistical research of special interest in the field of epidemiology and, in particular, in the modeling and analysis of infectious diseases, which merges, in a comparative analysis, with other regression models in the ML context and traditional spatio-temporal models, under an FDA perspective.

Our immediate future goal will be to extrapolate the results and empirical ranking of derived models on COVID-19 mortality for Spanish communities to other countries. This study will also be extended to the analysis and modeling of the spread of COVID-19, with the aim of making short- and long-term predictions on possible resurgences, as well as on the health needs to be covered to care for the infected population associated with such outbreaks.

BIBLIOGRAFÍA

-
- O. O. Aalen, O. Borgan, y H. K. Gjessing. *Survival and event history analysis: a process point of view*. Springer Science & Business Media, New-York, 2008.
- C. Abboud, O. Bonnefon, E. Parent, y S. Soubeyrand. Dating and localizing an invasion from post-introduction data and a coupled reaction-diffusion-absorption model. *Journal of Mathematical Biology*, 79:765–789, 2019.
- A. Adin, T. Goicoa, y M. D. Ugarte. Online relative risks/rates estimation in spatial and spatio-temporal disease mapping. *Computer Methods and Programs in Biomedicineh*, 172:103–116, 2019. URL <http://www.sciencedirect.com/science/article/pii/S0169260718318030>.
- R. Adler y J. Taylor. *Random Fields and Geometry*. Springer Science + Business Media LLC, 2007.
- C. Agostinelli. Robust model selection in regression via weighted likelihood methodology. *Statistics & Probability Letters*, 56:289–300, 2001.
- H. M. Alomari, M. P. Frías, N. N. Leonenko, M. D. Ruiz-Medina, L. Sakhno, y A. Torres. Asymptotic properties of parameter estimates for random fields with tapered data. *Electronic Journal of Statistics*, 11:3332–3367, 2017. URL <https://doi.org/10.1214/17-EJS1315>.
- E. Alpaydin. *Introduction to Machine Learning*. MIT Press, Cambridge, MA, 2004.
- C. Anderson, D. Lee, y N. Dean. Spatial clustering of average risks and risk trends in bayesian disease mapping. *Biometrical Journal*, 59:41–56, 2017. URL <https://doi.org/10.1002/bimj.201600018>.
- J. Angulo, H.-L. Yu, A. Langousis, A. Kolovos, J. Wang, Madrid A. E., y G. Christakos. Spatiotemporal infectious disease modeling: A bme-sir approach. *PLoS One*, 8(9):e72168, 2013. URL <https://doi.org/10.1371/annotation/4adf8407-a5b8-4f4b-877d-e8b944f0e6ee>.

-
- J. M. Angulo y M. D. Ruiz-Medina. On the orthogonal representation of generalized random fields. *Statistics & Probability Letters*, 31:145–153, 1997.
- V. V. Anh, J. M. Angulo, y M. D. Ruiz-Medina. Possible long-range dependence in fractional random fields. *Journal of Statistical Planning and Inference*, 80:95–110, 1999.
- V. V. Anh, N. N. Leonenko, y L. M. Sakhno. On a class of minimum contrast estimators for fractional stochastic processes and fields. *Journal of Statistical Planning and Inference*, 123:161–185, 2004a.
- V. V. Anh, N. N. Leonenko, y L. M. Sakhno. Quasi-likelihood-based higher-order spectral estimation of random fields with possible long-range dependence. Stochastic methods and their applications. *Journal of Applied Probability*, 41A:35–53, 2004b.
- V. V. Anh, N. N. Leonenko, y L. M. Sakhno. LM evaluation of bias in higher-order spectral estimation. *Theory of Probability and Mathematical Statistics*, 80:1–14, 2007a.
- V. V. Anh, N. N. Leonenko, y L. M. Sakhno. Minimum contrast estimation of random processes based on information of second and third orders. *Journal of Statistical Planning and Inference*, 137(4):1302–1331, 2007b.
- V. V. Anh, N. N. Leonenko, y L. M. Sakhno. Statistical inference based on the information of the second and third order. *Journal of Statistical Planning and Inference*, 137:1302–1331, 2007c.
- V. V. Anh, N. N. Leonenko, y L. M. Sakhno. Statistical inference using higher-order information. *Journal of Multivariate Analysis*, 98(4):706–742, 2007d.
- V. V. Anh y K. E. Lunney. Parameter estimation of random fields with longrange dependence. *Mathematical and Computer Modelling*, 21:67–77, 1995.
- A. Antoniadis y T. J. Spatinas. Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis*, 87:133–158, 2003.

-
- A. Aswi, S. M. Cramb, P. Moraga, y K. Mengersen. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: a systematic review. *Epidemiology and Infection*, 33:1–14, 2019. URL <https://doi.org/10.1017/S0950268818002807>.
- F. Avram. On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probability Theory and Related Fields*, 79:37–45, 1988.
- F. Avram, N. N. Leonenko, y L. M. Sakhno. Harmonic analysis tools for statistical inference in the spectral domain. *Lecture Notes in Statistics*, 200:59–70, 2010a.
- F. Avram, N. N. Leonenko, y L. M. Sakhno. On a szegö type limit theorem, the Hölder-Young-Brascamp-Lieb inequality, and the asymptotic theory of integrals and quadratic forms of stationary fields. *ESAIM: Probability and Statistics*, 14:210–255, 2010b.
- J-M. Azaïs y M. Wschebor. *Level Sets and Extrema of Random Processes and Fields*. John Wiley & Sons, Inc., 2009.
- A. Baddeley, P. Gregori, J. Mateu, R. Stoica, y D. Stoyan. *Case Studies in Spatial Point Process Modeling*. Springer, New York, 2006.
- A. Baddeley, J. Møller, y R. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–50, 2000.
- A. Baddeley y R. Turner. Practical maximum pseudolikelihood for spatial point processes. *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.
- A. Banks, J. Vincent, y C. Anyakoha. A review of particle swarm optimization. part i: background and development. *Natural Computing*, 6:467–484, 2007.
- M. Barstugan, U. Ozkaya, y S. Ozturk. Coronavirus (covid–19) classification using ct images by machine learning methods, 2020. Preimpreso en <https://arxiv.org/abs/2003.09424>.
- S. Basu y A. Dassios. A Cox process with log-normal intensity. *Insurance: Mathematics and Economics*, 31:297–302, 2002.

-
- S. Basu y G. C. Reinsel. Properties of the spatial unilateral first-order ARMA model. *Advances in Applied Probability*, 25(3):631–648, 1993.
- R. Bentkus. The error in estimating the spectral function of a stationary process. *Lietuvos Matematikos Rinkinys*, 12(1):55–71, 1972.
- R. Bentkus. Cumulants of estimates of the spectrum of a stationary time series. *Litovskii Matematicheskii Sbornik*, 16(4):37–61, 1976.
- E. Beretta, T. Hara, W. Ma, y Y. Takeuchi. Global asymptotically stability of an sir epidemic model with distributed time delay. *Nonlinear Analysis, Theory, Methods and Applications*, 47:4107–4115, 2001.
- J. Besag, J. York, y A. Mollie. Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59, 1991.
- J. E. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36:192–225, 1974. URL <https://www.jstor.org/stable/2984812>.
- S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, M. F. Myers, D. B. George, T. Jaenisch, G. R. W. Wint, C. P. Simmons, T. W. Scott, J. J. Farrar, y S. I. Hay. The global distribution and burden of dengue. *Nature*, 496:504–507, 2013. URL <https://doi.org/10.1038/nature12060>.
- R. Bivand, E. Pebesma, y V. Gómez-Rubio. *Applied Spatial Data Analysis with R*. Springer, 2013.
- R. Blanquero, E. Carrizosa, A. Jiménez-Cordero, y B. Martín-Barragán. Selection of time instants and intervals with support vector regression for multivariate functional data, 2020. Preimpreso en <https://www.researchgate.net/publication/327552293>.
- Y. Boissy, B. B. Bhattacharyya, X. Li, y G. D. Richardson. Parameter estimates for fractional autoregressive spatial process. *Annals of Statistics*, 33:2553–2567, 2005.

-
- B. M. Bolker y B. Grenfell. Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proceedings of the National Academy of Sciences*, 93:12648–12653, 1996.
- D. Bosq. *Linear Processes in Function Spaces*. Springer, 2000.
- D. Bosq y D. Blanke. *Inference and Predictions in Large Dimensions*. John Wiley & Sons, 2007.
- D. Bosq y M. D. Ruiz-Medina. Bayesian estimation in a high dimensional parameter framework. *Electronic Journal of Statistics*, 8:1604–1640, 2014.
- P. R. Bouzas, A. M. Aguilera, M. J. Valderrama, y N. Ruiz-Fuentes. On the structure of the stochastic process of mortgages in Spain. *Computational Statistics and Data Analysis*, 21:73–89, 2006a.
- P. R. Bouzas, M. J. Valderrama, A. M. Aguilera, y N. Ruiz-Fuentes. Modeling the mean of a doubly stochastic poisson process by functional data analysis. *Computational Statistics and Data Analysis*, 50:2655–2667, 2006b.
- D. R. Brillinger. The frequency analysis of relations between stationary spatial series. 1970.
- D. R. Brillinger. *Time Series: Data Analysis and Theory*. Holden Day, San Francisco, 1981.
- M. Cabrera y G. Taylor. Modelling spatio-temporal data of dengue fever using generalized additive mixed models. *Spatial and spatio-temporal Epidemiology*, 28:1–13, 2019. URL <https://doi.org/10.1016/j.sste.2018.11.006>.
- A. E. Carbajo, N. Schweigmann, S. I. Curto, A. De Garín, y R. Bejarán. Dengue transmission risk maps of Argentina. *Tropical Medicine and International Health*, 6:170–183, 2001. URL <https://doi.org/10.1046/j.1365-3156.2001.00693.x>.
- L. Cattarino, I. Rodríguez-Barraquer, N. Imai, D. A. T. Cummings, y N. M. Ferguson. Mapping global variation in dengue transmission intensity. *Science Translational Medicine*, 12(528):1–11, 2020. URL <https://doi.org/10.1126/scitranslmed.aax4144>.

-
- A. Chakraborty y A. E. Gelfand. Analyzing spatial point patterns subject to measurement error. *Bayesian Analysis*, 5:97–122, 2001.
- K. S. Chan y H. Tsai. Inference of seasonal long-memory aggregate time series. *Bernoulli*, 4(18):1448–1464, 2012.
- D. L. Chao, J. D. Bloom, B. F. Kochin, R. Antia, y I. M. Longini Jr. The global spread of drug-resistant influenza. *Journal of the Royal Society Interface*, 9:648–656, 1996.
- O. Chapelle, V. Vapnik, y Y. Bengio. Model selection for small sample regression. *Machine Learning*, 48:9–23, 2002.
- L.-Ch. Chien y L.-W. Chen. Meteorological impacts on the incidence of covid-19 in the u.s. *Stochastic Environmental Research and Risk Assessment*, 34:1675–1680, 2002.
- K. M. Choi, M. L. Serre, y G. Christakos. Efficient mapping of california mortality fields at different spatial scales. *Journal of Exposure Analysis and Environmental Epidemiology*, 13:120–133, 2003.
- G. Christakos. *Random Field Models in Earth Science*. Dover Publications, INC, Mineola, New York, 1992.
- G. Christakos. *Modern Spatiotemporal Geostatistics*. Oxford University Press, New York, 2000.
- G. Christakos. *Spatiotemporal Random Fields: Theory and Applications*. Elsevier, New York, 2017.
- G. Christakos, P. Bogaert, y M. L. Serre. *Temporal GIS: Advanced Functions for Field-Based Applications*. Springer Verlag, New York, 2001.
- G. Christakos y D. T. Hristopulos. *Spatiotemporal Environmental Health Modeling: A Tractatus Stochasticus*. Kluwer Academic Publisher, Boston, 1998.
- G. Christakos y R. A. Olea. New space-time perspectives on the propagation characteristics of the black death epidemic and its relation to bubonic plague. *Stochastic Environmental Research and Risk Assessment*, 19:307–314, 2005.

-
- C. F. Chung. Estimating a generalized long memory process. *Journal of Econometrics*, 73:237–259, 1996.
- P. Congdon. Representing spatial dependence and spatial discontinuity in ecological epidemiology: a scale mixture approach. *Stochastic Environmental Research and Risk Assessment*, 31:291—304, 2017.
- D. R. Cox. Some statistical methods connected with series of events. *Journal of Royal Statistical Society: Series B*, 17:129–164, 1955.
- D. R. Cox, J. Ingersoll, y S. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–408, 1985.
- R. M. Crujeiras. *Contributions to Spectral Spatial Statistics*. Tesis Doctoral, Universidad de Santiago de Compostela, 2006.
- G. Da Prato y J. Zabczyk. Stochastic equations in infinite dimensions. En *Encyclopedia of Mathematics and its Applications*, tomo 44. Cambridge University Press, 1992.
- R. Dahlhaus. Spectral analysis with tapered data. *Journal of Time Series Analysis*, 4(3):163–175, 1983.
- R Dahlhaus y H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 73(4):877–882, 1987.
- R. Dahlhaus y W. Polonik. Empirical spectral processes and nonparametric maximum likelihood estimation for time series. En H. Dehling, T. Mikosch, y M. Sorensen, eds., *Empirical Process Techniques for Dependent Data*. Springer, 2002.
- X. Dai y H. G. Müller. Principal component analysis for functional data on riemannian manifolds and spheres. *Annals of Statistics*, 46:3334–3361, 2018.
- S. Dalal y A. McIntosh. When to stop testing for large software systems with changing code. *IEEE Transactions on Software Engineering*, 20:318–323, 1994.
- D. Daley y D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. New York: Springer-Verlag, 2003.

-
- D. Daley y D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. New York: Springer-Verlag, 2008.
- D. J. Daley y D. Vere-Jones. *An Introduction to Theory of Point Processes*. Springer-Verlag, New York, 1988.
- J. Damon y S. Guillas. Estimation and simulation of autoregressive Hilbertian processes with exogenous variables. *Statistical Inference for Stochastic Processes*, 8:185–204, 2005.
- I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
- R. Dautray y J. L. Lions. Spectra theory and applications. En *Mathematical Analysis and Numerical Methods for Science and Technology*, tomo 3. Springer, New York, 1985.
- A. De Genero y A. Simonis. Properties of doubly stochastic Poisson process with affine intensity. *IEEE Transactions on Software Engineering*, 2011. ArXiv:1109.2884v2.
- P. Delicado, R. Giraldo, C. Comas, y J. Mateu. Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21:224–239, 2010.
- P. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press, Taylor & Francis, Boca Raton, 2013.
- P. Diggle, Y. Guan, A. Hart, F. Paize, y M. Stanton. Estimating individual-level risk in spatial epidemiology using spatially aggregated information on the population at risk. *Journal of the American Statistical Association*, 105:1394–1402, 2010a.
- P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. London: Academic Press, 1983.
- P. J. Diggle, I. Kaimi, y R. Abellana. Partial-likelihood analysis of spatio-temporal point-process data. *Biometrics*, 66:347–354, 2010b.
- P. J. Diggle y R. K. Milne. Bivariate cox processes: Some models for bivariate spatial point processes. *Journal of the Royal Statistical Society: Series B*, 45:11–21, 1983.

-
- P. J. Diggle, P. Moraga, B. Rowlingson, y B. M. Taylor. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statistica Science*, 28:542–563, 2013.
- P. Doukhan, J. León, y P. Soulier. Central and non central limit theorems for strongly dependent stationary Gaussian field. *Rebrape*, 10:205–223, 1996.
- Z. Du, X. Xu, Y. Wu, L. Wang, L. A. Cowling, y B. J. Meyers. Serial interval of covid-19 among publicly reported confirmed cases. *Emerging Infectious Diseases*, 26(6):1341–1343, 2020.
- D. Duffie y R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6:379–406, 1996.
- J. Dushoff, J. Plotkin, S. Levin, y D. Earn. Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 101:16915–16916, 2004.
- M. Elhia, A. Laaroussi, M. Rachik, Z. Rachik, y E. Labriji. Global stability of a susceptible-infected-recovered (sir) epidemic model with two infectious stages and treatment. *International Journal of Scientific Research*, 3:114–121, 2014.
- R. Espejo, N. N. Leonenko, A. Olenko, y M. D. Ruiz-Medina. On a class of minimum contrast estimators for Gegenbauer random fields. *TEST*, 24:657–680, 2015.
- E. L. Estallo, A. E. Carbajo, M. G. Grech, M. Frías-Céspedes, L. López, M. A. Lanfri, F. F. Ludueña Almeida, y W. R. Almirón. Spatio-temporal dynamics of dengue 2009 outbreak in córdoba city, argentina. *Acta Tropica*, 136:129–136, 2014. URL <https://doi.org/10.1016/j.actatropica.2014.04.024>.
- F. Ferraty y Y. Romain. *The Oxford Handbook of Functional Data Analysis*. Oxford University Press, 2010.
- F. Ferraty y P. Vieu. *Nonparameric Functional Data Analysis*. Springer, 2006.
- T. R. Fleming y D. P. Harrington. *Counting processes and survival analysis*. John Wiley & Sons, Inc., New-York, 1991.

-
- J. P. Fouque, G. Papanicolaou, y K. R. Sircar. *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge University Press, 2000.
- R. Fox y M. S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics*, 14:517–532, 1986.
- R. Fox y M. S. Taqqu. Central limit theorems for quadratic forms in random variables having long-range dependence. *Probability Theory and Related Fields*, 74:213–240, 1987.
- M. P. Frías, A. Torres-Signes, M.D. Ruiz-Medina, y J. Mateu. Spatial cox processes in an infinite-dimensional framework, 2020. Bajo segunda revisión en la revista *TEST*. Preimpreso en <https://arxiv.org/pdf/arXiv:1811.11139.pdf>.
- M. P. Frías y M. D. Ruiz-Medina. Computing functional estimators of spatio-temporal long-range dependence parameters in the spectral-wavelet domain. *Journal of Statistical Planning and Inference*, 141:2417–2427, 2011.
- M. P. Frías, M. D. Ruiz-Medina, y V. V. Anh. Wavelet-based estimation of anisotropic spatiotemporal long-range dependence. *Stochastic Analysis and Applications*, 31:359–380, 2013.
- M. Fuhrman. A note on the nonsymmetric Ornstein-Uhlenbeck process in Hilbert spaces. *Applied Mathematics Letters*, 8:19–22, 1995.
- C. Gaetan y X. Guyon. *Spatial Statistics and Modeling*. Springer, 2010.
- M. Gail, T. Santner, y C. Brown. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, 36:255–266, 1980.
- A. Gelman. Prior distributions for variance parameters in hierarchical models. *International Society for Bayesian Analysis*, 3:515–533, 2006.
- M. S. Ginovyan, A. A. Sahakyan, y M. S. Taqqu. The trace problem for Toeplitz matrices and operators and its impact in probability. *Probability Surveys*, 11:393–440, 2014.
- R. Giraldo, P. Delicado, y J. Mateu. Geostatistics for functional data: an ordinary kriging approach. *Environmental and Ecological Statistics*, 18:411–426, 2010.

-
- A. Goia y P. Vieu. An introduction to recent advances in high/infinite dimensional statistics. *Journal of Multivariate Analysis*, 146:1–6, 2016.
- T. Goicoa, M. D. Ugarte, J. Etxeberria, y A. F. Militino. Comparing CAR and P-spline models in spatial disease mapping. *Environmental and Ecological Statistics*, 19:573–599, 2012.
- F. B. Gonçalves y D. Gamerman. Exact bayesian inference in spatiotemporal cox processes driven by multivariate gaussian processes. *Journal of the Royal Statistical Society: Series B*, 80:157–175, 2018.
- J. A. González, F. J. Rodríguez-Cortés, O. Cronie, y J. Mateu. Spatio-temporal point process statistics: A review. *Spatial Statistics*, 18:505–544, 2016.
- J. Grandell. *Doubly Stochastic Process*. Springer-Verlag, New York, 1976.
- Y. Guan. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101:1502–1512, 2006.
- Y. Guan, M. Sherman, y J. A. Calvin. Assessing isotropy for spatial point processes. *Biometrics*, 62:119–125, 2006.
- Y. Guan, R. Waagepetersen, y C. M. Beale. Second-order analysis of inhomogeneous spatial point processes with proportional intensity functions. *Journal of the American Statistical Association*, 103:769–777, 2008.
- L. N. Guin y P. K. Mandal. Spatiotemporal dynamics of reaction–diffusion models of interacting populations. *Applied Mathematical Modelling*, 38:4417–4427, 2014.
- H. Guo, C. Y. Lim, y M. M. Meerschaert. Local whittle estimator for anisotropic random field. *Journal of Multivariate Analysis*, 100:993–1028, 2009.
- X. Guyon. Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, 69:95–105, 1982.
- X. Guyon. *Random Fields on a Network*. Springer-Verlag, New York, 1995.
- E. J. Hannan. The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10:130–145, 1973.

-
- T. Hastie, R. Tibshirani, y J. Friedman. *The elements of statistical learning*. Springer-Verlag, New-York, 2001.
- J. He, G. Chen, Y. Jiang, R. Jin, A. Shortridge, S. Agusti, M. Hea, J. Wua, C. M. Duarte, y G. Christakos. Comparative infection modeling and control of covid-19 transmission patterns in china, south korea, italy and iran. *Science of the Total Environment*, págs. 1–15, 2020. URL [doi:10.1016/j.scitotenv.2020.141447](https://doi.org/10.1016/j.scitotenv.2020.141447).
- C. Heil y F. Walnut. *Fundamental Papers in Wavelet Theory*. Princeton University Press, Oxford, 2006.
- C. C. Heyde y R. Gay. Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence. *Stochastic Processes and their Applications*, 45:169–182, 1993.
- S. Hörmann y P. Kokoszka. Weakly dependent functional data. *Annals of Statistics*, 38:1845–1884, 2010.
- S. Hörmann y P. Kokoszka. Functional time series. En C.R. Rao, ed., *Time Series Analysis: Methods and Applications*, tomo 30, págs. 157–186. Springer, 2012.
- S. Hörmann y P. Kokoszka. Consistency of the mean and the principal components of spatially distributed functional data. *Bernoulli*, 19(5A):1535–1558, 2013.
- S. Hörmann, P. Kokoszka, y G. Nisol. Testing for periodicity in functional time series. *Annals of Statistics*, 46(6A):2960–2984, 2018.
- L. Horváth y P. Kokoszka. *Inference for Functional Data with Applications*. New York: Springer, 2012.
- L. Horváth, P. Kokoszka, y R. Reeder. Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B*, 75:103–122, 2013.
- T. Hsing y R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons, 2015.
- A. Huppert y G. Katriel. Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection*, 19:999–1005, 2013.

-
- I. A. Ibragimov. On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theory of Probability and Its Applications*, 12(1):115–119, 1967.
- R. Ignaccolo, J. Mateu, y R. Giraldo. Kriging with external drift for functional data for air quality monitoring. *Stochastic Environmental Research and Risk Assessment*, 28:1171–1186, 2014.
- J. Illian, A. Penttinen, H. Stoyan, y D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. New York: John Wiley & Sons, 2008.
- ISCIIII. Instituto de Salud Carlos III (Centro Nacional de Epidemiología). 2020. URL <https://www.isciii.es/Paginas/Inicio.aspx>. Accedido el 30-05-2020.
- A. V. Ivanov, N. N. Leonenko, M. D. Ruiz-Medina, y B. M. Zhurakovsky. Estimation of harmonic component in regression with cyclically dependent errors. *Statistics: A Journal of Theoretical and Applied Statistics*, 49:156–186, 2015.
- A.V. Ivanov y V. Prikhodko. Asymptotic properties of Ibragimov's estimate of parameter of spectral density of noise in the nonlinear regression method. *Theory of Probability and Mathematical Statistics*, 93:50–66, 2015.
- B. Ivorra, M. R. Ferrández, M. Vela-Pérez, y A. M. Ramos. Mathematical modeling of the spread of the coronavirus disease 2019 (covid-19) taking into account the undetected infections. the case of china. *Communications in Nonlinear Science and Numerical Simulation*, 88:105–303, 2020.
- B. Ivorra, A. M. Ramos, y Ngom. Be-codis: A mathematical model to predict the risk of human diseases spread between countries. validation and application to the 2014 ebola virus disease epidemic. *Bulletin of Mathematical Biology*, 77:1668–1704, 2015.
- A. Jalilian, Y. Guan, J. Mateu, y R. Waagepetersen. Multivariate product-shot-noise cox point process models. *Biometrics*, 71:1022–1033, 2015.
- A. Jalilian, Y. Guan, y R. Waagepetersen. Decomposition of variance for spatial cox processes. *Scandinavian Journal of Statistics*, 40:119–137, 2013.

-
- A. Jalilian, Y. Guan, y R. Waagepetersen. Orthogonal series estimation of the pair correlation function of a spatial point process. *Statistica Sinica*, 29:769–787, 2019.
- M. Jaya, A. Atje Setiawan, E. Hermawan, y B. Ruchjana. Bayesian spatial modeling and mapping of dengue fever: A case study of dengue fever in the city of bandung indonesia. *International Journal of Applied Mathematics and Statistics*, 54(3):94–103, 2016. URL <http://www.ceser.in/ceserp/index.php/ijamas/article/view/4303>. Accedido el 10–08–2019.
- C. Ji, D. Jiang, y N. Shi. The behavior of an sir epidemic model with stochastic perturbation. *Stochastic Analysis and Applications*, 30:755–773, 2012.
- M. J. Keeling, D. A. Rand, y A. J. Morris. Correlation models for childhood epidemics. *Proceedings of the Royal Society of London*, 264:1149—1156, 1997.
- M. J. Keeling y P. Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, Princeton, 2008.
- M. Kelbert, N. N. Leonenko, y M. D. Ruiz-Medina. Fractional random fields associated with stochastic fractional heat equations. *Advances in Applied Probability*, 37:108–133, 2005.
- J. Kennedy y R. Eberhart. Particle swarm optimization. *Proceedings of ICNN'95-International Conference on Neural Networks*, 4:1942–1948, 1995.
- W. Kermack y A. McKendrick. Contributions to the mathematical theory of epidemics - i. *Proceedings of the Royal Society of Edinburgh Section A*, 115:700–721, 1927.
- M. A. Khan y A. Atangana. Modeling the dynamics of novel coronavirus (2019-ncov) with fractional derivative. *Alexandria Engineering Journal*, 2020. URL doi.org/10.1016/j.aej.2020.02.033.
- L. Kidziński, P. Kokoszka, y N. M. Jouzdani. Principal component analysis of periodically correlated functional time series. *Journal of Time Series Analysis*, 2018. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12283>.

-
- P. Kokoszka, R. Gabrys, y L. Horváth. Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, 105:1113–1125, 2010.
- E. Kopp. *From Measures to Itô Integrals*. Cambridge University Press, 2011.
- Y. V. Kozachenko y O. O. Pogorilyak. A method of modelling log Gaussian Cox process. *Theory of Probability and Mathematical Statistics*, 77:91–105, 2008.
- A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, y R. M. Eggo. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 2020. URL [doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4).
- Y. A. Kuznetsov y C. Piccardi. Bifurcation analysis of periodic seir and sir epidemic models. *Journal of Mathematical Biology*, 32:109–121, 1994.
- A. Laaroussi, M. Rachik, y M. Elhia. An optimal control problem for a spatiotemporal sir model. *International Journal of Dynamics and Control*, 6:384–397, 2018.
- D. Lando. On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120, 1998.
- A. Langousis y A. A. Carsteanu. Undersampling in action and at scale: application to the covid-19 pandemic. *Stochastic Environmental Research and Risk Assessment*, 34:1281–1283, 2020.
- A. B. Lawson. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Taylor and Francis, Boca Raton, Florida, USA, 1 ed^{ón}., 2008.
- M. Ledoux y M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete (3)*. Springer, Berlin, 1991.
- E. L. Lehmann y G. Casella. *Theory of Point Estimation*. Springer, 2 ed^{ón}., 1998.
- N. Leonenko. *Limit Theorems for Random Fields with Singular Spectrum*. Springer Science + Business Media, B. V., 1999.

-
- N. N. Leonenko y M. D. Ruiz-Medina. Scaling laws for the multidimensional burgers equation with quadratic external potentials. *Journal of Statistical Physics*, 124:191–205, 2006.
- N. N. Leonenko y E. Taufer. Disaggregation of spatial autoregressive processes. *Spatial Statistics*, 3:1–20, 2013.
- N.N. Leonenko, M.D. Ruiz-Medina, y M. Taqqu. Fractional elliptic, hyperbolic and parabolic random fields. *Electronic Journal of Probability*, 16:1134–1172, 2011.
- L. Li, J. Wang, Z. Cao, y E. Zhong. An information-fusion method to identify pattern of spatial heterogeneity for improving the accuracy of estimation. *Stochastic Environmental Research and Risk Assessment*, 22:689–704, 2008.
- W. K. Li y A. I. McLeod. Fractional time series modelling. *Biometrika*, 73:217–221, 1986.
- F. Lindgren y H. Rue. Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, 63:1–25, 2015.
- C. Lorenz, F. Chiaravalloti-Neto, M. de Oliveira Lage, J. A. Quintanilha, M. C. Parra, M. R. Dibo, E. A. Fávaro, M. M. Guirado, y M. L. Nogueira. Remote sensing for risk mapping of aedes aegypti infestations: Is this a practical task? *Acta Tropica*, 205:105398, 2020. URL <https://doi.org/10.1016/j.actatropica.2020.105398>.
- R. Lowe, Coelho C. Barcellos, C. and, T. Bailey, G. E. Coelho, R. Graham, T. Jupp, W. Ramalho, M. Carvalho, D. Stephenson, y J. Rodó. Dengue outlook for the world cup in brazil: an early warning model framework driven by real-time seasonal climate forecasts. *The Lancet Infectious Diseases*, 14(7):619–626, 2014. URL [https://doi.org/10.1016/S1473-3099\(14\)70781-9](https://doi.org/10.1016/S1473-3099(14)70781-9).
- C. Ludeña y M. Lavielle. The Whittle estimator for strongly dependent stationary Gaussian fields. *Scandinavian Journal of Statistics*, 26:433–450, 1999.
- C. Malesios, N. Demiris, P. Kostoulas, K. Dadousis, T. Koutroumanidis, y Z. Abas. Spatio-temporal modelling of foot-and-mouth disease outbreaks. *Epidemiology and Infection*, 144:2485–2493, 2016.

-
- D. Marinucci y G. Peccati. *Random fields on the sphere. Representation, limit theorems and cosmological applications*. London Mathematical Society Lecture Note Series 389. Cambridge University Press, Cambridge, 2011.
- D. Martínez-Bello, A. López-Quílez, y A. T. Prieto. Spatiotemporal modeling of relative risk of dengue disease in colombia. *Stochastic Environmental Research and Risk Assessment*, 32:1587–1601, 2018. URL <https://doi.org/10.1007/s00477-017-1461-5>.
- C. C. McCluskey. Complete global stability for an sir epidemic model with delay distributed or discrete. *Nonlinear Analysis: Real World Applications*, 11:55–59, 2010.
- J. P. Messina, O. J. Brady, N. Golding, M. U. G. Kraemer, G. R. W. Wint, S. E. Ray, D. M. Pigott, F. M. Shearer, K. Johnson, L. Earl, L. B. Marczak, S. Shirude, N. D. Weaver, M. Gilbert, R. Velayudhan, P. Jones, T. Jaenisch, T. W. Scott, R. C. Reiner Jr, y S. I. Hay. The current and future global distribution and population at risk of dengue. *Nature Microbiology*, 4:1508–1515, 2018. URL <https://doi.org/10.1038/s41564-019-0476>.
- J. L. Meza. Empirical bayes estimation smoothing of relative risks in disease mapping. *Journal of Statistics Planning and Inference*, 112:43–62, 2003. URL [https://doi.org/10.1016/S0378-3758\(02\)00322-1](https://doi.org/10.1016/S0378-3758(02)00322-1).
- F. A. Milner y R. Zhao. Sir model with directed spatial diffusion. *Mathematical Population Studies*, 15:160–181, 2008.
- M. Mohammady, H. Reza Pourghasemi, M. Amiri, y J. P. Tiefenbacher. Spatial modeling of susceptibility to subsidence using machine learning techniques. *Stochastic Environmental Research and Risk Assessment*, 2021. URL <https://doi.org/10.1007/s00477-020-01967-x>.
- J. Møller. Properties of spatial Cox process models. *Journal of Statistical Research of Iran*, 2:1–18, 2005.
- J. Møller, F. Safavimanesh, y J. G. Rasmussen. The cylindrical k-function and poisson line cluster point processes. *Biometrika*, 103:937–954, 2016.
- J. Møller, A. R. Syversveen, y R. Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25:451–482, 1998.

-
- J. Møller y H. Toftaker. Geometric anisotropic spatial point pattern analysis and cox processes. *Scandinavian Journal of Statistics*, 41(2):414–435, 2014.
- J. Møller y R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall, 2004.
- Mukhsar., A. Bahridin, S. Asrul, C. Edi, A. Pasrun, y A. A. Farah. Extended convolution model to bayesian spatiotemporal for diagnosing the dhf endemic locations. *Journal of Interdisciplinary Mathematics*, 19:233–244, 2016. URL <https://doi.org/10.1080/09720502.2015.1047591>.
- H. G. Müller y U. Stadtmüller. Generalized functional linear models. *Annals of Statistics*, 33:774–805, 2005.
- D. B. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45:7–38, 1990.
- P. Nguyen, P. E. Brown, y J. Stafford. Mapping cancer risk in Southwestern Ontario with changing census boundaries. *Biometrics*, 68(4):1228–1237, 2012.
- H. Nishiura, N. M. Linton, y A. R. Akhmetzhanov. Serial interval of novel coronavirus (covid-19) infections. *International Journal of Infectious Diseases*, 93:284–286, 2020.
- Y. Ogata y K. Katsura. Likelihood analysis of spatial inhomogeneity for marked point patterns. *Annals of the Institute of Statistical Mathematics*, 40:29–39, 1988.
- PAHO. Pan American Health Organization. 2019. URL <https://www.paho.org/hq/index.php?lang=en>. Accedido el 01–08–2019.
- D. Pak, K. Langohr, J. Ning, J. Cortés Martínez, G. Gómez-Melis, y Y. Shen. Modeling the coronavirus disease 2019 incubation period: impact on quarantine policy. *Mathematics*, 8(9), 2020. URL <https://www.mdpi.com/2227-7390/8/9/1631>.
- V. M. Panaretos y S. Tavakoli. Cramér-karhunen-loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, 123:2779–2807, 2013a.

-
- V. M. Panaretos y S. Tavakoli. Fourier analysis of stationary time series in function space. *Annals of Statistics*, 41:568–603, 2013b.
- M. E. Pastrana, R. Brito, N. R. Romero, C. S. de Oliveira, y J. P. Haddad. Spatial and statistical methodologies to determine the distribution of dengue in brazilian municipalities and relate incidence with the health vulnerability index. *Spatial and Spatio-temporal Epidemiology*, 11:143–151, 2014. URL <https://doi.org/10.1016/j.sste.2014.04.001>.
- S. Pathak, A. Maiti, y G. Samanta. Rich dynamics of an sir epidemic model. *Nonlinear Analysis: Modelling and Control*, 15:71–81, 2010.
- G. Peccati y M. S. Taqqu. *Wiener Chaos: Moments, Cumulants and Diagrams*. Springer, New York, 2011.
- K. M. Pepin, C. B. Leach, C. Marques-Toledo, K. H. Laass, K. S. Paixao, A. D. Luis, y C. T. Webb. Utility of mosquito surveillance data for spatial prioritization of vector control against dengue viruses in three brazilian cities. *Parasites and vectors*, 8(98):1–15, 2015. URL <https://doi.org/10.1186/s13071-015-0659-y>.
- A. M. Ramos, M. R. Ferrández, M. Vela-Pérez, y B. Ivorra. A simple but complex enough θ -sir type model to be used with covid-19 real data. application to the case of italy, 2020. Preimpreso en <https://doi.org/10.13140/RG.2.2.32466.17601>.
- J. O. Ramsay y B. W. Silverman. *Functional data analysis, 2nd ed.* New York: Springer, 2005.
- S. L. Rathbun y N. Cressie. A space-time survival point process for a longleaf pine forest in Southern Georgia. *Journal of the American Statistical Association*, 89:1164–1174, 1994.
- V. Reisen, A. L. Rodrigues, y W. Palma. Estimation of seasonal fractionally integrated processes. *Computational Statistics and Data Analysis*, 50:568–582, 2006.
- A. Remuzzi y G. Remuzzi. Covid-19 and italy: what next? *The Lancet*, 2020. URL [doi.org/10.1016/S0140-6736\(20\)30690-5](https://doi.org/10.1016/S0140-6736(20)30690-5).

-
- A. C. Restrepo, P. Baker, y A. C. Clements. National spatial and temporal patterns of notified dengue cases, colombia 2007–2010. *Tropical Medicine & International Health*, 19:863–871, 2014. URL <https://doi.org/10.1111/tmi.12325>.
- M. Riedle. *Séminaire de Probabilités XLIII. Lecture Notes in Mathematics, vol 2006.*, cap. Cylindrical Wiener processes. Springer, Berlin, Heidelberg, 2011.
- B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B*, 39:172–212, 1977.
- P. M. Robinson. Nonparametric spectrum estimation for spatial data. *Journal of Statistical Planning and Inference*, 137:1024–1034, 2007.
- P. M. Robinson y J. V. Sanz. Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis*, 97:1090–1120, 2006.
- A. Rodrigues y P. J. Diggle. Bayesian estimation and prediction for inhomogeneous spatio-temporal log-gaussian cox processes using low-rank models, with application to criminal surveillance. *Journal of the American Statistical Association*, 107:93–101, 2012.
- K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. Hyman, P. Yan, y G. Chowell. Real-time forecasts of the covid-19 epidemic in china from february 5th to february 24th. *Infectious Disease Modelling*, 5:256–263, 2020.
- L. Roques y O. Bonnefon. Modelling population dynamics in realistic landscapes with linear elements: A mechanistic-statistical reaction-diffusion approach. *PloS One*, 11(3):e0151217, 2016.
- L. Roques, S. Soubeyrand, y O. Bonnefon. A statistical-reaction-diffusion approach for analyzing expansion processes. *Journal of Theoretical Biology*, 274:43–51, 2011.
- M. Rostami, Y. Mohammadi, A. Jalilian, y B. Nazparvar. Modeling spatio-temporal variations of substance abuse mortality in iran using a log-gaussian cox point process. *Spatial and Spatio-Temporal Epidemiology*, 22:15–25, 2012.
- H. Rue, S. Martino, y N. Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B*, 71:319–392, 2009.

-
- M. D. Ruiz-Medina. Spatial autoregressive and moving average Hilbertian processes. *Journal of Multivariate Analysis*, 102:292–305, 2011.
- M. D. Ruiz-Medina. Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics*, 23:119–128, 2012. doi:10.1002/env.1143.
- M. D. Ruiz-Medina. Functional analysis of variance for Hilbert-valued multivariate fixed effect models. *Statistics*, 50:689–715, 2016.
- M. D. Ruiz-Medina y J. M. Angulo. Spatio-temporal filtering using wavelets. *Stochastic Environmental Research and Risk Assessment*, 16:241–266, 2002.
- M. D. Ruiz-Medina y R. Espejo. Spatial autoregressive functional plug-in prediction of ocean surface temperature. *Stochastic Environmental Research and Risk Assessment*, 26:335–344, 2012.
- M. D. Ruiz-Medina y R. Espejo. Integration of spatial functional interaction in the extrapolation of ocean surface temperature anomalies due to global warming. *Journal of Applied Earth Observation and Geoinformation*, 22:27–39, 2013.
- M. D. Ruiz-Medina, R. M. Espejo, M. D. Ugarte, y A. F. Militino. Functional time series analysis of spatio-temporal epidemiological data. *Stochastic Environmental Research and Risk Assessment*, 28:943–954, 2014.
- M. D. Ruiz-Medina, E. Romano, y R. Fernández-Pascual. Plug-in interval prediction for a special class of standard ARH(1) processes. *Journal of Multivariate Analysis*, 146:138–150, 2016.
- L. Sakhno. Bias control in the estimation of spectral functionals. *Theory of Stochastic Processes*, 13:225–233, 2007.
- L. Sakhno. *Minimum Contrast Method for Parameter Estimation in the Spectral Domain*, págs. 319–336. Springer International Publishing, Cham, 2014. ISBN 978-3-319-03512-3. URL http://dx.doi.org/10.1007/978-3-319-03512-3_18.
- S. Salap-Ayca y P. Jankowski. Analysis of the influence of parameter and scale uncertainties on a local multi-criteria land use evaluation model. *Stochastic Environmental Research and Risk Assessment*, 32:699–2719, 2018.

-
- M. Sekiguchi y E. Ishiwata. Global dynamics of a discretized sirs epidemic model with time delay. *Journal of Mathematical Analysis and Applications*, 371:195–202, 2010.
- L. Serra, M. Saez, J. Mateu, D. Varga, P. Juan, C. Díaz-Avalos, y H. Rue. Spatio-temporal log-gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994-2008. *Environmental and Ecological Statistics*, 21(3):531–563, 2014.
- D. Simpson, J. B. Illian, F. Lindgren, Sorbye S. H., y Rue H. Going off grid: computationally efficient inference for log-gaussian cox processes. *Biometrika*, 103(1):49–70, 2016.
- B. Sivakumar. Covid-19 and water. *Stochastic Environmental Research and Risk Assessment*, 2020. URL <https://doi.org/10.1007/s00477-020-01837-6>.
- A. Sánchez-Gómez, C. Amela, E. Fernández-Carrión, M. Martínez-Avilés, JM. Sánchez-Vizcaíno, y M. J. Sierra-Moros. Risk mapping of west nile virus circulation in spain, 2015. *Acta Tropica*, 169:163–169, 2017. URL <https://doi.org/10.1016/j.actatropica.2017.02.022>.
- A. R. Soltani y M. Hashemi. Periodically correlated autoregressive Hilbertian processes. *Statistical Inference for Stochastic Processes*, 14:177–188, 2011.
- D. Stoyan, W. S. Kendall, y J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, 1995.
- R. Sujath, J. M. Chatterjee, y A. E. Hassanien. A machine learning forecasting model for covid-19 pandemic in india. *Stochastic Environmental Research and Risk Assessment*, 34:959–972, 2020.
- Y. Takano y R. Miyashiro. Best subset selection via cross-validation criterion. *TOP*, 28:475–488, 2020.
- M. Taniguchi. Minimum contrast estimation for spectral densities of stationary processes. *Journal of the Royal Statistical Society: Series B*, 49:315–325, 1987.
- B. M. Taylor, R. Andrade-Pacheco, y H. J. Sturrock. Continuous inference for aggregated point process data. *Journal of the Royal Statistical Society: Series A*, 181:1125–1150, 2018.

-
- E. Tornatore, S. M. Buccellato, y P. Vetro. Stability of a stochastic sir system. *Physica A: Statistical Mechanics and its Applications*, 354:111–126, 2005.
- A. Torres, M. P. Frías, y M. D. Ruiz-Medina. Log-gaussian Cox processes in infinite-dimensional spaces. *Theory of Probability and Mathematical Statistics*, 95:157–177, 2016. URL <https://doi.org/10.1090/tpms/1028>.
- A. Torres-Signes y J. A. Dip. A Bayesian functional methodology for dengue risk mapping in Latin America and the Caribbean. *Acta Tropica*, 215:105788, 2021. URL <https://doi.org/10.1016/j.actatropica.2020.105788>.
- A. Torres-Signes, M. P. Frías, J. Mateu, y M. D. Ruiz-Medina. A spatial functional count model for heterogeneity analysis in time. *Stochastic Environmental Research and Risk Assessment*, 2021a. URL <https://doi.org/10.1007/s00477-020-01951-5>.
- A. Torres-Signes, M. P. Frías, y M. D. Ruiz-Medina. Covid–19 mortality analysis from soft-data multivariate curve regression and machine learning, 2021b. Preimpreso en <https://arxiv.org/pdf/arXiv:2008.06344.pdf>.
- H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland, Amsterdam, 1978.
- M. D. Ugarte, T. Goicoa, J. Etxeberria, y A. F. Militino. A P-spline ANOVA type model in space-time disease mapping. *Stochastic Environmental Research and Risk Assessment*, 26(6):835–845, 2012.
- M. D. Ugarte, T. Goicoa, B. Ibáñez, y A. F. Militino. Evaluating the performance of spatio–temporal Bayesian models in disease mapping. *Environmetrics*, 20(6):647–665, 2009a.
- M. D. Ugarte, T. Goicoa, y A. F. Militino. Empirical Bayes and fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics and Data Analysis*, 53:2938–2949, 2009b.
- M. D. Ugarte, T. Goicoa, y A. F. Militino. Spatio–temporal modelling of mortality risks using penalized splines. *Environmetrics*, 21(3/4):270–289, 2010.
- G. Vicente, T. Goicoa, y M. D. Ugarte. Bayesian inference in multivariate spatio-temporal areal models using inla: analysis of gender-based violence in small

-
- areas. *Stochastic Environmental Research and Risk Assessment*, 34:1421–1440, 2020.
- E. Volz. Sir dynamics in random networks with heterogeneous connectivity. *Journal of Mathematical Biology*, 56:293–310, 2008.
- R. Waagepetersen y D. Y. Guan. Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society: Series B*, 71(3):685–702, 2009.
- R. Waagepetersen, Y. Guan, A. Jalilian, y J. Mateu. Analysis of multispecies point patterns by using multivariate log-gaussian cox processes. *Journal of the Royal Statistical Society: Series C*, 65:77–96, 2016.
- C. Wang, P. W. Horby, F. Hayden, y G. F. Gao. A novel coronavirus outbreak of global health concern. *Lancet*, 395:470–473, 2020.
- R. K. Wasiur, B. Choiy, E. Kenahz, y G. A. Rempa. Survival dynamical systems for the population-level analysis of epidemics. Url <https://arxiv.org/pdf/1901.00405v1.pdf>, 2019. Accedido el 10-07-2020.
- G. Webb. A reaction-diffusion model for a deterministic diffusive epidemic. *Journal of Mathematical Analysis and Applications*, 84:150—161, 1981.
- G. Wei, P. Clifford, y J. Feng. Population death sequences and Cox processes driven by interacting Feller diffusions. *Journal of Physics A: Mathematical and General*, 35:9–31, 2002.
- X. WeiLin, Z. WeiGuo, y Z. XiLi. Minimum contrast estimator for fractional Ornstein-Uhlenbeck processes. *Science China Mathematics*, 55(7):1497–1511, 2012.
- S. Wu, H. G. Müller, y Z. Zhang. Functional data analysis for point processes with rare events. *Statistica Sinica*, 23:1–23, 2013.
- Y. Xu, L. Allena, y A. Perelson. Stochastic model of an influenza epidemic with drug resistance. *Journal of Theoretical Biology*, 248:179–193, 2007.
- Q. Yao y P. J. Brockwell. Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli*, 12(3):403–429, 2006.

-
- H. Yu, J. M. Angulo, M. Cheng, J. Wu, y G. Christakos. An online spatio temporal prediction model for dengue fever epidemic in kaohsiung (taiwan). *Biometrical Journal*, 56:428–440, 2014. URL <https://doi.org/10.1002/bimj.201200270>.
- H. L. Yu y G. Christakos. Modeling and estimation of heterogeneous spatio-temporal attributes under conditions of uncertainty. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1):366–376, 2011.
- J. Yu, D. Jiang, y N. Shi. Global stability of two-group sir model with random perturbation. *Journal of Mathematical Analysis and Applications*, 360:235–244, 2009.
- Y. R. Yue y J. M. Loh. Bayesian nonparametric estimation of pair correlation function for inhomogeneous spatial point processes. *Journal of Nonparametric Statistics*, 25(2):463–474, 2013.
- F. Zhang, Z. Li, y F. Zhang. Global stability of an sir epidemic model with constant infectious period. *Applied Mathematics and Computation*, 199:285–291, 2008.
- W.-b. Zhang, Y. Ge, M. Liu, P. M. Atkinson, J. Wang, Zhang X., y Z. Tian. Risk assessment of the step-by-step return-to-work policy in beijing following the covid-19 epidemic peak. *Stochastic Environmental Research and Risk Assessment*, 2020. URL <https://doi.org/10.1007/s00477-020-01929-3>.
- F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, Ji. Xu, S. Tu, Y. Zhang, H. Chen, y B. Cao. Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The Lancet*, 2020. URL [doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3).
- T. Zhou, Z. Fu, y B. Wang. Epidemic dynamics on complex networks. *Progress in Natural Science*, 16:452–457, 2006.
- G. Zhu, J. Liu, Q. Tan, y B. Shi. Inferring the spatio-temporal patterns of dengue transmission from surveillance data in guangzhou, china. *PLoS neglected tropical diseases*, 10(4):1–20, 2016. URL <https://doi.org/10.1371/journal.pntd.0004633>.

