







# Modelización de la función de covarianza en procesos espacio-temporales: análisis y aplicaciones

UNIVERSITAT DE VALÈNCIA  
REGISTRE GENERAL  
ENTRADA

24 ENE. 2008

N.º.....5369.....  
HORA.....  
OFICINA AUXILIAR NÚM. 17

UNIVERSITAT DE VALÈNCIA  
BIBLIOTECA CIÈNCIES

N.º Registro .....

DATA 29 - II - 2008

SIGNATURA TD-11 257

N.º LITIG: j20486637

*24 ene*

Memoria presentada por  
Francisco Martínez Ruiz  
para optar al grado de Doctor

Dirigida por  
Dr. Jorge Mateu Mahiques  
Dr. Francisco Montes Suay



UMI Number: U603104

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U603104

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346



Jorge Mateu Mahiques, Catedrático de Estadística e Investigación Operativa del Departament de Matemàtiques de la Universitat Jaume I de Castelló, Francisco Montes Suay, Catedrático de Estadística e Investigación Operativa del Departament d'Estadística i Investigació Operativa de la Universitat de València,

CERTIFICAN que la presente memoria, *Modelización de la función de covarianza en procesos espacio-temporales: análisis y aplicaciones*, ha sido realizada bajo su dirección en el Departament de Matemàtiques y Departament d'Estadística i Investigació Operativa por Don Francisco Martínez Ruiz y constituye su Tesis para optar al Grado de Doctor en Matemáticas.

Y para que conste, en cumplimiento de la legislación vigente, la presenta ante la Facultat de Matemàtiques de la Universitat de València, a 24 de febrero de 2008.



Jorge Mateu Mahiques



Francisco Montes Suay



## Agradecimientos

Esta memoria es el fruto de la investigación llevada a cabo durante los últimos años en el Departament de Matemàtiques de la Universitat Jaume I de Castelló y en el Departament d'Estadística i Investigació Operativa de la de la Universitat de València.

En primer lugar quiero mostrar mi más sincera gratitud a mis tutores, Jorge Mateu y Paco Montes, sin cuya dirección y constante apoyo no habría sido posible realizar este proyecto. Como fruto de todos estos años no sólo ha quedado esta memoria de investigación, sino una profunda y sincera amistad de la que estoy todavía más orgulloso. También quiero agradecer al profesor Noel Cressie su hospitalidad y amistad durante mi estancia en la Ohio State University para asistir a su curso de doctorado Spatial Statistics.

Por último, quiero dirigir una especial dedicatoria a mi familia. A mis padres, por su apoyo y sacrificio durante todos estos años. Y en especial a vosotras, Maite y Lucía, por vuestro cariño y paciencia, especialmente durante estos años de duro trabajo.



# Índice general

<b>1. Introducción</b>	<b>15</b>
1.1. Objetivo y estructura . . . . .	19
<b>2. Análisis de datos geoestadísticos</b>	<b>21</b>
2.1. Introducción . . . . .	21
2.2. Definiciones básicas . . . . .	22
2.2.1. Estacionariedad e isotropía . . . . .	23
2.2.2. El covariograma . . . . .	26
2.2.3. El variograma . . . . .	29
2.2.4. El correlograma . . . . .	32
2.2.5. Forma general de estas funciones . . . . .	32
2.3. Estimación del variograma y del covariograma . . . . .	33
2.3.1. Estimador clásico . . . . .	33
2.3.2. Estimador robusto . . . . .	35
2.4. Principales modelos de variogramas y covariogramas isotrópicos . . . . .	35
2.4.1. Modelo de efecto pepita . . . . .	36
2.4.2. Modelo lineal . . . . .	36
2.4.3. Modelo esférico . . . . .	37
2.4.4. Modelo exponencial . . . . .	37
2.4.5. Modelo gaussiano . . . . .	38
2.4.6. Modelo racional cuadrático . . . . .	38
2.4.7. Modelo ondulado . . . . .	39
2.4.8. Modelo potencial . . . . .	39
2.4.9. Modelo de Matérn . . . . .	39
2.4.10. Combinación convexa de modelos. Modelos con efecto pepita . . . . .	40
2.4.11. Combinación lineal (no necesariamente convexa) de modelos . . . . .	42
2.5. Estimación de los parámetros del variograma . . . . .	42
2.5.1. Estimación por mínimos cuadrados . . . . .	43
2.5.2. Estimación mediante máxima verosimilitud . . . . .	43
2.5.3. Estimación mediante verosimilitud compuesta ponderada . . . . .	45
2.6. Predicción espacial . . . . .	47
2.6.1. Kriging ordinario . . . . .	49
2.6.2. Kriging universal . . . . .	53

---

2.6.3. Kriging suavizado mediante la mediana . . . . .	55
2.7. Comprobación de las hipótesis del modelo . . . . .	55
2.7.1. Técnicas exploratorias . . . . .	55
2.7.2. Test de independencia . . . . .	57
2.7.3. Test de validación de modelos de variograma . . . . .	59
2.7.4. Diagnóstico mediante validación cruzada . . . . .	60
<b>3. Geoestadística espacio-temporal . . . . .</b>	<b>63</b>
3.1. Introducción . . . . .	63
3.2. Definiciones . . . . .	65
3.3. Propiedades de la función de covarianza y del variograma espacio-temporal . . . . .	71
3.4. Estimación del variograma y del covariograma . . . . .	76
3.5. Modelización de procesos espacio-temporales . . . . .	77
3.5.1. El filtro de Kalman . . . . .	79
3.6. Ajuste y evaluación de los modelos espacio-temporales . . . . .	82
3.6.1. Ajuste mediante mínimos cuadrados . . . . .	82
3.6.2. Ajuste mediante máxima-verosimilitud . . . . .	82
3.6.3. Ajuste mediante verosimilitud compuesta ponderada . . . . .	84
3.6.4. Validación y comparación de modelos . . . . .	84
3.7. Predicción . . . . .	86
3.7.1. Kriging simple . . . . .	88
3.7.2. Kriging ordinario . . . . .	89
3.7.3. Kriging universal . . . . .	89
3.8. Análisis en presencia de no estacionariedad . . . . .	90
<b>4. Modelos de covarianza espacio-temporales . . . . .</b>	<b>93</b>
4.1. Introducción . . . . .	93
4.2. Modelo métrico . . . . .	96
4.3. Modelo suma . . . . .	97
4.4. Modelo producto . . . . .	99
4.5. Modelos propuestos por Cressie y Huang . . . . .	101
4.6. Modelo suma-producto . . . . .	105
4.7. Modelos propuestos por Gneiting . . . . .	109
4.8. Modelos producto y suma-producto integrado . . . . .	114
4.9. Modelos de mixturas propuestos por Ma . . . . .	119
4.9.1. Generación mediante mixturas escalares . . . . .	120
4.9.2. Generación mediante mixturas de potencias positivas . . . . .	120
4.10. Modelos generados mediante combinaciones lineales propuestos por Ma . . . . .	123
4.11. Modelos propuestos por Stein . . . . .	126
4.12. Modelo suma de productos generalizado . . . . .	128
4.13. Modelos que no satisfacen la simetría completa . . . . .	146
4.14. Modelos no estacionarios . . . . .	146
4.14.1. Mixtura de procesos estacionarios localmente ortogonales . . . . .	147
4.14.2. Modelos no estacionarios propuestos por Ma . . . . .	148

4.14.3. Modelos no estacionarios propuestos por Porcu y Mateu . . . . .	151
<b>5. Comparación de modelos mediante simulación</b>	<b>153</b>
5.1. Introducción . . . . .	153
5.2. Modelos espacio-temporales considerados . . . . .	155
5.2.1. Modelo espacio-temporal dinámico (Huang y Cressie, 1996) . . . . .	156
5.2.2. Modelo espacio-temporal no separable (Cressie y Huang, 1999) . . . . .	157
5.2.3. Modelo espacio-temporal no separable (Gneiting, 2002) . . . . .	158
5.2.4. Modelo suma-producto (De Cesare et al., 2001a) . . . . .	159
5.2.5. Modelo suma de productos generalizado . . . . .	160
5.3. Estudio de simulación . . . . .	161
5.4. Resultados . . . . .	165
5.5. Conclusiones . . . . .	173
<b>6. Modelización de datos reales</b>	<b>177</b>
6.1. Introducción . . . . .	177
6.2. Análisis espacio-temporal del acuífero del cuaternario de Xàbia . . . . .	178
6.3. A comparative analysis of different spatial sampling schemes: Modelling of SSRB data . . . . .	187
6.4. Model comparison and selection for stationary space-time models . . . . .	200
6.4.1. Modelización de la SSRB con el modelo suma de productos generalizado . . . . .	220
6.5. Spatiotemporal modeling and prediction of solar radiation . . . . .	221
6.6. Dynamic Life Tables: a Geostatistical Approach . . . . .	234
<b>7. Resumen y futuras líneas de investigación</b>	<b>251</b>
7.1. Resumen . . . . .	251
7.2. Futuras líneas de trabajo . . . . .	252
<b>Bibliografía</b>	<b>255</b>



# Índice de figuras

2.1. Forma general del variograma y covariograma de un proceso espacial homogéneo . . . . .	33
2.2. Representación gráfica de las principales familias de semivariogramas . . . . .	41
3.1. Relaciones entre los diferentes tipos de funciones de covarianza espacio-temporales. . . . .	69
4.1. Diferentes representaciones del modelo métrico (4.3). . . . .	97
4.2. Diferentes representaciones del modelo suma (4.5). . . . .	98
4.3. Diferentes representaciones del modelo producto (4.9). . . . .	101
4.4. Diferentes representaciones del modelo de Cressie y Huang (4.20). . . . .	106
4.5. Diferentes representaciones del modelo suma-producto (4.29). . . . .	110
4.6. Diferentes representaciones del modelo de Gneiting (4.32). . . . .	115
4.7. Ejemplo del modelo suma de productos generalizado (4.77) con covarianzas negativas. . . . .	144
4.8. Diferentes representaciones del modelo suma de productos generalizado (4.77). . . . .	145
5.1. Intervalo de confianza al 95 % para la media de los valores del <i>EPCM</i> basado en las 100 réplicas de datos generados con el Modelo I para las 16 combinaciones de parámetros utilizadas. . . . .	165
5.2. Intervalo de confianza al 95 % para la media de los valores del <i>EPCM</i> basado en las 100 réplicas de datos generados con el Modelo II para las 16 combinaciones de parámetros utilizadas. . . . .	166
5.3. Intervalo de confianza al 95 % para la media de los valores del <i>EPCM</i> basado en las 100 réplicas de datos generados con el Modelo III para las 16 combinaciones de parámetros utilizadas. . . . .	167
5.4. Intervalo de confianza al 95 % para la media de los valores del <i>EPCM</i> basado en las 100 réplicas de datos generados con el Modelo IV para las 16 combinaciones de parámetros utilizadas. . . . .	168
5.5. Intervalo de confianza al 95 % para la media de los valores del <i>EPCM</i> basado en las 100 réplicas de datos generados con el Modelo V para las 16 combinaciones de parámetros utilizadas. . . . .	169
5.6. Ejemplo de función de covarianza teórica del Modelo I y funciones ajustadas con los cinco modelos considerados. . . . .	171



5.7. Ejemplo de función de covarianza teórica del Modelo II y funciones ajustadas con los cinco modelos considerados. . . . .	171
5.8. Ejemplo de función de covarianza teórica del Modelo III y funciones ajustadas con los cinco modelos considerados. . . . .	172
5.9. Ejemplo de función de covarianza teórica del Modelo IV y funciones ajustadas con los cinco modelos considerados. . . . .	172
5.10. Ejemplo de función de covarianza teórica del Modelo V y funciones ajustadas con los cinco modelos considerados. . . . .	173

# Índice de tablas

4.1. Ejemplos de funciones completamente monótonas. . . . .	111
4.2. Ejemplos de funciones de Bernstein. . . . .	112
4.3. Valores de $m_t$ , $M_t$ , $m_e$ y $M_e$ para $n = 2$ en el caso en que se han combinado funciones de covarianza gaussianas y de Matérn en función de sus parámetros. . . . .	142
5.1. Media del EPCM para todas las simulaciones y todas las combinaciones de cada uno de los modelos. . . . .	170
6.1. Media y error estándar del EPCM para los cinco modelos ajustados a los datos de SSRB sobre las 50 rejillas . . . . .	220



# Capítulo 1

## Introducción

Hasta hace bien poco, en la mayoría de problemas reales con los que nos enfrentábamos se suponía como hipótesis de partida que las observaciones del fenómeno objeto de estudio estaban tomadas bajo idénticas condiciones y que además cada observación se tomaba de forma independiente al resto. Los datos observados bajo esta hipótesis se dice que forman una muestra aleatoria del proceso objeto de estudio, es decir, son independientes e idénticamente distribuidos. En esta situación, podemos aplicar procedimientos de estadística clásica para modelizar el proceso y estimar los parámetros del mismo a partir de los datos observados.

No obstante, en la mayoría de fenómenos físicos presentes en la naturaleza esta hipótesis de partida es poco realista, pues las observaciones realizadas no satisfacen la condición de independencia sino que existe una dependencia entre ellas presente en todas las direcciones, que se hace más débil a medida que las localizaciones de los datos están más dispersas. Por ejemplo, si estamos interesados en estudiar la contaminación por  $SO_2$  en una determinada ciudad a partir de observaciones realizadas en un conjunto de estaciones de medida ubicadas en la misma, esperamos que las medidas de dos estaciones próximas entre sí sean más parecidas que las de dos estaciones distantes.

El uso de modelos de correlación entre observaciones, en función de la distancia existente entre ellas es una técnica ampliamente utilizada en el análisis de series temporales, donde tenemos un conjunto de medidas de un determinado fenómeno tomadas en diferentes tiempos (o localizaciones temporales) que no consideramos como una muestra aleatoria, sino que se modelizan mediante un modelo de correlación entre las observaciones que depende de la distancia (temporal) entre ellas.

La extensión de estas técnicas al ámbito espacial es todavía reciente, ya que hemos de tener en cuenta que las primeras herramientas de la estadística espacial fueron introducidas hacia los años 50. Su aplicación en la última década del siglo XX ha supuesto importantes avances en ciencias como la epidemiología, la geología, la ecología, la astronomía el tratamiento de imágenes, la criminología, la climatología y, en general, en todas aquellas disciplinas que trabajen con fenómenos en los que la influencia de la localización de los valores debe ser tomada en cuenta.

En todas estas ciencias y disciplinas científicas, la necesidad de analizar características

de procesos espaciales es clara y evidente. Por ejemplo, en cualquier análisis estadístico riguroso en ciencias ambientales puede decirse que la determinación de la estructura espacial subyacente es necesaria. El reto de la investigación estadística en la última década ha estado centrado en el desarrollo, comprobación y evaluación de los métodos más apropiados para la cuantificación de la dependencia espacial de los datos.

La generalización de estas técnicas a procesos en los que, además de tener en cuenta su dependencia espacial y temporal, se considera también su evolución espacio-temporal de forma conjunta, ha dado lugar a la definición de modelos espacio-temporales, de mucha mayor complejidad. La modelización espacio-temporal es un campo de estudio de gran desarrollo en la actualidad por las mejoras que puede suponer respecto a los análisis meramente espaciales y temporales del proceso.

La mayoría de la teoría actual sobre estadística espacial ha sido desarrollada a partir del trabajo pionero en la industria de la minería, que se proponía resolver el problema de la estimación de la cantidad de mineral en una región a partir de las concentraciones de dicho mineral en un conjunto finito de localizaciones muestreadas. Estos trabajos, cuyo origen cabe situar en Sudáfrica, alcanzaron pleno desarrollo en la Escuela de Minas de Fontainebleau, dando lugar al término *geoestadística*, con el que se ahora se designa la ciencia que estudia los fenómenos que fluctúan en el espacio y/o tiempo, ofreciendo una colección de herramientas estadísticas para la descripción y modelización de la variabilidad espacial (y temporal). Las aplicaciones actuales de la geoestadística en problemas medioambientales van mucho más allá de sus usos originales en la industria minera, aunque mucha de la terminología clásica definida en sus inicios sigue siendo ampliamente usada hoy en día. Uno de los trabajos pioneros en este campo fue el llevado a cabo por el ingeniero de minas sudafricano Krige (1951), que presentó las ecuaciones básicas para lograr una interpolación lineal óptima con variables espacialmente correladas. Su nombre fue inmortalizado por Matheron (1962, 1963, 1971) en una serie de libros y trabajos, en los que introdujo el término francés *krigeage* o su equivalente en inglés *kriging* para designar la técnica propuesta por Krige.

Un aproximación diferente a la estadística espacial surgió con el estudio de fenómenos relacionados con la agricultura. Las variables implicadas, bajo ningún concepto se podían suponer independientes (Fairfield Smith, 1938), lo que condujo al desarrollo de nuevos métodos estadísticos (Papadakis, 1937; Bartlett, 1938). Esto conllevó nuevas aproximaciones dadas por Whittle (1954) y Bartlett (1976, 1978). El desarrollo moderno de los métodos espaciales surgió de la mano de Besag (1974).

No es necesario un profundo análisis de la aplicabilidad de este tipo de técnicas, para darse cuenta de su importancia para la resolución de problemas reales en muy diversos contextos. La breve relación que sigue, que no pretende ser exhaustiva, recoge algunos trabajos aparecidos en la literatura referidos a los más diversos contextos científicos, con especial mención a los relacionados con problemas medioambientales.

Chilès (1976), Verly y Sullivan (1985) o Chilès y Liao (1993) son algunos ejemplos de trabajos relacionados con estudios *geoestadísticos en minería*. Trabajos relacionados con estudios *geoestadísticos del suelo* son Burgess y Webster (1980a,b,c, 1981), Tillotson y Nielsen (1984), Trangmar (1985), Cressie y Horton (1987), Oliver (1987), Haas (1990), Laslett y McBratney (1990), Webster y Oliver (1990), Goovaerts (1992), Atteia et al.

(1994), Goovaerts y Webster (1994), Goovaerts (1994), Russo et al. (1994), Webster et al. (1994), Goovaerts y Journel (1995), Haas (1995), Haas et al. (1995), Persicani (1995), Jacques et al. (1996), Vanderborcht et al. (1996), Smith y Williams (1996), Chien et al. (1997), Goovaerts (1997b), Kanesvskey et al. (1997), Juang y Lee (1998) o Papritz y Moyeed (1998). Entre los trabajos relacionados con estudios *geoestadísticos sobre hidrología* podemos citar Andrieu et al. (1989), Azimi-Zonooz et al. (1989), De Kwaadsteniet (1990), Seo et al. (1990a,b), Kitanidis (1983, 1997), Bardossy et al. (1997) o Dirks et al. (1998), entre otros. En particular, son abundantes las referencias que analizan la distribución espacial de las precipitaciones (Seo et al., 1990a,b; Savaliev et al., 1998).

El paradigma básico de la geoestadística predictiva es la caracterización del valor desconocido  $z$  como una variable aleatoria  $Z$ , siendo la incertidumbre asociada a  $z$  definida por la correspondiente distribución de probabilidad. A lo largo de este trabajo, y siguiendo la notación habitual en este contexto, denotaremos por mayúsculas las variables aleatorias, y por minúsculas los correspondientes valores muestreados o conocidos. En el contexto de la geoestadística, la variable aleatoria  $Z$  muestra una importante dependencia de la localización espacial, por lo que será denotada por  $Z(\mathbf{s})$ . Es conveniente hacer notar que en este momento  $\mathbf{s}$  puede denotar tanto un vector de localizaciones espaciales como una única coordenada espacial. Ejemplos de cantidades o medidas reales que pueden ser modelizadas por variables aleatorias dependientes de la localización espacial incluyen, tanto propiedades petrofísicas (porosidad, permeabilidad, concentración de metales o polución) como geográficas (elevaciones topográficas, densidades de población). También encontramos ejemplos de variables aleatorias discretas como propiedades geológicas, clases de insectos o especies de exfoliados.

La función de distribución acumulada (cdf) de  $Z(\mathbf{s})$  viene definida por

$$F(\mathbf{s}; z) = P\{Z(\mathbf{s}) \leq z\}. \quad (1.1)$$

En la práctica es habitual trabajar con información referida a  $n$  datos (posiblemente vecinos), por lo que la variable aleatoria suele denotarse por  $Z(\mathbf{s}_i) = z(\mathbf{s}_i)$ ,  $i = 1, \dots, n$ . En este caso, podemos trabajar con la función de distribución acumulada condicional (ccdf) dada por

$$F(\mathbf{s}; z | n) = P\{Z(\mathbf{s}) \leq z | n\}. \quad (1.2)$$

En geoestadística es importante modelizar el grado de correlación o dependencia entre una cierta variable  $Z(\mathbf{s}_i)$ ,  $i = 1, \dots, n$  y posibles covariables consideradas en otros puntos de muestreo  $Y(\mathbf{v}_j)$ ,  $j = 1, \dots, n$ . El concepto de función aleatoria permite tal modelización. Una función aleatoria define un grupo de variables aleatorias definidas sobre el campo de interés.

Buenas referencias bibliográficas en el contexto de la teoría de la probabilidad y variables aleatorias son Anderson (1984), DeGroot (1989) o Tanner (1996), entre otras muchas. Asimismo, referencias básicas sobre modelización geoestadística, que cualquier lector debe disponer para una correcta, completa y necesaria información sobre este tema, pueden ser Journel y Huijbregts (1978), Isaaks y Srisvastava (1989), Cressie (1993), Goovaerts (1997a), Wackernagel (1998), Chilès y Delfiner (1999) o Stein (1999).

Tanto en el caso de una cdf univariante como multivariante, para el cálculo de distintos estadísticos (por ejemplo, algunos de sus momentos como la media, varianza o covarianzas) sería necesario disponer de muestreos repetitivos. Por ejemplo, muestreos repetitivos de la variable  $Z(\mathbf{s})$  serían en teoría necesarios para evaluar la cdf empírica  $F(\mathbf{s}; z) = P\{Z(\mathbf{s}) \leq z\}$ . Sin embargo, en la mayoría de las aplicaciones sólo dispondremos de un muestreo en cada localización espacial  $\mathbf{s}$ , por lo que necesitaremos hacer uso de alguna condición de repetibilidad, que quedará garantizada bajo la hipótesis de estacionariedad. Una correcta decisión sobre la hipótesis de estacionariedad es esencial para una buena representatividad de las herramientas geoestadísticas. La estacionariedad es una propiedad intrínseca a los modelos; así, la propiedad de la estacionariedad puede cambiar si la escala de estudio cambia o si una mayor cantidad de datos está disponible. Si el fin del estudio es global, las características locales pueden obtenerse mediante medias, consecuentemente a mayor número de datos, mayor será la precisión obtenida por parte de las técnicas estadísticas.

Una variable aleatoria  $Z(\mathbf{s})$ ,  $\mathbf{s} \in D$ , se dice que es *estacionaria*, o que tiene la propiedad de la estacionariedad, en la región  $D$ , si la cdf multivariante es invariante bajo cualquier traslación  $C$  efectuada sobre las localizaciones, esto es:

$$F(\mathbf{s}_1, \dots, \mathbf{s}_k; z_1, \dots, z_k) = F(\mathbf{s}_1 + C, \dots, \mathbf{s}_k + C; z_1, \dots, z_k) \quad (1.3)$$

para cualquier vector de traslación  $C$ .

En lo que respecta a la modelización espacio-temporal, se ha manifestado en los últimos años una gran demanda de modelos suficientemente realistas que describan la evolución de procesos medioambientales en el espacio y tiempo, modelos que puedan capturar simultáneamente el comportamiento de ambas componentes. Si éstas fueran analizadas de forma separada, se corre el riesgo de obviar información importante.

Considerado el estudio simultáneo de ambas componentes de importancia crucial, hay que decir que el acercamiento científico a este problema es variable dependiendo del punto de vista y origen del investigador. Por ejemplo, es de destacar la dicotomía entre modelización determinista o estocástica. Así, la predicción en climatología ha sido tradicionalmente tratada como un problema determinista. Sin embargo, y siguiendo a Gneiting (2003), para una adecuada predicción determinista hace falta conocer perfectamente el estado presente de la atmósfera y las leyes físicas involucradas en estos procesos. Pero la realidad es que la incertidumbre juega un papel importante: tramas incompletas de observaciones, errores en las medidas y localizaciones, conocimiento incompleto de las leyes físicas, etc. Por tanto, está claro que probabilistas y estadísticos pueden jugar un papel importante en este contexto, puesto que los procedimientos estadísticos representan una adecuada alternativa para tratar convenientemente la incertidumbre.

La geoestadística espacio-temporal hace referencia al conjunto de técnicas geoestadísticas que analizan, describen y modelizan procesos espaciales con evolución temporal. Como es sabido, los procedimientos de interpolación basados en kriging dependen de la elección de la autocovarianza asociada al campo espacio-temporal. Por tanto, la perspectiva geoestadística se basa en la obtención de covarianzas espacio-temporales permisibles que analicen de forma adecuada las interacciones espacio-tiempo. En otras palabras, necesitamos modelos de covarianza espacio-temporal no separables asociados a campos aleatorios

estacionarios y no estacionarios. Este ha sido y es actualmente uno de los retos más importantes para la comunidad estadística que trabaja en este campo científico.

En la literatura moderna podemos encontrar diferentes aproximaciones a la construcción de covarianzas no separables. Así, por ejemplo, Kyriakidis y Journel (1999) se centraron en extender los métodos espaciales o temporales al caso espacio-temporal, considerando dicha dependencia de forma separada. Estas extensiones pueden clasificarse en dos principales categorías: (a) las que extienden directamente los modelos multivariantes espaciales o temporales, y (b) las que consideran un proceso espacio-temporal univariante.

Dimitrakopoulos y Luo (1994) propusieron algunos modelos geoméricamente anisotrópicos para la situación espacio-temporal. Guttorp et al. (1994) propusieron covarianzas separables obtenidas a través del producto tensorial de una covarianza espacial y de otra temporal. En este contexto de separabilidad, Rouhani y Hall (1989) propusieron el variograma suma. Otros métodos, como el de Sampson y Guttorp (1992), se basaron en el hecho de que la tendencia pudiera recoger completamente la variabilidad temporal para obtener así un proceso puramente espacial. En el contexto no separable, los esfuerzos se han centrado en obtener clases generales de covarianzas espacio-temporales, entre ellos podemos citar a Jones y Zhang (1997), Cressie y Huang (1999), Christakos (2000), De Cesare et al. (2001b), Gneiting (2002), Ma (2002, 2005c), Stein (2005b) o Fernández-Casal et al. (2003). La mayoría de estas contribuciones tratan covarianzas estacionarias suponiendo isotropía en el espacio y tiempo. En particular, Cressie y Huang (1999) propusieron un método espectral para obtener covarianzas espacio-temporales, y Gneiting (2002) generalizó este método mediante el uso de funciones completamente monótonas y funciones de Bernstein. Stein (2005b) pone énfasis en el método espectral y en el hecho de que la densidad espectral espacio-temporal debe ser suficientemente suave lejos del origen. Finalmente, Fernández-Casal et al. (2003) extienden el método de Shapiro-Botha para dar lugar a variogramas espacio-temporales flexibles. En el caso no estacionario, podemos encontrar importantes contribuciones de la mano de Christakos (2002), Fuentes y Smith (2001), Fuentes (2002b) o Kolovos et al. (2004), entre otros.

A pesar de contar con un buen número de referencias bibliográficas en la propuesta de nuevos modelos de covarianzas espacio-temporales, queda un largo camino por recorrer en el análisis de la capacidad predictiva de muchos de estos nuevos modelos. Y todavía más, qué hacer cuando son varios los modelos alternativos que a priori ajustan a un mismo banco de datos. Qué decisión tomar, y en qué fundamentos científicos podemos basarnos para dicha selección, son preguntas usuales e importantes que un investigador debe afrontar ante la modelización práctica. Y es en estas líneas en la que nuestro trabajo pretende aportar alguna respuesta.

## 1.1. Objetivo y estructura

El objetivo del trabajo llevado a cabo a lo largo de estos años, y que se recoge en la presente memoria, es doble. De una parte contribuir a la librería de funciones de modelos aportando un nuevo modelo de función de covarianza espacio-temporal; de otra una presentación del estado del arte por lo que respecta a las funciones de covarianza espacio-

temporal, completado con un exhaustivo trabajo de comparación de los diferentes modelos que pensamos ha de ser de gran utilidad para los investigadores en este campo, particularmente a la hora de aplicar los modelos a situaciones reales. Nosotros mismos hemos hecho uso de estos resultados en algunas de las aplicaciones, ya publicadas, que se reproducen en uno de los capítulos. De acuerdo con cuanto se ha dicho, hemos estructurado la memoria de la siguiente forma:

**Capítulo 2.-** Describe brevemente los principales conceptos y resultados utilizados en el análisis de datos espaciales, muchos de los cuales se generalizarán en los capítulos siguientes con el fin de poder aplicarlos a procesos espacio-temporales.

**Capítulo 3.-** Presenta las principales herramientas utilizadas en el análisis de procesos espacio-temporales, definiendo los principales conceptos involucrados en su estudio y sus propiedades, así como los diferentes procedimientos de ajuste y predicción. Se completa el Capítulo con un resumen de los diferentes procedimientos utilizados en procesos no estacionarios.

**Capítulo 4.-** En el capítulo anterior se han introducido, entre otros conceptos, la función de covarianza espacio-temporal y el semivariograma, que son herramientas fundamentales a la hora de modelizar la estructura de dependencia espacio-temporal. En este capítulo se estudian con detalle los principales modelos de covarianza espacio-temporales descritos en la literatura. Además se propone un nuevo modelo, al que llamaremos suma de productos generalizado, que genera modelos espacio-temporales como combinaciones lineales de procesos espaciales y temporales, con la importante ventaja de permitir que la covarianza resultante tome valores negativos o valores que oscilen entre negativos y positivos. La validez de dicho modelo se trata a través de varios resultados teóricos.

**Capítulo 5.-** Constituye este capítulo, junto con el anterior, el núcleo de la memoria que presentamos. Contiene un exhaustivo estudio de simulación para comparar la capacidad predictiva de los principales modelos de función de covarianza utilizados en la modelización estadística espacio-temporal. El estudio aquí presentado completa el ya realizado en su día y que ha sido publicado recientemente (Huang et al., 2007).

**Capítulo 6.-** La aplicación de las técnicas descritas en los capítulo precedentes no es un tema menor y ha merecido nuestra atención desde el comienzo de esta investigación. Prueba de ello son las publicaciones que se reproducen en este último capítulo, 4 de ellas ya publicadas o aceptadas en revistas especializadas y la última bajo revisión. Tratan de diversos temas, desde el análisis de un acuífero en Xàbia (Alicante) a la modelización de las tablas de mortalidad dinámicas, pasando por un estudio de la radiación solar en superficie en una zona del Noroeste de la provincia de Valencia.

**Capítulo 7.-** Este último capítulo resume, a modo de conclusión, las aportaciones del presente trabajo y señala alguna de sus posibles evoluciones como futuras líneas de investigación.

Una exhaustiva y actualizada bibliografía cierra la presente memoria.

## Capítulo 2

# Análisis de datos geoestadísticos

### 2.1. Introducción

La estadística espacial puede definirse como aquella rama de la estadística que estudia los fenómenos aleatorios con implantación espacial, ofreciendo una colección de herramientas que permitan su descripción y modelización. Según la naturaleza de los procesos espaciales analizados podemos distinguir tres ramas dentro de la estadística espacial: la geoestadística, los procesos espaciales sobre retículos y los procesos espaciales puntuales.

Cressie (1993) muestra una formulación general que permite la modelización de todas estas posibilidades. Sea  $s$  una localización cualquiera del espacio euclídeo  $d$ -dimensional  $\mathbb{R}^d$ , supongamos que estamos interesados en analizar un determinado fenómeno de interés que toma un valor aleatorio  $Z(s)$  en cada localización  $s$ . Si ahora permitimos que  $s$  varíe sobre un determinado conjunto  $D \subseteq \mathbb{R}^d$ , tendremos el proceso aleatorio  $\{Z(s); s \in D\}$ , que es el objeto de estudio de la estadística espacial. La geoestadística estudiará aquellos fenómenos en los que el índice espacial  $s$  varíe de forma continua sobre toda la región de estudio  $D$ .

El origen de la geoestadística podemos situarlo en la industria minera, en estudios relacionados con el problema de la estimación de la cantidad de mineral en una determinada región a partir de las concentraciones de dicho mineral observadas en un conjunto finito de localizaciones muestreadas. Estos estudios, que se iniciaron en Sudáfrica y alcanzaron su madurez en la Escuela de Minas de Fontainebleau, dieron nombre al término geoestadística. El uso actual de la geoestadística va mucho más allá de estas primeras aplicaciones, aunque mucha de la terminología originaria sigue siendo ampliamente utilizada. Uno de los trabajos pioneros en este campo se debe al ingeniero de minas sudafricano Danie G. Krige, que presenta en Krige (1951) las ecuaciones básicas para una interpolación lineal óptima con variables espacialmente correladas. Su trabajo empírico fue formalizado por el ingeniero francés Matheron (1962, 1963, 1971), en una serie de trabajos en los que introduce el término francés *krigeage* o su equivalente en inglés *kriging* para denotar la técnica de interpolación introducida por Krige. En la actualidad, la geoestadística se ha convertido en una herramienta de gran aplicación en numerosas disciplinas como la geología, la ecología, la astronomía, la ciencia atmosférica o el análisis de imágenes, entre otros.

Algunas referencias básicas sobre modelización geoestadística son Journel y Huijbregts (1978), Ripley (1981), Cliff y Ord (1981), Isaaks y Srisvastava (1989), Cressie (1993), Goovaerts (1997a), Wackernagel (1998), Diggle et al. (1998), Chilès y Delfiner (1999), Stein (1999) o Banerjee et al. (2004).

En este capítulo vamos a describir brevemente los principales conceptos y resultados utilizados en el análisis de datos espaciales, los cuales generalizaremos en los capítulos siguientes para aplicarlos a procesos espacio-temporales. La Sección 2.2 introduce los conceptos más relevantes que utilizaremos en el análisis espacial, como son la estacionariedad, la isotropía, el covariograma o el variograma. La Sección 2.3 muestra los principales estimadores del variograma y covariograma, así como sus principales propiedades. La Sección 2.4 introduce los modelos de covariograma y variograma más utilizados en el análisis de fenómenos espaciales. La Sección 2.5 se dedica a profundizar en el ajuste de los modelos anteriores mediante las principales técnicas de estimación. En la Sección 2.6 se introducen las herramientas de predicción espacial más relevantes. Finalmente, la Sección 2.7 muestra algunos procedimientos de validación de las hipótesis asumidas en la modelización geoestadística, así como herramientas de diagnóstico de los modelos ajustados.

## 2.2. Definiciones básicas

Como ya hemos comentado en la introducción, a lo largo de todo este capítulo nuestro objetivo será estudiar el proceso aleatorio  $\{Z(\mathbf{s}); \mathbf{s} \in D\}$ , siendo  $D \subseteq \mathbb{R}^d$  (en general  $d = 2$ , aunque no necesariamente). Vamos a suponer que  $D$  es una determinada región fija y continua de estudio, y que el índice espacial  $\mathbf{s}$  varía de forma continua en  $D$ , es decir, existe un número infinito de posibles localizaciones en las que podemos observar el proceso. El proceso objeto de estudio  $Z(\mathbf{s})$  podría representar, por ejemplo, la temperatura media diaria observada en una determinada localización  $\mathbf{s}$ .

Vamos a suponer a lo largo de todo este capítulo que, para cada localización  $\mathbf{s} \in D$ , existe la media del proceso que denotaremos por

$$\mu(\mathbf{s}) = E(Z(\mathbf{s})) < \infty, \quad (2.1)$$

y que además el proceso es de segundo orden, esto es

$$\text{Var}(Z(\mathbf{s})) < \infty. \quad (2.2)$$

**Definición 2.1** Decimos que el proceso  $Z(\mathbf{s})$  es Gaussiano si para cualquier conjunto de localizaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \in D$ , el vector aleatorio  $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$  sigue una distribución normal multivariante.

**Definición 2.2** Sea  $Z(\mathbf{s})$  un proceso estocástico de segundo orden. Se define su función de covarianza como

$$C(\mathbf{s}_i, \mathbf{s}_j) = C(Z(\mathbf{s}_i), Z(\mathbf{s}_j)), \forall \mathbf{s}_i, \mathbf{s}_j \in D. \quad (2.3)$$

### 2.2.1. Estacionariedad e isotropía

Generalmente en la práctica sólo disponemos de un conjunto  $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$  de observaciones del proceso aleatorio  $\{Z(\mathbf{s}); \mathbf{s} \in D\}$  obtenidas sobre un conjunto de localizaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ , que pueden distribuirse de forma regular sobre una rejilla o de forma irregular sobre la región de estudio  $D \subseteq \mathbb{R}^d$ . Por lo tanto, sólo disponemos de una única realización incompleta del proceso aleatorio que queremos analizar, por lo que será necesario asumir algún tipo de hipótesis simplificadora de la naturaleza del proceso que asegure cierta regularidad en los datos y permita hacer estimaciones e inferencias del modelo a partir de los datos observados. Esta condición es la de estacionariedad, que permite que el proceso *se repita* a sí mismo en el espacio, proporcionando la *replicación* necesaria para la estimación e inferencia del modelo. La elección del tipo de estacionariedad es a menudo una decisión de modelización, por lo que puede cambiar si la escala de estudio cambia o si disponemos de una mayor cantidad de datos. A continuación veremos los principales tipos de estacionariedad que generalmente asumiremos en los procesos a analizar.

**Definición 2.3** Decimos que el proceso  $Z(\mathbf{s})$  es estrictamente estacionario (o estacionario en sentido fuerte) si, para cualquier conjunto de localizaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \in D$ , la función de distribución conjunta de las variables aleatorias  $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$  permanece invariable ante una traslación. Sea  $F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(z_1, \dots, z_n) = P(Z(\mathbf{s}_1) \leq z_1, \dots, Z(\mathbf{s}_n) \leq z_n)$  la función de distribución conjunta, entonces se cumple que

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(z_1, \dots, z_n) = F_{\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_n + \mathbf{h}}(z_1, \dots, z_n), \forall \mathbf{h} \in \mathbb{R}^d.$$

Esta condición es demasiado restrictiva para la mayoría de los fenómenos observados en la naturaleza, por lo que necesitamos algún tipo de relajación de la misma, como la estacionariedad de segundo orden o la estacionariedad intrínseca.

**Definición 2.4** Decimos que un proceso espacial  $Z(\mathbf{s})$  es estacionario de segundo orden (o estacionario en sentido débil) (o simplemente estacionario) si

1. La función media existe y no depende de la localización, esto es,  $\mu(\mathbf{s}_i) = \mu, \forall \mathbf{s}_i \in D$ .
2. La función de covarianza existe y sólo depende de la distancia entre las localizaciones involucradas, esto es,  $C(\mathbf{s}_i, \mathbf{s}_j) = C(\mathbf{h}), \forall \mathbf{s}_i, \mathbf{s}_j \in D$ , siendo  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$  el vector distancia entre dichas localizaciones. La función  $C(\cdot)$  recibe el nombre de covariograma (o autocovarianza).

De la definición se deduce que si un proceso de segundo orden es estrictamente estacionario, entonces es estacionario de segundo orden. El recíproco es falso en general, aunque se cumple para los procesos gaussianos, que quedan completamente caracterizados por su media y su covariograma.

La estacionariedad de segundo orden implica que la varianza del proceso no depende de la localización, es decir, que

$$\text{Var}(Z(\mathbf{s})) = C(\mathbf{0}) = \sigma^2, \forall \mathbf{s} \in D,$$

donde  $C(\mathbf{0})$  recibe el nombre de varianza a priori del proceso.

**Definición 2.5** Decimos que el proceso  $Z(\mathbf{s})$  es intrínsecamente estacionario si

1. La función media existe y no depende de la localización, esto es,  $\mu(\mathbf{s}_i) = \mu, \forall \mathbf{s}_i \in D$ .
2. La varianza de la diferencia de dos variables aleatorias para dos localizaciones cualesquiera depende únicamente de la distancia entre las localizaciones involucradas, esto es,  $\text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = 2\gamma(\mathbf{h}), \forall \mathbf{s}_i, \mathbf{s}_j \in D$ , con  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ . La función  $2\gamma(\cdot)$  recibe el nombre de variograma, mientras que  $\gamma(\cdot)$  se conoce como semivariograma.

Esta condición es la menos restrictiva de las tres, ya que dado un proceso estacionario  $Z(\mathbf{s})$  de segundo orden con covariograma  $C(\cdot)$ , entonces

$$\text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = \text{Var}(Z(\mathbf{s}_i)) + \text{Var}(Z(\mathbf{s}_j)) - 2\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = 2C(\mathbf{0}) - 2C(\mathbf{s}_i - \mathbf{s}_j),$$

por lo que el proceso  $Z(\mathbf{s})$  es intrínsecamente estacionario con variograma

$$2\gamma(\mathbf{h}) = 2C(\mathbf{0}) - 2C(\mathbf{h}). \quad (2.4)$$

El recíproco no siempre es cierto. Por ejemplo, el movimiento browniano isotrópico  $d$ -dimensional es un proceso intrínsecamente estacionario  $\{W(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$  con variograma  $2\gamma(\mathbf{h}) = 2\|\mathbf{h}\|, \forall \mathbf{h} \in \mathbb{R}^d$ . Para este proceso, el covariograma  $\text{Cov}(W(\mathbf{u}), W(\mathbf{v})) = \frac{1}{2}(\|\mathbf{u}\| + \|\mathbf{v}\| - \|\mathbf{u} - \mathbf{v}\|), \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ , no es función únicamente de la distancia  $\|\mathbf{u} - \mathbf{v}\|$ , por lo que no es un proceso estacionario de segundo orden. Luego la clase de los procesos estacionarios de segundo orden está estrictamente contenida en la clase de los procesos intrínsecamente estacionarios.

Para que un proceso intrínsecamente estacionario lo sea también de segundo orden, deberá tener un semivariograma acotado, esto es, con  $\lim_{h \rightarrow \infty} \gamma(\mathbf{h}) = M < +\infty$ , en cuyo caso su covariograma existe y es igual a  $C(\mathbf{h}) = M - \gamma(\mathbf{h})$ .

El hecho de que este tipo de estacionariedad sea la menos restrictiva de las tres la convierte en la más utilizada en la modelización geoestadística, ya que permite su aplicación sobre un mayor espectro de procesos espaciales. Además, como veremos en la Sección 2.4, existe un amplio abanico de formas paramétricas de semivariogramas muy flexibles que permiten el ajuste de multitud de situaciones prácticas.

**Definición 2.6** Decimos que el proceso  $Z(\mathbf{s})$  es isotrópico si la dependencia espacial del proceso entre dos localizaciones cualesquiera depende únicamente de la distancia existente entre ellas, y no de su localización. En caso contrario decimos que el proceso es anisotrópico.

**Definición 2.7** Decimos que el proceso  $Z(\mathbf{s})$  es homogéneo si es intrínsecamente estacionario e isotrópico.

Si  $Z(\mathbf{s})$  es un proceso homogéneo, entonces su semivariograma es una función que, para cada par de localizaciones, depende únicamente de la longitud del vector distancia entre ellas, esto es,  $\gamma(\mathbf{h}) = \gamma(h), \forall \mathbf{h} \in \mathbb{R}^d$ , siendo  $h \equiv \|\mathbf{h}\|$ . En cambio si un proceso intrínsecamente estacionario  $Z(\cdot)$  es anisotrópico, la dependencia entre  $Z(\mathbf{s})$  y  $Z(\mathbf{s} + \mathbf{h})$

será función tanto de la magnitud como de la dirección de  $\mathbf{h}$ , por lo que el variograma no será únicamente una función de la distancia entre dos localizaciones espaciales. Las anisotropías están causadas por procesos físicos subyacentes que se comportan de forma diferente en el espacio. Por ejemplo, los procesos físicos se comportan generalmente de forma diferente en la dirección vertical (donde actúa el campo gravitatorio) que en las direcciones horizontales. Otro ejemplo es el análisis de cualquier proceso físico relacionado con la penetración salina, en el que el comportamiento en la dirección de la línea de costa suele ser diferente que en la perpendicular a ella. Hay varias formas de trabajar con procesos anisotrópicos, considerándolos como generalizaciones más o menos directas de procesos isotrópicos. Veamos a continuación las más usuales.

**Definición 2.8** Decimos que el proceso  $Z(\mathbf{s})$  tiene anisotropía geométrica si su variograma es de la forma

$$2\gamma(\mathbf{h}) = 2\gamma_0(\|\mathbf{A}\mathbf{h}\|), \mathbf{h} \in \mathbb{R}^d,$$

siendo  $\gamma_0$  un semivariograma isotrópico y  $\mathbf{A}$  una matriz  $d \times d$  que representa una determinada transformación lineal en  $\mathbb{R}^d$ .

En este caso el proceso original no es isotrópico en el espacio original, pero en cambio si lo es sobre un espacio transformado como, por ejemplo, uno en el que se estire o estreche una determinada coordenada. Observamos que la anisotropía geométrica es una generalización de la isotropía, tomando  $\mathbf{A}$  la matriz identidad.

Tenemos que dados  $Z_1(\cdot), \dots, Z_n(\cdot)$ ,  $n$  procesos intrínsecamente estacionarios independientes, entonces  $Z_1(\cdot) + \dots + Z_n(\cdot)$  es un proceso intrínsecamente estacionario con semivariograma dado por  $\gamma(\mathbf{h}) = \gamma_1(\mathbf{h}) + \dots + \gamma_n(\mathbf{h})$ , siendo  $\gamma_i(\mathbf{h})$  el semivariograma del proceso  $Z_i(\cdot)$ . Esta propiedad permite definir la siguiente generalización de la anisotropía geométrica.

**Definición 2.9** Decimos que el proceso  $Z(\mathbf{s})$  tiene anisotropía zonal si su variograma es de la forma

$$2\gamma(\mathbf{h}) = \sum_{i=1}^n 2\gamma_0(\|\mathbf{A}_i\mathbf{h}\|)$$

siendo  $\gamma_0$  un semivariograma isotrópico y  $\mathbf{A}_1, \dots, \mathbf{A}_n$  matrices  $d \times d$ .

Otro tipo de tratamiento de la anisotropía es la de suponer que, dado el proceso original  $Z(\mathbf{s})$ , existe una función no lineal  $g(\mathbf{s})$ , de forma que el proceso  $Z(g(\mathbf{s}))$  es un proceso isotrópico estacionario. Esta idea permite analizar tanto la anisotropía como la no estacionariedad, como podemos ver en Sampson y Guttorp (1992).

En ocasiones trabajamos con procesos físicos en los que la hipótesis de estacionariedad no podrá ser admitida, por lo que muchas de las técnicas de la geoestadística clásica no serán directamente aplicables. En los últimos años han surgido gran número de métodos para modelizar este tipo de procesos no estacionarios. Probablemente el más estudiado es el propuesto por Sampson y Guttorp (1992), que presentan un procedimiento de estimación no paramétrica para la estructura de covarianza espacial no estacionaria. Haas (1995)

introduce una técnica de kriging de ventanas móviles para la estimación en procesos no estacionarios. Higdon et al. (1999) proponen una alternativa usando una representación de medias móviles de un proceso gaussiano. Nychka y Saltzman (1998) y Holland et al. (1999) desarrollan métodos que extienden la técnica de funciones ortogonales empíricas, muy utilizada por los meteorólogos. Otro modelo para procesos no estacionarios es el propuesto por Fuentes (2001, 2002a,b) y desarrollado también en Fuentes y Smith (2001). En este modelo, se considera que el proceso es localmente un campo aleatorio estacionario e isotrópico, que representaremos con un modelo cuyos parámetros variarán a lo largo de la región de estudio, lo que permite la realización de predicciones sobre el campo aleatorio no estacionario con una única realización del proceso.

### 2.2.2. El covariograma

Dado un proceso estacionario de segundo orden  $Z(\cdot)$ , hemos definido su covariograma como

$$C(\mathbf{h}) = \text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)), \quad (2.5)$$

con  $\mathbf{s}_i, \mathbf{s}_j \in D$  y  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$  el vector distancia entre dichas localizaciones. De su definición se deduce fácilmente que  $C(\mathbf{h}) = C(-\mathbf{h})$ . Además, por la desigualdad de Cauchy-Schwartz se cumple que  $|C(\mathbf{h})| \leq C(\mathbf{0})$ ,  $\forall \mathbf{h} \in \mathbb{R}^d$ .

La función de covarianza  $C(\cdot)$  de un proceso estacionario de segundo orden debe ser definida positiva, esto es, debe cumplir

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0, \quad (2.6)$$

para cualquier número finito de localizaciones espaciales  $\{\mathbf{s}_i, i = 1, \dots, n\}$  y de números reales  $\{a_i, i = 1, \dots, n\}$ . Esto es evidente de la definición de covariograma, ya que

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) = \text{Var} \left( \sum_{i=1}^n a_i Z(\mathbf{s}_i) \right) \geq 0.$$

El recíproco también es cierto, ya que dada una función  $C(\cdot)$  definida positiva existe un proceso estocástico estacionario de segundo orden con dicha función de covarianza (ver, por ejemplo, Cressie (1993)). Luego la clase de funciones definidas positivas en  $\mathbb{R}^d$  y la clase de funciones de covarianza estacionarias son idénticas.

**Teorema 2.1 (Teorema de Bochner (1933))** *Una función  $C$  definida sobre  $\mathbb{R}^d$  es continua y definida positiva si y sólo si es la transformada de Fourier de una medida acotada no negativa  $F$  en  $\mathbb{R}^d$ , es decir, si*

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} e^{i\mathbf{h}'\mathbf{x}} dF(\mathbf{x}), \quad \forall \mathbf{h} \in \mathbb{R}^d. \quad (2.7)$$

Veamos a continuación algunas especializaciones de la representación espectral anterior. Si  $C$  es además integrable, entonces  $F$  es absolutamente continua con densidad

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathfrak{R}^d} e^{i\mathbf{h}'\mathbf{x}} C(\mathbf{h}) d\mathbf{h}, \quad \mathbf{x} \in \mathfrak{R}^d.$$

En particular,  $C$  y  $f$  forman una pareja de transformadas de Fourier. Se cumple que una función  $C : \mathfrak{R}^d \rightarrow \mathfrak{R}$  integrable y continua es definida positiva si y sólo si su transformada de Fourier inversa,

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathfrak{R}^d} e^{i\mathbf{h}'\mathbf{x}} C(\mathbf{h}) d\mathbf{h}, \quad \mathbf{x} \in \mathfrak{R}^d,$$

es una función no negativa.

Una función  $C : \mathfrak{R}^d \rightarrow \mathfrak{R}$  es continua y definida positiva si y sólo si es la transformada coseno de una medida simétrica acotada no negativa  $F$  en  $\mathfrak{R}_+^d$ , es decir, si

$$C(\mathbf{h}) = \int_{\mathfrak{R}_+^d} \cos(\mathbf{h}'\mathbf{x}) dF(\mathbf{x}), \quad \forall \mathbf{h} \in \mathfrak{R}^d.$$

Desde un punto de vista geoestadístico, la restricción a funciones continuas utilizada en el Teorema de Bochner no suele ser suficientemente satisfactoria, ya que los modelos de covarianza ajustados a datos experimentales suelen presentar una discontinuidad en el origen que recibe el nombre de efecto pepita. En la Sección 2.4 se define la covarianza de un proceso de efecto pepita como

$$C(\mathbf{h}) = \mathbf{1}_{c_0} = \begin{cases} c_0 & , \text{ si } \|\mathbf{h}\| = 0 \\ 0 & , \text{ si } \|\mathbf{h}\| > 0, \end{cases} \quad (2.8)$$

con  $c_0 \geq 0$ . El teorema de Riesz (1934) nos asegura que dada una función  $C$  definida positiva y medible, entonces  $C = C_c + C_0$ , donde  $C_c$  es una función continua y definida positiva y  $C_0$  es una función definida positiva e igual a cero casi por todas partes respecto a la medida de Lebesgue.

Veamos a continuación algunas propiedades de las funciones de covarianza que nos permitirán construir nuevos modelos:

- (P1) Sea  $C(\cdot)$  una función de covarianza definida sobre  $\mathfrak{R}^d$ , entonces  $C(\cdot)$  es también una función de covarianza definida sobre  $\mathfrak{R}^l$ , para  $l \leq d$ . En cambio, el recíproco no será cierto en general.
- (P2) Sea  $C(\cdot)$  una función de covarianza definida sobre  $\mathfrak{R}^d$  y  $b > 0$ , entonces  $b \cdot C(\cdot)$  es también una función de covarianza definida sobre  $\mathfrak{R}^d$ .
- (P3) Sean  $C_1(\cdot)$  y  $C_2(\cdot)$  dos funciones de covarianza definidas sobre  $\mathfrak{R}^d$ , entonces  $C_1(\cdot) + C_2(\cdot)$  es también una función de covarianza definida sobre  $\mathfrak{R}^d$ .
- (P4) Primera propiedad de estabilidad (Chilès y Delfiner, 1999). Sean  $C_1(\cdot)$  y  $C_2(\cdot)$  dos funciones de covarianza sobre  $\mathfrak{R}^d$ , entonces  $C_1(\cdot) \cdot C_2(\cdot)$  es también una función de covarianza definida sobre  $\mathfrak{R}^d$ .

(P5) Sean  $C_1(\cdot)$  y  $C_2(\cdot)$  dos funciones de covarianza en  $\mathfrak{R}^d$ , entonces la convolución  $C(\mathbf{h}) = \int C_1(\mathbf{h} - \mathbf{t})C_2(\mathbf{t})d\mathbf{t}$  es también una función de covarianza sobre  $\mathfrak{R}^d$ .

(P6) Segunda propiedad de estabilidad (Chilès y Delfiner, 1999).

Sea  $\mu(a)$  una medida acotada no negativa definida sobre  $U \subset \mathfrak{R}$  y sea  $C(\mathbf{s}_1, \mathbf{s}_2|a)$  una función de covarianza en  $\mathfrak{R}^d$  para cada  $a \in V \subset U$  integrable sobre el conjunto  $V \subset U$ , entonces  $C(\mathbf{s}_1, \mathbf{s}_2) = \int_V C(\mathbf{s}_1, \mathbf{s}_2|a)d\mu(a)$  es una función de covarianza en  $\mathfrak{R}^d$ .

(P7) Sea  $\{C_n(\cdot), n = 1, \dots\}$  una sucesión de funciones de covarianza en  $\mathfrak{R}^d$  convergente a  $C(\cdot)$  (es decir, con  $\lim_{n \rightarrow \infty} C_n(\mathbf{h}) = C(\mathbf{h}), \forall \mathbf{h} \in \mathfrak{R}^d$ ), entonces  $C(\cdot)$  es una función de covarianza sobre  $\mathfrak{R}^d$ .

De (P2) y (P3) se tiene que el conjunto de funciones de covarianza definidas sobre  $\mathfrak{R}^d$  forman un cono convexo.

Como veremos en los capítulos siguientes, un concepto que utilizaremos será el de separabilidad. Sea  $Z(\cdot)$  un proceso estocástico definido sobre  $\mathfrak{R}^d$  que puede descomponerse en  $d$  procesos estocásticos  $Z_i(\cdot)$  independientes definidos sobre  $\mathfrak{R}$ , esto es, de forma que  $Z(\mathbf{s}) = \prod_{i=1}^d Z_i(s_i)$ . Entonces su covariograma  $C(\cdot)$  puede expresarse como

$$C(\mathbf{h}) = \prod_{i=1}^d C_i(h_i), \forall \mathbf{h} = (h_1, \dots, h_d)' \in \mathfrak{R}^d, \quad (2.9)$$

donde cada  $C_i$  será el covariograma de los procesos  $Z_i(\cdot)$ . En ese caso decimos que  $C(\mathbf{h})$  es un covariograma separable. Puede comprobarse fácilmente que, si  $C_i(\cdot)$  son covariogramas válidos en  $\mathfrak{R}$ , entonces  $C(\cdot)$  definido por (2.9) es también un covariograma válido en  $\mathfrak{R}^d$ .

Los modelos isotrópicos son de gran importancia, ya que no sólo permiten el tratamiento de los procesos estacionarios e isotrópicos, sino que además proporcionan la base sobre la que construir modelos más complejos anisotrópicos o no estacionarios. Por ejemplo, los modelos de anisotropía geométrica se construyen a partir de funciones isotrópicas, o el tratamiento de la no estacionariedad de Sampson y Guttorp (1992) consiste en construir modelos de covarianza no estacionarios mediante una deformación de las coordenadas espaciales en un nuevo sistema de coordenadas sobre la que se modeliza una estructura de covarianza isotrópica.

Schoenberg (1938) demuestra que una función  $C : \mathfrak{R}^d \rightarrow \mathfrak{R}$  continua e isotrópica es definida positiva (y por tanto un modelo válido de covarianza) si y sólo si es de la forma

$$C(h) = \int_0^{+\infty} \Omega_d(\omega h) dF(\omega), \quad h \geq 0, \quad (2.10)$$

donde esta integral es considerada en el sentido de integral de Riemann-Stieltjes,  $F$  es una medida acotada no negativa sobre  $\mathfrak{R}_+$ ,  $J_\nu(\cdot)$  es la función de Bessel de primera especie de orden  $\nu$  (ver, por ejemplo, Abramowitz y Stegun (1965)) y  $\Omega_d : [0, +\infty[ \rightarrow \mathfrak{R}$  se define como

$$\Omega_d(t) = \Gamma(d/2) \left(\frac{2}{t}\right)^{(d-2)/2} J_{(d-2)/2}(t). \quad (2.11)$$

Para los casos más comunes en la práctica geoestadística en que  $d = 1, 2$  o  $3$ , la función (2.11) se reduce a  $\Omega_1(t) = \cos(t)$ ,  $\Omega_2(t) = J_0(t)$  y  $\Omega_3(t) = \sin(t)/t$ . Este teorema caracteriza las funciones de covarianza continuas isotrópicas sobre  $\mathfrak{R}^d$  como mixturas escalares de las funciones anteriores, que dependen de la dimensión espacial  $d$ . Se pueden consultar más detalles en Gneiting y Sasvári (1999).

Se puede comprobar que cualquier función de covarianza continua e isotrópica  $C(\cdot)$  es de la forma  $C(\mathbf{h}) = \varphi(\|\mathbf{h}\|^2)$ ,  $\mathbf{h} \in \mathfrak{R}^d$ , donde  $\varphi(\cdot)$  es una función completamente monótona en  $[0, +\infty[$ . En la Sección 4.7 se analizan con más detalle este tipo de funciones, de gran importancia en la modelización de la covarianza de los procesos espaciales y espacio-temporales.

Gneiting y Sasvári (1999) muestran que para dimensiones de  $\mathfrak{R}^d$ , con  $d \geq 2$ , y para funciones definidas positivas e isotrópicas, las únicas funciones que se anulan casi por todas partes que aparecen en el Teorema de Riesz son los efectos pepita (2.8). Esto es, sea  $C$  una función definida positiva, isotrópica y medible en  $\mathfrak{R}^d$ , con  $d \geq 2$ , entonces puede descomponerse como  $C = C_c + 1_{c_0}$ , donde  $c_0 \geq 0$  y  $C_c$  es una función definida positiva y continua.

La verificación de la validez de un determinado modelo de covariograma mediante la condición (2.6) suele ser muy dificultosa. En Schlather (1999) podemos encontrar un conjunto de condiciones necesarias y suficientes que permiten el contraste de la validez de los modelos propuestos de forma más sencilla.

### 2.2.3. El variograma

Hemos definido el variograma de un proceso intrínsecamente estacionario  $Z(\cdot)$  como la función

$$2\gamma(\mathbf{h}) = \text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)), \quad (2.12)$$

con  $\mathbf{s}_i, \mathbf{s}_j \in D$  y  $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$  el vector distancia entre dichas localizaciones. De (2.12) se deduce fácilmente que  $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$  y que  $\gamma(\mathbf{0}) = 0$ . Obsérvese que el variograma, al contrario del covariograma, no depende de la media del proceso, lo que como veremos tendrá implicaciones en la estimación de ambos.

Una condición necesaria que debe cumplir el variograma es que debe ser una función condicionalmente definida negativa, esto es,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0 \quad (2.13)$$

para cualquier conjunto finito de localizaciones espaciales  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \in \mathfrak{R}^d$  y para cualquier conjunto de números reales  $\{a_1, \dots, a_n\} \in \mathfrak{R}$  con  $\sum_{i=1}^n a_i = 0$ . Esto es evidente de su

definición, ya que dados  $\{a_1, \dots, a_n\} \in \mathfrak{R}$  con  $\sum_{i=1}^n a_i = 0$ , entonces

$$\begin{aligned}
& \sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(\mathbf{s}_i - \mathbf{s}_j) = \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) \\
&= \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Var}(Z(\mathbf{s}_i)) + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Var}(Z(\mathbf{s}_j)) - 2 \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \\
&= \sum_{i=1}^n a_i \text{Var}(Z(\mathbf{s}_i)) \sum_{j=1}^n a_j + \sum_{j=1}^n a_j \text{Var}(Z(\mathbf{s}_j)) \sum_{i=1}^n a_i - 2 \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \\
&= 0 + 0 - 2 \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = -\text{Var}\left(\sum_{i=1}^n a_i Z(\mathbf{s}_i)\right) \leq 0.
\end{aligned}$$

Algunos autores definen la estacionariedad intrínseca de un proceso espacial  $\{Z(\mathbf{s}), \mathbf{s} \in D\}$  a partir de la estacionariedad de segundo orden del proceso de los incrementos  $\{Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}), \mathbf{s} \in D\}$ , esto es, si los dos primeros momentos  $E(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))$  y  $E(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2$  no dependen de la localización espacial  $\mathbf{s}$ . En ese caso, definen el variograma no centrado como  $2\gamma(\mathbf{h}) = E(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2$ , mientras que denominan variograma centrado a la expresión  $2\gamma(\mathbf{h}) = \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))$ . En el caso en que  $E(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 0$ , que es lo que nosotros consideramos en la definición de proceso intrínsecamente estacionario, ambas definiciones coinciden y no es necesario por tanto hacer distinciones.

Otra condición que debe satisfacer un variograma (Matheron (1971)) es que debe tener un ritmo de crecimiento inferior al de  $h^2$ , esto es

$$\lim_{h \rightarrow \infty} \frac{2\gamma(\mathbf{h})}{h^2} = 0. \quad (2.14)$$

Recíprocamente, se cumple que dada una función  $2\gamma : \mathfrak{R}^d \rightarrow \mathfrak{R}$  condicionalmente definida negativa, con  $2\gamma(\mathbf{0}) = 0$  y que verifica la condición (2.14), entonces existe un proceso aleatorio  $Z(\cdot)$  gaussiano intrínsecamente estacionario con variograma  $2\gamma(\cdot)$ .

Veamos a continuación un importante teorema de caracterización de los variogramas de los procesos intrínsecamente estacionarios.

**Teorema 2.2** *Sea  $2\gamma : \mathfrak{R}^d \rightarrow \mathfrak{R}$  una función con  $2\gamma(\mathbf{0}) = 0$ , entonces  $2\gamma(\cdot)$  es condicionalmente definida negativa si y sólo si para todo  $t > 0$ ,  $\exp(-t\gamma(\cdot))$  es una función de covarianza.*

Este resultado permite establecer analogías muy útiles entre las funciones de covarianza y los variogramas, y verificar de forma sencilla la validez de determinados variogramas. El teorema anterior, bajo la condición adicional de continuidad, puede encontrarse por ejemplo en Cressie (1993) o Chilès y Delfiner (1999). Gneiting et al. (2001) muestran una generalización que no exige la continuidad, lo que permite su aplicación sobre aquellos

modelos de variograma obtenidos a partir de la suma de un modelo continuo y un efecto pepita, de uso muy común en la modelización de procesos reales.

Como consecuencia de los resultados vistos para los covariogramas y de las equivalencias anteriores, se tiene que cualquier semivariograma medible  $\gamma$  en  $\mathfrak{R}^d$  puede descomponerse como

$$\gamma(\mathbf{h}) = \gamma_c(\mathbf{h}) + \gamma_0(\mathbf{h}) = \log \frac{C_c(\mathbf{0})}{C_c(\mathbf{h})} + \log \frac{1}{C_c(\mathbf{0})} \frac{C_c(\mathbf{h})}{C_c(\mathbf{h}) + C_0(\mathbf{h})}, \quad \mathbf{h} \in \mathfrak{R}^d, \quad (2.15)$$

donde  $2\gamma_c$  es un variograma continuo y  $2\gamma_0$  es un variograma constante casi por todas partes.

El teorema de Bochner también permite la siguiente caracterización espectral de los variogramas.

**Teorema 2.3** *Sea  $2\gamma$  una función real sobre  $\mathfrak{R}^d$  con  $\gamma(\mathbf{0}) = 0$ , entonces es un variograma si y sólo si es de la forma*

$$\gamma(\mathbf{h}) = Q(\mathbf{h}) + \int_{\mathfrak{R}^d} \frac{1 - \cos(\mathbf{u}'\mathbf{h})}{\|\mathbf{u}\|^2} dF(\mathbf{u}), \quad \mathbf{h} \in \mathfrak{R}^d,$$

donde  $Q(\cdot) \geq 0$  es una forma cuadrática y  $F$  es una función medible no negativa en  $\mathfrak{R}^d$  sin masa en el origen y que satisface

$$\int_{\mathfrak{R}^d} \frac{1}{1 + \|\mathbf{u}\|^2} dF(\mathbf{u}) < +\infty.$$

Como ocurría con los covariogramas, la familia de semivariogramas también forma un cono convexo, por lo que dados  $\gamma_1(\cdot)$  y  $\gamma_2(\cdot)$  dos semivariogramas válidos en  $\mathfrak{R}^d$  y  $b > 0$ , entonces  $\gamma(\cdot) = \gamma_1(\cdot) + \gamma_2(\cdot)$  y  $b \cdot \gamma(\cdot)$  son semivariogramas válidos en  $\mathfrak{R}^d$ . Esta propiedad nos permitirá construir nuevos modelos de semivariograma válidos como suma o producto escalar de otros modelos válidos.

Si el proceso estocástico  $Z(\cdot)$  definido sobre  $\mathfrak{R}^d$  puede descomponerse como  $Z(\mathbf{s}) = \prod_{i=1}^d Z_i(s_i)$ , con  $Z_i(\cdot)$   $d$  procesos independientes definidos sobre  $\mathfrak{R}$ , entonces su semivariograma puede expresarse como

$$\gamma(\mathbf{h}) = \sum_{i=1}^d \gamma_i(h_i), \quad \forall \mathbf{h} = (h_1, \dots, h_d)' \in \mathfrak{R}^d, \quad (2.16)$$

siendo  $\gamma_i$  el semivariograma del proceso  $Z_i(\cdot)$ , con  $i = 1, \dots, d$ . Este modelo de semivariograma dado por (2.16) recibe el nombre de semivariograma de estructura anidada, y como veremos ha sido utilizado en la modelización de procesos espacio-temporales.

Al contrario de lo que ocurría con el covariograma, el producto de semivariogramas no dará lugar a modelos válidos de semivariograma.

En el caso de los variogramas isotrópicos, en los que  $2\gamma(\mathbf{h}) = 2\gamma(h)$ , la representación espectral vista en el teorema 2.3 se reduce a la integral de Bessel, y la parte discontinua de la descomposición del variograma que hemos visto en (2.15) es necesariamente un efecto pepita.

### 2.2.4. El correlograma

Sea  $Z(\cdot)$  un proceso estacionario de segundo orden con función de covarianza  $C(\cdot)$ . Tenemos que  $C(\mathbf{0}) = \text{Cov}(Z(\mathbf{s}), Z(\mathbf{s})) = \text{Var}(Z(\mathbf{s}))$ , por lo que  $C(\mathbf{0}) > 0$  a no ser que  $Z(\cdot)$  sea un proceso constante en  $D$ . Definimos el correlograma (o función de autocorrelación) como

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

De la definición se desprende que  $\rho(\mathbf{h}) = \rho(-\mathbf{h})$  y que  $\rho(\mathbf{0}) = 1$ .

### 2.2.5. Forma general de estas funciones

El semivariograma representa un índice del cambio que una variable muestra con la distancia. Su forma describe el patrón de variación espacial en términos de su magnitud y forma general. Generalmente el semivariograma crece con la distancia, ya que en la mayoría de procesos físicos existen mayores similitudes en los valores observados en localizaciones próximas, que disminuyen al aumentar la distancia.

En ocasiones, este crecimiento del semivariograma con la distancia se estabiliza alrededor de un determinado valor  $c_s > 0$ , que es una cota superior de la función (esto es,  $c_s = \lim_{\mathbf{h} \rightarrow \infty} \gamma(\mathbf{h})$ ). En este caso decimos que el variograma es de tipo transitivo o acotado, y el valor alrededor del cual se estabiliza recibe el nombre de meseta o varianza a priori (*sill*, en inglés), que es igual por (2.4) a  $C(\mathbf{0})$ , siendo  $C(\cdot)$  el covariograma del proceso. Llamaremos rango o alcance (*range*, en inglés) al valor  $\mathbf{h}_r$  en el que el semivariograma alcanza su meseta, esto es, la distancia para la que  $\gamma(\mathbf{h}_r) = c_s$ , y que también representa el valor a partir del cual el covariograma se anula. Observar que el rango marca el límite de la dependencia espacial de la propiedad. Para algunos semivariogramas transitivos, la meseta  $c_s$  sólo se alcanza asintóticamente en el límite, por lo que estrictamente hablando el variograma tendrá rango infinito. En este caso utilizaremos el término de rango efectivo, que definimos como la distancia en la que el semivariograma alcanza el 95 % de su meseta.

Otro tipo posible de semivariogramas son aquellos que presentan un incremento sin límites, y por tanto no son acotados. En este caso el proceso que representan no es estacionario de segundo orden, y no podemos por tanto definir su covariograma asociado.

Sabemos que el semivariograma es una función que debe cumplir que  $\gamma(\mathbf{0}) = 0$ , pero en la práctica suele ocurrir que  $\lim_{\mathbf{h} \rightarrow \mathbf{0}} \gamma(\mathbf{h}) = c_0 > 0$ , donde  $c_0$  recibe el nombre de pepita (*nugget*, en inglés). Esta discontinuidad en el origen puede estar causada por variaciones de pequeña escala (que sólo tienen sentido en procesos que no son  $L_2$ -continuos), o por errores de medida (es decir, que si realizamos varias observaciones en una misma localización, los valores observados fluctúan alrededor de un determinado valor, que es el valor real). En la práctica, sólo se habrán observado un conjunto de datos  $\{z(\mathbf{s}_i), i = 1, \dots, n\}$ , por lo que no podemos conocer nada del comportamiento del variograma a distancias menores de mín  $\{\|\mathbf{s}_i - \mathbf{s}_j\|, 1 \leq i < j \leq n\}$  y solemos determinar el valor de  $c_0$  extrapolando el comportamiento del variograma a distancias cercanas a cero. En este caso, definimos la meseta parcial (*partial sill*, en inglés) como  $c_s - c_0$ .

Si el proceso  $Z(\cdot)$  es isotrópico, entonces  $2\gamma(\mathbf{h}) = 2\gamma(h)$ , es decir, el variograma depende únicamente de la distancia entre dos localizaciones y no de la dirección. En la Figura 2.1

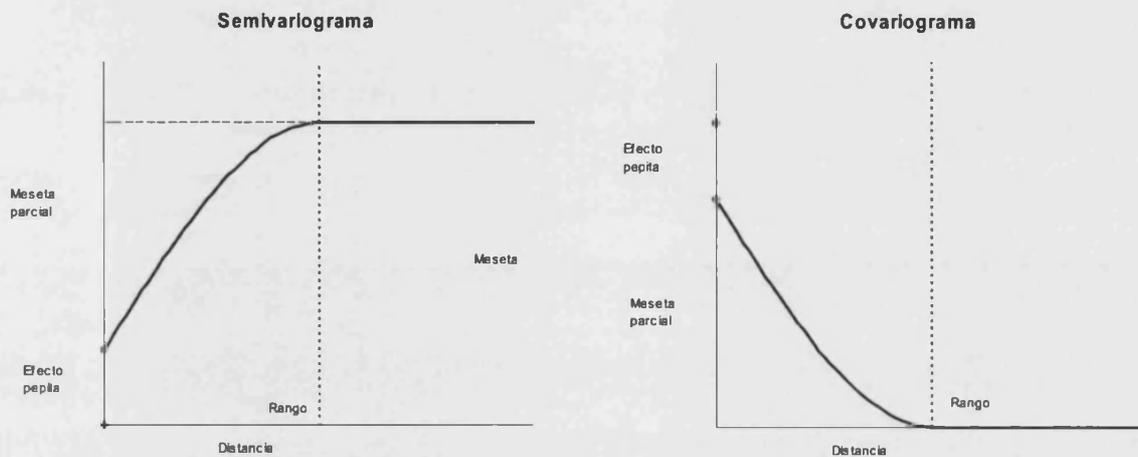


Figura 2.1: Forma general del variograma y covariograma de un proceso espacial homogéneo

se muestra la forma típica del variograma de un proceso homogéneo y de su covariograma asociado, donde podemos observar la interpretación de los parámetros introducidos anteriormente.

## 2.3. Estimación del variograma y del covariograma

Dado un proceso espacial  $Z(\cdot)$  intrínsecamente estacionario, veamos como obtener una estimación del variograma  $2\gamma(\cdot)$  (y del covariograma  $C(\cdot)$  si el proceso es además estacionario de segundo orden) a partir de los valores observados sobre un conjunto de localizaciones  $\{s_1, \dots, s_n\}$ . Del conjunto de estimadores propuestos en la literatura para la estimación de estas medidas de variabilidad espacial, veremos a continuación el estimador clásico propuesto por Matheron (1962) o el estimador robusto de Cressie y Hawkins (1980).

### 2.3.1. Estimador clásico

La estimación del variograma más sencilla es la obtenida mediante el estimador del método de los momentos, que recibe el nombre de estimador clásico del variograma. Tenemos que, bajo la hipótesis de estacionariedad intrínseca y por tanto de media del proceso constante, se cumple que

$$2\gamma(\mathbf{h}) = \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = E[(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2].$$

Si los puntos de muestreo  $\{s_1, \dots, s_n\}$  estuviesen localizados sobre una rejilla regular,

el estimador del método de los momentos vendría definido por

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, \quad (2.17)$$

donde  $N(\mathbf{h})$  denota todos aquellos pares  $(\mathbf{s}_i, \mathbf{s}_j)$  para los que  $\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}$  y  $|N(\mathbf{h})|$  denota el cardinal de  $N(\mathbf{h})$ . Obsérvese que no es necesario estimar la media  $\mu$  del proceso.

Debido a que (2.17) es esencialmente una media muestral, tiene todas las desventajas asociadas comúnmente a este tipo de estimadores como la no robustez. Se trata de un estimador no paramétrico que es óptimo cuando se dispone de una malla regular de muestreo que sea representativa y la distribución es normal. No obstante, en la práctica el empleo de este estimador produce en ocasiones variogramas experimentales erráticos, debido a desviaciones del caso ideal para la aplicación del mismo, como son distribuciones alejadas de la normalidad, heterocedasticidad, desviaciones en el muestreo o existencia de valores atípicos.

En el caso general en que los puntos de muestreo no estén sobre una rejilla regular, necesitamos aplicar algún tipo de agrupación de los datos para suavizar el variograma experimental obtenido. Para ello se utiliza la siguiente expresión

$$2\gamma^+(\mathbf{h}) = \text{ave}\{(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, (i, j) \in N(\mathbf{h}), \mathbf{h} \in T(\mathbf{h}(l))\}$$

donde la región  $T(\mathbf{h}(l))$  es alguna región de tolerancia en  $\mathbb{R}^d$  alrededor de  $\mathbf{h}(l)$ , con  $l = 1, \dots, K$ , y donde  $\text{ave}\{\cdot\}$  representa la media (posiblemente ponderada) de los elementos en  $\{\cdot\}$ . Generalmente las regiones  $T(\mathbf{h}(l))$  se escogen disjuntas, y formadas por todos los pares de localizaciones situadas a una distancia  $h \in [h(l) - d_l/2, h(l) + d_l/2]$ , con  $d_l$  la tolerancia espacial utilizada, y sin tener en cuenta la dirección del vector de distancia, con lo que tenemos una estimación del variograma experimental omnidireccional. Si en lugar de utilizar todos los pares de puntos cuya distancia tiene una longitud determinada se utilizan únicamente los que además forman un ángulo  $\theta \in [\alpha - d_\alpha/2, \alpha + d_\alpha/2]$ , con  $d_\alpha$  la tolerancia angular, obtenemos una estimación del variograma experimental direccional en la dirección  $\alpha$ .

Con el fin de que el variograma empírico refleje convenientemente la realidad del proceso, su estimación debe estar basada en un número suficiente de valores, por lo que suele calcularse únicamente para valores de  $\mathbf{h}$  con  $h < d_{max}/2$ , siendo  $d_{max}$  la distancia máxima observada. Respecto al número de intervalos a considerar dentro del intervalo de estudio, siempre que sea posible es conveniente seleccionar al menos 10 intervalos y con al menos 30 pares de localizaciones en cada uno de ellos, con el fin de retener la resolución espacial garantizando a la vez la estabilidad del estimador. Con el fin de conocer la forma y magnitud del variograma, suele representarse un gráfico con los valores estimados en cada región de tolerancia frente a la distancia promedio de todos los pares dentro de dicha región. Si tenemos en cuenta la dirección en la distancia entre las localizaciones, tendremos diferentes estimaciones direccionales.

Para la covarianza, el estimador obtenido por el método de los momentos sería

$$\hat{C}(\mathbf{h}) = |N(\mathbf{h})| \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} (Z(\mathbf{s}_i) - \hat{Z})(Z(\mathbf{s}_j) - \hat{Z}),$$

donde  $\hat{Z} = \frac{1}{n} \sum_i Z(\mathbf{s}_i)$  es un estimador de la media  $\mu$  del proceso, y  $N(\mathbf{h})$  se define como antes.

Señalemos que  $2\hat{\gamma}(\mathbf{h}) \neq 2(\hat{C}(\mathbf{0}) - \hat{C}(\mathbf{h}))$ , por lo que los estimadores no cumplen la propiedad (2.4) que verifican el variograma y el covariograma del proceso, aunque para  $|N(\mathbf{h})|/n \approx 1$  la diferencia entre ambas cantidades es pequeña. En Cressie (1993) podemos encontrar un amplio desarrollo de estos estimadores, comparados entre sí respecto a sus sesgos y a sus distribuciones asintóticas.

### 2.3.2. Estimador robusto

Como ya hemos comentados anteriormente, aunque el método de estimación clásico presenta la ventaja de su facilidad de cálculo, tiene algunos inconvenientes prácticos como que no es robusto frente valores extremos de  $Z(\mathbf{s})$ . Cressie y Hawkins presentan una variación de (2.17) de mayor robustez como estimador insesgado del variograma, y que se define como

$$2\bar{\gamma}(\mathbf{h}) = \frac{1}{0,457 + 0,494/|N(\mathbf{h})|} \left\{ \frac{1}{|N(\mathbf{h})|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(\mathbf{h})} |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{1/2} \right\}^4. \quad (2.18)$$

Este estimador se basa en la siguiente propiedad. Sea  $Z(\cdot)$  un proceso intrínsecamente estacionario gaussiano, entonces dadas dos localizaciones  $\mathbf{s}_1, \mathbf{s}_2$  cualesquiera en  $D$ , la distribución de  $(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2$  es de la forma  $2\gamma(\mathbf{s}_1 - \mathbf{s}_2)\chi_1^2$ , donde la distribución de  $\chi_1^2$  es muy sesgada. Sin embargo, si  $X \sim \chi_1^2$ , entonces  $X^{1/4}$  tiene una distribución casi simétrica y por tanto las medias muestrales de  $|Z(\mathbf{s}_1) - Z(\mathbf{s}_2)|^{1/2}$  se comportarán mejor que las de  $\{Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\}^2$ . Los coeficientes de la expresión (2.18) se introducen para asegurar la insesgaredad del estimador propuesto. En Cressie (1993) podemos encontrar un análisis detallado de este estimador comparado con el anterior, así como otras variantes robustas para la estimación empírica del variograma.

## 2.4. Principales modelos de variogramas y covariogramas isotrópicos

En la sección anterior hemos visto algunos estimadores del variograma o covariograma de los procesos espaciales. El problema es que estas estimaciones no podemos utilizarlas directamente en la práctica geoestadística, ya que no satisfacen en general la condición de ser condicionalmente definidas negativas que deben verificar los variogramas (o definidas positivas para los covariogramas). Su uso tendría efectos no deseables, como la obtención de varianzas negativas en la predicción espacial mediante kriging. Es por ello que, en lugar de utilizar directamente las predicciones, ajustaremos a las estimaciones obtenidas anteriormente uno de los modelos válidos de variograma o covariograma que veremos en esta sección.

A continuación vamos a introducir algunos de los principales modelos de semivariograma isotrópicos más utilizados en la práctica geoestadística, junto con su covariogramas

y correlogramas asociados en el caso en que estén definidos. Como hemos visto, para obtener los correspondientes variogramas bastará con multiplicar por 2 cada uno de los semivariogramas que se muestran a continuación. Estos modelos, además de constituir una herramienta esencial del tratamiento de los datos geoestadísticos, servirán como base para la posterior construcción de modelos espacio-temporales que trataremos en los capítulos siguientes. Como hemos dicho, todos los variogramas y covariogramas que se mostrarán en esta sección son isotrópicos (esto es, dependen de la distancia  $\mathbf{h}$  entre localizaciones únicamente por su módulo  $h = \|\mathbf{h}\|$ ), ya que son el punto de arranque sobre los que construir modelos más complejos.

En concreto, vamos a describir el modelo de efecto pepita, el modelo lineal, el modelo esférico, el modelo exponencial, el modelo gaussiano, el modelo racional cuadrático, el modelo ondulado, el modelo potencial y el modelo de Matérn. En Schlather (1999) pueden consultarse detalles de estos y otros modelos isotrópicos utilizados en la práctica geoestadística, como el modelo coseno, el modelo de Bessel, el modelo de Cauchy o el modelo exponencial de potencia, entre otros. También veremos los procedimientos más utilizados para construir nuevos modelos mediante la combinación de estos, aplicando las propiedades que hemos visto en las secciones 2.2.2 y 2.2.3.

#### 2.4.1. Modelo de efecto pepita

Este modelo es válido en  $\mathfrak{R}^d$ , para cualquier  $d \geq 1$ . Es la estructura más básica de semivariograma indicando falta de estructura espacial. Corresponde a un ruido blanco, es decir, un proceso espacial puramente aleatorio sin ninguna correlación entre las muestras con independencia de la distancia que las separe, y viene dado por la siguiente expresión

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ si } h = 0 \\ c_0 & , \text{ si } h > 0, \end{cases} \quad (2.19)$$

siendo  $\theta = c_0 \geq 0$  el efecto pepita del proceso. Aplicando (2.4) obtenemos que su covariograma asociado viene dado por

$$C(\mathbf{h}) = \begin{cases} c_0 & , \text{ si } h = 0 \\ 0 & , \text{ si } h > 0. \end{cases} \quad (2.20)$$

La función de autocorrelación asociada a este modelo vendría dada por

$$\rho(\mathbf{h}) = \begin{cases} 1 & , \text{ si } h = 0 \\ 0 & , \text{ si } h > 0. \end{cases}$$

#### 2.4.2. Modelo lineal

Es un modelo válido en  $\mathfrak{R}^d$ , para cualquier dimensión  $d \geq 1$ , que viene dado por la siguiente expresión

$$\gamma(\mathbf{h}) = b_l h, \quad (2.21)$$

siendo  $b_l > 0$  un parámetro de forma. Esta expresión define un modelo no acotado que no corresponde con un proceso estacionario de segundo orden, y que no tendrá por tanto ni

covariograma ni correlograma asociados. En la Figura 2.2 podemos ver dos semivariogramas lineales con diferente  $b_l$  (1 en negro y 2 en rojo), que como vemos marca la pendiente de la recta.

### 2.4.3. Modelo esférico

Es un modelo válido únicamente en  $\mathfrak{R}^1, \mathfrak{R}^2$  y  $\mathfrak{R}^3$ , cuyo semivariograma asociado es

$$\gamma(h) = \begin{cases} c_s \left( \frac{3}{2} \frac{h}{a_s} - \frac{1}{2} \left( \frac{h}{a_s} \right)^3 \right) & , \text{ si } 0 \leq h < a_s \\ c_s, & , \text{ si } h \geq a_s, \end{cases} \quad (2.22)$$

con  $c_s \geq 0$  la meseta y  $a_s > 0$  el rango del modelo. En la Figura 2.2 se muestran tres integrantes de esta familia, con diferentes mesetas y rangos (1 y 0,7 en negro, 2 y 0,7 en rojo, 1 y 0,3 en verde, respectivamente). La forma en todos los casos es la de una función creciente y acotada que alcanza su valor máximo  $c_s$  a una distancia igual al rango  $a_s$ . Aplicando (2.4) es fácil comprobar que su covariograma y su correlograma asociados vienen dados por

$$C(h) = \begin{cases} c_s \left( 1 - \left( \frac{3}{2} \frac{h}{a_s} - \frac{1}{2} \left( \frac{h}{a_s} \right)^3 \right) \right) & , \text{ si } 0 \leq h < a_s \\ 0 & , \text{ si } h \geq a_s, \end{cases} \quad (2.23)$$

$$\rho(h) = \begin{cases} 1 - \frac{3}{2} \frac{h}{a_s} + \frac{1}{2} \left( \frac{h}{a_s} \right)^3 & , \text{ si } 0 \leq h < a_s \\ 0 & , \text{ si } h \geq a_s. \end{cases}$$

### 2.4.4. Modelo exponencial

Es un modelo válido en  $\mathfrak{R}^d$ , para cualquier dimensión  $d \geq 1$ . Se utiliza mucho en la práctica por la sencillez de su formulación y por permitir rápidas simulaciones gracias a su propiedad markoviana, y su expresión viene dada por

$$\gamma(h) = c_e(1 - \exp(-h/a_e)), \quad (2.24)$$

siendo  $c_e \geq 0$  la meseta y  $3a_e > 0$  el rango efectivo del modelo (es decir, aquella distancia a la que se alcanza un 95 % de la meseta). Como vemos, este modelo tiene una expresión analítica más sencilla que el modelo esférico y además es válido para cualquier dimensión  $d$ . También es un modelo acotado, pero al contrario que ocurría con el modelo esférico, alcanza su meseta sólo de forma asintótica, por lo que no tiene rango finito y se define el rango efectivo del modelo como hemos visto. En la Figura 2.2 se muestran tres ejemplos de esta familia, en los que hemos variado su meseta  $c_s$  y su rango  $a_s$  (1 y 0,3 en negro, 2 y 0,3 en rojo, 1 y 0,8 en verde, respectivamente). Como vemos, los dos parámetros marcan el crecimiento máximo de la función y su velocidad. Cerca del origen tanto el modelo esférico como el exponencial presentan un comportamiento lineal, que es característico de aquellos

fenómenos que muestran una dependencia espacial a corta distancia que disminuye linealmente con la misma. El covariograma y correlograma asociados a este modelo responden a las expresiones siguientes

$$C(\mathbf{h}) = c_e \exp(-h/a_e), \quad (2.25)$$

$$\rho(\mathbf{h}) = \exp(-h/a_e).$$

#### 2.4.5. Modelo gaussiano

Es un modelo válido en  $\mathfrak{R}^d$ , para cualquier  $d \geq 1$ , con la siguiente expresión

$$\gamma(\mathbf{h}) = c_g (1 - \exp(-h^2/a_g^2)), \quad (2.26)$$

siendo  $c_g \geq 0$  la meseta y  $\sqrt{3}a_g$  el rango efectivo. Este modelo da lugar a realizaciones muy suaves de procesos espaciales y su comportamiento parabólico cerca del origen es característico de fenómenos altamente regulares, que difícilmente se observan en la práctica. Como ocurría con el modelo exponencial, el modelo gaussiano es un modelo acotado donde la meseta se alcanza únicamente de forma asintótica. En la Figura 2.2 se muestran tres ejemplos de esta familia en los que hemos variado su meseta  $c_g$  y rango  $a_g$  (1 y 0,3 en negro, 2 y 0,3 en rojo, 1 y 0,8 en verde, respectivamente). Su función de covarianza y su correlograma asociados son

$$C(\mathbf{h}) = c_g \exp(-h^2/a_g^2), \quad (2.27)$$

$$\rho(\mathbf{h}) = \exp(-h^2/a_g^2).$$

#### 2.4.6. Modelo racional cuadrático

Es un modelo válido en  $\mathfrak{R}^d$ , para cualquier dimensión  $d \geq 1$ , con la siguiente expresión

$$\gamma(\mathbf{h}) = \frac{c_r h^2}{1 + h^2/a_r}, \quad (2.28)$$

con  $c_r \geq 0$  la meseta y  $a_r > 0$  el rango. En la Figura 2.2 se muestran tres ejemplos de esta familia, en los que hemos variado la meseta  $c_r$  y el rango  $a_r$  (3 y 0,3 en negro, 6 y 0,3 en rojo, 3 y 0,8 en verde, respectivamente). El covariograma y correlograma asociados a este modelo vienen dados por

$$C(\mathbf{h}) = c_r \left( 1 - \frac{h^2}{1 + h^2/a_r} \right), \quad (2.29)$$

$$\rho(\mathbf{h}) = 1 - \frac{h^2}{1 + h^2/a_r}.$$

### 2.4.7. Modelo ondulado

Este modelo también se conoce como modelo de efecto hoyo, y es válido únicamente en  $\mathfrak{R}$ ,  $\mathfrak{R}^2$  y  $\mathfrak{R}^3$ . Responde a la siguiente expresión

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ si } h = 0 \\ c_w \left(1 - \frac{\sin(h/a_w)}{h/a_w}\right) & , \text{ si } h > 0, \end{cases}$$

siendo  $c_w \geq 0$  y  $a_w > 0$  sus parámetros. Obsérvese que este es un ejemplo de semivariograma que no es monótonamente creciente, que es útil para procesos con un comportamiento periódico donde existe una sucesión entre zonas ricas y pobres. En la Figura 2.2 se muestran tres ejemplos de esta familia donde hemos variado sus parámetros  $c_w$  y  $a_w$  (1,5 y 0,05 en negro, 0,5 y 0,05 en rojo, 1,5 y 0,20 en verde, respectivamente). Su función de covarianza y su correlograma asociado son

$$C(\mathbf{h}) = \begin{cases} c_w & , \text{ si } h = 0 \\ c_w \frac{\sin(h/a_w)}{h/a_w} & , \text{ si } h > 0, \end{cases}$$

$$\rho(\mathbf{h}) = \begin{cases} 1 & , \text{ si } h = 0 \\ \frac{\sin(h/a_w)}{h/a_w} & , \text{ si } h > 0. \end{cases}$$

### 2.4.8. Modelo potencial

Es un modelo válido en  $\mathfrak{R}^d$ , para cualquier  $d \geq 1$ , que responde a la siguiente expresión

$$\gamma(\mathbf{h}) = c_p h^\lambda, \quad h \geq 0, \quad (2.30)$$

donde  $c_p > 0$  es una pendiente positiva y  $0 \leq \lambda < 2$  un factor de crecimiento, que marca el comportamiento del semivariograma cerca del origen. Este modelo no es válido para valores de  $\lambda \geq 2$  al no cumplir la condición (2.13). Es una generalización del modelo lineal ( $\lambda = 1$ ), y no tiene ni covariograma ni correlograma asociados por no ser el proceso que genera estacionario de segundo orden. En la Figura 2.2 se muestran tres ejemplos de esta familia en los que se han variado sus parámetros  $c_p$  y  $\lambda$  (2 y 1,5 en negro, 1 y 1,5 en rojo, 2 y 0,5 en verde, respectivamente).

### 2.4.9. Modelo de Matérn

Este modelo fue obtenido por Whittle (1954) a partir de ecuaciones diferenciales estocásticas para el caso  $\nu = 1$ , y por Matérn (1960) en su forma general. Es por ello que también se conoce como modelo de Whittle-Matérn. Inicialmente fue descartado en la práctica a favor de modelos más sencillos, pero la flexibilidad que ha demostrado en la modelización de una gran cantidad de datos experimentales lo ha convertido en la actualidad en uno de los más utilizados. La expresión de su semivariograma viene dada por

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ si } h = 0 \\ c_m \left(1 - \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}h}{a_m}\right)^\nu \mathcal{K}_\nu \left(\frac{2\sqrt{\nu}h}{a_m}\right)\right) & , \text{ si } h > 0, \end{cases} \quad (2.31)$$

siendo  $c_m \geq 0$ ,  $\nu > 0$  y  $a_m > 0$  los parámetros del modelo. En la Figura 2.2 se muestran cinco ejemplos de esta familia en los que se hemos variado sus parámetros  $c_m$ ,  $\nu$  y  $a_m$  (4, 1 y 2 en negro; 4, 0,4 y 2 en rojo; 2, 1 y 2 en verde; 4, 1 y 1 en azul oscuro; 2, 0,5 y 0,3 en azul claro, respectivamente). El parámetro  $\nu$  controla la suavidad del proceso estocástico subyacente (cuanto mayor es  $\nu$ , más suave es el proceso espacial resultante), mientras que  $a_m$  es un parámetro de escala espacial.  $\Gamma(\cdot)$  es la función Gamma y  $\mathcal{K}_\nu(\cdot)$  es la función de Bessel modificada de segunda especie de orden  $\nu$  (ver, por ejemplo, Abramowitz y Stegun (1965)). Esta familia es muy útil al permitir al usuario especificar el grado de diferenciabilidad del proceso estocástico subyacente. La función de covarianza asociada a estos procesos viene dada por

$$C(\mathbf{h}) = \begin{cases} c_m & , \text{ si } h = 0 \\ \frac{c_m}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}h}{a_m}\right)^\nu \mathcal{K}_\nu\left(\frac{2\sqrt{\nu}h}{a_m}\right) & , \text{ si } h > 0. \end{cases} \quad (2.32)$$

Su función de autocorrelación asociada vendría dada por

$$\rho(\mathbf{h}) = \begin{cases} 1 & , \text{ si } h = 0 \\ \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}h}{a_m}\right)^\nu \mathcal{K}_\nu\left(\frac{2\sqrt{\nu}h}{a_m}\right) & , \text{ si } h > 0. \end{cases}$$

En ocasiones este modelo se reparametriza como

$$C(\mathbf{h}) = \begin{cases} c_m & , \text{ si } h = 0 \\ \frac{c_m}{2^{\nu-1}\Gamma(\nu)} (bh)^\nu \mathcal{K}_\nu(bh) & , \text{ si } h > 0, \end{cases} \quad (2.33)$$

donde  $b = \frac{2\sqrt{\nu}}{a_m} > 0$  es un parámetro de escala espacial.

El modelo de Matérn es muy general y flexible, e incluye tanto el modelo exponencial ( $\nu = 1/2$ ) como el modelo gaussiano ( $\nu \rightarrow \infty$ ). Para  $\nu = 3/2$  obtenemos también una expresión cerrada de su semivariograma, dada por

$$\gamma(\mathbf{h}) = c_m [1 - (1 + \phi h) \exp(-h/a_m)], \quad h \geq 0.$$

#### 2.4.10. Combinación convexa de modelos. Modelos con efecto pepita

Como hemos visto en la Sección 2.2.3, la combinación lineal convexa de covariogramas o semivariogramas producen modelos válidos. Esto permite definir nuevos semivariogramas o covariogramas como la suma de dos de los modelos vistos anteriormente.

En particular, tiene gran utilidad práctica la suma de un modelo de efecto pepita con cualquiera de los otros modelos vistos anteriormente, lo que permite modelizar discontinuidades en el origen debidas a variaciones a pequeña escala o a errores de medida. Por ejemplo, podemos definir el modelo exponencial con efecto pepita como

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ si } h = 0 \\ c_0 + c_e(1 - \exp(-h/a_e)) & , \text{ si } h > 0, \end{cases} \quad (2.34)$$

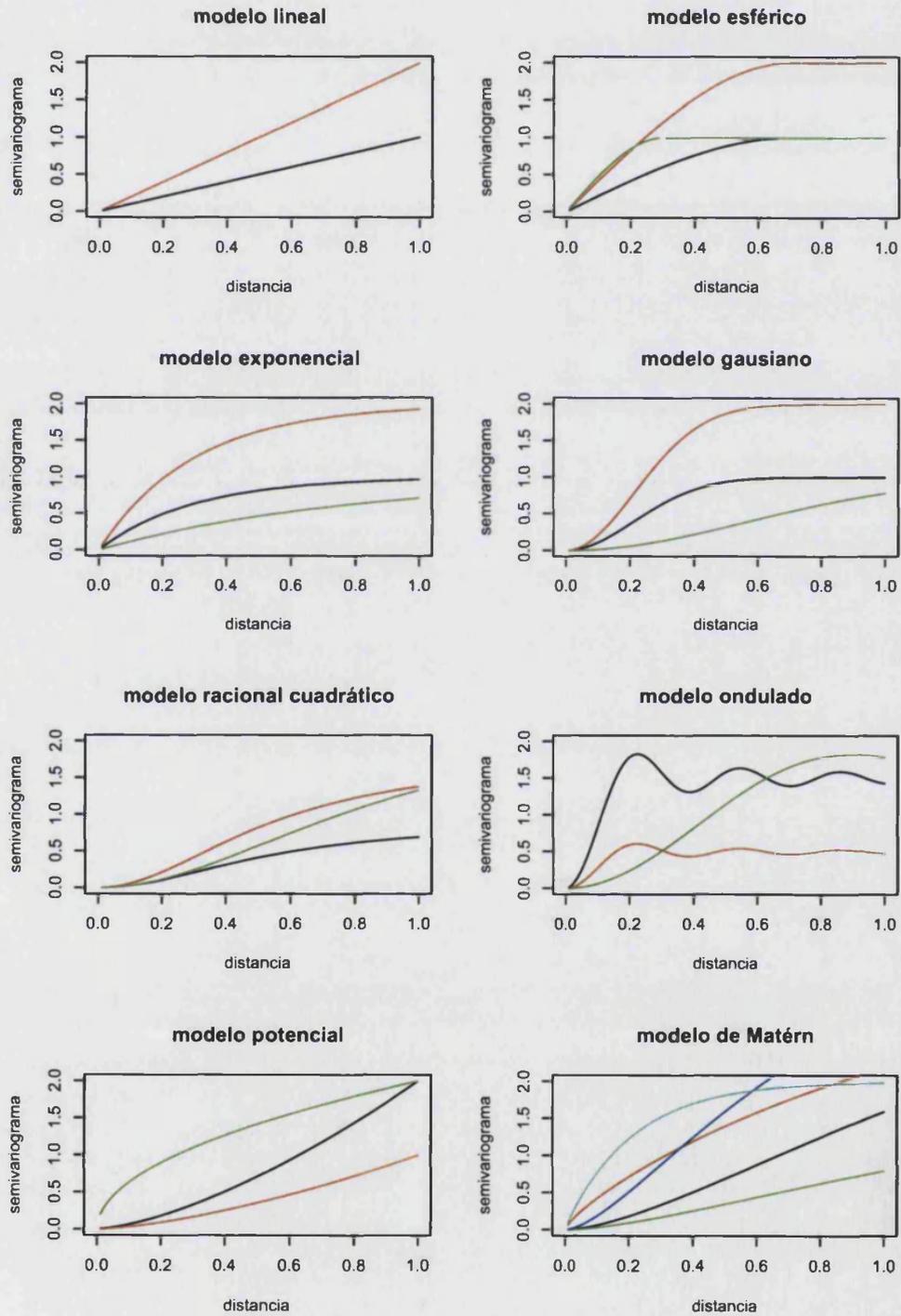


Figura 2.2: Representación gráfica de las principales familias de semivariogramas

donde  $c_0 \geq 0$  es el efecto pepita,  $c_e \geq 0$  la meseta parcial (y por tanto  $c_0 + c_e$  la meseta del modelo) y  $3a_e > 0$  el rango efectivo del semivariograma. Como ya hemos comentado en la Sección 2.2.5, la pepita marcará el comportamiento del semivariograma cerca del origen.

Con frecuencia encontraremos también estructuras anidadas, formadas por la suma de semivariogramas con diferentes rangos. Estos modelos sirven para describir componentes con variaciones espaciales a diferentes escalas que se deben, por tanto, a factores de naturaleza diferente.

#### 2.4.11. Combinación lineal (no necesariamente convexa) de modelos

Todos los modelos que hemos visto son modelos no negativos, es decir, con  $C(\mathbf{h}) \geq 0$  y  $\gamma(\mathbf{h}) \geq 0$ , para todo  $\mathbf{h} \in \mathbb{R}^d$ . No obstante, desde el punto de vista práctico existen determinados fenómenos naturales que requieren para su modelización de modelos que puedan tomar valores negativos, o que incluso oscilen entre valores positivos y negativos mientras tienden a cero cuando la distancia espacial aumenta. Por ello, durante los últimos años se han investigado las combinaciones lineales de modelos isotrópicos que permitan valores negativos de las funciones anteriores. En la Sección 4.12 se muestran los rangos de variación posibles del parámetro  $\theta$  para los que los modelos de la forma  $C(\mathbf{h}) = \theta C_1(\mathbf{h}) + (1 - \theta)C_2(\mathbf{h})$  son funciones de covarianza válidas en  $\mathbb{R}^d$  para diferentes posibles funciones de covarianza espaciales  $C_1(\mathbf{h})$  y  $C_2(\mathbf{h})$ . La ventaja es que estos procedimientos permitirán construir funciones de covarianza que, para determinados pesos negativos, darán lugar a valores negativos o valores oscilantes alrededor de cero. En Mateu et al. (2007b) podemos encontrar una aplicación considerando una combinación lineal de funciones de covarianza del tipo Matérn (2.32) para modelizar la temperatura de una zona termal de Grecia.

### 2.5. Estimación de los parámetros del variograma

Sea  $Z(\cdot)$  un proceso observado sobre un conjunto de localizaciones  $\{s_1, \dots, s_n\}$ . Sean  $\hat{\gamma}(\mathbf{h}_j)$  los valores estimados del semivariograma a partir de los datos aplicando alguno de los métodos que hemos visto en la Sección 2.3. Aunque son muchas las buenas propiedades de estos estimadores, carecen de la propiedad de ser semidefinidos positivos, con lo que sería posible que algunas predicciones espaciales derivadas a partir de tales estimadores presenten varianzas negativas. La forma más común de evitar esta dificultad es reemplazando el semivariograma empírico por algún modelo paramétrico  $\gamma(\mathbf{h}; \theta)$  de los que hemos presentado anteriormente que se aproxime a la dependencia espacial encontrada por el semivariograma empírico, y del que sabemos cumple la condición de ser semidefinido positivo. Obsérvese que, en general, no es necesario restringirse a modelos isotrópicos, aunque suelen ser los primeros que son considerados. El objetivo será elegir de entre todos los semivariogramas posibles  $\{\gamma(\mathbf{h}; \theta), \theta \in \Theta\}$  aquél que mejor se ajuste a las observaciones realizadas, obteniendo con ello un modelo de semivariograma que más tarde será utilizado en el proceso de predicción espacial. En esta sección veremos los principales métodos de estimación.

### 2.5.1. Estimación por mínimos cuadrados

La estimación por *mínimos cuadrados ordinarios* (Ordinary Least Squares, OLS) consiste en obtener el valor  $\hat{\theta}$  que minimiza

$$\sum_{j=1}^n (\hat{\gamma}(\mathbf{h}_j) - \gamma(\mathbf{h}_j; \theta))^2 = (\hat{\gamma} - \gamma(\theta))'(\hat{\gamma} - \gamma(\theta)), \quad (2.35)$$

siendo  $\hat{\gamma} = (\hat{\gamma}(\mathbf{h}_1), \dots, \hat{\gamma}(\mathbf{h}_n))'$  y  $\gamma(\theta) = (\gamma(\mathbf{h}_1; \theta), \dots, \gamma(\mathbf{h}_n; \theta))'$ . Un problema que presenta este procedimiento es que en este caso las estimaciones están correladas y tienen varianzas diferentes.

Una solución es aplicar *mínimos cuadrados generalizados* (Generalized Least Squares, GLS), que consiste en minimizar

$$(\hat{\gamma} - \gamma(\theta))'V(\theta)^{-1}(\hat{\gamma} - \gamma(\theta)). \quad (2.36)$$

siendo  $V(\theta)$  la matriz de varianzas-covarianzas de  $\hat{\gamma}$ , que depende del valor  $\theta$  desconocido y cuyos elementos pueden ser además difíciles de estimar.

Un compromiso entre las dos anteriores es la estimación por *mínimos cuadrados ponderados* (Weighted Least Squares, WLS), que consiste en minimizar

$$\sum_{j=1}^n w_j (\hat{\gamma}(\mathbf{h}_j) - \gamma(\mathbf{h}_j; \theta))^2 = (\hat{\gamma} - \gamma(\theta))'W(\theta)^{-1}(\hat{\gamma} - \gamma(\theta)) \quad (2.37)$$

siendo  $W(\theta)$  una matriz diagonal cuyos elementos son las varianzas de  $\hat{\gamma}$ , que pueden aproximarse bajo la hipótesis de que el proceso es gaussiano y que las estimaciones son incorreladas por  $2\gamma(\mathbf{h}_j; \theta)^2/N(\mathbf{h}_j)$ , con  $N(\mathbf{h}_j)$  el número de localizaciones a distancia  $\mathbf{h}_j$ . Por tanto, los pesos de (2.37) vendrían dados por  $w_j = N(\mathbf{h}_j)/(2\gamma(\mathbf{h}_j; \theta)^2)$ .

En general, los tres estimadores OLS, WLS y GLS aparecen en orden creciente de eficiencia pero decreciente en simplicidad, siendo la estimación por mínimos cuadrados ponderados la más utilizada en la práctica estadística debido a la facilidad de su implementación y a las ventajas computacionales que presenta. No obstante, presenta inconvenientes prácticos como, por ejemplo, que depende de las estimaciones del semivariograma, que son correladas y muy sensibles a la selección de las distancias y las regiones de tolerancia utilizadas para su cálculo.

### 2.5.2. Estimación mediante máxima verosimilitud

Supongamos que el proceso espacial que estamos analizando es un proceso gaussiano que puede descomponerse como

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}),$$

donde la media del proceso  $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\beta$  es una función lineal de un conjunto de  $q$  regresores  $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), \dots, x_q(\mathbf{s}))'$  y  $\delta(\mathbf{s})$  representa el error espacial. En este caso es bastante sencillo obtener la forma exacta de la verosimilitud y maximizarla numéricamente.

Sea  $\mathbf{Z}$  un vector de observaciones del proceso de dimensión  $n$ ,  $\mathbf{X}$  la matriz de diseño de las  $q$  variables observadas de dimensión  $q \times n$  (que generalmente incluirá una columna de unos para una media constante y estacionaria),  $\boldsymbol{\beta}$  el vector de dimensión  $q$  de parámetros desconocidos y  $\boldsymbol{\Sigma}$  la matriz de varianzas-covarianzas de las observaciones. La estimación mediante *máxima verosimilitud* consiste en obtener de forma simultánea los valores de  $\boldsymbol{\beta}$  y  $\boldsymbol{\theta}$  que maximizan la función de distribución conjunta normal multivariante, o lo que es lo mismo, que minimizan -2 veces la log-verosimilitud

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + n \log(2\pi). \quad (2.38)$$

Minimizar la ecuación anterior puede ser computacionalmente demasiado costoso si tenemos muchos parámetros en  $\boldsymbol{\beta}$ , como ocurre a menudo. Generalmente,  $\boldsymbol{\theta}$  sólo incluye tres parámetros (la pepita, la meseta parcial y el rango). Podemos lograr un ahorro computacional minimizando (2.38) en dos fases: en una primera etapa suponemos que  $\boldsymbol{\theta}$  es conocido, y por tanto  $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$  también, con lo que la estimación de  $\boldsymbol{\beta}$  se obtendría mediante el estimador de mínimos cuadrados generalizado  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{z}$ . Podemos ahora sustituir este estimador en la expresión (2.38), obteniendo

$$L_{ML}(\boldsymbol{\theta}) = (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + n \log(2\pi). \quad (2.39)$$

Ahora la expresión (2.39) a minimizar únicamente depende de  $\boldsymbol{\theta}$ , lo que hace el proceso más sencillo y permite obtener  $\hat{\boldsymbol{\theta}}$ . Para obtener  $\hat{\boldsymbol{\beta}}$  bastará con sustituir esta estimación en la expresión del estimador de mínimos cuadrados generalizado. Aunque la estimación por máxima verosimilitud es asintóticamente insesgada, presenta un sesgo considerable para muestras pequeñas porque los elementos de la diagonal de  $(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}$  son demasiado pequeños, lo que produce infraestimaciones de los parámetros  $\boldsymbol{\beta}$ .

La estimación por *máxima verosimilitud restringida* (REML) tiene muchas mejores propiedades de sesgo que la anterior. La función a minimizar en este caso viene dada por la expresión

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + \log(|\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}|) + (n-p) \log(2\pi). \quad (2.40)$$

Como antes, vamos a sustituir el parámetro desconocido  $\boldsymbol{\beta}$  por el estimador de mínimos cuadrados generalizado  $\hat{\boldsymbol{\beta}}$  obtenido a partir de unos valores iniciales de  $\boldsymbol{\theta}$ , con lo que la función a minimizar sería

$$L(\boldsymbol{\theta}) = (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + \log(|\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}|) + (n-p) \log(2\pi). \quad (2.41)$$

Después de minimizar esta expresión obtenemos una estimación  $\hat{\boldsymbol{\theta}}$ , que sustituiremos en la expresión del estimador de mínimos cuadrados generalizado para obtener  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}\mathbf{z}$ .

Cuando la matriz de diseño  $\mathbf{X}$  contiene únicamente una columna de unos y  $\boldsymbol{\beta}$  un único parámetro que representa la media general del proceso, entonces el variograma empírico es insesgado respecto al variograma teórico, y el método de mínimos cuadrados funciona suficientemente bien. En cambio, cuando  $\boldsymbol{\beta}$  tiene múltiples parámetros necesitaremos ajustar un variograma a los residuos del modelo de regresión y las estimaciones dejan de ser

insesgadas. Aunque los procedimientos de máxima verosimilitud se han desarrollado bajo la hipótesis de que los datos provienen de una distribución normal multivariante, desviaciones en la distribución del error del modelo no afectan demasiado a sus estimaciones con estos métodos por ser suficientemente robustos. Otra ventaja es que no se necesitan condiciones previas en el proceso de estimación, tales como distancias máximas y regiones de tolerancia. Además, con estas técnicas podemos utilizar procedimientos de diagnóstico clásicos para los modelos ajustados (Faraway, 2005) como el valor de menos dos veces la log-verosimilitud asociada ( $-2LL$ ), el criterio de información de Akaike (AIC) (Akaike, 1973, 1974) o el criterio de información bayesiano (BIC) (Schwarz, 1978).

El principal problema de los métodos de máxima verosimilitud es que necesitan de la evaluación del determinante y la inversa de la matriz  $\Sigma$ , lo que puede dar lugar a problemas computacionales con muestras de tamaño considerable. Por ejemplo, si utilizamos la descomposición de Cholesky y tenemos  $n$  observaciones del proceso, la aplicación de los métodos de máxima verosimilitud requieren  $\mathcal{O}(n^3)$  operaciones, lo que es prohibitivo en problemas con muchos datos, que es una situación bastante usual en problemas espacio-temporales como veremos en los capítulos siguientes.

### 2.5.3. Estimación mediante verosimilitud compuesta ponderada

La *verosimilitud compuesta* (CL) es un término introducido por Lindsay (1988) para referirse a un procedimiento de estimación que busca reducir la complejidad computacional de los métodos basados en el cálculo de la función de máxima verosimilitud, utilizando para ello una función de pseudo-verosimilitud basada en la suma de verosimilitudes de sucesos marginales o condicionales.

Para adaptar esta técnica al caso espacial vamos a seguir el desarrollo descrito en Curreiro y Lele (1999), que introducen una forma particular de verosimilitud compuesta para la estimación de la covarianza a partir de la nube de valores del semivariograma (Sección 2.7.1) y demuestran que el procedimiento resultante mantiene las propiedades más interesantes de la estimación WLS o de las de máxima verosimilitud. En particular, su aplicación no depende de la elección de las distancias y regiones de tolerancia en la estimación del variograma, es robusto frente a la no especificación de la distribución de los datos, y su orden de computación está entre el orden de la estimación WLS y el de los métodos de máxima verosimilitud.

Sea  $Z(\mathbf{s})$  un proceso estocástico gaussiano intrínsecamente estacionario con semivariograma  $\gamma(\cdot; \boldsymbol{\theta})$ , un modelo paramétrico que depende de un vector paramétrico  $p$ -dimensional desconocido  $\boldsymbol{\theta} \in \Theta$ . Las diferencias del proceso serán gaussianas, con

$$R(\mathbf{r}, \mathbf{s}) = Z(\mathbf{r}) - Z(\mathbf{s}) \sim N(0, 2\gamma(\mathbf{r} - \mathbf{s}; \boldsymbol{\theta})). \quad (2.42)$$

En ese caso, las  $n(n-1)/2$  diferencias correspondientes a las distancias experimentales observadas,  $R_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ , con  $i, j = 1, \dots, n$ ,  $i < j$ , seguirán una distribución gaussiana de media cero y varianza dada por el variograma de la distancia existente entre ellas. Sea  $L(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$  la expresión analítica de la densidad gaussiana asociada a cada  $R_{ij}$ , entonces para cada par  $(i, j)$ , la log-verosimilitud negativa  $cl(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}) = -\log L(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$

vendría dada, eliminando las constantes que no dependen de  $\theta$ , por la función

$$cl(\mathbf{s}_i, \mathbf{s}_j; \theta) = \log(\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)) + \frac{R_{ij}^2}{2\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)}. \quad (2.43)$$

Si sumamos todas estas funciones de log-verosimilitud negativas, obtenemos la siguiente función de verosimilitud compuesta

$$CL(\theta) = \sum_{i < j} cl(\mathbf{s}_i, \mathbf{s}_j; \theta). \quad (2.44)$$

El valor  $\hat{\theta}$  se obtiene minimizando (2.44) respecto a  $\theta \in \Theta$ . Observar que este método no requiere de la inversión de ninguna matriz de varianzas-covarianzas, y su orden de cálculo es de  $\mathcal{O}(n^2)$  operaciones.

En Bevilacqua et al. (2007) los autores presentan una versión ponderada de la verosimilitud compuesta (2.44), donde los pesos son seleccionados siguiendo un criterio de optimalidad basado en la inversa de la matriz de Información de Godambe (Godambe, 1960) asociada a la función de estimación de este procedimiento. Definimos la *verosimilitud compuesta ponderada* como

$$WCL(\theta, h) = \sum_{i < j} cl(\mathbf{s}_i, \mathbf{s}_j; \theta) \omega_{ij}, \quad (2.45)$$

donde los pesos  $\omega_{ij}$  valen uno para pares de localizaciones situadas a una distancia  $\|\mathbf{s}_i - \mathbf{s}_j\| \leq h$  y cero en otro caso, y donde  $h \in H = \{h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|\} \subset \mathfrak{R}$  es el conjunto de distancias euclídeas empíricas. Este método puede usarse como un compromiso entre las dificultades computacionales asociadas al uso de la máxima verosimilitud, y la pérdida de eficiencia producida por el uso de las técnicas de estimación por mínimos cuadrados o por la verosimilitud compuesta clásica.

Sea  $\nabla\gamma(\cdot, \cdot; \theta)$  el vector de derivadas parciales de  $\gamma(\cdot, \cdot; \theta)$  respecto al vector  $\theta$ , se tiene que la función de estimación asociada a este procedimiento vendría dada por

$$s(\theta, h) = \sum_{i < j} \frac{\nabla\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)}{\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)} \left( 1 - \frac{R_{ij}^2}{2\gamma(\mathbf{s}_i - \mathbf{s}_j; \theta)} \right) \omega_{ij}. \quad (2.46)$$

La inversa de la matriz de información de Godambe asociada a esta expresión sería:

$$G(\theta, h) = H(\theta, h)^{-1} J(\theta, h) H(\theta, h)^{-1}, \quad (2.47)$$

donde

$$H(\theta, h) = E[\nabla s(\theta, h)] = \sum_{i < j} \frac{\nabla\gamma_{ij} \nabla\gamma'_{ij}}{\gamma_{ij}^2} \omega_{ij},$$

y

$$J(\theta, h) = E[s(\theta, h)^2] = \frac{1}{2} \sum_{i < j} \sum_{l < k} \frac{\nabla\gamma_{ij} \nabla\gamma'_{lk}}{\gamma_{ij}^2 \gamma_{lk}^2} (\gamma_{il} - \gamma_{jl} - \gamma_{jk} + \gamma_{ik})^2 \omega_{ij} \omega_{lk},$$

siendo  $\gamma_{xy} = \gamma(\mathbf{s}_x - \mathbf{s}_y; \boldsymbol{\theta})$ , con  $x, y = 1, \dots, n$ ,  $x < y$ . El valor  $\hat{\boldsymbol{\theta}}$  se obtiene minimizando  $WCL(\boldsymbol{\theta}, \mathbf{h}^*)$ , donde  $\mathbf{h}^*(\boldsymbol{\theta}) = \min_h \text{tr}(G(\boldsymbol{\theta}, h))$  ( $h^*(\boldsymbol{\theta}) = \min_h G(\boldsymbol{\theta}, h)$  para el caso escalar con  $p = 1$ ).

Como podemos ver, en este procedimiento necesitamos estimar  $\mathbf{h}^*(\boldsymbol{\theta})$  para llevar a cabo la optimización de  $WCL(\boldsymbol{\theta}, \mathbf{h}^*)$ . En Bevilacqua et al. (2007) los autores proponen algunas reglas prácticas para llevar a cabo esta estimación. También realizan un estudio de simulación y una aplicación sobre datos reales para mostrar la eficacia de este procedimiento. En Mateu et al. (2007b) podemos encontrar una aplicación del método de verosimilitud compuesta en el análisis de datos espaciales reales.

## 2.6. Predicción espacial

Sea  $\{Z(\mathbf{s}); \mathbf{s} \in D \subset \mathbb{R}^d\}$  un proceso espacial del que se han observado un conjunto de valores  $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$  sobre  $n$  localizaciones situadas en una determinada región que queremos estudiar. El objetivo de la predicción espacial será, a partir de las observaciones realizadas, obtener una estimación de  $g(Z(B))$ , siendo  $g(Z(\cdot))$  alguna característica de interés del proceso  $Z(\cdot)$  y  $B$  una región de interés en  $D$ .

Generalmente nuestro objetivo será estimar el valor del propio proceso  $Z(\cdot)$  sobre una determinada localización  $\mathbf{s}_0 \in D$ , con lo que  $g(Z(\cdot)) \equiv Z(\cdot)$  y  $B \equiv \{\mathbf{s}_0\}$ . No obstante, en ocasiones puede ser de interés estimar otras cantidades que no son el propio proceso (como ocurre, por ejemplo, al estimar mapas de probabilidad) o sobre regiones que no son un único punto (como en la estimación sobre un bloque).

Aunque existen multitud de procedimientos de predicción espacial, en esta sección nos vamos a centrar en el kriging, que es el término genérico adoptado en geoestadística para dar nombre a una metodología de interpolación basada en una familia de algoritmos de regresión generalizados por mínimos cuadrados que utilizan las propiedades de segundo orden del proceso  $Z(\cdot)$ . Recibe este nombre en reconocimiento a los trabajos pioneros de Danie G. Krige (1951).

A lo largo de esta sección vamos a suponer que el proceso estudiado  $Z(\cdot)$  puede descomponerse como

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}), \quad (2.48)$$

donde  $\mu(\mathbf{s}) = E(Z(\mathbf{s}))$  es la media (o tendencia o variación a gran escala) del proceso y representa la parte determinista del modelo, mientras que  $\delta(\mathbf{s}) = Z(\mathbf{s}) - \mu(\mathbf{s})$  es el error espacial (o variación a pequeña escala) y representa la parte aleatoria, con  $E(\delta(\mathbf{s})) = 0$ . El proceso del error espacial  $\delta(\cdot)$  puede descomponerse como

$$\delta(\mathbf{s}) = W(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad (2.49)$$

siendo  $W(\cdot)$  la variación a pequeña escala que es un proceso intrínsecamente estacionario con media cero y cuyo variograma tiene un rango (si existe) que es mayor que el mínimo de las distancias observadas,  $\eta(\cdot)$  la variación a microescala que es un proceso intrínsecamente estacionario con media cero independiente de  $W(\cdot)$  y cuyo variograma tiene un rango menor que el mínimo de las distancias observadas, y  $\varepsilon(\cdot)$  el error de medida o ruido que es un proceso de ruido blanco con media cero independiente de  $W(\cdot)$  y  $\eta(\cdot)$  y con

$Var(\varepsilon(\mathbf{s})) = c_{ME}$ . También suele utilizarse la siguiente descomposición del proceso  $Z(\cdot)$  como

$$Z(\mathbf{s}) = S(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad (2.50)$$

donde el proceso  $S(\cdot)$  es una versión filtrada del proceso original a la que se le ha eliminado el error de medida, con  $S(\cdot) = \mu(\cdot) + W(\cdot) + \eta(\cdot)$ . Aunque podemos obtener alguna estimación del error de medida si tenemos observaciones repetidas en determinadas localizaciones, o investigar el comportamiento de la variación a microescala si observamos valores a distancias menores que las que presentan los datos, en general las descomposiciones anteriores no serán únicas. En un caso práctico, los procesos  $\mu(\cdot)$  y  $\delta(\cdot)$  no serán directamente observados, sino que lo que en realidad observaremos es una realización del proceso  $Z(\cdot)$  sobre un conjunto de  $n$  localizaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ .

En ocasiones, como ya señalábamos anteriormente, se modeliza la media del proceso,  $\mu(\mathbf{s})$ , como una función lineal de un conjunto de  $k$  regresores o variables explicativas  $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}) \dots x_k(\mathbf{s}))$ , esto es

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta},$$

siendo  $\boldsymbol{\beta}$  el vector de coeficientes de este modelo de regresión. Las covariables a considerar pueden ser tanto superficies polinómicas de las coordenadas en las que se han observado los valores del proceso (generalmente de orden uno o dos) y variables regionalizadas observadas en las localizaciones analizadas con relación con el proceso considerado.

Vamos a partir del problema de la predicción del proceso original  $Z(\cdot)$  sobre una determinada localización  $\mathbf{s}_0 \in D$ , esto es, de la predicción puntual sin considerar error de medida, y luego generalizaremos los resultados obtenidos para estos problemas más generales.

El predictor que el kriging utiliza para la predicción de  $Z(\mathbf{s}_0)$  es un predictor lineal de los datos

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad (2.51)$$

donde  $\lambda_i$  es el peso asignado a cada uno de los datos que intervienen en el sumatorio. De todos los posibles predictores lineales, la técnica de kriging selecciona aquel predictor lineal que sea insesgado ( $E[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)] = 0$ ) y además sea óptimo en el sentido que minimice la varianza del error cuadrático medio de la predicción ( $\min Var[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)]$ ). Por este motivo el predictor resultante recibe el nombre de BLUP (Best Linear Unbiased Predictor).

Si denotamos por  $\sigma_K^2(\mathbf{s}_0)$  el error cuadrático medio de la predicción, bajo la condición de insesgadez vendrá dado por

$$\sigma_K^2(\mathbf{s}_0) = Var \left[ \left( Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0) \right) \right] = E \left( Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0) \right)^2. \quad (2.52)$$

A continuación vamos a describir algunos de los tipos de kriging más utilizados en la práctica geoestadística, como son el kriging ordinario (en el que suponemos que la media es constante pero desconocida), el kriging universal (que se basa en la suposición de que la media es desconocida y no es una constante, aunque sí una función lineal de un conjunto

de variables que dependen de la localización) y el kriging suavizado mediante la mediana (que se basa en la suposición de que la media es desconocida y no constante, pero que puede descomponerse como suma de un efecto debido a la primera coordenada y un efecto debido a la segunda). En Cressie (1993) podemos encontrar un mayor desarrollo de cada uno de estos métodos y otros procedimientos de gran aplicación como el kriging indicador, el kriging lognormal, el kriging transgaussiano, el kriging robusto o el cokriging para casos en los que el fenómeno observado es multinomial.

### 2.6.1. Kriging ordinario

Supongamos que el proceso espacial  $Z(\cdot)$  es intrínsecamente estacionario con variograma  $2\gamma(\cdot)$ . Tenemos entonces que  $\delta(\cdot)$  también es un proceso intrínsecamente estacionario con el mismo variograma que  $Z(\cdot)$ , ya que

$$\begin{aligned} 2\gamma_\delta(\mathbf{h}) &= \text{Var}(\delta(\mathbf{s}_1) - \delta(\mathbf{s}_2)) = \text{Var}(Z(\mathbf{s}_1) - \mu(\mathbf{s}_1) + Z(\mathbf{s}_2) - \mu(\mathbf{s}_2)) = \\ &= \text{Var}(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = 2\gamma(\mathbf{h}). \end{aligned} \quad (2.53)$$

En este caso consideramos que la tendencia del proceso  $\mu(\mathbf{s})$  es desconocida pero constante a lo largo de toda el área de estudio  $D$ , es decir,  $\mu(\mathbf{s}) = \mu, \forall \mathbf{s} \in D$ . La insesgadez del predictor

$$\hat{Z}_{KO}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad (2.54)$$

implica que los pesos  $\lambda_i$  deben sumar uno, ya que

$$E[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)] = E[Z(\mathbf{s}_0)] - E\left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right] = \mu - \sum_{i=1}^n \lambda_i \mu = 0 \leftrightarrow \sum_{i=1}^n \lambda_i = 1. \quad (2.55)$$

Nuestro objetivo es calcular los pesos  $\lambda_i$  de (2.54) que minimizan (2.52) bajo la restricción de insesgadez dada en (2.55). Si definimos  $m$  el multiplicador de Lagrange asociado a la restricción de insesgadez, este problema es equivalente a minimizar la expresión

$$E \left( Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 - 2m \left( \sum_{i=1}^n \lambda_i - 1 \right),$$

donde

$$\begin{aligned} \left( Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 &= \left( \sum_{i=1}^n \lambda_i Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 \\ &= \left( \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_0) - Z(\mathbf{s}_i)) \right)^2 \\ &= \left( \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (Z(\mathbf{s}_0) - Z(\mathbf{s}_i))(Z(\mathbf{s}_0) - Z(\mathbf{s}_j)) \right)^2. \end{aligned}$$

Se tiene que

$$\begin{aligned} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2 &= (Z(\mathbf{s}_i) - Z(\mathbf{s}_0) + Z(\mathbf{s}_0) - Z(\mathbf{s}_j))^2 = \\ &= (Z(\mathbf{s}_i) - Z(\mathbf{s}_0))^2 + (Z(\mathbf{s}_j) - Z(\mathbf{s}_0))^2 + 2(Z(\mathbf{s}_i) - Z(\mathbf{s}_0))(Z(\mathbf{s}_j) - Z(\mathbf{s}_0)). \end{aligned}$$

Luego

$$\begin{aligned} E \left( Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 &= E \left( \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (Z(\mathbf{s}_0) - Z(\mathbf{s}_i))(Z(\mathbf{s}_0) - Z(\mathbf{s}_j)) \right)^2 \\ &= 2 \sum_{i=1}^n \lambda_i E[Z(\mathbf{s}_0) - Z(\mathbf{s}_i)]^2 / 2 - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 / 2. \end{aligned}$$

Así, la expresión a minimizar sería

$$2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) - 2m \left( \sum_{i=1}^n \lambda_i - 1 \right). \quad (2.56)$$

Si derivamos respecto a  $\lambda_i$  y  $m$  e igualamos a cero, tenemos el siguiente sistema de ecuaciones lineales

$$\begin{cases} -\sum_{j=1}^n \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + \gamma(\mathbf{s}_0 - \mathbf{s}_i) - m = 0 & i = 1, \dots, n \\ \sum_{i=1}^n \lambda_i = 1, \end{cases} \quad (2.57)$$

que puede expresarse de forma matricial como

$$\begin{pmatrix} \Gamma & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma} \\ 1 \end{pmatrix}, \quad (2.58)$$

siendo  $\Gamma$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ ,  $\boldsymbol{\gamma}$  el vector dimensión  $n$  cuyo elemento  $i$ -ésimo es  $\gamma(\mathbf{s}_0 - \mathbf{s}_i)$ ,  $\mathbf{1}$  el vector unidad de dimensión  $n$  y  $\boldsymbol{\lambda}$  el vector con los coeficientes del predictor del kriging simple.

Resolviendo (2.58), llegamos a que

$$\boldsymbol{\lambda}' = \left( \boldsymbol{\gamma} + \mathbf{1} \frac{(1 - \mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma})}{\mathbf{1}'\Gamma^{-1}\mathbf{1}} \right)' \Gamma^{-1} \quad (2.59)$$

$$m = -\frac{(1 - \mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma})}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}. \quad (2.60)$$

En este caso, el error cuadrático medio de la predicción que hemos minimizado (y que también se conoce como varianza del kriging) vendría dado por la siguiente expresión

$$\sigma_{KO}^2(\mathbf{s}_0) = \boldsymbol{\gamma}'\Gamma^{-1}\boldsymbol{\gamma} - \frac{(\mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma} - 1)^2}{\mathbf{1}'\Gamma^{-1}\mathbf{1}} = 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0 - \mathbf{s}_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j). \quad (2.61)$$

Al aplicar esta técnica, hemos de tener en cuenta ciertos aspectos computacionales y operativos. Obsérvese que los pesos del estimador (2.54) dependen de la localización sobre la que se desee predecir el valor del fenómeno observado, por lo que deben ser calculados de nuevo para cada predicción. Aunque en general se debe resolver el sistema (2.58) con todos los datos observados, en la práctica  $n$  suele ser bastante grande y esto implica que la matriz a invertir es demasiado grande, por lo que suele limitarse en la práctica el número de localizaciones que intervienen en la predicción considerando únicamente aquellas situadas a una distancia menor al rango del variograma estimado. Los pesos  $\lambda_0$  se calculan mediante la resolución de un sistema lineal de  $n+1$  ecuaciones con  $n+1$  incógnitas, por lo que podemos encontrarnos con todos aquellos problemas inherentes a este tipo de sistemas, como son las matrices mal condicionadas o numéricamente no invertibles.

Todas estas expresiones pueden escribirse en términos de la función de covarianza en el caso de tener un proceso estacionario de segundo orden. Supongamos que tenemos un proceso  $Z(\cdot)$  estacionario de segundo orden de media cero y covariograma  $C(\cdot)$ . En este caso

$$\begin{aligned} \left( Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 &= \left( Z(\mathbf{s}_0) - \mu + \mu - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 \\ &= (Z(\mathbf{s}_0) - \mu)^2 + \left( \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) - \mu \right)^2 - 2(Z(\mathbf{s}_0) - \mu) \left( \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) - \mu \right) \\ &= (Z(\mathbf{s}_0) - \mu)^2 + \left( \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i) - \mu) \right)^2 - 2(Z(\mathbf{s}_0) - \mu) \left( \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i) - \mu) \right) \\ &= (Z(\mathbf{s}_0) - \mu)^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (Z(\mathbf{s}_i) - \mu) (Z(\mathbf{s}_j) - \mu) - 2 \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_0) - \mu) (Z(\mathbf{s}_i) - \mu). \end{aligned}$$

Tomando esperanzas tenemos que

$$E \left( Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 = C(\mathbf{0}) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i - \mathbf{s}_j) - 2 \sum_{i=1}^n \lambda_i C(\mathbf{s}_0 - \mathbf{s}_i).$$

Luego la expresión a minimizar sería

$$C(\mathbf{0}) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i - \mathbf{s}_j) - 2 \sum_{i=1}^n \lambda_i C(\mathbf{s}_0 - \mathbf{s}_i) - 2m \left( \sum_{i=1}^n \lambda_i - 1 \right).$$

Si derivamos como antes respecto a  $\lambda_i$  y  $m$  e igualamos a cero, tenemos un sistema de ecuaciones lineales que puede expresarse en forma matricial como

$$\begin{pmatrix} \Sigma & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ m \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ 1 \end{pmatrix},$$



siendo  $\Sigma$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $C(\mathbf{s}_i - \mathbf{s}_j)$  y  $\mathbf{c}$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $C(\mathbf{s}_0 - \mathbf{s}_i)$ . El predictor del kriging simple sería en este caso

$$\hat{Z}_{KO}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad (2.62)$$

con

$$\lambda' = \left( \mathbf{c} + \mathbf{1} \frac{(1 - \mathbf{1}'\Sigma^{-1}\mathbf{c})}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \right)' \Sigma^{-1} \quad (2.63)$$

$$m = \frac{(1 - \mathbf{1}'\Sigma^{-1}\mathbf{c})}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \quad (2.64)$$

La varianza del kriging vendría dada por

$$\sigma_{KO}^2(\mathbf{s}_0) = C(\mathbf{0}) - \lambda'\mathbf{c} + \frac{(1 - \mathbf{1}'\Sigma^{-1}\mathbf{c})}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \quad (2.65)$$

Como hemos comentado anteriormente, en determinadas aplicaciones prácticas nuestro interés no será conocer el proceso directamente observado,  $Z(\cdot)$ , sino una versión del mismo libre de ruido. Supongamos que

$$Z(\mathbf{s}) = S(\mathbf{s}) + \epsilon(\mathbf{s}),$$

donde  $S(\cdot)$  es el proceso realmente de interés libre del error de medida, y  $\epsilon(\cdot)$  es un proceso de ruido blanco que representa el error de medida, con media cero y varianza  $\sigma_\epsilon^2 > 0$ . En este caso nuestro objetivo será predecir  $S(\mathbf{s}_0) = W(\mathbf{s}_0) + \eta(\mathbf{s}_0)$ , que es el valor del fenómeno investigado en la localización  $\mathbf{s}_0$  una vez eliminado el error de medida (debido, por ejemplo, al instrumento de medición). El estimador que utilizaremos será

$$\hat{S}(\mathbf{s}_0) = \sum_{i=1}^n \nu_i Z(\mathbf{s}_i),$$

donde los valores óptimos de los coeficientes  $\nu_i$  (es decir, que dan lugar a un estimador insesgado y que minimiza el error cuadrático medio de la predicción) se obtienen resolviendo el sistema de ecuaciones

$$\begin{pmatrix} \Gamma & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu} \\ m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma}^* \\ 1 \end{pmatrix},$$

donde  $\Gamma$  es como antes la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ ,  $\boldsymbol{\nu}$  el vector con los coeficientes del predictor del kriging simple, pero donde ahora  $\boldsymbol{\gamma}^*$  es un vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $\gamma^*(\mathbf{s}_0 - \mathbf{s}_i)$ , que se define como  $\gamma_Z(\mathbf{s}_0 - \mathbf{s}_i)$  para cualquier localización no observada y como  $c_{ME}$  para aquellas localizaciones observadas, por lo que no es igual a cero en este último caso como ocurría antes. En este caso, el predictor definido ya no es un interpolador exacto sobre las localizaciones muestreadas (esto es,  $\hat{S}(\mathbf{s}_i) \neq Z(\mathbf{s}_i)$ ) como ocurría en el caso anterior. El error cuadrático medio de la predicción vendría dado por

$$\sigma^2(\mathbf{s}_0) = \sum_{i=1}^n \nu_i \gamma^*(\mathbf{s}_0 - \mathbf{s}_i) + m - c_{ME}.$$

Aunque las estimaciones sobre puntos no observados coinciden en ambos casos ( $\widehat{Z}(s_0) = \widehat{S}(s_0)$ ), no ocurre lo mismo con su error cuadrático medio, que serán iguales únicamente en el caso en que  $c_{ME} = 0$ . Aunque tanto  $\widehat{Z}(s_0)$  como  $\widehat{S}(s_0)$  son discontinuos alrededor de  $s_0$ , el primero es mucho más rugoso mientras que el segundo da estimaciones mucho más suaves.

En todas las ecuaciones anteriores hemos supuesto que los datos y el predictor tenían el mismo soporte o nivel de agregación, localizaciones puntuales sobre la región de estudio. Pero en ocasiones lo que se desea es, para una determinada región de interés  $B \subseteq D$  cuya localización y geometría es conocida y cuyo volumen d-dimensional es  $|B|$ , predecir el valor del proceso  $Z(B)$ , a partir de los valores observados en un conjunto de  $n$  localizaciones  $\{s_1, \dots, s_n\}$ . Las ecuaciones del kriging quedarían en este caso

$$\begin{pmatrix} \Gamma & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\nu} \\ m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\gamma} \\ 1 \end{pmatrix},$$

donde ahora  $\boldsymbol{\gamma}$  es un vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es

$$\gamma(B - s_i) = \int_B \gamma(\mathbf{v} - s_i) d\mathbf{v} / |B|, \quad i = 1, \dots, n.$$

### 2.6.2. Kriging universal

El kriging universal generaliza el kriging ordinario, permitiendo que el valor medio del proceso no sea constante sino una combinación lineal de funciones conocidas o covariables ligadas a las mismas localizaciones. De esta forma, el kriging universal incorpora términos de regresión y correlación espacial.

La hipótesis de partida es que el proceso objeto de estudio  $Z(\mathbf{s})$  puede descomponerse como

$$Z(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta} + \delta(\mathbf{s}), \quad (2.66)$$

donde  $\mathbf{x}(\mathbf{s}) = (1, x_1(\mathbf{s}), \dots, x_p(\mathbf{s}))'$  es un vector que contiene variables explicativas ligadas a la localización espacial  $\mathbf{s}$ ,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$  un vector de parámetros desconocidos y  $\delta(\cdot)$  un proceso de error espacial intrínsecamente estacionario de media cero y variograma  $2\gamma(\cdot)$ .

El vector de datos  $\mathbf{Z} = (Z(s_1), \dots, Z(s_n))'$  puede escribirse como

$$\mathbf{Z} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\delta}, \quad (2.67)$$

siendo  $\mathbf{X}$  la matriz  $n \times (p+1)$  cuya primera columna está formada por unos y las otras por variables explicativas cualesquiera. Si la matriz de diseño  $\mathbf{X}$  contiene únicamente funciones polinomiales de las coordenadas espaciales de la localización  $\mathbf{s}$ , entonces el método de predicción recibe el nombre de kriging universal con tendencia interna.

Como hicimos en el kriging ordinario, queremos predecir el valor de  $Z(s_0)$  a partir del conjunto de observaciones  $\mathbf{Z}$ . Para ello, utilizaremos un predictor lineal de la forma

$$\widehat{Z}(s_0) = \boldsymbol{\lambda}' \mathbf{Z} = \sum_{i=1}^n \lambda_i Z(s_i), \quad (2.68)$$

donde los coeficientes  $\lambda_i$  se obtienen al imponer que el estimador resultante sea insesgado y de mínima varianza. La insesgaredad de (2.68) implica que  $\lambda'X = \mathbf{x}(s_0)'$ , ya que en este caso

$$E[\widehat{Z}(s_0)] = E[\lambda'Z] = \lambda'E[Z] = \lambda'X\beta = \mathbf{x}(s_0)'\beta = E[Z(s_0)]. \quad (2.69)$$

Obsérvese que el kriging universal es una generalización del kriging ordinario, ya que para  $X = \mathbf{1}$  y  $\beta = \mu$  tenemos el modelo  $Z(s) = \mu + \delta(s)$  que asumíamos en aquel y además la condición (2.69) se reduce a (2.55).

Procediendo como en el caso del kriging ordinario, llegamos al siguiente sistema de ecuaciones lineales

$$\begin{pmatrix} \Gamma & X \\ X' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \lambda \\ m \end{pmatrix} = \begin{pmatrix} \gamma \\ \mathbf{x}(s_0) \end{pmatrix},$$

siendo  $\Gamma$  una matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $\gamma(\mathbf{s}_i - \mathbf{s}_j)$ ,  $\gamma$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $\gamma(\mathbf{s}_0 - \mathbf{s}_i)$  y  $m$  un vector de  $p+1$  multiplicadores de Lagrange.

Resolviendo este sistema de ecuaciones, tenemos que

$$\begin{aligned} \lambda' &= (\gamma + X(X'\Gamma^{-1}X)^{-1}(\mathbf{x}(s_0) - X'\Gamma^{-1}\gamma))'\Gamma^{-1} \\ m &= -(X'\Gamma^{-1}X)^{-1}(\mathbf{x}(s_0) - X'\Gamma^{-1}\gamma). \end{aligned} \quad (2.70)$$

El error cuadrático de la predicción sería, en este caso

$$\sigma^2(s_0) = \gamma'\Gamma^{-1}\gamma - (\mathbf{x}(s_0) - X'\Gamma^{-1}\gamma)'\Gamma^{-1}X(\mathbf{x}(s_0) - X'\Gamma^{-1}\gamma). \quad (2.71)$$

Podemos expresar también todas estas ecuaciones en función de la función de covarianza del proceso  $C(\cdot)$ . En este caso

$$\begin{aligned} \lambda' &= (\mathbf{c} + X(X'\Sigma^{-1}X)^{-1}(\mathbf{x}(s_0) - X'\Sigma^{-1}\mathbf{c}))'\Sigma^{-1} \\ m &= -(X'\Sigma^{-1}X)^{-1}(\mathbf{x}(s_0) - X'\Sigma^{-1}\mathbf{c}), \end{aligned} \quad (2.72)$$

siendo  $\Sigma$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $C(\mathbf{s}_i - \mathbf{s}_j)$  y  $\mathbf{c}$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $C(\mathbf{s}_0 - \mathbf{s}_i)$ .

Respecto a la estimación óptima de los parámetros de la media  $\beta$ , como los datos  $Z$  satisfacen un modelo lineal general con  $E(Z) = X\beta$  y  $Var(Z) = \Sigma$ , puede obtenerse por mínimos cuadrados generalizados como

$$\widehat{\beta}_{gls} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Z,$$

con  $Var(\widehat{\beta}_{gls}) = (X'\Sigma^{-1}X)^{-1}$ . A partir de estas expresiones, podemos obtener intervalos de confianza para estos parámetros o combinaciones lineales de los mismos. Se deduce que aun cuando para estimar los parámetros de la media de forma óptima es necesario conocer las covarianzas involucradas en  $\Sigma$ , para la predicción óptima de  $Z(s_0)$  únicamente hemos de conocer los variogramas implicados en  $\Gamma$ .

### 2.6.3. Kriging suavizado mediante la mediana

El kriging suavizado mediante la mediana (median-polish kriging) constituye una aproximación diferente al problema de predicción espacial. En lugar de intentar una identificación óptima del proceso estocástico, este método está basado en la reconstrucción robusta de una superficie irregular. La idea básica es, como en los métodos anteriores, la descomposición del proceso espacial  $Z(\mathbf{s})$  como suma de una función media desconocida  $\mu(\mathbf{s})$  y una estructura de error  $\delta(\mathbf{s})$ , pero ahora la función media se supondrá suma de componentes direccionales. Por ejemplo, para  $\mathcal{R}^2$  suponemos que

$$\mu(\mathbf{s}) = a + c(x) + r(y), \quad \mathbf{s} = (x, y) \in D. \quad (2.73)$$

Supongamos que los valores  $\{\mathbf{s}_i, i = 1, \dots, n\}$  han sido observados sobre una red regular, es decir, son de la forma  $\{(x_l, y_k), l = 1, \dots, p, k = 1, \dots, q\}$ . Podemos entonces denotar la media del proceso en términos de los índices de fila y columna

$$\mu(\mathbf{s}_i) = a + r_k + c_l, \quad \mathbf{s}_i = (x_l, y_k) \in D. \quad (2.74)$$

En este caso, la estimación de esta superficie puede realizarse mediante el algoritmo del suavizado mediante la mediana basado en el análisis robusto de la varianza, que es un procedimiento iterativo que proporciona una estimación de los parámetros anteriores  $a$ ,  $r_k, k = 1, \dots, q$  y  $c_l, l = 1, \dots, p$ . En Cressie (1993) podemos encontrar los detalles teóricos del algoritmo de ajuste de este procedimiento y sus propiedades. El número de observaciones en cada nodo no necesariamente debe ser uno, sino que puede aplicarse para casos en que tengamos cero observaciones en algún nodo o un número de observaciones superior a uno.

En el caso de que las observaciones no estén dispuestas sobre una red regular, un método aproximado puede aplicarse sobre un mapa de baja resolución. Para ello definimos un grid  $\{(x_l, y_k), l = 1, \dots, p, k = 1, \dots, q\}$  y asignamos cada localización muestreada  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  al nodo más cercano del mismo, con lo que tenemos un valor  $Z(x_l, y_k)$  obtenido a partir de las observaciones  $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$  para esta nueva resolución.

## 2.7. Comprobación de las hipótesis del modelo

### 2.7.1. Técnicas exploratorias

Las técnicas del análisis exploratorio de datos (EDA) son un conjunto de herramientas ampliamente utilizadas en la estadística clásica, tanto en la descripción previa de los datos como en la comprobación de la validez de los modelos estadísticos considerados. En Tukey (1977) podemos consultar las principales técnicas del EDA, que son tanto gráficas (diagramas de cajas, diagramas de dispersión, histogramas, diagramas de tallos y hojas,...) como cuantitativas (suavizado mediante la mediana, suavizado resistente,...). Todas estas técnicas también son muy útiles para el análisis de datos espaciales, aunque hay que tener en cuenta que en este caso las observaciones no son independientes y que los valores atípicos deben ser considerados junto a su localización, para ver su valor respecto a los valores cercanos o respecto a covariables que se deban tener en consideración. Son muy útiles

en este punto los paquetes que permiten la vinculación de los gráficos con la localización espacial de los datos, de forma que al seleccionar un área determinada se ofrezcan los gráficos de resumen de dicha región. En una primera aproximación al fenómeno a investigar suele realizarse un histograma o un diagrama de tallos y hojas de los valores observados para analizar su distribución o la presencia de posibles valores atípicos, aunque siempre teniendo en cuenta lo dicho anteriormente sobre este tipo de gráficos en presencia de dependencia de los datos.

Desde el punto de vista espacial, el primer paso antes de cualquier intento de análisis de los datos es la realización de un mapa de las localizaciones en las que se han observado los datos. Esto permite conocer su disposición espacial sobre la superficie de estudio, observando si se distribuyen de forma regular o por el contrario se concentran en determinadas zonas, lo que puede tener gran trascendencia, como en cualquier muestreo, en los resultados del análisis efectuado. Es importante tener en cuenta también la dispersión de las localizaciones muestreadas respecto a covariables que puedan influir en el fenómeno analizado, para evitar concentraciones que afecten las conclusiones finales.

A continuación podemos representar los valores observados junto con las localizaciones asociadas mediante un gráfico tridimensional, como un diagrama de dispersión de los  $Z(\mathbf{s}_i)$  frente a los  $\mathbf{s}_i$  (en el que se suele añadir una línea vertical desde cada valor observado al plano de localizaciones para facilitar su observación) o una superficie suavizada tridimensional que pase por los valores del fenómeno observados. Estos gráficos permiten evaluar el proceso espacial de forma global, observando si existen zonas con valores mayores del proceso o con mayor variabilidad.

Como ya hemos visto en las secciones anteriores, los procesos espaciales  $Z(\mathbf{s})$  suelen descomponerse como suma de una tendencia global  $\mu(\mathbf{s})$  y un error espacial  $\delta(\mathbf{s})$ . Para explorar la tendencia global del proceso, son muy útiles los gráficos de los valores observados del proceso proyectados sobre cada una de las coordenadas consideradas. Estos valores pueden resumirse agrupando las coordenadas observadas en subintervalos y representando mediante diagramas de cajas los valores del proceso correspondientes. Estos gráficos permiten observar como varían los valores de  $Z(\cdot)$  en función de cada una de las coordenadas, para ver si es asumible la hipótesis de estacionariedad con media constante, o por el contrario, hay que ajustar previamente una superficie de tendencia en función de las coordenadas o de un polinomio de las mismas. También puede ocurrir que dispongamos de información auxiliar regionalizada que ayude a la modelización del fenómeno observado, y que por tanto sea conveniente su inclusión en el modelo de la tendencia global.

Para estudiar el comportamiento del proceso a pequeña escala, podemos generar en primer lugar una nube del semivariograma, esto es, representar los pares  $(\|\mathbf{s}_i - \mathbf{s}_j\|, (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2)$ . Generalmente este diagrama de dispersión tiene demasiado ruido, aunque puede permitir detectar valores anómalos. Con el fin de reducir algo este ruido, suele construirse también el semivariograma empírico, tal y como hemos visto en la Sección 2.3, y que permite evaluar el tipo de dependencia espacial presente en los datos. Todos estos análisis deben llevarse a cabo sobre los residuos, a no ser que podamos asumir una media constante en los datos.

También son muy útiles los diagramas de dispersión para diferentes distancias  $h$ , esto es, representar los pares  $(Z(\mathbf{s}_i), Z(\mathbf{s}_i + \mathbf{h}))$  para una distancia fija  $h$ . La comparación de

estos gráficos podría revelar la presencia de isotropía o quizá de no estacionariedad.

Supongamos que definimos, para cada localización, una determinada vecindad. En ese caso podemos calcular, para cada una de ellas, la media y varianza muestral, e incluso un coeficiente de correlación muestral. Los gráficos de estos estadísticos pueden ser muy informativos. Así, el diagrama de las medias puede darnos una idea de cómo es la estructura de la media sobre la región analizada. El de las varianzas y el de los coeficientes de correlación pueden darnos información sobre la no estacionariedad. Si la media no puede considerarse constante, entonces deberemos realizar estos gráficos con los residuos para obtener información válida para el análisis espacial.

Para evaluar la posible anisotropía del proceso analizado, es muy útil calcular diferentes semivariogramas direccionales, generalmente para los ángulos de  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  y  $135^\circ$ , y con una tolerancia angular de  $22,5^\circ$ , con el fin de evaluar la dependencia espacial en diferentes direcciones. Hay que tener en cuenta que las conclusiones de este análisis deben estar relacionadas con los tamaños muestrales que intervienen en la estimación de las estimaciones de cada uno de los semivariogramas direccionales, ya que sino tenemos muchos puntos los cálculos pueden basarse en pocos elementos y dar lugar a estimaciones muy erráticas, de las que es peligroso extraer conclusiones definitivas. También podemos utilizar un diagrama de rosas para evaluar la posibilidad de anisotropía geométrica, ya que en ese caso observaremos formas elípticas en este gráfico. Otra posibilidad es construir un gráfico tridimensional del valor del semivariograma en función de la distancia direccional para estudiar direcciones más continuas o aquellas en las que la dependencia espacial es más acentuada.

### 2.7.2. Test de independencia

Hemos visto que el variograma es una herramienta muy utilizada en el análisis de datos espaciales, y que su forma proporciona información de gran utilidad sobre el tipo de correlación espacial presente en los datos. Dibalasi y Bowman (2001) muestran un uso complementario del variograma, como herramienta para evaluar la existencia o no de correlación espacial en los datos antes de iniciar el estudio geoestadístico. Este tipo de prueba es de gran interés, ya que en el caso de que no exista suficiente evidencia para considerar que existe cualquier tipo de correlación espacial, podemos adoptar modelos más simples basados en errores independientes.

Supongamos que tenemos un proceso espacial  $Z(\cdot)$  gaussiano, intrínsecamente estacionario e isotrópico que puede descomponerse como

$$Z(\mathbf{s}) = \mu + \delta(\mathbf{s}), \quad (2.75)$$

donde  $\mu = E[Z(\mathbf{s})]$  es la media constante del proceso y  $\delta(\cdot)$  es un proceso de error gaussiano con media cero y variograma  $2\gamma(\cdot)$  igual al del proceso original.

Si no existe correlación espacial en el proceso observado, entonces  $\gamma(h) = \sigma^2$  es una constante. Luego cualquier test estadístico que trate de contrastar la independencia en el modelo deberá evaluar la evidencia, expresada en términos del variograma empírico obtenido a partir de los datos observados, de que éste no cambia con la distancia.

En la Sección 2.3 hemos visto los principales estimadores del variograma propuestos en la literatura. El estimador clásico dado por (2.17) se basa en el cuadrado de las diferencias  $R_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$ , aunque como ya dijimos da lugar a un estimador muy influenciado por la presencia de observaciones atípicas en los datos. Cressie y Hawkins sugieren una modificación robusta del anterior dada por (2.18), basada en  $|R_{ij}|^{1/2}$ , que tiene la ventaja adicional de presentar una distribución aproximadamente normal. Debido a la isotropía del proceso, bastará con considerar únicamente los  $n(n-1)/2$  pares  $(h_{ij}, d_{ij}) = (\|\mathbf{s}_i - \mathbf{s}_j\|, |R_{ij}|^{1/2})$  para  $i, j = 1, \dots, n$ , con  $i < j$ . A partir de estos pares, utilizaremos regresión no paramétrica para crear una estimación suavizada del variograma correspondiente, que vendrá definida por

$$\hat{\gamma}^*(h) = \sum_{i < j} \omega_{ij} d_{ij},$$

donde los pesos  $\omega_{ij}$  suman uno y disminuyen con la distancia de  $h_{ij}$  al punto de estimación  $h$ . Existen gran variedad de formas de regresión no paramétrica disponibles, aunque aquí se utiliza la regresión lineal local, en la que para cada distancia de interés  $h$  se define la estimación del semivariograma como la solución  $\hat{\alpha}$  al problema de mínimos cuadrados

$$\min_{\alpha, \beta} \sum_{i < j} (d_{ij} - \alpha - \beta(h_{ij} - h))^2 \frac{1}{b} \phi\left(\frac{h_{ij} - h}{b}\right),$$

donde  $\phi$  denota la función de densidad de la distribución normal estándar y  $b$  denota un parámetro de amplitud de banda.

Como hemos comentado anteriormente, si los errores son independientes el semivariograma  $\gamma(h)$  es constante. Cualquier evidencia significativa de cambio en la curva de regresión no paramétrica  $\hat{\gamma}^*(h)$  con la distancia  $h$  indicaría la presencia de correlación espacial. En regresión lineal, las comparaciones entre modelos se realizan generalmente mediante las correspondientes sumas de cuadrados residuales. Aplicando esta idea a nuestro problema, compararemos la variación de las diferencias observadas  $d_{ij}$  alrededor de su media  $\bar{d}$ , y la variación alrededor de los valores  $\bar{d}_{ij} = \hat{\gamma}^*(h_{ij})$  obtenidos a partir de la curva de regresión no paramétrica. Para ello definiremos el estadístico de contraste

$$T = \frac{\sum_{i < j} (d_{ij} - \bar{d})^2 - \sum_{i < j} (d_{ij} - \bar{d}_{ij})^2}{\sum_{i < j} (d_{ij} - \bar{d}_{ij})^2},$$

que puede expresarse en forma matricial como

$$T = \frac{\mathbf{d}' A \mathbf{d}}{\mathbf{d}' B \mathbf{d}},$$

donde  $B = (I - W)'(I - W)$  y  $A = I - L - B$ , siendo  $\mathbf{d}$  un vector que contiene los  $d_{ij}$ ,  $I$  la matriz identidad,  $L$  una matriz cuyos elementos coinciden con el valor  $2/(n(n-1))$  y  $W$  una matriz cuyas filas están formadas por los pesos  $\omega_{ki}$  utilizados en la construcción de  $\bar{d}_{ij}$ . En Diblasi y Bowman (2001) podemos encontrar la distribución del estadístico de contraste anterior bajo la hipótesis nula, así como el p-valor asociado al contraste con dicho estadístico.

Hay que tener en cuenta que el procedimiento descrito anteriormente produce, para una muestra de  $n$  localizaciones espaciales, un conjunto de diferencias de tamaño  $n^* = n(n - 1)/2$ , y unas matrices involucradas en el cálculo del estadístico de contraste de dimensión  $n^* \times n^*$ . Esto da lugar a lógicos problemas computacionales para muestras grandes o con muchas irregularidades en su distribución espacial, que pueden solucionarse considerando una versión agrupada del estadístico de contraste anterior basada en la siguiente expresión

$$T = \frac{\sum_{i=1}^N n_i^* (\bar{d}_i - \bar{d})^2 - \sum_{i=1}^N n_i^* (\bar{d}_i - \bar{\bar{d}}_i)^2}{\sum_{i=1}^N n_i^* (\bar{d}_i - \bar{\bar{d}}_i)^2},$$

donde el conjunto de  $n^*$  posibles distancias se ha agrupado en  $N$  categorías, siendo  $\bar{d}_i$  la media de los  $d_{ij}$  observados y  $\bar{\bar{d}}_i$  la media suavizada en cada agrupación, y siendo  $n_i^*$  el número de diferencias que intervienen en el cálculo de cada una de estas medias, con  $\sum_{i=1}^N n_i^* = n^*$ .

### 2.7.3. Test de validación de modelos de variograma

Maglione y Diblasi (2001, 2004) presentan una generalización del test anterior para contrastar la validez de un determinado modelo de variograma para un determinado variograma experimental obtenido a partir de un conjunto de observaciones. Como en el caso anterior, vamos a suponer que tenemos un proceso espacial  $Z(\cdot)$  gaussiano, intrínsecamente estacionario e isotrópico que puede descomponerse como suma de una media constante y un proceso de error  $\delta(\cdot)$  gaussiano con media cero y variograma  $2\gamma(\cdot)$  igual al del proceso original. Como ya hemos comentado anteriormente, vamos a utilizar las variables  $|R_{ij}|^{1/2}$ , que presentan las ventajas de robustez y distribución aproximadamente normal. Supongamos que se desea contrastar que el variograma del proceso  $2\gamma(\cdot)$  es igual a un determinado modelo paramétrico  $2\gamma_0(\cdot)$ , que denominaremos hipótesis nula del contraste. Bajo dicha hipótesis nula, se puede comprobar que  $E[|R_{ij}|^{1/2}] = \sqrt{2}\Gamma(3/4)(\gamma_0(h_{ij}))^{1/4}/\sqrt{\pi}$ . Vamos a definir las variables  $S_{ij} = |R_{ij}|^{1/2} - E[|R_{ij}|^{1/2}]$ ,  $i, j = 1, \dots, n$ , con  $i < j$ . El estadístico de contraste definido para este test viene dado por

$$T = \frac{\sum_{i<j} (S_{ij} - a\gamma_0(h_{ij})^{1/4}\bar{S})^2 - \sum_{i<j} (S_{ij} - \tilde{S}_{ij})^2}{\sum_{i<j} (S_{ij} - \tilde{S}_{ij})^2},$$

donde  $\bar{S}$  es la media de los  $S_{ij}$ ,  $a = \sqrt{2}\Gamma(3/4)/\sqrt{\pi}$  y  $\tilde{S}_{ij} = \sum_{l<t} \omega_{lt}(h_{ij})S_{lt}$  es el valor suavizado mediante regresión local, y los pesos de la estimación se han definido mediante

$$\omega_{lt}(h_{ij}) = \frac{\exp\left(-\left(\frac{h_{ij}-h_{lt}}{b}\right)^2\right)}{\sum_{l<t} \exp\left(-\left(\frac{h_{ij}-h_{lt}}{b}\right)^2\right)},$$

donde  $b$  denota un parámetro de ancho de banda que controla el grado de suavizado en  $S_{ij}$ . Obsérvese que la suma de cuadrados  $\sum_{i<j} (S_{ij} - a\gamma_0(h_{ij})^{1/4}\bar{S})^2$  puede verse como una medida de la distancia al cuadrado entre la variabilidad recogida por  $S_{ij}$  y la curva

$2\gamma_0$  en escala de raíz cuarta en  $h_{ij}$ , donde la constante  $a$  permite relacionar  $a\gamma_0^{1/4}$  con el valor esperado de  $|R_{ij}|^{1/2}$ . La otra suma de cuadrados  $\sum_{i<j}(S_{ij} - \tilde{S}_{ij})^2$  puede verse como una medida entre cada  $S_{ij}$  y el variograma subyacente representado por  $\tilde{S}_{ij}$ . En Maglione y Diblasi (2004) se muestra la distribución de este estadístico de contraste y el p-valor asociado al test que define, así como unas bandas de referencia muy útiles desde el punto de vista exploratorio.

Hay que tener en cuenta que para el contraste de un modelo de variograma constante (es decir, de independencia espacial), los dos test no ofrecen exactamente el mismo p-valor, aunque ambos valores están muy próximos entre sí. Como ocurría con el test de independencia, la eficiencia computacional está directamente relacionada con el número de distancias distintas existentes entre las localizaciones muestreadas, y con la regularidad de la distribución de las mismas. Aunque se parte de la hipótesis de que el proceso es gaussiano, el test continua siendo razonablemente válido aunque no se satisfaga dicha condición, en especial si las localizaciones se distribuyen de manera regular.

#### 2.7.4. Diagnóstico mediante validación cruzada

La validación cruzada es una técnica que permite evaluar la capacidad predictiva del modelo seleccionado. Supongamos que se ha observado el valor del proceso  $Z(\mathbf{s})$  sobre un conjunto de localizaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ . Sean  $\hat{Z}(\mathbf{s}_i)$  y  $\hat{\sigma}(\mathbf{s}_i)$  el valor predicho y el error típico de la predicción para la localización  $\mathbf{s}_i$  obtenida a partir de las observaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_{i-1}, \mathbf{s}_{i+1}, \dots, \mathbf{s}_n\}$ . Ver Hoef (2002) muestra diferentes estadísticos de resumen que permiten comparar los valores observados con los valores obtenidos mediante validación cruzada, de los que a continuación se citan algunos de los más utilizados.

**Error de predicción estandarizado medio.** Sirve para contrastar la insesgidez de la predicción y debe valer aproximadamente 0. Viene dado por

$$VC1 = \frac{\sum_{i=1}^n (\hat{Z}(\mathbf{s}_i) - z(\mathbf{s}_i)) / \hat{\sigma}(\mathbf{s}_i)}{n}. \quad (2.76)$$

**Raíz del error de predicción estandarizado cuadrático medio.** Depende de la precisión de la desviación típica del error de predicción y debe valer aproximadamente 1. Su expresión es

$$VC2 = \sqrt{\frac{\sum_{i=1}^n ((\hat{Z}(\mathbf{s}_i) - z(\mathbf{s}_i)) / \hat{\sigma}(\mathbf{s}_i))^2}{n}}. \quad (2.77)$$

**Raíz del error de predicción cuadrático medio.** Es una medida global de la bondad de la predicción. Viene dado por

$$VC3 = \sqrt{\frac{\sum_{i=1}^n (\hat{Z}(\mathbf{s}_i) - z(\mathbf{s}_i))^2}{n}}. \quad (2.78)$$

Lo ideal sería que todos los estadísticos anteriores satisfagan las propiedades óptimas, es decir, que  $VC1$  sea prácticamente cero,  $VC2$  prácticamente uno y  $VC3$  sea un valor

pequeño. No obstante, en la práctica una mejoría en uno de ellos puede empeorar los otros, por lo que es recomendable hacer un análisis integral de todos estos estadísticos. Suele ser muy útil observar el valor que cada localización muestreada aporta al sumatorio, lo que permite identificar valores atípicos, sospechosos o anomalías de otra naturaleza. Su eliminación y el recálculo del semivariograma suele producir mejorías significativas.

El proceso anterior de eliminar un caso y ajustar con los restantes es un caso particular del método Jackknife. Si disponemos de suficientes datos espacialmente distribuidos de forma homogénea del proceso espacial  $Z(s)$ , podemos entonces dividir los datos en dos subconjuntos, uno de modelización y otro de validación, sobre el que analizaremos las diferencias entre los valores observados y las predicciones obtenidas mediante el modelo ajustado con el primer subconjunto.



## Capítulo 3

# Geoestadística espacio-temporal

### 3.1. Introducción

En los estudios geoestadísticos analizamos procesos de la naturaleza que se desarrollan en el espacio, pero para los que suele ser habitual disponer de un determinado seguimiento temporal. Por ejemplo, si estamos estudiando la contaminación atmosférica por partículas en suspensión en una determinada ciudad, dispondremos de un conjunto de estaciones de control distribuidas espacialmente por la ciudad en las que se medirá la concentración del contaminante, y para las que será habitual disponer de un seguimiento temporal de mediciones en cada una de ellas. Como veremos a lo largo de los capítulos siguientes, en la modelización y predicción de un determinado fenómeno se obtienen importantes beneficios si, en lugar de considerar únicamente su distribución espacial para un determinado instante temporal de interés (proceso meramente espacial) o la evolución temporal del proceso sobre una localización determinada (proceso meramente temporal), se considera la evolución conjunta del proceso en el espacio-tiempo.

Aunque los fenómenos físicos objeto de estudio generalmente están sujetos a determinados modelos determinísticos que expresan las leyes naturales que los rigen, estos procedimientos se suelen nutrir de un elevado número de parámetros de entrada difíciles de determinar a partir de muestreos limitados e incluso indirectos. Esta complejidad intrínseca de los fenómenos espacio-temporales analizados, junto con la imposibilidad de recoger de forma completa la información asociada a los mismos, tienen como consecuencia que presenten una importante variabilidad. Esto hace de las técnicas estadísticas una herramienta de gran utilidad para el estudio de este tipo de fenómenos, pues consideran modelos estocásticos que incluyen de forma explícita la incertidumbre subyacente en los datos. Seleccionados de forma adecuada, estos modelos permitirán obtener predicciones del proceso de gran calidad para tiempos futuros o sobre localizaciones situadas en la región de interés, incluyendo además una estimación del posible error asociado a dicha predicción.

No obstante, la utilización de modelos estadísticos en el estudio de fenómenos espacio-temporales lleva asociado un conjunto de problemas inherentes a la naturaleza del problema analizado. En particular, necesitaremos manipular grandes conjuntos de datos obteni-

dos con una elevada resolución espacial y periodicidad temporal, lo que hará indispensable la utilización de métodos computacionales intensivos para su manejo. También será necesario disponer de procedimientos estadísticos suficientemente flexibles que permitan el ajuste de las diferentes situaciones analizadas.

Una prueba de la gran utilidad de la modelización estadística espacio-temporal es el espectacular crecimiento que se ha producido en los últimos años en el número de aplicaciones prácticas en una gran variedad de ramas de la ciencia, como el control de la contaminación medioambiental, la climatología, las ciencias sociales, la geología, la biología, la medicina, la arqueología o cualquier otra rama científica en la que sea de interés el estudio de fenómenos que se desarrollen conjuntamente en el espacio y en el tiempo. Un ejemplo de este crecimiento exponencial podemos encontrarlo en que Cressie (1993), uno de los libros de referencia de la estadística espacial, dedicaba únicamente 3 páginas al análisis de datos espacio-temporales. En Kyriakidis y Journel (1999), Christakos (2000), Gneiting y Schlather (2002), Mateu et al. (2003), Kolovos et al. (2004), Banerjee et al. (2004), Sahu y Mardia (2005) o Chen et al. (2006) podemos encontrar un resumen de las principales técnicas de modelización espacio-temporal, junto con numerosas aplicaciones prácticas para multitud de fenómenos de la naturaleza.

Kyriakidis y Journel (1999) llevan a cabo un exhaustivo análisis de los diferentes enfoques utilizados en la literatura para el análisis de procesos espacio-temporales. Concretamente establecen dos metodologías diferenciadas: aquellas que extienden las herramientas del análisis de series temporales considerando los datos como realizaciones de vectores de series temporales localizadas en el espacio o extienden las herramientas del análisis espacial considerando vectores de procesos espaciales que evolucionan en el tiempo, o aquellas que consideran los datos como una única realización de un proceso estocástico espacio-temporal que evoluciona conjuntamente en ambas componentes. En este trabajo, los autores analizan las relaciones existentes entre ambos procedimientos y las ventajas y desventajas que supone la adopción de cada uno de ellos. Nosotros nos centraremos en la segunda vertiente, que evita determinados problemas de aplicabilidad presentes en la primera perspectiva, y que considera extensiones de las herramientas geoestadísticas espaciales para incluir la dimensión temporal adicional.

El dominio de variación de los procesos espacio-temporales generalmente será  $\mathfrak{R}^d \times \mathfrak{R}$ , donde  $\mathfrak{R}^d$  es el dominio de variación espacial y  $\mathfrak{R}$  el dominio de variación temporal. Aunque desde el punto de vista matemático  $\mathfrak{R}^d \times \mathfrak{R} \equiv \mathfrak{R}^{d+1}$ , desde el punto de vista físico no tendrá sentido considerar las componentes espacial y temporal del mismo modo, debido a las importantes diferencias existentes entre ambos ejes de coordenadas. Así, mientras el eje temporal presenta una ordenación intrínseca (pues existe el pasado, el presente y el futuro), no ocurre lo mismo con las coordenadas espaciales. Por ejemplo, para el eje temporal las observaciones generalmente se toman en una dirección (el pasado) y se trata de extrapolar su valor para el futuro, mientras que para el eje espacial solemos interpolar el valor en una localización no observada. Tampoco las escalas y las unidades de distancia en el tiempo y en el espacio son generalmente comparables. Por ello, aunque la identidad anterior permite la generalización de muchas de las herramientas que hemos visto en el Capítulo 2, deberemos siempre mantener una distinción entre las coordenadas espaciales y temporales de cada localización para tener en cuenta la naturaleza diferencial de ambos

ejes.

Como veremos a lo largo de este capítulo y en los siguientes, la principal dificultad del análisis de procesos espacio-temporales será seleccionar aquel modelo de función de covarianza que mejor se ajuste a nuestros datos y comprobar su validez. En los primeros trabajos la dificultad que encerraba esta validación hizo que se utilizasen modelos estacionarios separables, que como veremos no tendrán en consideración la posible interacción espacio-temporal. Para solucionar este problema se han ido introduciendo en la literatura estadística diferentes familias de covarianza estacionarias no separables, que permiten la inclusión de la interacción entre el espacio y el tiempo en los procesos analizados. En los últimos años estamos asistiendo a una generalización constante de los modelos existentes, que ganan con ello en flexibilidad y aplicabilidad. También se han definido nuevos modelos que permitan el estudio de procesos en los que las hipótesis de simetría o de estacionariedad no sean admisibles. En el Capítulo 4 llevaremos a cabo un exhaustivo análisis de los diferentes modelos propuestos en la literatura para la función de covarianza de un proceso espacio-temporal.

En este capítulo vamos a describir las principales herramientas utilizadas en el análisis de procesos espacio-temporales. La Sección 3.2 resume los principales conceptos involucrados en el estudio de los procesos espacio-temporales. La Sección 3.3 describe las propiedades más importantes de la función de covarianza y del semivariograma de los procesos espacio-temporales. La Sección 3.4 ofrece un resumen de las principales técnicas para la estimación de las funciones anteriores. La Sección 3.5 introduce las técnicas de modelización espacio-temporal más utilizadas en la práctica geoestadística. La Sección 3.6 se centra en los diferentes procedimientos de ajuste de los parámetros involucrados en los modelos espacio-temporales. En la Sección 3.7 se generalizan las técnicas de predicción espacial al caso espacio-temporal. Por último, la Sección 3.8 ofrece un resumen de los diferentes procedimientos utilizados en procesos no estacionarios.

## 3.2. Definiciones

El objetivo de esta sección es introducir aquellos conceptos más relevantes en el análisis de procesos estocásticos espacio-temporales. Como veremos, mucho del material que trataremos serán generalizaciones del caso espacial que hemos visto en el Capítulo 2.

Nuestro principal objetivo será el análisis de un determinado proceso estocástico espacio-temporal  $\{Z(\mathbf{s}, t), \mathbf{s} \in D, t \in T\}$ , definido como un conjunto de variables aleatorias generalmente dependientes,  $Z(\mathbf{s}, t)$ , una para cada localización espacial  $\mathbf{s} \in D$  y cada tiempo  $t \in T$ , donde el índice espacial varía en  $D \subseteq \mathbb{R}^d$  y el índice temporal en  $T \subseteq \mathbb{R}$  o  $\mathbb{Z}$ . Este proceso objeto de estudio podría ser, por ejemplo, la contaminación por  $SO_2$ , el nivel de ruido, la temperatura o la velocidad del viento registrados en una determinada localización  $\mathbf{s}$  para un instante temporal  $t$ .

En la mayoría de aplicaciones  $Z(\mathbf{s}, t)$  será una variable aleatoria univariante, donde la coordenada espacial  $\mathbf{s}$  variará en  $D \subset \mathbb{R}^2$  y denotará un par de coordenadas espaciales (la latitud y longitud de la localización, o cualquier transformación de estas), y donde  $T \in \mathbb{Z}^+$  será un conjunto discreto de posibles valores temporales considerados. Como

ocurría con el caso espacial, el proceso  $Z(\cdot, \cdot)$  puede representar directamente el fenómeno objeto de estudio, una transformación del mismo o incluso los residuos obtenidos al ajustar un modelo de tendencia espacio-temporal.

Supondremos a lo largo de todo el capítulo que tenemos un proceso espacio-temporal  $Z(\cdot, \cdot)$  sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , que cumple la propiedad de regularidad, esto es, que  $Var(Z(\mathbf{s}; t)) < \infty, \forall (\mathbf{s}, t) \in \mathfrak{R}^d \times \mathfrak{R}$ , con lo que tenemos garantizada la existencia de sus dos primeros momentos. Podemos, por tanto, definir bajo este supuesto la función media y la función de covarianza del proceso como

$$\mu(\mathbf{s}, t) = E[Z(\mathbf{s}, t)] \quad (3.1)$$

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) \quad (3.2)$$

para cualquier par de localizaciones  $(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2) \in \mathfrak{R}^d \times \mathfrak{R}$ . Generalmente el proceso suele modelizarse como

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \delta(\mathbf{s}, t), \quad (3.3)$$

donde la componente  $\mu(\mathbf{s}, t)$  es la media del proceso, que denominaremos tendencia o variación a gran escala, y donde  $\delta(\mathbf{s}, t)$  es la componente del error espacio-temporal o variación a pequeña escala, que expresa las fluctuaciones del proceso espacio-temporal alrededor de la media. En la Sección 3.5 profundizaremos en esta descomposición, estudiando las diferentes alternativas a considerar en la modelización de la tendencia del proceso. En esta sección nos vamos a centrar en el análisis de la estructura de variación espacio-temporal, introduciendo las diferentes herramientas a utilizar en su modelización, así como las diferentes hipótesis simplificadoras que en la práctica suelen realizarse.

Como ocurría para los procesos espaciales, el proceso estocástico  $Z(\cdot, \cdot)$  estará completamente caracterizado por su función de distribución acumulada,

$$F(\mathbf{s}, t; z) = P(Z(\mathbf{s}, t) \leq z), \forall z, (\mathbf{s}, t) \in \mathfrak{R}^d \times \mathfrak{R}.$$

Como en la práctica generalmente dispondremos de una única realización del proceso a estudiar, cualquier inferencia de las leyes que lo rigen requerirá de la admisión de determinadas leyes simplificadoras relacionadas con la regularidad del proceso, como son la estacionariedad, la separabilidad, la simetría completa o el soporte compacto, que veremos más adelante. Bajo estas hipótesis de regularidad, podremos analizar juntas todas las observaciones separadas por un vector de distancia espacio-temporal  $(\mathbf{h}, u)$  obteniendo con ello la replicación necesaria para el análisis.

**Definición 3.1** Decimos que el proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene función de covarianza espacialmente estacionaria si, para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2)$  en  $\mathfrak{R}^d \times \mathfrak{R}$ , la covarianza  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  depende únicamente de la distancia entre las localizaciones espaciales  $\mathbf{s}_1 - \mathbf{s}_2$  y de los tiempos  $t_1$  y  $t_2$ .

**Definición 3.2** Decimos que el proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene función de covarianza temporalmente estacionaria si, para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2)$  en  $\mathfrak{R}^d \times \mathfrak{R}$ , la covarianza  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  depende únicamente de la distancia entre los tiempos  $t_1 - t_2$  y de las localizaciones espaciales  $\mathbf{s}_1$  y  $\mathbf{s}_2$ .

**Definición 3.3** Si el proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene función de covarianza estacionaria tanto espacial como temporalmente, entonces decimos que tiene función de covarianza estacionaria. En ese caso, la función de covarianza puede expresarse como

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = C(\mathbf{h}, u),$$

siendo  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  y  $u = t_1 - t_2$  las distancias espacial y temporal, respectivamente.

Observemos que si un proceso tiene una función de covarianza estacionaria, entonces su varianza no depende de la localización, ya que

$$\text{Var}(Z(\mathbf{s}, t)) = C(\mathbf{0}, 0) = \sigma^2, \forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R},$$

donde  $C(\mathbf{0}, 0) \geq 0$  recibe el nombre de varianza a priori del proceso.

Hemos de tener en cuenta que dada una función de covarianza espacio-temporal estacionaria  $C(\mathbf{h}, u)$  definida sobre  $\mathbb{R}^d \times \mathbb{R}$  y una determinada distancia espacial  $\mathbf{h} \in \mathbb{R}^d$  (o una determinada distancia temporal  $u \in \mathbb{R}$ , la función  $C_{\mathbf{h}}(u)$  definida sobre  $\mathbb{R}$  (o la función  $C_u(\mathbf{h})$  definida sobre  $\mathbb{R}^d$ ) no es necesariamente una función de covarianza temporal (o espacial), ya que no necesariamente cumplirá la condición (2.6) de ser definida positiva. Por ejemplo, si consideramos  $C(\mathbf{h}, u) = C_e(\mathbf{h})C_t(u)$ , con  $C_e$  un función de covarianza espacial y  $C_t$  una función de covarianza temporal, entonces es muy sencillo comprobar que dado  $\mathbf{h} \in \mathbb{R}^d$  cualquiera, la función  $C_{\mathbf{h}}(u)$  verifica la condición (2.6) si y sólo si  $C_e(\mathbf{h}) \geq 0$ . No obstante, por ser una varianza, esta condición se cumplirá lógicamente para  $u = 0$  (y también para  $\mathbf{h} = \mathbf{0}$ ), por lo que  $C(\cdot, 0)$  será una función de covarianza espacial (y del mismo modo,  $C(\mathbf{0}, \cdot)$  será una función de covarianza temporal).

**Definición 3.4** Decimos que un proceso espacio-temporal  $Z(\mathbf{s}, t)$  es estacionario de segundo orden (o estacionario en sentido débil, o simplemente estacionario) si tiene media constante y función de covarianza estacionaria.

**Definición 3.5** Decimos que el proceso espacio-temporal  $Z(\mathbf{s}, t)$  es un proceso Gaussiano si para cualquier conjunto de localizaciones espacio-temporales  $\{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)\}$ , el vector aleatorio  $\mathbf{Z} = (Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n))'$  siguen una distribución normal multivariante.

**Definición 3.6** Un proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene función de covarianza separable si existe una función de covarianza puramente espacial  $C_s(\cdot)$  y una función de covarianza puramente temporal  $C_t(\cdot)$  tales que

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = C_s(\mathbf{s}_1, \mathbf{s}_2)C_t(t_1, t_2),$$

para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2) \in \mathbb{R}^d \times \mathbb{R}$ . Si esta descomposición no es posible, diremos que la función de covarianza es no separable.

Esta descomposición de la función de covarianza como producto de una función de covarianza puramente espacial y otra puramente temporal permite, como veremos más adelante,

una inferencia y estimación más eficiente desde el punto de vista computacional. Es por ello que este modelo separable ha sido uno de los más utilizados en la práctica geoestadística, incluso en situaciones en las que esta hipótesis no es justificable por la propia naturaleza del proceso analizado.

**Definición 3.7** *Un proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene función de covarianza completamente simétrica si*

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = C((\mathbf{s}_1, t_2), (\mathbf{s}_2, t_1)), \quad (3.4)$$

para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2) \in \mathbb{R}^d \times \mathbb{R}$ .

Algunos procesos atmosféricos, medioambientales y geofísicos están a menudo bajo la influencia de corrientes de aire o marinas que dan lugar a una falta de simetría completa. Este tipo de efectos de transporte son bien conocidos en la literatura y han sido tratados en Gneiting (2002), Stein (2005b) o Huang y Hsu (2004), entre otros. Observemos que la separabilidad es un caso particular de la simetría completa, por lo que cualquier test construido para contrastar la simetría completa puede utilizarse para rechazar la separabilidad.

Esta condición de simetría completa (3.4) se reduce para las funciones de covarianza espacio-temporales estacionarias a que cumpla

$$C(\mathbf{h}, u) = C(\mathbf{h}, -u) = C(-\mathbf{h}, u) = C(-\mathbf{h}, -u), \forall (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

Observemos que por la simetría de la función de covarianza, para las funciones estacionarias siempre se cumplirá que  $C(\mathbf{h}, u) = C(-\mathbf{h}, -u)$  y  $C(\mathbf{h}, -u) = C(-\mathbf{h}, u)$ , por lo que la verificación de la simetría completa se reducirá a comprobar que  $C(\mathbf{h}, u) = C(\mathbf{h}, -u)$  (o que  $C(-\mathbf{h}, u) = C(-\mathbf{h}, -u)$ ).

**Definición 3.8** *Un proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene función de covarianza con soporte compacto si, para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2) \in \mathbb{R}^d \times \mathbb{R}$ , la covarianza  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  tiende a cero cuando la distancia espacial o temporal es suficientemente grande.*

Los procesos con covarianzas de soporte compacto son atractivos desde el punto de vista computacional, pues permiten una simulación y predicción espacio-temporal eficiente. Por esta razón, algunos autores imponen esta condición para el kriging sobre conjuntos de datos en los que la elevada resolución espacial dificulta su análisis.

La Figura 3.1, tomada de Gneiting et al. (2005), representa de forma esquemática la relación existente entre las covarianzas separables, completamente simétricas, estacionarias y de soporte compacto, dentro del conjunto general de funciones de covarianza espacio-temporales (estacionarias o no estacionarias). Como vemos, una función de covarianza separable puede ser estacionaria o no estacionaria, y lo mismo ocurre con las funciones de covarianza completamente simétricas. También se observa que las funciones de covarianza separables son siempre completamente simétricas, aunque el recíproco no es cierto en general.

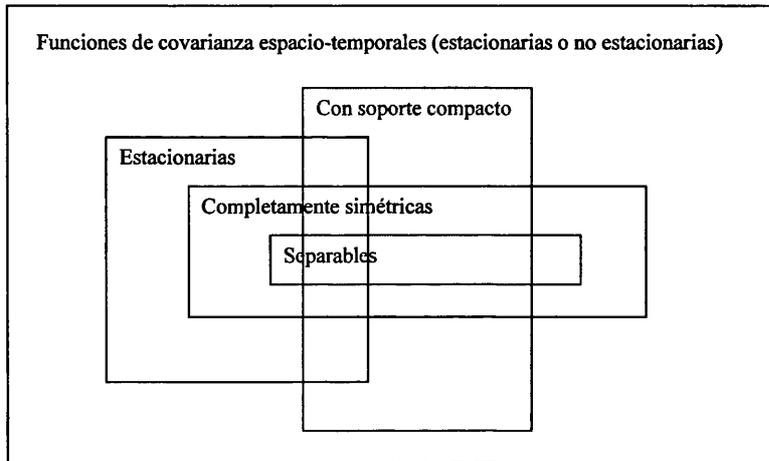


Figura 3.1: Relaciones entre los diferentes tipos de funciones de covarianza espacio-temporales.

**Definición 3.9** Decimos que una función de covarianza espacio-temporal estacionaria  $C(\mathbf{h}, u)$  definida sobre  $\mathbb{R}^d \times \mathbb{R}$  verifica la hipótesis de Taylor (1938) si existe un vector de velocidad  $\mathbf{v} \in \mathbb{R}^d$  tal que

$$C(\mathbf{0}, u) = C(\mathbf{v}u, 0), \quad \forall u \in \mathbb{R}.$$

Como vemos, esta hipótesis de Taylor se centra únicamente en la existencia de un determinado tipo de relación entre las funciones de covarianza marginales  $C(\mathbf{0}, u)$  y  $C(\mathbf{h}, 0)$  del proceso. Se ha comprobado que esta hipótesis es plausible para determinados fenómenos naturales de la dinámica de fluidos, la meteorología o la hidrología. Cox e Isham (1988) estudian el cumplimiento de esta hipótesis para diferentes modelos de covarianza espacio-temporales.

**Definición 3.10** Un proceso espacio-temporal estacionario  $Z(\mathbf{s}, t)$  tiene función de covarianza espacialmente isotrópica si

$$C(\mathbf{h}, u) = C(\|\mathbf{h}\|, u), \quad \forall(\mathbf{s}, t) \in \mathbb{R} \times \mathbb{R}.$$

**Definición 3.11** Un proceso espacio-temporal estacionario  $Z(\mathbf{s}, t)$  tiene función de covarianza temporalmente isotrópica (o simétrica) si

$$C(\mathbf{h}, u) = C(\mathbf{h}, |u|), \quad \forall(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}.$$

Observemos que si la función de covarianza de un proceso estacionario es espacial y temporalmente isotrópica, entonces es completamente simétrica.

Como ocurría con los procesos espaciales, en ocasiones modelizamos la estructura de segundo orden de un proceso espacio-temporal utilizando variogramas en lugar de funciones de covarianzas. Se define el variograma espacio-temporal como la función

$$2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \text{Var}(Z(\mathbf{s}_1, t_1) - Z(\mathbf{s}_2, t_2)),$$

mientras que la mitad de esta cantidad recibe el nombre de semivariograma. En el caso en que el proceso tenga media constante, entonces

$$2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = E[Z(\mathbf{s}_1, t_1) - Z(\mathbf{s}_2, t_2)]^2.$$

Siempre que sea posible definir la función de covarianza y el variograma, se relacionarán mediante la siguiente expresión

$$2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \text{Var}(Z(\mathbf{s}_1, t_1)) + \text{Var}(Z(\mathbf{s}_2, t_2)) - 2C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) \quad (3.5)$$

**Definición 3.12** Decimos que el proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene variograma intrínsecamente estacionario espacialmente si, para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2)$  en  $\mathbb{R}^d \times \mathbb{R}$ , el variograma  $2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  depende únicamente de la distancia entre las localizaciones espaciales  $\mathbf{s}_1 - \mathbf{s}_2$  y de los tiempos  $t_1$  y  $t_2$ .

**Definición 3.13** Decimos que el proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene variograma intrínsecamente estacionario temporalmente si, para cualquier par de localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2)$  en  $\mathbb{R}^d \times \mathbb{R}$ , el variograma  $2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  depende únicamente de la distancia entre los tiempos  $t_1 - t_2$  y de las localizaciones espaciales  $\mathbf{s}_1$  y  $\mathbf{s}_2$ .

**Definición 3.14** Si el proceso espacio-temporal  $Z(\mathbf{s}, t)$  tiene variograma intrínsecamente estacionario tanto espacial como temporalmente, entonces decimos que tiene variograma intrínsecamente estacionario. En ese caso, el variograma puede expresarse como

$$2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = 2\gamma(\mathbf{h}, u),$$

siendo  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  y  $u = t_1 - t_2$  la distancia espacial y temporal, respectivamente. Las restricciones  $2\gamma(\cdot, 0)$  y  $2\gamma(0, \cdot)$  reciben el nombre de variogramas puramente espacial y puramente temporal, respectivamente.

**Definición 3.15** Decimos que un proceso espacio-temporal  $Z(\mathbf{s}, t)$  es intrínsecamente estacionario si tiene media constante y variograma intrínsecamente estacionario.

La estacionariedad intrínseca es menos restrictiva que la estacionariedad de segundo orden, ya que dado un proceso estacionario de segundo orden  $Z(\mathbf{s}, t)$  con función de covarianza  $C(\cdot, \cdot)$ , entonces por (3.5) se cumple que

$$\begin{aligned} \text{Var}(Z(\mathbf{s}_1, t_1) - Z(\mathbf{s}_2, t_2)) &= \text{Var}(Z(\mathbf{s}_1, t_1)) + \text{Var}(Z(\mathbf{s}_2, t_2)) \\ &\quad - 2\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) \\ &= 2C(\mathbf{0}, \mathbf{0}) - 2C(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2), \end{aligned}$$

por lo que el proceso  $Z(\mathbf{s}, t)$  será intrínsecamente estacionario con semivariograma

$$\gamma(\mathbf{h}, u) = C(\mathbf{0}, 0) - C(\mathbf{h}, u). \quad (3.6)$$

La expresión (3.6) nos muestra la relación entre el semivariograma y la función de covarianza en los procesos estacionarios. El recíproco no será cierto en general, ya que la equivalencia se cumple bajo la condición de que el variograma  $2\gamma(\mathbf{h}, u)$  sea acotado, lo que no es necesariamente cierto para los variogramas intrínsecamente estacionarios.

Mientras que la separabilidad es una propiedad que tiene sentido para las funciones de covarianza espacio-temporales, no lo tiene para el variograma espacio-temporal, ya que el producto de variogramas no asegura un semivariograma válido. En cambio, si es posible generalizar otras propiedades como la de simetría completa, o la isotropía espacial o temporal.

Otra función muy utilizada en la modelización de la dependencia espacio-temporal de los procesos estacionarios es la función de correlación.

**Definición 3.16** Sea  $Z(\mathbf{s}, t)$  un proceso estacionario de segundo orden con varianza a priori  $\sigma^2 = C(\mathbf{0}, 0) > 0$ , definimos la función de correlación del proceso como

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

Es evidente que si  $\rho(\mathbf{h}, u)$  es una función de correlación sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , entonces sus marginales  $\rho(\mathbf{0}, u)$  y  $\rho(\mathbf{h}, 0)$  serán, respectivamente, funciones de correlación espacial sobre  $\mathfrak{R}^d$  y temporal sobre  $\mathfrak{R}$ .

### 3.3. Propiedades de la función de covarianza y del variograma espacio-temporal

En la sección anterior hemos visto la definición de las funciones de covarianza y del variograma para procesos espacio-temporales. Aunque para los procesos espaciales el variograma suele ser la herramienta utilizada en la descripción de la variabilidad espacial del proceso, para los procesos espacio-temporales la función de covarianza es la más utilizada en la modelización de la variabilidad del proceso espacio-temporal. No obstante, hay que tener en cuenta que siempre que sea posible definir ambas funciones, por (3.5) podremos pasar fácilmente de una a otra.

Una condición necesaria y suficiente para que una función  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  de valores reales definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  sea una función de covarianza es que sea simétrica  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = C((\mathbf{s}_2, t_2), (\mathbf{s}_1, t_1))$  y que sea definida positiva, esto es,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \geq 0, \quad (3.7)$$

para cualquier  $n \in \mathbb{N}$ , y para cualesquiera  $(\mathbf{s}_i, t_i) \in \mathfrak{R}^d \times \mathfrak{R}$  y  $a_i \in \mathfrak{R}$ ,  $i = 1, \dots, n$ . La condición (3.7) es suficiente si permitimos que la función de covarianza tome valores

complejos. En las aplicaciones prácticas, la verificación de esta condición (3.7) para la función de covarianza propuesta es un proceso mucho más costoso incluso que lo era para los procesos meramente espaciales. En el Capítulo 4 introduciremos los principales modelos utilizados en la práctica geoestadística, estudiando su validez y principales características.

Análogamente, una condición necesaria y suficiente para que una función de valores reales  $\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  no negativa definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  sea un semivariograma es que sea una función simétrica  $\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \gamma((\mathbf{s}_2, t_2), (\mathbf{s}_1, t_1))$  y que sea condicionalmente definida negativa, esto es,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \leq 0, \quad (3.8)$$

para cualquier  $n \in \mathbb{N}$ , y cualesquiera  $(\mathbf{s}_i, t_i) \in \mathfrak{R}^d \times \mathfrak{R}$  y  $a_i \in \mathfrak{R}$ ,  $i = 1, \dots, n$ , con  $\sum_1^n a_i = 0$ .

Como consecuencia de los resultados de Schoenberg (1938), tenemos el siguiente teorema de caracterización del semivariograma espacio-temporal.

**Teorema 3.1** Sea  $\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  una función definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , con  $\gamma((\mathbf{s}, t), (\mathbf{s}, t)) = 0$ , para todo  $(\mathbf{s}, t) \in \mathfrak{R}^d \times \mathfrak{R}$ , entonces son equivalentes:

- $\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  es un semivariograma sobre  $\mathfrak{R}^d \times \mathfrak{R}$ .
- $\exp(-\theta \gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)))$  es una función de covarianza sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , para cualquier  $\theta > 0$ .
- $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \gamma((\mathbf{s}_1, t_1), (\mathbf{0}, 0)) + \gamma((\mathbf{s}_2, t_2), (\mathbf{0}, 0)) - \gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  es una función de covarianza sobre  $\mathfrak{R}^d \times \mathfrak{R}$ .

Para el caso estacionario, de los resultados anteriores se deduce que una función de valores reales  $C(\mathbf{h}, u)$  definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  es una función de covarianza estacionaria si y sólo si es una función par ( $C(\mathbf{h}, u) = C(-\mathbf{h}, -u)$ ) y definida positiva, esto es,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{h}_i, u_i) \geq 0, \quad (3.9)$$

para cualquier  $n \in \mathbb{N}$ , y para cualesquiera  $(\mathbf{h}_i, u_i) \in \mathfrak{R}^d \times \mathfrak{R}$  y  $a_i \in \mathfrak{R}$ ,  $i = 1, \dots, n$ .

Del mismo modo, una función de valores reales  $\gamma(\mathbf{h}, u)$  no negativa definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  es un semivariograma intrínsecamente estacionario si y sólo si es una función par ( $\gamma(\mathbf{h}, u) = \gamma(-\mathbf{h}, -u)$ ) y condicionalmente definida negativa, esto es,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(\mathbf{h}_i, u_i) \leq 0, \quad (3.10)$$

para cualquier  $n \in \mathbb{N}$ , y cualesquiera  $(\mathbf{h}_i, u_i) \in \mathfrak{R}^d \times \mathfrak{R}$  y  $a_i \in \mathfrak{R}$ ,  $i = 1, \dots, n$ , con  $\sum_1^n a_i = 0$ .

El Teorema 3.1 también tendrá un equivalente estacionario, según el cual dada una función  $\gamma(\mathbf{h}, u)$  definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , con  $\gamma(\mathbf{0}, 0) = 0$ , entonces son equivalentes:

- $\gamma(\mathbf{h}, u)$  es un semivariograma intrínsecamente estacionario sobre  $\mathfrak{R}^d \times \mathfrak{R}$ .
- $\exp(-\theta\gamma(\mathbf{h}, u))$  es una función de covarianza estacionaria sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , para cualquier  $\theta > 0$ .
- $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \gamma(\mathbf{s}_1, t_1) + \gamma(\mathbf{s}_2, t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2)$  es una función de covarianza sobre  $\mathfrak{R}^d \times \mathfrak{R}$ .

Observemos que muchas de las propiedades que vimos en el Capítulo 2 para las funciones de covarianza y los semivariogramas espaciales pueden generalizarse para el caso espacio-temporal, considerándolas como funciones sobre  $\mathfrak{R}^{d+1} \equiv \mathfrak{R}^d \times \mathfrak{R}$ . Por ejemplo, se tendrá que  $\gamma(\mathbf{0}, 0) = 0$  o que  $|C(\mathbf{h}, u)| \leq C(\mathbf{0}, 0)$ ,  $\forall (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}$ .

Es interesante considerar la generalización al caso espacio-temporal de algunas de las propiedades que vimos para los procesos espaciales, y que permitirán la construcción de nuevas funciones de covarianza a partir de funciones ya conocidas. En concreto, sea  $C_1(\mathbf{h}, u)$  una función de covarianza definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  y  $b > 0$ , entonces

$$C(\mathbf{h}, u) = b \cdot C_1(\mathbf{h}, u) \quad (3.11)$$

es también una función de covarianza definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ . Además, sean  $C_1(\mathbf{h}, u)$  y  $C_2(\mathbf{h}, u)$  dos funciones de covarianza definidas sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , entonces

$$C(\mathbf{h}, u) = C_1(\mathbf{h}, u) + C_2(\mathbf{h}, u) \quad (3.12)$$

es una función de covarianza definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ . También se cumple la primera propiedad de estabilidad, según la cual dadas dos funciones de covarianza  $C_1(\mathbf{h}, u)$  y  $C_2(\mathbf{h}, u)$  sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , entonces

$$C(\mathbf{h}, u) = C_1(\mathbf{h}, u) \cdot C_2(\mathbf{h}, u) \quad (3.13)$$

es también una función de covarianza definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ . Por la segunda propiedad de estabilidad, sea  $\mu(a)$  una medida acotada no negativa definida sobre  $U \subset \mathfrak{R}$  y sea  $C(\mathbf{h}, u|a)$  una función de covarianza en  $\mathfrak{R}^d \times \mathfrak{R}$ , para cada  $a \in V \subset U$ , integrable sobre el conjunto  $V \subset U$ , entonces

$$C(\mathbf{h}, u) = \int_V C(\mathbf{h}, u|a) d\mu(a) \quad (3.14)$$

es una función de covarianza en  $\mathfrak{R}^d \times \mathfrak{R}$ . Otra propiedad interesante que permite construir nuevas funciones de covarianza es que su validez se mantiene en el límite, esto es, si  $\{C_n(\mathbf{h}, u), n = 1, \dots\}$  es una sucesión de funciones de covarianza sobre  $\mathfrak{R}^d \times \mathfrak{R}$  convergente a  $C(\mathbf{h}, u)$ ,  $\forall (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}$ , entonces

$$C(\mathbf{h}, u) = \lim_{n \rightarrow \infty} C_n(\mathbf{h}, u) \quad (3.15)$$

es una función de covarianza sobre  $\mathfrak{R}^d \times \mathfrak{R}$ .

Es sencillo comprobar que, dada una función de covarianza espacial estacionaria  $C_e(\mathbf{h})$  definida sobre  $\mathfrak{R}^d$  y una función de covarianza temporal estacionaria  $C_t(u)$  definida sobre  $\mathfrak{R}$ , entonces las funciones

$$\begin{aligned} C(\mathbf{h}, u) &= C_e(\mathbf{h} + \theta u) \\ C(\mathbf{h}, u) &= C_t(u + \theta' \mathbf{h}) \end{aligned}$$

con  $\theta \in \mathfrak{R}^d$ , son funciones de covarianza estacionarias sobre  $\mathfrak{R}^d \times \mathfrak{R}$ . También es cierto para un vector aleatorio, esto es, dado  $\theta$  un vector aleatorio sobre  $\mathfrak{R}^d$ , entonces

$$\begin{aligned} C(\mathbf{h}, u) &= E_{\theta} [C_e(\mathbf{h} + \theta u)] \\ C(\mathbf{h}, u) &= E_{\theta} [C_t(u + \theta' \mathbf{h})] \end{aligned} \quad (3.16)$$

son funciones de covarianza estacionarias sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , siempre que las esperanzas anteriores existan.

El teorema de Bochner (1933) establece que una función  $C(\cdot, \cdot)$  continua definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  es una función de covarianza estacionaria si y sólo si es de la forma

$$C(\mathbf{h}, u) = \int \int e^{i(\omega' \mathbf{h} + \tau u)} dF(\omega, \tau), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (3.17)$$

donde la función  $F$  es la función de distribución de una medida finita y no negativa definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , que recibe el nombre de función de distribución espectral. Por lo tanto, la clase de las funciones de covarianza espacio-temporales estacionarias sobre  $\mathfrak{R}^d \times \mathfrak{R}$  es idéntica a la clase de transformadas de Fourier de funciones de distribución de medidas finitas y no negativas sobre ese dominio.

Si la función  $C$  es además integrable, entonces la función de distribución espectral  $F$  es absolutamente continua, y la representación (3.17) se simplifica a

$$C(\mathbf{h}, u) = \int \int e^{i(\omega' \mathbf{h} + \tau u)} f(\omega, \tau) d\omega d\tau, \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (3.18)$$

donde  $f$  es una función integrable, continua y no negativa, que recibe el nombre de función de densidad espectral. La función de covarianza  $C$  y la función de densidad espectral  $f$  forman entonces una pareja de transformadas de Fourier, y

$$f(\omega, \tau) = (2\pi)^{-d-1} \int \int e^{-i(\omega' \mathbf{h} + \tau u)} C(\mathbf{h}, u) d\mathbf{h} du.$$

En la Definición 3.6 hemos visto que una función de covarianza espacio-temporal estacionaria  $C(\cdot, \cdot)$  era separable si existían dos funciones de covarianza  $C_s$  y  $C_t$  estacionarias puramente espacial y temporal, respectivamente, tales que  $C(\mathbf{h}, u) = C_s(\mathbf{h}) \cdot C_t(u)$ ,  $\forall (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}$ . Aplicando la descomposición espectral anterior, esto será equivalente a que su función de distribución espectral se pueda descomponer como producto de una función de distribución espectral sobre el dominio espacial y una función de distribución espectral sobre el dominio temporal.

La descomposición (3.17) puede especializarse para covarianzas completamente simétricas. Sea  $C(\cdot, \cdot)$  una función continua definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , entonces  $C(\cdot, \cdot)$  es una función de covarianza estacionaria completamente simétrica si y sólo si admite la descomposición

$$C(\mathbf{h}, u) = \int \int \cos(\boldsymbol{\omega}'\mathbf{h}) \cos(\tau u) dF(\boldsymbol{\omega}, \tau), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (3.19)$$

donde  $F$  es la función de distribución espectral simétrica y no negativa definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ . Si además la función de densidad espectral  $f$  existe, entonces  $f$  es completamente simétrica, esto es,

$$f(\boldsymbol{\omega}, \tau) = f(\boldsymbol{\omega}, -\tau) = f(-\boldsymbol{\omega}, \tau) = f(-\boldsymbol{\omega}, -\tau), \quad \forall (\boldsymbol{\omega}, \tau) \in \mathfrak{R}^d \times \mathfrak{R}$$

Todos los resultados anteriores son aplicables sobre funciones continuas. En la práctica, las funciones de covarianza espacio-temporales estacionarias ajustadas a menudo se ven afectadas por un efecto pepita, esto es, una discontinuidad en el origen. En el caso espacio-temporal, el efecto pepita puede ser únicamente espacial, únicamente temporal o espacio-temporal, por lo que vendrá dado por

$$C(\mathbf{h}, u) = a\delta_{(\mathbf{h}, u)=(0,0)} + b\delta_{\mathbf{h}=0} + c\delta_{u=0}, \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (3.20)$$

con  $a, b, c$  constantes no negativas y  $\delta$  la función indicatriz. El producto y la suma de las funciones de covarianza continuas con funciones de covarianza de efecto pepita del tipo (3.20) darán lugar a modelos de covarianza válidos.

En Cressie y Huang (1999) se ofrece un teorema de caracterización de la clase de funciones de covarianza espacio-temporales estacionarias bajo la hipótesis adicional de integrabilidad. Sea  $C(\cdot, \cdot)$  una función continua, acotada, simétrica e integrable definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ , entonces  $C(\cdot, \cdot)$  es una función de covarianza estacionaria si y sólo si dado  $u \in \mathfrak{R}$ ,

$$C_{\boldsymbol{\omega}}(u) = \int e^{-i\boldsymbol{\omega}'\mathbf{h}} C(\mathbf{h}, u) d\mathbf{h}, \quad (3.21)$$

es una función de covarianza para todo  $\boldsymbol{\omega} \in \mathfrak{R}^d$  excepto a lo sumo en un conjunto de medida de Lebesgue nula. En Gneiting (2002) se generaliza este resultado para  $C$  definida en  $\mathfrak{R}^d \times \mathfrak{R}^l$ , del que el enunciado anterior es un caso particular para  $l = 1$ .

Tanto la función de covarianza como la función de densidad espectral son importantes herramientas para caracterizar campos aleatorios estacionarios espacio-temporales. Matemáticamente hablando, ambas funciones están íntimamente relacionadas como un par de transformadas de Fourier. Además, la función de densidad espectral es particularmente útil en aquellas situaciones en las que no tenemos una expresión explícita de la función de covarianza. Stein (2005b) muestra una familia paramétrica de densidades espectrales que satisfacen algunas propiedades muy interesantes para la modelización de procesos espacio-temporales. En particular, muestra la conveniencia de utilizar funciones de covarianza que sean suaves lejos del origen, lo que puede comprobarse verificando si sus densidades espectrales tienen derivadas de determinados órdenes. Estos resultados son utilizados para construir funciones de covarianza que son infinitamente diferenciables lejos del origen y que permiten diferentes grados de suavizado del proceso en el espacio y en el tiempo. Todas estas propiedades serán analizadas con detalle en el Capítulo 4.

### 3.4. Estimación del variograma y del covariograma

Para obtener una estimación empírica de la función de covarianza o del variograma de un proceso espacio-temporal podemos generalizar de forma natural los procedimientos vistos en la Sección 2.3 para procesos espaciales. Sea  $Z(\cdot, \cdot)$  un proceso intrínsecamente estacionario observado sobre un conjunto de  $n$  localizaciones espacio-temporales  $\{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)\}$ , podemos obtener una estimación de su variograma  $2\gamma(\cdot, \cdot)$  (y de su función de covarianza  $C(\cdot, \cdot)$  si el proceso es además estacionario de segundo orden) a partir de los valores observados utilizando el estimador clásico propuesto por Matheron o el estimador robusto de Cressie y Hawkins.

#### Estimador clásico.

Es el estimador obtenido aplicando el método de los momentos, que para el variograma del proceso viene dado por

$$2\hat{\gamma}(\mathbf{h}(l), u(k)) = \frac{1}{|N(\mathbf{h}(l), u(k))|} \sum_{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j) \in N(\mathbf{h}(l), u(k))} (Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2, \quad (3.22)$$

donde  $N(\mathbf{h}(l), u(k)) = \{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j) : \mathbf{s}_i - \mathbf{s}_j \in T(\mathbf{h}(l)), t_i - t_j \in T(u(k))\}$ , siendo  $T(\mathbf{h}(l))$  una región de tolerancia en  $\mathfrak{R}^d$  alrededor de  $\mathbf{h}(l)$  y  $T(u(k))$  una región de tolerancia en  $\mathcal{T}$  alrededor de  $u(k)$ , y  $|N(\mathbf{h}(l), u(k))|$  el número de elementos distintos en  $N(\mathbf{h}(l), u(k))$ , con  $l = 1, \dots, L$  y  $k = 1, \dots, K$ .

Generalmente las regiones  $T(\mathbf{h}(l))$  y  $T(u(k))$  se escogen de forma que den como resultado conjuntos disjuntos, y con un número de elementos suficientes para dar lugar a estimaciones estables. Si tiene sentido asumir la hipótesis de isotropía espacial para el proceso espacio-temporal analizado, entonces se suele escoger la región de tolerancia espacial alrededor de cada uno de los valores  $h(l)$  como  $[h(l) - d_l/2, h(l) + d_l/2]$ , siendo  $d_l$  la tolerancia espacial utilizada. También es usual que la componente temporal tome valores en  $\mathbb{Z}$ , en cuyo caso estimamos el variograma empírico para  $u(k) = 0, 1, \dots$ , obtenidos como las sucesivas diferencias de los tiempos en los que el proceso es observado.

Para la covarianza, el estimador obtenido por el método de los momentos sería

$$\hat{C}(\mathbf{h}(l), u(k)) = |N(\mathbf{h}(l), u(k))| \sum_{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j) \in N(\mathbf{h}(l), u(k))} (Z(\mathbf{s}_i, t_i) - \hat{Z})(Z(\mathbf{s}_j, t_j) - \hat{Z}),$$

donde  $\hat{Z} = \frac{1}{n} \sum_i Z(\mathbf{s}_i, t_i)$  es un estimador de la media  $\mu$  del proceso, y  $N(\mathbf{h}(l), u(k))$  se define como antes.

#### Estimador robusto

Aunque el método de estimación clásico presenta la ventaja de su facilidad de cálculo, tiene algunos inconvenientes prácticos, como la falta de robustez frente a valores extremos de  $Z(\mathbf{s})$ . Para evitar este problema, Cressie y Hawkins definen el estimador del variograma

siguiente

$$2\hat{\gamma}_u(\mathbf{h}(l), u(k)) = \left( \frac{1}{|N(\mathbf{h}(l), u(k))|} \sum_{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j) \in N(\mathbf{h}(l), u(k))} |Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)|^{1/2} \right)^4 \times \left( 0.457 + \frac{0.494}{|N_u(\mathbf{h}(l), u(k))|} \right)^{-1}, \quad (3.23)$$

donde  $N(\mathbf{h}(l), u(k))$  y  $|N(\mathbf{h}(l), u(k))|$  se definen como en el caso anterior.

Como ocurría en el caso espacial, estos estimadores de la función de covarianza o del variograma del proceso no satisfacen en general la condición de ser definidos positivos o condicionalmente definidas negativas, respectivamente. Por ello, en la práctica seleccionaremos un modelo paramétrico de funciones de covarianza o de variogramas que sepanos dan lugar a modelos válidos, y estimaremos aquellos valores de los parámetros que dan lugar a una función de covarianza o un variograma que mejor se ajuste a los valores empíricos obtenidos con estos estimadores (3.22) o (3.23). En el Capítulo 4 estudiaremos en profundidad los principales modelos de funciones de covarianza y de variograma utilizados en la literatura, con el fin de disponer de una batería de modelos lo suficientemente amplia y flexible que permita el análisis de cualquier fenómeno natural.

### 3.5. Modelización de procesos espacio-temporales

En la modelización práctica, el investigador debe tomar la importante decisión de cuál es el modelo espacio-temporal que mejor ajusta sus datos empíricos. En esta sección vamos a introducir algunas de las estrategias de modelización más utilizadas. En Kyriakidis y Journel (1999), Banerjee et al. (2004) o Chen et al. (2006) podemos encontrar una revisión completa y exhaustiva de las diferentes técnicas de modelización de procesos espacio-temporales.

Sea  $\{Z(\mathbf{s}, t), \mathbf{s} \in D, t \in T\}$  un determinado proceso espacio-temporal de interés. Supongamos que este proceso se ha observado sobre un conjunto de localizaciones espacio-temporales  $\{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)\} \in D \times T$ , obteniendo un conjunto de valores del proceso  $\{Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n)\}$ . Como hemos visto en (3.3), el proceso espacio-temporal  $Z(\mathbf{s}, t)$  suele descomponerse como

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \delta(\mathbf{s}, t), \quad (\mathbf{s}, t) \in D \times T, \quad (3.24)$$

donde  $\mu(\mathbf{s}, t) = E[Z(\mathbf{s}, t)]$  y  $\delta(\mathbf{s}, t)$  es un proceso estocástico de media cero que caracteriza la dependencia espacio-temporal y que modeliza las fluctuaciones espacio-temporales de  $Z(\mathbf{s}, t)$  alrededor de su media  $\mu(\mathbf{s}, t)$ .

En ocasiones supondremos que el proceso  $Z(\mathbf{s}, t)$  ha sido observado con un determinado error de medida, con lo que

$$Z(\mathbf{s}, t) = Y(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t), \quad (3.25)$$

siendo  $Y(\mathbf{s}, t)$  el verdadero proceso de interés en nuestro estudio al que se le ha filtrado el error de medida asociado a su observación, y  $\varepsilon(\mathbf{s}, t)$  un proceso de ruido blanco de media

cero con distribución normal y varianza  $\sigma_\varepsilon > 0$ . En principio,  $\sigma_\varepsilon$  podría variar con el tiempo, aunque en la mayoría de aplicaciones la consideraremos constante aplicando el principio de parsimonia.

La Geoestadística clásica se basa en considerar procesos espacio-temporales  $Z(\mathbf{s}, t)$  sujetos a algún tipo de estacionariedad. Como hemos visto en la Sección 3.2, la estacionariedad de segundo orden del proceso indicaba que su media era constante y que la covarianza del proceso sobre dos localizaciones cualesquiera dependía únicamente de la distancia espacial y temporal existente entre ellas. Del mismo modo, definíamos la estacionariedad intrínseca del proceso si tenía media constante y la varianza de la diferencia del proceso entre dos localizaciones cualesquiera dependía únicamente de la distancia espacial y temporal existente entre ellas. No obstante, los procesos reales rara vez verificarán directamente estas condiciones de estacionariedad al actuar sobre grandes regiones espaciales y sobre tiempos distintos, lo que obligará a considerar modelos no estacionarios en su modelización. Generalmente la no estacionariedad se limitará a la componente de tendencia  $\mu(\mathbf{s}, t)$ , que en lugar de considerarse constante se modelará como una función del espacio, del tiempo, o de ambos, mientras que la componente  $\delta(\mathbf{s}, t)$  se considerará estacionaria.

Kyriakidis y Journel (1999) distinguen dos tipos de modelos espacio-temporales, según se considere la media  $\mu(\mathbf{s}, t)$  como una componente determinista o como un proceso estocástico. En el primer caso, la media  $\mu(\mathbf{s}, t)$  se considerará como una función de las coordenadas espacio-temporales  $\mathbf{s}$  y  $t$ , mientras que en el segundo caso  $\mu(\mathbf{s}, t)$  será un proceso espacio-temporal no estacionario que modelizará la distribución espacio-temporal del proceso de la media, y que suele considerarse incorrelado con  $\delta(\mathbf{s}, t)$ , para todo  $(\mathbf{s}, t) \in D \times T$ . En ambos casos, el proceso estocástico  $\delta(\mathbf{s}, t)$  de media cero suele considerarse estacionario. Todas estas componentes pueden considerarse como combinaciones de componentes puramente espaciales o temporales. Así, por ejemplo, en determinadas situaciones puede adoptarse una descomposición para  $\mu(\mathbf{s}, t)$  igual a  $\mu_1(\mathbf{s})$ , a  $\mu_2(t)$ , a  $\mu_1(\mathbf{s}) + \mu_2(t)$ , o a  $\mu_1(\mathbf{s})\mu_2(t)$ , siendo  $\mu_1(\mathbf{s})$  y  $\mu_2(t)$  funciones que dependen únicamente de las coordenadas espaciales y temporales, respectivamente. Observemos que todas estas descomposiciones son siempre decisiones subjetivas que tomaremos durante la modelización, ya que únicamente disponemos de datos observados para el proceso  $Z(\mathbf{s}, t)$  pero no para su media ni para los residuos correspondientes.

Generalmente consideraremos una media determinista para el proceso que descompondremos como

$$\mu(\mathbf{s}, t) = \sum_{k=1}^p \beta_k f_k(\mathbf{s}, t), \quad (\mathbf{s}, t) \in D \times T, \quad (3.26)$$

con lo que  $\mu(\mathbf{s}, t)$  se expresa como función de  $p$  funciones conocidas  $f_k(\mathbf{s}, t)$  elegidas durante la modelización para ajustar la media observada de los datos, y donde  $\beta_k$  son  $p$  coeficientes desconocidos que ajustaremos. Por ejemplo, podemos considerar funciones periódicas que dependan de  $t$  para capturar variaciones periódicas sobre el eje temporal, y funciones polinomiales o continuas a trozos que dependan de  $\mathbf{s}$  para modelizar variaciones suaves o discontinuidades en el espacio. Dimitrakopoulos y Luo (1997) proporcionan diferentes expresiones para la tendencia permisibles en el contexto espacio-temporal. Los coeficientes de (3.26) pueden considerarse fijos, o dependientes del espacio y del tiempo,  $\beta_k = \beta_k(\mathbf{s}, t)$ .

Si  $\mathcal{T} = \mathbb{Z}$ , entonces puede considerarse  $\beta_k(\mathbf{s}, t) = \beta_k(t)$  si hay suficientes medidas en el tiempo para sugerir un cambio temporal de los coeficientes de la relación (3.26). Del mismo modo, pueden considerarse coeficientes  $\beta_k(\mathbf{s}, t) = \beta_k(\mathbf{s})$  que varíen espacialmente.

Observemos que nunca dispondremos de datos para las medias ni para los residuos en ninguna localización, por lo que la estimación de los residuos asociados al proceso y obtenidos al abstraer la media ajustada al proceso original dependerán de la forma de la tendencia considerada y del procedimiento utilizado para ajustar dicha tendencia, por lo que dependerán de una decisión tomada durante la modelización. Pero mientras no dispongamos de información auxiliar, consideraremos los residuos del modelo admitido como un conjunto de observaciones de la variabilidad espacio-temporal no incluida en el modelo de tendencia considerado, y que usaremos en la estimación de las funciones de covarianza o del variograma correspondientes, a los que ajustaremos modelos válidos que utilizaremos durante la predicción del proceso en cualquier nueva localización.

Todos estos modelos basados en la descomposición del proceso espacio-temporal como suma de una tendencia determinista y unos efectos aleatorios con respecto al tiempo, al espacio, o al espacio-tiempo, presentan problemas cuando los datos provienen de un proceso de grandes dimensiones con una estructura dinámica complicada. En estos casos, el filtro de Kalman suele ser una herramienta de gran utilidad, dando lugar a un modelo dinámico desde el punto de vista temporal y descriptivo desde el punto de vista espacial, y que analizaremos a continuación.

### 3.5.1. El filtro de Kalman

El estudio y análisis de los procesos espacio-temporales puede llevarse a cabo mediante una formulación espacio-estado dinámica, lo que convierte al filtro de Kalman (Kalman, 1960) en una forma natural de llevar a cabo predicciones en este contexto. Aunque la comunidad científica viene utilizando el filtro de Kalman para realizar predicciones espacio-temporales desde hace más de dos décadas (Ghil et al., 1981), su uso por parte de los estadísticos es mucho más reciente. Fue durante los años 90 cuando varios investigadores constataron de forma independiente el gran potencial de la inclusión de un aspecto temporal dinámico en la modelización estadística espacio-temporal. Entre los trabajos que analizan este aspecto podemos citar Goodall y Mardia (1994), Huang y Cressie (1996), Mardia et al. (1998), Meiring et al. (1998), Berke (1998) o Wikle y Cressie (1999), entre otros. Podemos encontrar una completa descripción del uso del filtro de Kalman en el análisis espacio-temporal en Cressie y Wikle (2002) o en Chen et al. (2006).

A continuación vamos a describir brevemente el modelo presentado por Huang y Cressie (1996), que utiliza en la modelización una estructura de covarianza separable obtenida a partir de una función de covarianza espacial arbitraria y de una función de covarianza temporal con estructura autorregresiva de orden  $p$ ,  $AR(p)$ . Este modelo es bastante general, cubriendo muchas de las covarianzas espacio-temporales utilizadas en la literatura. Como veremos, la inferencia y la predicción se implementarán mediante la técnica del filtro de Kalman, que presenta numerosas ventajas desde el punto de vista computacional.

Supongamos que estamos interesados en el análisis de un proceso espacio-temporal Gaussiano  $\{Y(\mathbf{s}, t), \mathbf{s} \in D \subseteq \mathbb{R}^d, t \in \mathbb{Z}\}$  estacionario con media cero y función de covarianza

espacio-temporal  $C(\mathbf{h}, u) = \text{Cov}(Y(\mathbf{s}_1, t_1), Y(\mathbf{s}_2, t_2))$ , siendo  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  la distancia espacial y  $u = t_1 - t_2$  la distancia temporal. Supongamos que la observación del proceso tiene asociado un error de medida con una estructura de ruido blanco, esto es, que el proceso espacio-temporal observado es

$$\mathbf{Z}_t(\mathbf{s}) = Y_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \quad (3.27)$$

donde  $\varepsilon_t(\mathbf{s})$  es un proceso Gaussiano de media cero y varianza  $\sigma_\varepsilon^2$ . Supongamos que se han observado los valores de este proceso sobre un conjunto de  $n$  localizaciones espaciales y durante  $T$  tiempos, con lo que tendremos un conjunto de vectores  $\mathbf{Z}_t = (Z_t(\mathbf{s}_1), \dots, Z_t(\mathbf{s}_n))'$ , con  $t = 1, \dots, T$ .

Vamos a suponer que el proceso evoluciona en el tiempo de manera markoviana, de forma que sólo los valores observados durante los  $p$  tiempos anteriores influyen sobre el valor observado en el tiempo actual. Para ello linealizamos esta dependencia temporal como un proceso autorregresivo de orden  $p$ ,

$$Y_t(\mathbf{s}) = \alpha_1 Y_{t-1}(\mathbf{s}) + \dots + \alpha_p Y_{t-p}(\mathbf{s}) + \eta_t(\mathbf{s}), \quad (3.28)$$

donde las constantes  $(\alpha_1, \dots, \alpha_p)'$  se eligen de forma que se logre estacionariedad temporal, y donde  $\{\eta_t(\mathbf{s}), \mathbf{s} \in D, t \in T\}$  es un proceso Gaussiano estacionario con media cero tal que  $\eta_{t-1}(\cdot)$  y  $\eta_{t-2}(\cdot)$  son independientes para todo  $t_1 \neq t_2$ ,  $\eta_t(\cdot)$  y  $Y_{t-1}(\cdot)$  son independientes para todo  $t$ , y  $\varepsilon_{t-1}(\cdot)$  y  $\eta_{t-2}(\cdot)$  son independientes para todo  $t_1 \neq t_2$ .

El filtro de Kalman es un algoritmo recursivo que permite inferencias de las variables de estado de un modelo espacio-estado definido por (3.27) y (3.28). Sea  $\hat{Y}_{t'|t}(\mathbf{s}) = E[Y_{t'}(\mathbf{s}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t]$ , puede demostrarse que es una función lineal de  $\mathbf{Z}_t$  y de  $\hat{Y}_{t'-1|t-1}(\mathbf{s})$  para cualquier  $t'$ , por lo que el filtro de Kalman permite adaptar la variable estado cuando una nueva observación está disponible. Con el procedimiento que se describe a continuación veremos como obtener el predictor espacio-temporal óptimo de  $\hat{Y}_t(\mathbf{s}_0)$ , no sólo para aquellas localizaciones para las que tenemos observaciones, sino también para aquellas en las que no se dispone de observaciones.

Vamos a denotar

$$\begin{aligned} \hat{Z}_{t'|t}(\mathbf{s}) &= E[Z_{t'}(\mathbf{s}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t] \\ \hat{Z}_{t'|t} &= E[Z_{t'} | \mathbf{Z}_1, \dots, \mathbf{Z}_t] \\ \Sigma_{t_1, t_2|t}(\mathbf{s}, \mathbf{r}) &= \text{Cov}[Y_{t_1}(\mathbf{s}), Y_{t_2}(\mathbf{r}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t] \\ \Sigma_{t'|t}(\mathbf{s}, \mathbf{r}) &= \text{Cov}[Y_{t'}(\mathbf{s}), Y_{t'}(\mathbf{r}) | \mathbf{Z}_1, \dots, \mathbf{Z}_t] \\ \Sigma_{t'|t}(\mathbf{s}, \mathbf{r}) &= \text{Cov}[(Y_{t'-p+1}(\mathbf{s}), \dots, Y_{t'}(\mathbf{s}))', (Y_{t'-p+1}(\mathbf{r}), \dots, Y_{t'}(\mathbf{r}))' | \mathbf{Z}_1, \dots, \mathbf{Z}_t] \\ \mathbf{F}_t &= \text{Var}[\mathbf{Z}_t | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] + \sigma_\varepsilon^2 \mathbf{I}_n \\ \mathbf{G}_t(\mathbf{s}) &= \text{Cov}[\mathbf{Z}_t, (Y_{t-p+1}(\mathbf{s}), \dots, Y_t(\mathbf{s}))' | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] \\ \boldsymbol{\alpha} &= (\alpha_p, \alpha_{p-1}, \dots, \alpha_1)' \\ \mathbf{A} &= (\mathbf{e}'_2, \dots, \mathbf{e}'_p, \boldsymbol{\alpha}')' \end{aligned}$$

donde  $\mathbf{e}_i$  es un vector con  $p-1$  ceros y un uno en la posición  $i$ , con  $i = 2, \dots, p$ .

Todas las distribuciones multivariantes, conjuntas o condicionadas, son Gaussianas. Así, la distribución de  $(Y_{t-p+1}(\mathbf{s}), \dots, Y_t(\mathbf{s}), Y_{t-p+1}(\mathbf{r}), \dots, Y_t(\mathbf{r}), \mathbf{Z}_t)'$  condicionada a  $(\mathbf{Z}_1, \dots, \mathbf{Z}_{t-1})$  es Gaussiana multivariante con media

$$\left( \hat{Y}_{t-p+1|t-1}(\mathbf{s}), \dots, \hat{Y}_{t|t-1}(\mathbf{s}), \hat{Y}_{t-p+1|t-1}(\mathbf{r}), \dots, \hat{Y}_{t|t-1}(\mathbf{r}), \mathbf{Z}'_{t|t-1} \right)',$$

y con matriz de covarianzas

$$\begin{pmatrix} \Sigma_{t|t-1}(\mathbf{s}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) & (\mathbf{G}_t(\mathbf{s}))' \\ \Sigma_{t|t-1}(\mathbf{r}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{r}, \mathbf{r}) & (\mathbf{G}_t(\mathbf{r}))' \\ \mathbf{G}_t(\mathbf{s}) & \mathbf{G}_t(\mathbf{r}) & \mathbf{F}_t \end{pmatrix}$$

Por tanto, la distribución de  $(Y_{t-p+1}(\mathbf{s}), \dots, Y_t(\mathbf{s}), Y_{t-p+1}(\mathbf{r}), \dots, Y_t(\mathbf{r}))'$  condicionada a  $(\mathbf{Z}_1, \dots, \mathbf{Z}_t)$  será Gaussiana multivariante con media

$$\left( \hat{Y}_{t-p+1|t-1}(\mathbf{s}), \dots, \hat{Y}_{t|t-1}(\mathbf{s}), \hat{Y}_{t-p+1|t-1}(\mathbf{r}), \dots, \hat{Y}_{t|t-1}(\mathbf{r}) \right)' + (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r}))' \mathbf{F}_t^{-1} (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r})),$$

y con matriz de covarianzas

$$\begin{pmatrix} \Sigma_{t|t-1}(\mathbf{s}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) \\ \Sigma_{t|t-1}(\mathbf{r}, \mathbf{s}) & \Sigma_{t|t-1}(\mathbf{r}, \mathbf{r}) \end{pmatrix} - (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r}))' \mathbf{F}_t^{-1} (\mathbf{G}_t(\mathbf{s}), \mathbf{G}_t(\mathbf{r})).$$

Tenemos por tanto que

$$\begin{pmatrix} \hat{Y}_{t-p+1|t}(\mathbf{s}) \\ \vdots \\ \hat{Y}_{t|t}(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} \hat{Y}_{t-p+1|t-1}(\mathbf{s}) \\ \vdots \\ \hat{Y}_{t|t-1}(\mathbf{s}) \end{pmatrix} + (\mathbf{G}_t(\mathbf{s}))' \mathbf{F}_t^{-1} (\mathbf{Z}_t - \hat{\mathbf{Z}}_{t|t-1}), \quad (3.29)$$

$$\Sigma_{t|t}(\mathbf{s}, \mathbf{r}) = \Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) - (\mathbf{G}_t(\mathbf{s}))' \mathbf{F}_t^{-1} \mathbf{G}_t(\mathbf{r}), \quad (3.30)$$

donde

$$\begin{aligned} \hat{\mathbf{Z}}_{t|t-1}(\mathbf{s}) &= E[\mathbf{Z}_t(\mathbf{s}) | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] = E[Y_t(\mathbf{s}) | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] \\ &= \boldsymbol{\alpha}' \left( \hat{Y}_{t-p|t-1}(\mathbf{s}), \dots, \hat{Y}_{t-1|t-1}(\mathbf{s}) \right)', \end{aligned} \quad (3.31)$$

$$\Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) = \mathbf{A} \Sigma_{t-1|t-1}(\mathbf{s}, \mathbf{r}) \mathbf{A}' + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & C\eta(\mathbf{s}, \mathbf{r}) \end{pmatrix}, \quad (3.32)$$

$$\mathbf{F}_t = \text{Var}[\mathbf{Z}_t | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] = \text{Var}[Y_t | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] + \sigma_\varepsilon^2 \mathbf{I}, \quad (3.33)$$

$$\begin{aligned} \mathbf{G}_t(\mathbf{s}) &= \text{Cov}[\mathbf{Z}_t, (Y_{t-p+1}(\mathbf{s}), \dots, Y_t(\mathbf{s}))' | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}] \\ &= \text{Cov}[Y_t, (Y_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}))' | \mathbf{Z}_1, \dots, \mathbf{Z}_{t-1}]. \end{aligned} \quad (3.34)$$

Sea  $(\mathbf{s}_0, t_0)$  una nueva localización sobre la que deseamos predecir el valor del proceso de interés a partir de los valores observados  $\mathbf{Z}_1, \dots, \mathbf{Z}_{t_0}$ . Sean  $(\hat{Y}_{-p+1|0}(\mathbf{s}_0), \dots, \hat{Y}_{-1|0}(\mathbf{s}_0))$ ,



$\hat{Y}_{0|0}(\mathbf{s}_0)$ ' y  $\Sigma_{0|0}(\mathbf{s}, \mathbf{r})$  unas predicciones iniciales de  $(Y_{-p+1}(\mathbf{s}_0), \dots, Y_{-1}(\mathbf{s}_0), Y_0(\mathbf{s}_0))'$  y de la matriz de varianzas y covarianzas entre  $(Y_{-p+1}(\mathbf{s}), \dots, Y_{-1}(\mathbf{s}), Y_0(\mathbf{s}))' - (\hat{Y}_{-p+1|0}(\mathbf{s}), \dots, \hat{Y}_{-1|0}(\mathbf{s}), \hat{Y}_{0|0}(\mathbf{s}))'$  y  $(Y_{-p+1}(\mathbf{r}), \dots, Y_{-1}(\mathbf{r}), Y_0(\mathbf{r}))' - (\hat{Y}_{-p+1|0}(\mathbf{r}), \dots, \hat{Y}_{-1|0}(\mathbf{r}), \hat{Y}_{0|0}(\mathbf{r}))'$ , respectivamente. Entonces mediante un procedimiento recursivo consistente en la actualización de las ecuaciones (3.29)-(3.34) obtendremos el predictor de  $(Y_{t_0-p+1}(\mathbf{s}_0), \dots, Y_{t_0}(\mathbf{s}_0))'$  de menor error cuadrático medio, y su matriz de errores cuadráticos medios asociada  $\Sigma_{t_0|t_0}(\mathbf{s}_0, \mathbf{s}_0)$ .

Debido a las condiciones del proceso, una elección razonable para el predictor inicial que sirve de partida del procedimiento anterior es considerar  $(\hat{Y}_{-p+1|0}(\mathbf{s}), \dots, \hat{Y}_{-1|0}(\mathbf{s}), \hat{Y}_{0|0}(\mathbf{s}))'$  =  $E[(Y_{-p+1}(\mathbf{s}), \dots, Y_{-1}(\mathbf{s}), Y_0(\mathbf{s}))'] = \mathbf{0}$  y  $\Sigma_{0|0}(\mathbf{s}, \mathbf{r})$  la matriz  $p \times p$  cuyo elemento  $(i, j)$  es  $C_{i-j}^Y(\mathbf{s}, \mathbf{r})$ .

Observemos que el predictor de  $Y_{t_0+1}(\mathbf{s}_0)$  que tiene error cuadrático medio mínimo y dicho error cuadrático medio asociado vienen dados por

$$\begin{aligned}\hat{Y}_{t_0+1|t_0}(\mathbf{s}_0) &= \alpha_1 \hat{Y}_{t_0|t_0}(\mathbf{s}_0) + \alpha_2 \hat{Y}_{t_0-1|t_0}(\mathbf{s}_0) + \dots + \alpha_p \hat{Y}_{t_0-p+1|t_0}(\mathbf{s}_0), \\ \Sigma_{t_0+1|t_0}(\mathbf{s}_0) &= \alpha' \Sigma_{t_0|t_0}(\mathbf{s}_0, \mathbf{s}_0) \alpha + C^\eta(\mathbf{s}_0, \mathbf{s}_0),\end{aligned}$$

donde  $\hat{Y}_{t_0|t_0}(\mathbf{s}_0), \hat{Y}_{t_0-1|t_0}(\mathbf{s}_0), \dots, \hat{Y}_{t_0-p+1|t_0}(\mathbf{s}_0)$  y  $\Sigma_{t_0|t_0}(\mathbf{s}_0)$  se obtienen mediante el procedimiento anterior.

## 3.6. Ajuste y evaluación de los modelos espacio-temporales

### 3.6.1. Ajuste mediante mínimos cuadrados

Cualquiera de los procedimientos de estimación mediante mínimos cuadrados que vimos en la Sección 2.5.1 para el caso espacial, pueden ser fácilmente generalizados para el caso espacio-temporal. Supongamos que el proceso que estamos analizando se ha descompuesto de forma que, o bien el proceso original o bien el proceso de los residuos una vez modelizada su media, es un proceso intrínsecamente estacionario. En ese caso, podemos estimar los parámetros que definen el modelo de semivariograma seleccionado utilizando mínimos cuadrados ordinarios, generalizados o ponderados. Esta última opción es la más utilizada, y consiste en minimizar

$$\sum_{i=1}^n w_i (\hat{\gamma}(\mathbf{h}_i, u_i) - \gamma(\mathbf{h}_i, u_i | \boldsymbol{\theta}))^2 = (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta}))' W(\boldsymbol{\theta})^{-1} (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta})) \quad (3.35)$$

donde  $W(\boldsymbol{\theta})$  es una matriz diagonal cuyos elementos son las varianzas de  $\hat{\boldsymbol{\gamma}}$ , esto es,  $w_j = \text{Var}(\hat{\boldsymbol{\gamma}}(\mathbf{h}_j, u_j))^{-1} = N(\mathbf{h}_j, u_j) / (2\gamma(\mathbf{h}_j, u_j | \boldsymbol{\theta}))^2$ , siendo  $N(\mathbf{h}_j, u_j)$  el número de localizaciones a distancia espacial  $\mathbf{h}_j$  y distancia temporal  $u_j$ .

### 3.6.2. Ajuste mediante máxima-verosimilitud

Si el proceso espacio-temporal analizado es Gaussiano, entonces podemos utilizar un método más eficiente para la estimación de los parámetros del modelo, el método de máxima

verosimilitud. Consideremos la descomposición (3.24) del proceso  $Z(\mathbf{s}, t)$  como

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \delta(\mathbf{s}, t), \quad (\mathbf{s}, t) \in D \times T,$$

donde  $\mu(\mathbf{s}, t) = E[Z(\mathbf{s}, t)]$  y  $\delta(\mathbf{s}, t)$  es un proceso estocástico de media cero que caracteriza la dependencia espacio-temporal y que modeliza las fluctuaciones espacio-temporales de  $Z(\mathbf{s}, t)$  alrededor de su media  $\mu(\mathbf{s}, t)$ . Vamos a suponer que la media del proceso se ha modelizado según (3.26), con lo que  $\mu(\mathbf{s}, t) = \beta' \mathbf{f}(\mathbf{s}, t)$ , siendo  $\mathbf{f}(\mathbf{s}, t)$  un vector de  $p$  funciones que en general dependen de la localización espacio-temporal y  $\beta$  el vector de  $p$  coeficientes asociado.

Supongamos que se han observado  $n$  datos del proceso  $Z(\mathbf{s}, t)$  de interés sobre un conjunto de localizaciones espacio-temporales  $\{(\mathbf{s}_1, t_1), \dots, (\mathbf{s}_n, t_n)\}$ . Entonces, bajo el supuesto de normalidad

$$\mathbf{Z} = (Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n))' \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

siendo  $\boldsymbol{\mu} = (\mu(\mathbf{s}_1, t_1), \dots, \mu(\mathbf{s}_n, t_n)) = \beta' \mathbf{X}$ , con  $\mathbf{X}$  la matriz  $p \times n$  cuya columna  $i$ -ésima es  $\mathbf{f}(\mathbf{s}_i, t_i)$ , y  $\boldsymbol{\Sigma}$  la matriz  $n \times n$  cuyo elemento  $(i, j)$  es  $C(Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j))$ .

Supongamos que se ha seleccionado un determinado modelo para la función de covarianza asociada al proceso que depende de un vector de parámetros  $\boldsymbol{\theta}$ , con lo que tendremos una expresión para la matriz de varianzas-covarianzas  $\boldsymbol{\Sigma}$  que dependerá del vector  $\boldsymbol{\theta}$  ( $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ ). El logaritmo de la función de verosimilitud de los parámetros del modelo  $\beta$  y  $\boldsymbol{\theta}$  asociada a los datos  $\mathbf{z} = (z(\mathbf{s}_1, t_1), \dots, z(\mathbf{s}_n, t_n))'$  vendrá dado por

$$L(\beta, \boldsymbol{\theta} | \mathbf{z}) = -\frac{1}{2}(\mathbf{z} - \mathbf{X}\beta)' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\beta) - \frac{1}{2} \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) - \frac{n}{2} \log(2\pi).$$

La estimación por máxima verosimilitud, que vimos en la Sección 2.5.2 para el caso espacial, consiste en obtener de forma simultánea los valores de  $\beta$  y  $\boldsymbol{\theta}$  que maximizan la función anterior, o lo que es lo mismo, que minimizan  $-2$  veces la log verosimilitud

$$-2L(\beta, \boldsymbol{\theta} | \mathbf{z}) = (\mathbf{z} - \mathbf{X}\beta)' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\beta) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + n \log(2\pi). \quad (3.36)$$

En el caso en que asumamos el modelo (3.25), es decir, supongamos un error de medida durante la observación del proceso consistente en un ruido blanco de media cero y varianza  $\sigma_{\epsilon}$ , entonces en la expresión anterior a los elementos de la diagonal de la matriz  $\boldsymbol{\Sigma}$  les sumaremos la varianza  $\sigma_{\epsilon}$ . Si puede asumirse estacionariedad en la media, la expresión a minimizar quedaría

$$-2L(\beta, \boldsymbol{\theta} | \mathbf{z}) = \mathbf{z}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{z} + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + n \log(2\pi).$$

Como vimos en la Sección 2.5.2 para el caso espacial, podemos utilizar la máxima verosimilitud restringida, que consiste en obtener de forma simultánea los valores de  $\beta$  y  $\boldsymbol{\theta}$  que minimizan

$$-2L(\beta, \boldsymbol{\theta} | \mathbf{z}) = (\mathbf{z} - \mathbf{X}\beta)' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\beta) + \log(|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|) + \log(|\mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \mathbf{X}|) + (n - q) \log(2\pi),$$

siendo  $q$  el número total de parámetros asociados al modelo.

### 3.6.3. Ajuste mediante verosimilitud compuesta ponderada

En la Sección 2.5.3 hemos introducido un procedimiento de estimación de los parámetros del modelo mediante verosimilitud compuesta ponderada, que trataba de obtener estimaciones de calidad comparable a las obtenidas mediante máxima verosimilitud pero con menor coste de computación. Hemos visto que para un conjunto de  $n$  datos experimentales, los métodos de máxima verosimilitud exigen la inversión y cálculo del determinante de la matriz de varianzas-covarianzas, lo que conlleva un coste computacional de  $\mathcal{O}(n^3)$  operaciones. En el caso de los estudios espacio-temporales, en los es habitual observar el valor del proceso sobre un conjunto de  $n$  localizaciones durante un conjunto de  $t$  instantes temporales, el número de datos  $nt$  hará que el coste de los procedimientos de máxima verosimilitud sea en muchas ocasiones prohibitivo, lo que nos llevará a procedimientos de menor coste computacional como el ajuste mediante verosimilitud compuesta. Veamos a continuación de forma resumida cómo podemos generalizar el procedimiento espacial de la Sección 2.5.3 al caso espacio-temporal, siguiendo el desarrollo descrito en Bevilacqua et al. (2007).

Sea  $Z(\mathbf{s}, t)$  un proceso estocástico espacio-temporal gaussiano intrínsecamente estacionario del que se han observado un conjunto de valores  $\{Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n)\}$  sobre la región espacio-temporal objeto de estudio. Las diferencias del proceso en las localizaciones muestreadas seguirán una distribución gaussiana

$$R(\mathbf{r}, \mathbf{s}, t, q) = Z(\mathbf{r}, t) - Z(\mathbf{s}, q) \sim N(0, 2\gamma(\mathbf{r} - \mathbf{s}, t - q; \boldsymbol{\theta})),$$

siendo  $\gamma(\cdot, \cdot; \boldsymbol{\theta})$  un modelo válido de semivariograma donde  $\boldsymbol{\theta} \in \Theta$  es el vector paramétrico  $p$ -dimensional desconocido a estimar.

En este procedimiento consideraremos la suma ponderada de las log-verosimilitudes negativas marginales de las diferencias anteriores  $CL(\boldsymbol{\theta}, h, u)$ , obteniendo la verosimilitud compuesta ponderada

$$WCL(\boldsymbol{\theta}, h, u) = \sum_{i < j} cl(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j; \boldsymbol{\theta}) \omega_{ij}, \quad (3.37)$$

donde los pesos  $\omega_{ij}$  valen uno para los pares de localizaciones espacio-temporales situadas a una distancia espacial  $\|\mathbf{s}_i - \mathbf{s}_j\| \leq h$  y a una distancia temporal  $\|t_i - t_j\| \leq u$ , y cero en otro caso. El valor  $\hat{\boldsymbol{\theta}}$  se obtiene minimizando  $WCL(\boldsymbol{\theta}, h^*, u^*)$  respecto a  $\boldsymbol{\theta} \in \Theta$ , siendo  $(h^*, u^*) = \min_{h, u} tr(G(\boldsymbol{\theta}, h, u))$ , con  $G(\boldsymbol{\theta}, h, u)$  la inversa de la matriz de información de Godambe asociada a la función de estimación ponderada. La búsqueda simultánea de la distancia espacio-temporal óptima  $(h^*, u^*)$  puede ser complicada, en especial si tenemos muchos parámetros en el modelo de covarianza seleccionado. Una solución práctica puede ser analizar de forma separada la marginal espacial y la marginal temporal del modelo y encontrar  $h^*$  y  $u^*$  de forma independiente. En Bevilacqua et al. (2007) pueden encontrarse más detalles sobre este procedimiento de estimación.

### 3.6.4. Validación y comparación de modelos

Supongamos que se han observado los valores del proceso de interés  $Z(\mathbf{s}, t)$  sobre un determinado conjunto de  $n$  localizaciones espacio-temporales. Si el tamaño de este conjunto

lo permite, podemos dividirlo en dos subconjuntos, uno de modelización formado por  $n - k$  localizaciones y otro de validación formado por las  $k$  localizaciones restantes. A partir del conjunto de modelización, ajustaremos un modelo espacio-temporal a los  $n - k$  datos observados. A partir de dicho modelo obtendremos, para cada una de las  $k$  localizaciones  $(s_i, t_i)$  del conjunto de validación  $\{Z(s_1, t_1), \dots, Z(s_k, t_k)\}$ , la predicción  $\hat{Z}(s_i, t_i)$  y el error de predicción asociado  $\sigma^2(s_i, t_i)$  obtenidos mediante alguno de los procedimientos de kriging que veremos en la Sección 3.7. Una vez obtenidas las  $k$  predicciones y sus errores de predicción asociados, podemos calcular los siguientes estadísticos de resumen que permitirán la evaluación de la capacidad predictiva del modelo ajustado:

- Error de la predicción medio:

$$EPM = \frac{1}{k} \sum_1^k (\hat{Z}(s_i, t_i) - Z(s_i, t_i)).$$

- Error de la predicción estandarizado medio:

$$EPEM = \frac{1}{k} \sum_1^k (\hat{Z}(s_i, t_i) - Z(s_i, t_i)) / \sigma(s_i, t_i).$$

- Error de la predicción estandarizado cuadrático medio:

$$EPECM = \frac{1}{k} \sum_1^k ((\hat{Z}(s_i, t_i) - Z(s_i, t_i)) / \sigma(s_i, t_i))^2.$$

- Porcentaje de cubrimiento (suponiendo normalidad):

$$PCOV = \frac{1}{k} \sum_1^k 1_{[\hat{Z}(s_i, t_i) - 1,96\sigma^2(s_i, t_i), \hat{Z}(s_i, t_i) + 1,96\sigma^2(s_i, t_i)]}(Z(s_i, t_i)).$$

donde  $1_{[a,b]}(x)$  es la función característica del intervalo  $[a, b]$ . También podemos calcular las raíces cuadradas de los valores  $EPCM$  y  $EPECM$  para expresarlos en las mismas unidades que el proceso observado.

Generalmente no dispondremos de suficientes observaciones para definir un conjunto de modelización y un conjunto de validación de tamaño adecuado para que cada uno cumpla con su cometido. En ese caso podemos utilizar las herramientas de validación cruzada, que son un caso particular del método Jackknife, y que vimos en la Sección 2.7.4 para el caso espacial. La validación cruzada es una técnica que permite evaluar la capacidad predictiva del modelo seleccionado calculando, para cada una de las observaciones  $Z(s_i, t_i)$ , los valores predichos  $\hat{Z}_{-i}(s_i, t_i)$  y los errores cuadráticos asociados a dichas predicciones  $\sigma_{-i}^2(s_i, t_i)$ , obtenidos a partir del conjunto de todas las observaciones restantes  $\{(s_1, t_1) \dots, (s_{i-1}, t_{i-1}), (s_{i+1}, t_{i+1}), \dots, (s_n, t_n)\}$ . A continuación se citan algunos de los diferentes estadísticos de resumen que permiten comparar los diferentes modelos considerados.

- Error de la predicción estandarizado medio, que sirve para contrastar la insesgadez de la predicción y debe valer aproximadamente 0.

$$VC1 = \frac{1}{k} \sum_1^k (\hat{Z}(s_i, t_i) - Z(s_i, t_i)) / \sigma(s_i, t_i).$$

- Raíz del error de la predicción estandarizado cuadrático medio, que depende de la precisión de la desviación típica del error de predicción y debe valer aproximadamente 1.

$$VC2 = \sqrt{\frac{1}{k} \sum_1^k ((\hat{Z}(\mathbf{s}_i, t_i) - Z(\mathbf{s}_i, t_i)) / \sigma(\mathbf{s}_i, t_i))^2}.$$

- Raíz del error de la predicción cuadrático medio, que es una medida global de la bondad de la predicción.

$$VC3 = \sqrt{\frac{1}{k} \sum_1^k (\hat{Z}(\mathbf{s}_i, t_i) - Z(\mathbf{s}_i, t_i))^2}.$$

Aunque lo deseable es que todos los estadísticos anteriores satisfagan las propiedades óptimas, en la práctica una mejoría en uno de ellos puede empeorar otro, por lo que es recomendable hacer un análisis integral de todos estos estadísticos. Suele ser muy útil observar el valor que cada localización muestreada aporta al sumatorio, lo que permite identificar valores atípicos, sospechosos o anomalías de otra naturaleza. Su eliminación y reajuste del modelo considerado puede producir mejorías significativas.

En el caso en que se haya asumido que el proceso objeto de estudio es Gaussiano, podemos comparar los modelos ajustados utilizando el criterio de información de Akaike (AIC) (Akaike, 1973, 1974) o el criterio de información bayesiano (BIC) (Schwarz, 1978). No obstante, hay que indicar que en la actualidad se conoce poco sobre sus propiedades en el contexto espacio-temporal.

El criterio de información de Akaike es un valor de compromiso entre la complejidad del modelo estimado y la bondad del ajuste conseguido. Se define como

$$AIC = 2q - 2L(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{Z}), \quad (3.38)$$

siendo  $\mathbf{Z}$  el vector de datos observado,  $q$  el número total de parámetros del modelo y  $L$  el logaritmo de la función de verosimilitud. Buscaremos modelos que tengan valores de (3.38) pequeños.

El criterio de información bayesiano (también conocido como criterio de información de Schwarz) también permite la evaluación de la calidad predictiva de un modelo, y se define como

$$AIC = q \log(n) - 2L(\boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{Z}), \quad (3.39)$$

siendo  $n$  el número total de observaciones,  $q$  el número total de parámetros del modelo y  $L$  el logaritmo de la función de verosimilitud. Como antes, buscaremos modelos que tengan valores de (3.39) pequeños. Este criterio BIC penaliza el número de parámetros incluidos en el modelo final más que el AIC.

### 3.7. Predicción

Sea  $\{Z(\mathbf{s}, t), \mathbf{s} \in D, t \in T\}$  un proceso estocástico definido sobre una determinada región espacial  $D \subset \mathbb{R}^d$  que se analiza para un conjunto temporal  $T \subset \mathcal{T}$ , siendo  $\mathcal{T} = \mathbb{R}$  o  $\mathbb{Z}$ .

Supondremos a lo largo de toda esta sección que hemos observado el valor del proceso sobre un conjunto de  $n$  localizaciones espacio-temporales  $\{(s_1, t_1), \dots, (s_n, t_n)\}$ . Generalmente nuestro objetivo será la predicción del valor del proceso sobre una nueva localización  $(s_0, t_0)$ ,  $Z(s_0, t_0)$ , donde  $s_0$  será una determinada localización espacial en  $D$  y  $t_0$  será un tiempo de interés en  $T$ . Supondremos que el proceso  $Z(\cdot, \cdot)$  satisface la condición de regularidad,  $Var(Z(s, t)) < \infty$ ,  $\forall (s, t) \in D \times T$ , con lo que tanto su media  $\mu(s_1, t_1) = E(Z(s_1, t_1))$  como su función de covarianza  $Cov(Z(s_1, t_1), Z(s_2, t_2))$  existen, para cualesquiera  $(s_1, t_1), (s_2, t_2) \in D \times T$ . Supondremos que el proceso  $Z(s, t)$  admite la descomposición (3.24), es decir, puede expresarse como suma de una tendencia  $\mu(s, t)$  que expresa la media del proceso y un proceso estocástico  $\delta(s, t)$  de media cero que supondremos estacionario y que captura la variabilidad espacio-temporal respecto a dicha media.

En esta sección vamos a generalizar los procedimientos de kriging espacial que vimos en la Sección 2.6, considerando aquel predictor  $\hat{Z}(s_0, t_0)$  definido como combinación lineal de las observaciones de forma que sea insesgado y minimice el error cuadrático medio de la predicción. La forma de este predictor será, por tanto,

$$\hat{Z}(s_0, t_0) = \mu(s_0, t_0) + \sum_{i=1}^n \lambda_i (Z(s_i, t_i) - \mu(s_i, t_i)), \quad (3.40)$$

donde los pesos  $\lambda_i$  asociados a cada una de las observaciones  $Z(s_i, t_i)$  del proceso quedarán determinados por las condiciones de insesgaredad y optimalidad.

Para que el predictor (3.40) sea insesgado, deberá cumplir

$$E[Z(s_0, t_0) - \hat{Z}(s_0, t_0)] = 0. \quad (3.41)$$

En este caso, el error cuadrático medio  $\sigma_K^2(s_0, t_0)$  asociado a la predicción vendrá dado por

$$\sigma_K^2(s_0, t_0) = Var[Z(s_0, t_0) - \hat{Z}(s_0, t_0)] = E(Z(s_0, t_0) - \hat{Z}(s_0, t_0))^2. \quad (3.42)$$

Como vimos para el caso espacial, existen diferentes modalidades de kriging dependiendo de las hipótesis que adoptemos sobre la media  $\mu(s, t)$  del proceso. Así, mientras que en el kriging simple supondremos que la media del proceso es conocida, en el kriging ordinario supondremos que la media es desconocida pero constante, y en el kriging universal supondremos que la media es desconocida pero puede expresarse como función lineal de un conjunto de variables que dependen de la localización espacio-temporal. Como ya hemos comentado, la mayor parte de las ecuaciones expuestas a continuación no son más que simples generalizaciones de las vistas en la Sección 2.6 en las que incorporamos la componente temporal del proceso.

### 3.7.1. Kriging simple

Supongamos que la media  $\mu(\mathbf{s}, t)$  del proceso es conocida. En este caso, el estimador (3.40) es siempre insesgado, ya que

$$\begin{aligned} E[\hat{Z}(\mathbf{s}_0, t_0)] &= E\left[\mu(\mathbf{s}_0, t_0) + \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i, t_i) - \mu(\mathbf{s}_i, t_i))\right] \\ &= \mu(\mathbf{s}_0, t_0) + \sum_{i=1}^n \lambda_i (E[Z(\mathbf{s}_i, t_i)] - \mu(\mathbf{s}_i, t_i)) = \mu(\mathbf{s}_0, t_0) = E[Z(\mathbf{s}_0, t_0)]. \end{aligned}$$

Tenemos en este caso que

$$\begin{aligned} (Z(\mathbf{s}_0, t_0) - \hat{Z}(\mathbf{s}_0, t_0))^2 &= \left(Z(\mathbf{s}_0, t_0) - \mu(\mathbf{s}_0, t_0) - \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i, t_i) - \mu(\mathbf{s}_i, t_i))\right)^2 \\ &= (Z(\mathbf{s}_0, t_0) - \mu(\mathbf{s}_0, t_0))^2 + \left(\sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i, t_i) - \mu(\mathbf{s}_i, t_i))\right)^2 \\ &\quad - 2(Z(\mathbf{s}_0, t_0) - \mu(\mathbf{s}_0, t_0)) \left(\sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i, t_i) - \mu(\mathbf{s}_i, t_i))\right) \\ &= (Z(\mathbf{s}_0, t_0) - \mu(\mathbf{s}_0, t_0))^2 \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (Z(\mathbf{s}_i, t_i) - \mu(\mathbf{s}_i, t_i))(Z(\mathbf{s}_j, t_j) - \mu(\mathbf{s}_j, t_j)) \\ &\quad - 2 \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_0, t_0) - \mu(\mathbf{s}_0, t_0))(Z(\mathbf{s}_i, t_i) - \mu(\mathbf{s}_i, t_i)). \end{aligned}$$

Luego tomando esperanzas, el error cuadrático medio (3.42) a minimizar quedaría

$$C(\mathbf{0}, 0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) - 2 \sum_{i=1}^n \lambda_i C(\mathbf{s}_0 - \mathbf{s}_i, t_0 - t_i).$$

Derivando respecto de  $\lambda_i$  e igualando a cero, tenemos un sistema de ecuaciones cuya solución ofrece los pesos a considerar en el estimador del kriging simple

$$\lambda' = \mathbf{c}\Sigma^{-1},$$

siendo  $\Sigma$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)$  y  $\mathbf{c}$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $C(\mathbf{s}_0 - \mathbf{s}_i, t_0 - t_i)$ . Por tanto, el predictor de  $Z(\mathbf{s}_0, t_0)$  lineal óptimo (en el sentido de que minimiza el error de predicción cuadrático medio) sería

$$\hat{Z}(\mathbf{s}_0, t_0) = \mu(\mathbf{s}_0, t_0) + \mathbf{c}'\Sigma^{-1}(\mathbf{Z} - \boldsymbol{\mu}), \quad (3.43)$$

con  $\mathbf{Z} = (Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n))'$  y  $\boldsymbol{\mu} = (\mu(\mathbf{s}_1, t_1), \dots, \mu(\mathbf{s}_n, t_n))'$ . Su error cuadrático medio vendrá dado por

$$\sigma_{SK}^2(\mathbf{s}_0, t_0) = \mathbf{c}'\Sigma^{-1}\mathbf{c}. \quad (3.44)$$

### 3.7.2. Kriging ordinario

Supongamos que la media del proceso  $\mu(s, t) = \mu$  es una constante desconocida. En este caso, el proceso  $Z(s, t)$  será estacionario con función de covarianza  $C(s_1 - s_2, t_1 - t_2) = Cov(Z(s_1, t_1), Z(s_2, t_2)) = Cov(\delta(s_1, t_1), \delta(s_2, t_2)) = C_\delta(s_1 - s_2, t_1 - t_2)$ .

En este caso, la condición (3.41) de insesgadez del predictor impone que los pesos  $\lambda_i$  deben sumar uno ( $\sum_{i=1}^n \lambda_i = 1$ ), y el predictor asociado a este kriging vendrá dado por

$$\hat{Z}(s_0, t_0) = \sum_{i=1}^n \lambda_i Z(s_i, t_i). \quad (3.45)$$

Nuestro objetivo será calcular aquellos pesos  $\lambda_i$  que minimizan (3.42) bajo la restricción dada por la condición de insesgadez. Aplicando el mismo procedimiento visto en la Sección 2.6.1 para el caso espacial, llegamos a que el predictor buscado se obtiene considerando

$$\lambda' = \left( c + 1 \frac{(1 - 1' \Sigma^{-1} c)}{1' \Sigma^{-1} 1} \right)' \sigma^{-1}, \quad (3.46)$$

siendo  $\Sigma$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $C(s_i - s_j, t_i - t_j)$  y  $c$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $C(s_0 - s_i, t_0 - t_i)$ . La varianza del error de predicción vendrá dada en este caso por

$$\sigma_{KS}^2(s_0, t_0) = C(0, 0) - \lambda' c + \frac{(1 - 1' \Sigma^{-1} c)}{1' \Sigma^{-1} 1}. \quad (3.47)$$

Como en el caso espacial, todas estas expresiones pueden escribirse en términos del variograma del proceso si este es además intrínsecamente estacionario. Sea  $2\gamma(h, u)$  el variograma del proceso, los pesos  $\lambda_i$  a considerar vendrían dados por

$$\lambda' = \left( \gamma + 1 \frac{(1 - 1' \Gamma^{-1} \gamma)}{1' \Gamma^{-1} 1} \right)' \Gamma^{-1}, \quad (3.48)$$

siendo  $\Gamma$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $\gamma(s_i - s_j, t_i - t_j)$  y  $\gamma$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $\gamma(s_0 - s_i, t_0 - t_i)$ . Observemos que si el objetivo es únicamente la predicción del proceso en el punto  $(s_0, t_0)$ , mediante el procedimiento descrito anteriormente no será necesario estimar la media del proceso. En este caso, el error cuadrático medio de la predicción que hemos minimizado vendría dado por la siguiente expresión

$$\sigma_{KS}^2(s_0, t_0) = \gamma' \Gamma^{-1} \gamma - \frac{(1' \Gamma^{-1} \gamma - 1)^2}{1' \Gamma^{-1} 1}. \quad (3.49)$$

### 3.7.3. Kriging universal

En este caso vamos a suponer que la media del proceso  $\mu(s, t)$ , aunque desconocida, es una combinación lineal de funciones conocidas o covariables ligadas a las localizaciones espacio-temporales. Esto es

$$\mu(s, t) = \sum_{i=0}^p \beta_i x_i(s, t) = x(s, t)' \beta,$$

donde  $\mathbf{x}(\mathbf{s}, t)$  es el vector formado por los valores de las  $p + 1$  variables explicativas consideradas en la localización espacio-temporal  $(\mathbf{s}, t)$ , y  $\boldsymbol{\beta}$  es un vector de  $p + 1$  parámetros desconocidos. Denotaremos por  $X$  la matriz  $n \times (p + 1)$  cuyo elemento  $(i, j)$ -ésimo es  $x_{j-1}(\mathbf{s}_i, t_i)$ .

Como en los casos anteriores, queremos predecir el valor de  $Z(\mathbf{s}_0, t_0)$  a partir del conjunto de observaciones  $\mathbf{Z}$  utilizando el predictor lineal (3.40), escogiendo los pesos  $\lambda_i$  de forma que sea insesgado y que minimice el error cuadrático medio asociado. La condición de insesgadura (3.41) implica en este caso que  $\boldsymbol{\lambda}'X = \mathbf{x}(\mathbf{s}_0, t_0)'$ . El predictor asociado a este tipo de kriging viene dado por

$$\hat{Z}(\mathbf{s}_0, t_0) = \boldsymbol{\lambda}'\mathbf{Z},$$

con  $\boldsymbol{\lambda}' = (\mathbf{c} + X(X'\boldsymbol{\Sigma}^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_0, t_0) - X'\boldsymbol{\Sigma}^{-1}\mathbf{c}))'\boldsymbol{\Sigma}^{-1}$ , siendo  $\boldsymbol{\Sigma}$  la matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)$  y  $\mathbf{c}$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $C(\mathbf{s}_0 - \mathbf{s}_i, t_0 - t_i)$ . El error cuadrático medio asociado a este predictor viene dado por

$$\sigma_{UK}^2(\mathbf{s}_0, t_0) = C(\mathbf{0}, 0) - \mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c} + (\mathbf{x}(\mathbf{s}_0, t_0) - X'\boldsymbol{\Sigma}^{-1}\mathbf{c})'(X'\boldsymbol{\Sigma}^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_0, t_0) - X'\boldsymbol{\Sigma}^{-1}\mathbf{c}).$$

Como en el caso espacial, todas estas expresiones pueden escribirse en términos del variograma del proceso si este es además intrínsecamente estacionario. Sea  $2\gamma(\mathbf{h}, u)$  el variograma del proceso, los pesos  $\lambda_i$  a considerar vendrían dados por

$$\boldsymbol{\lambda}' = (\boldsymbol{\gamma} + X(X'\boldsymbol{\Gamma}^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_0, t_0) - X'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}))'\boldsymbol{\Gamma}^{-1},$$

siendo  $\boldsymbol{\Gamma}$  una matriz  $n \times n$  cuyo elemento  $(i, j)$ -ésimo es  $\gamma(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)$  y  $\boldsymbol{\gamma}$  el vector de dimensión  $n$  cuyo elemento  $i$ -ésimo es  $\gamma(\mathbf{s}_0 - \mathbf{s}_i, t_0 - t_i)$ . El error cuadrático de la predicción sería, en este caso

$$\sigma_{UK}^2(\mathbf{s}_0, t_0) = \boldsymbol{\gamma}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma} - (\mathbf{x}(\mathbf{s}_0, t_0) - X'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})'(X'\boldsymbol{\Gamma}^{-1}X)^{-1}(\mathbf{x}(\mathbf{s}_0, t_0) - X'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}).$$

Respecto a la estimación óptima de los parámetros de la media  $\boldsymbol{\beta}$ , como los datos  $\mathbf{Z}$  satisfacen un modelo lineal general con  $E(\mathbf{Z}) = X\boldsymbol{\beta}$  y  $Var(\mathbf{Z}) = \boldsymbol{\Sigma}$ , puede obtenerse por mínimos cuadrados generalizados como

$$\hat{\boldsymbol{\beta}}_{gls} = (X'\boldsymbol{\Sigma}^{-1}X)^{-1}X'\boldsymbol{\Sigma}^{-1}\mathbf{Z},$$

con  $Var(\hat{\boldsymbol{\beta}}_{gls}) = (X'\boldsymbol{\Sigma}^{-1}X)^{-1}$ . A partir de estas expresiones, podemos obtener intervalos de confianza para estos parámetros o combinaciones lineales de los mismos. Se deduce que aun cuando para estimar los parámetros de la media de forma óptima es necesario conocer las covarianzas involucradas en  $\boldsymbol{\Sigma}$ , para la predicción óptima de  $Z(\mathbf{s}_0, t_0)$  únicamente hemos de conocer los variogramas implicados en  $\boldsymbol{\Gamma}$ .

### 3.8. Análisis en presencia de no estacionariedad

Como hemos visto anteriormente, la mayoría de los procedimientos descritos se basan en la hipótesis de que el proceso espacio-temporal analizado es estacionario. No obstante, esta

hipótesis es difícil de asumir en procesos heterogéneos observados sobre grandes superficies espaciales o con una evolución temporal considerable.

En caso de que no se admita la hipótesis de estacionariedad, generalmente este hecho se restringe a la media del proceso, que en lugar de considerarse constante se hace depender de la localización espacio-temporal. Para trabajar con estos procesos, se aplica a su tendencia alguno de los modelos espacio-temporales que hemos visto en la Sección 3.5, con lo que obtenemos unos residuos que suelen considerarse ya estacionarios, y a los que se les aplican las técnicas generales que hemos visto en este capítulo. No obstante, la no estacionariedad puede también deberse a la componente que modeliza la variación espacio-temporal del proceso respecto a su media.

Incluso en el caso de que se admita la estacionariedad del proceso, existen otras hipótesis del mismo que simplificarán su análisis, como son la separabilidad o la simetría completa. En la Sección 3.2 hemos definido cada una de ellas y hemos introducido las pruebas existentes para su adecuación al proceso espacio-temporal de interés.



## Capítulo 4

# Modelos de covarianza espacio-temporales

### 4.1. Introducción

En el capítulo anterior hemos estudiado las principales propiedades de los procesos espacio-temporales. Entre las características de este tipo de procesos, hemos introducido la función de covarianza espacio-temporal y el semivariograma espacio-temporal, que son las dos funciones que generalmente se utilizan en la modelización de su estructura de dependencia espacio-temporal. Mientras que en los procesos espaciales generalmente se utilizaba el semivariograma en el estudio de la dependencia espacial, para los procesos espacio-temporales suele ser la función de covarianza la herramienta seleccionada. En este capítulo vamos a estudiar los principales modelos de covarianza espacio-temporales descritos en la literatura para el estudio de fenómenos espacio-temporales de interés.

A lo largo de todo el capítulo consideremos  $\{Z(\mathbf{s}, t), \mathbf{s} \in D, t \in T\}$  un proceso estocástico espacio-temporal, cuyas coordenadas espaciales tomarán valores en  $D \subseteq \mathbb{R}^d$  y cuyas coordenadas temporales toman valores en  $T \subseteq \mathcal{T}$ , con  $\mathcal{T} \subseteq \mathbb{Z}$  o  $\mathbb{R}$ . Veremos a continuación de forma breve algunas definiciones y propiedades de interés que han sido analizadas con detalle en el Capítulo 3.

Si un proceso  $Z(\mathbf{s}, t)$  cumple la propiedad de regularidad (es decir,  $Var(Z(\mathbf{s}, t)) < \infty, \forall \mathbf{s} \in D, \forall t \in T$ ), entonces podíamos definir su función media como  $\mu(\mathbf{s}, t) \equiv E(Z(\mathbf{s}, t))$ , y su función de covarianza como  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2))$ , para cualquier par de localizaciones  $(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2) \in D \times T$ . También podíamos en ese caso definir su variograma asociado como  $2\gamma((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = Var(Z(\mathbf{s}_1, t_1) - Z(\mathbf{s}_2, t_2))$ . Además, siempre que se puedan definir ambas funciones, se relacionaban mediante la expresión (3.5).

Además, si el proceso espacio-temporal  $Z(\mathbf{s}, t)$  era estacionario de segundo orden (o simplemente estacionario), tanto su media  $\mu(\mathbf{s}, t)$  como su función de covarianza  $C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  no dependían de las localizaciones  $(\mathbf{s}_1, t_1)$  y  $(\mathbf{s}_2, t_2)$ , sino únicamente de la distancia espacial  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  y temporal  $u = t_1 - t_2$  existentes entre ellas,  $\forall \mathbf{s}_1, \mathbf{s}_2 \in D, \forall t_1, t_2 \in T$ . En este caso, decíamos que la función de covarianza del proceso era estacionaria,

y podía expresarse como  $C((s_1, t_1), (s_2, t_2)) = C(h, u)$ .

También se ha definido en el capítulo anterior la estacionariedad intrínseca del proceso, en cuyo caso tanto la media  $\mu(s, t)$  como el variograma  $2\gamma((s_1, t_1), (s_2, t_2))$  del proceso no dependían de las localizaciones  $(s_1, t_1)$  y  $(s_2, t_2)$ , sino únicamente de la distancia espacial  $h = s_1 - s_2$  y temporal  $u = t_1 - t_2$  existentes entre ellas,  $\forall s_1, s_2 \in D, \forall t_1, t_2 \in T$ . En ese caso decíamos que el variograma era intrínsecamente estacionario, y podía expresarse como  $2\gamma((s_1, t_1), (s_2, t_2)) = 2\gamma(h, u)$ .

Una función de valores reales  $C$  definida sobre  $\mathfrak{R}^d \times T$  es una función de covarianza asociada a un proceso espacio-temporal estacionario si y sólo si es definida positiva, esto es, si y sólo si

$$\sum_{i=1}^k \sum_{j=1}^k a_i a_j C(s_i - s_j, t_i - t_j) \geq 0, \quad (4.1)$$

para cualesquiera  $(s_i, t_i) \in \mathfrak{R}^d \times T$  y  $a_i \in \mathfrak{R}$ , para  $i = 1, \dots, k$ . También se dice que  $C$  es una función de covarianza espacio-temporal permisible. El teorema de Bochner establece que para las funciones continuas, esta condición es equivalente a que sea la transformada de Fourier de una medida finita no negativa. Además, el cumplimiento de esta condición nos asegurará que el error cuadrático medio asociado a la predicción mediante kriging sea no negativo, como hemos visto en el Capítulo 3.

En la práctica, observaremos generalmente el valor del proceso de interés sobre  $n$  localizaciones espacio-temporales, con lo que obtendremos un conjunto de valores  $\{Z(s_1, t_1), \dots, Z(s_n, t_n)\}$ , a partir de las cuales estimaremos los valores de cualquiera de las dos funciones anteriores utilizadas en la descripción de la variabilidad del proceso. El problema, como ya hemos visto en el capítulo anterior, es que ninguna de las estimaciones obtenidas verificará en general la propiedad (4.1). Por esta razón, en lugar de utilizar directamente la estimación de la función de covarianza o del variograma empírico, ajustaremos a estos valores experimentales un modelo de función de covarianza o de variograma paramétrico que sepamos a priori que cumple esta propiedad (4.1).

Todo esto convierte la obtención y estudio de modelos de funciones de covarianza o semivariogramas válidos sobre  $\mathfrak{R}^d \times T$  en un problema de especial interés. El objetivo es poner a disposición del investigador una batería de modelos lo más amplia y flexible posible que permitan la modelización de la mayor parte de procesos físicos de manera eficiente y operativa. Prueba de ello es la creciente y abundante literatura generada en el último decenio sobre este aspecto.

En las primeras aproximaciones a la modelización de la función de covarianza espacio-temporal se utilizaron meras generalizaciones de los modelos estacionarios disponibles para el análisis espacial. Así, durante los primeros estudios era habitual modelizar la covarianza espacio-temporal mediante modelos métricos, que eran modelos geométricamente anisotrópicos y que podemos encontrar, por ejemplo, en Dimitrakopoulos y Luo (1994). Otra opción consistía en separar la dependencia espacial y temporal del proceso, obteniendo modelos separables, ya sea mediante suma (modelo suma) como mediante producto (modelo producto) de una covarianza espacial y una covarianza temporal. Con el fin de incluir la interacción espacio-temporal, presente en la mayor parte de procesos físicos, se han ido proponiendo en los últimos años diferentes modelos estacionarios

no separables. Entre ellos podemos citar los modelos obtenidos por Jones y Zhang (1997), Cressie y Huang (1999), Brown et al. (2000), De Cesare et al. (2001a,b, 2002), De Iaco et al. (2001, 2002a,b, 2003), Myers et al. (2002), Gneiting (2002), Ma (2002, 2003a,b, 2005a,b,c), Stein (2005b), Fernández-Casal et al. (2003) o Kolovos et al. (2004), entre otros. También podemos encontrar diferentes modelos espacio-temporales estacionarios no separables que no son espacialmente anisotrópicos o temporalmente asimétricos, como los descritos en Fernández-Casal et al. (2003), Stein (2005b), Mohammadian y Kent (2006), Porcu et al. (2006) o Mateu et al. (2007a), entre otros. Podemos citar además algunas aproximaciones al problema de la modelización de funciones de covarianza no estacionarias, como las realizadas en Ma (2002, 2003c), Fuentes et al. (2005), Stein (2005a), Chen et al. (2006), Porcu et al. (2007) o Porcu y Mateu (2007).

En la Sección 4.2 analizaremos el modelo métrico, presente en las primeras aproximaciones al problema de la modelización espacio-temporal aunque con muchas limitaciones prácticas en cuanto a su aplicabilidad sobre numerosos procesos naturales. La Sección 4.3 introduce el modelo suma, que es un modelo de covarianza espacio-temporal que considera la covarianza total como suma de una covarianza meramente espacial y una covarianza meramente temporal, y que como veremos presentará problemas de validez bajo determinadas condiciones. A continuación, la Sección 4.4 se dedica al estudio del modelo producto, un modelo separable de gran sencillez con numerosas ventajas computacionales, pero que no incluye la posible interacción espacio-temporal en el modelo final. La Sección 4.5 describe el procedimiento propuesto por Cressie y Huang (1999), que permite la construcción de funciones de covarianza estacionarias continuas e integrables que en general serán no separables, permitiendo por tanto la inclusión en el modelo de la interacción entre las componentes espacial y temporal, necesaria para el ajuste de la mayoría de problemas prácticos. La Sección 4.6 presenta el modelo suma-producto definido por De Cesare et al. (2001a,b, 2002) y De Iaco et al. (2001, 2002a), un modelo en general no separable construido a partir una sencilla combinación lineal de funciones de covarianza espaciales y temporales. La Sección 4.7 presenta el procedimiento introducido por Gneiting (2002), con el que se construyen funciones de covarianza estacionarias bajo un contexto más general, y que no se limitará a las funciones de covarianza integrables. A continuación, la Sección 4.8 presenta los modelos producto integrado y el modelo suma-producto integrado introducidos por De Iaco et al. (2002b) y generalizados por Ma (2003a), mientras que la Sección 4.9 abarca la generación de modelos espacio-temporales estacionarios mediante mixturas de funciones de covarianza espaciales y temporales presentada en Ma (2002, 2003b). La Sección 4.10 introduce el procedimiento de generación de modelos estacionarios mediante combinaciones lineales de covarianza y variogramas espacio-temporales presentado en Ma (2005a,b). En la Sección 4.11 se presenta el modelo estacionario generado mediante el procedimiento introducido por Stein (2005a) y que funciona en el dominio espectral. En la Sección 4.12 presentamos un nuevo procedimiento que permite la generación de modelos de funciones de covarianza espacio-temporal estacionaria mediante combinaciones lineales de covarianzas separables obtenidas como producto de covarianzas espaciales y temporales. Como veremos en esta sección, los coeficientes que utilizaremos en la definición de las combinaciones lineales generadoras de estos modelos no tendrán que ser necesariamente positivos sino que podrán tomar valores dentro de un determinado rango de variación,

lo que permitirá definir covarianzas con valores negativos o con valores que oscilen entre negativos y positivos, con importantes ventajas prácticas en la modelización de numerosos procesos. Hay que hacer notar que las cotas superiores e inferiores sobre los coeficientes de las combinaciones lineales que presentamos, y que aseguran la validez de nuestro modelo, suponen un avance de interés porque pueden dar lugar a covarianzas negativas u oscilantes, con las repercusiones prácticas que ya hemos comentado. A continuación, la Sección 4.13 introduce algunos modelos estacionarios que no satisfacen la propiedad de simetría completa, muy útiles para la modelización de determinados procesos físicos. Por último, la Sección 4.14 describe brevemente algunas de las principales aproximaciones a la modelización de procesos espacio-temporales no estacionarios.

## 4.2. Modelo métrico

Este modelo de funciones de covarianza espacio-temporal estacionarias consiste en adoptar una anisotropía geométrica sobre el espacio  $\mathfrak{R}^d \times \mathfrak{R}$ , esto es, definir una métrica en el espacio-tiempo que permita utilizar directamente los modelos isotrópicos utilizados en el análisis espacial, considerando

$$C(\mathbf{h}, u) = C^0 \left( (a^2 \|\mathbf{h}\|^2 + b^2 |u|^2)^{1/2} \right), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R} \quad (4.2)$$

donde  $a, b \in \mathfrak{R}$  son las constantes que definen la métrica espacio-temporal. Podemos encontrar aplicaciones de este modelo en Armstrong et al. (1993) o en Dimitrakopoulos y Luo (1994), entre otros. Este modelo supone una misma estructura de dependencia en el espacio y en el tiempo, y sólo se permiten cambios en el rango de ambas funciones de covarianza. Es por ello que, a pesar de que genera un modelo muy sencillo conceptualmente y eficaz desde el punto de vista computacional, tiene muchas limitaciones de aplicabilidad sobre procesos reales.

**Ejemplo 4.1** Sea  $C^0(\cdot)$  una función de covarianza perteneciente al modelo exponencial (2.25), el modelo métrico asociado vendría dado por

$$C(\mathbf{h}, u) = \sigma^2 \exp \left( -(a^2 \|\mathbf{h}\|^2 + b^2 |u|^2)^{1/2} / c \right), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R} \quad (4.3)$$

La Figura 4.1 muestra la función de covarianza asociada a este modelo (4.3) para dos combinaciones diferentes de parámetros. El primer ejemplo representa la función de covarianza espacio-temporal obtenida con  $\sigma^2 = 5$ ,  $a = 1$ ,  $b = 1$  y  $c = 2$ , mientras que el segundo ejemplo corresponde a la función de covarianza espacio-temporal obtenida con  $\sigma^2 = 5$ ,  $a = 1$ ,  $b = 2$  y  $c = 2$ . Se han incluido además mediante líneas discontinuas las representaciones de las funciones de covarianza marginales asociadas a cada función, utilizando el color verde para la covarianza marginal temporal  $C(\mathbf{0}, u)$  y el color rojo para la covarianza marginal espacial  $C(\mathbf{h}, 0)$ .

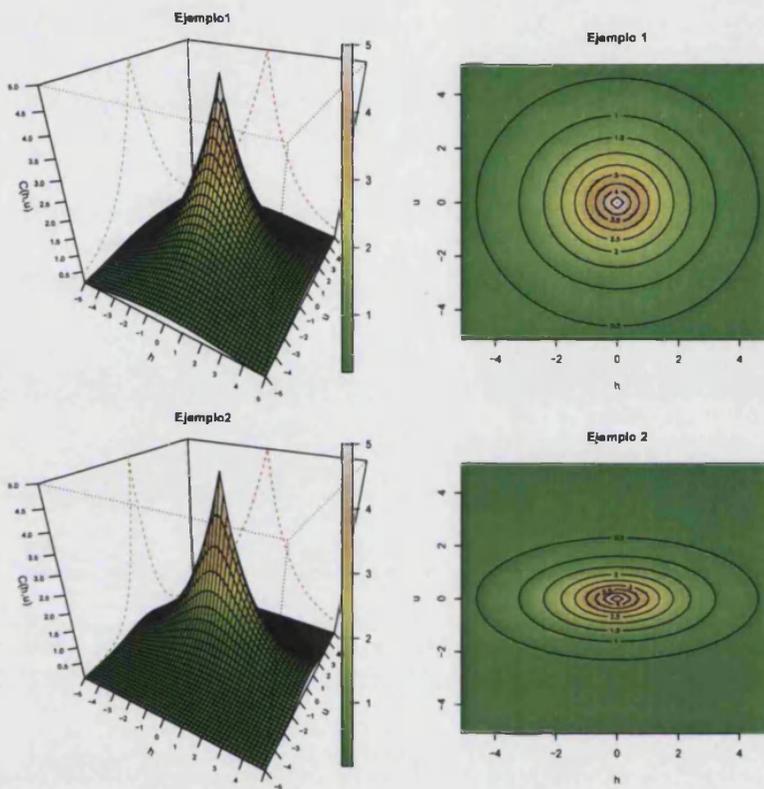


Figura 4.1: Diferentes representaciones del modelo métrico (4.3).

### 4.3. Modelo suma

Este modelo estacionario de covarianzas fue introducido por Rouhani y Hall (1989), y consiste en considerar la covarianza espacio-temporal del proceso como la suma de una covarianza meramente espacial y otra covarianza meramente temporal

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = C_e(\mathbf{s}_1, \mathbf{s}_2) + C_t(t_1, t_2),$$

siendo  $C_e$  una función de covarianza espacial definida sobre  $\mathbb{R}^d$  y  $C_t$  una función de covarianza temporal definida sobre  $\mathbb{R}$ . Si consideramos dos funciones de covarianzas  $C_e$  y  $C_t$  estacionarias, entonces el modelo suma viene dado por

$$C(\mathbf{h}, u) = C_e(\mathbf{h}) + C_t(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \quad (4.4)$$

El semivariograma asociado a este modelo (4.4) vendría dado por  $\gamma(\mathbf{h}, u) = \gamma_e(\mathbf{h}) + \gamma_t(u)$ , siendo  $\gamma_e$  un semivariograma espacial sobre  $\mathbb{R}^d$  y  $\gamma_t$  un semivariograma temporal sobre  $\mathbb{R}$ .

Este modelo ha sido muy utilizado en el análisis geoestadístico espacio-temporal debido fundamentalmente a su sencillez. El principal problema que presenta es que la suma de un modelo de covarianza espacial y otro temporal no es en general una función definida positiva, sino únicamente semidefinida positiva. Esto puede provocar que la matriz de

coeficientes de las ecuaciones de kriging no sea invertible para determinados conjuntos de localizaciones (Myers y Journel (1990), Rouhani y Myers (1990)).

**Ejemplo 4.2** *Un ejemplo de utilización de este modelo lo podemos encontrar en Cressie y Majure (1997), donde los autores estudian la evolución espacio-temporal del logaritmo de la concentración de nitratos en un río de Texas. Para ello modelizan la variación a pequeña escala con un variograma suma de dos variogramas estacionarios, uno espacial y otro temporal, ambos exponenciales, dado por*

$$C(\mathbf{h}, u) = \sigma_e^2 \exp(-\|\mathbf{h}\|/c_e) + \sigma_t^2 \exp(-|u|/c_t), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R} \quad (4.5)$$

La Figura 4.2 muestra dos casos particulares del modelo (4.5) para diferentes valores de los parámetros que lo definen. El primer ejemplo muestra la función de covarianza espacio-temporal obtenida para  $\sigma_e^2 = 2.5$ ,  $c_e = 2$ ,  $\sigma_t^2 = 2.5$  y  $c_t = 2$ , mientras que el segundo ejemplo es la covarianza espacio-temporal generada para  $\sigma_e^2 = 2.5$ ,  $c_e = 0.5$ ,  $\sigma_t^2 = 2.5$  y  $c_t = 2$ .

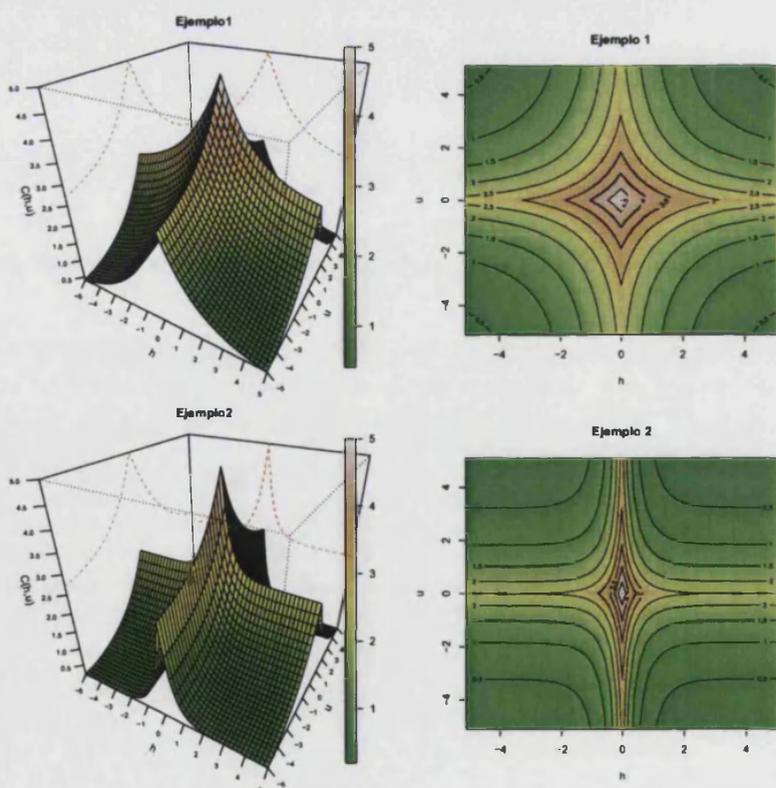


Figura 4.2: Diferentes representaciones del modelo suma (4.5).

#### 4.4. Modelo producto

Este modelo de funciones de covarianza espacio-temporal estacionarias ha sido uno de las más utilizados en la bibliografía de modelización espacio-temporal (Rodríguez-Iturbe y Mejía (1974) o De Cesare et al. (1997), entre otros). Se basa en considerar funciones de covarianza espacio-temporales separables, que como vimos en la Definición 3.6 venían dadas por el producto de una función de covarianza espacial  $C_e$  sobre  $\mathbb{R}^d$  y una función de covarianza estacionaria temporal  $C_t$  sobre  $\mathbb{R}$ , esto es,

$$C((s_1, t_1), (s_2, t_2)) = C_e(s_1, s_2)C_t(t_1, t_2).$$

Si consideramos dos funciones de covarianzas  $C_e$  y  $C_t$  estacionarias, entonces el modelo producto dará lugar a una función de covarianza estacionaria definida por

$$C(\mathbf{h}, u) = C_e(\mathbf{h})C_t(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \quad (4.6)$$

Aplicando la relación (3.6) entre el semivariograma y la función de covarianza, es sencillo comprobar que el semivariograma asociado a este modelo (4.6) viene dado por

$$\gamma(\mathbf{h}, u) = C_t(0)\gamma_e(\mathbf{h}) + C_e(\mathbf{0})\gamma_t(u) - \gamma_e(\mathbf{h})\gamma_t(u), \quad (4.7)$$

siendo  $C_e(\mathbf{0})$  la meseta de  $\gamma_e$  y  $C_t(0)$  la meseta de  $\gamma_t$ . Por tanto, la separabilidad de una función de covarianzas no implica que el variograma asociado sea separable. No obstante, mientras que la suma o el producto de un variograma espacial y otro temporal no es necesariamente un modelo válido de variograma, la combinación de ambas expresiones mediante el modelo anterior sí hereda esta propiedad. Para este modelo (4.7),  $\gamma(\mathbf{h}, 0) = C_t(0)\gamma_e(\mathbf{h})$  y  $\gamma(\mathbf{0}, u) = C_e(\mathbf{0})\gamma_t(u)$  son los semivariogramas marginales espacial y temporal, respectivamente.

Con este modelo de covarianza, (4.6), el proceso conjunto puede verse como el producto de dos procesos estacionarios independientes, uno que ocurre en el espacio y otro que ocurre en el tiempo. En la práctica no observaremos estos dos procesos separadamente, sino únicamente el proceso conjunto. En términos de este proceso conjunto, la condición de separabilidad (4.6) puede expresarse como

$$C(\mathbf{h}, u) = \frac{C(\mathbf{h}, 0)C(\mathbf{0}, u)}{C(\mathbf{0}, 0)}, \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

siendo  $C(\mathbf{0}, 0) = C_e(\mathbf{0})C_t(0) = \sigma^2 > 0$  la varianza a priori del proceso.

La principal ventaja de los modelos separables se produce desde el punto de vista computacional, ya que se reducen enormemente el número de parámetros involucrados a estimar durante la modelización. No obstante, presenta una importante limitación desde el punto de vista práctico, ya que no incluye la posible interacción entre las componentes espacial y temporal del proceso, presente en la mayoría de aplicaciones prácticas, suponiendo que la estructura espacial es la misma para todos los tiempos y que la estructural temporal es la misma en todas las localizaciones. Esta falta de interacción entre el espacio y el tiempo es lo que hace que este modelo sea poco atractivo desde el punto de vista práctico.

Mitchell et al. (2005) ofrece una caracterización alternativa de este modelo utilizando el producto de Kronecker. Sean  $\mathbf{U}$  y  $\mathbf{V}$  las matrices de varianzas-covarianzas de dimensión  $s$  y  $p$  para el espacio y el tiempo, respectivamente, y sea  $\mathbf{\Sigma}$  la matriz de varianzas-covarianzas de dimensión  $s \times p$  del proceso espacio-temporal. Diremos que la covarianza del proceso conjunto es separable si y sólo si

$$\mathbf{\Sigma} = \mathbf{U} \otimes \mathbf{V}, \quad (4.8)$$

que es una descomposición única salvo producto por una constante, ya que  $a\mathbf{U} \otimes (1/a)\mathbf{V} = \mathbf{U} \otimes \mathbf{V}$ , para cualquier  $a \neq 0$ . En (4.8),  $\otimes$  representa el producto de Kronecker de ambas matrices. Como ya hemos dicho, una de las principales ventajas de la separabilidad es el ahorro computacional que supone. Por ejemplo, en un problema con 50 localizaciones observadas durante 100 tiempos en el que queremos predecir el valor del proceso sobre una nueva localización no observada, la inversión de la matriz de varianzas-covarianzas  $\mathbf{\Sigma}$  necesaria para dicha predicción implica para un modelo no separable la inversión de una matriz  $5000 \times 5000$ , mientras que para el modelo separable por las propiedades del producto de Kronecker bastará con la inversión de una matriz  $50 \times 50$  y de una matriz  $100 \times 100$ , ya que  $\mathbf{\Sigma}^{-1} = (\mathbf{U} \otimes \mathbf{V})^{-1} = \mathbf{U}^{-1} \otimes \mathbf{V}^{-1}$ .

Debido a estas ventajas computacionales, existen en la literatura diferentes pruebas diagnósticas para contrastar la idoneidad de esta hipótesis en el análisis de un determinado proceso espacio-temporal de interés. Algunos de los test propuestos en la literatura son específicos para determinados modelos, como el propuesto por Brown et al. (2000) para modelos generados mediante determinadas funciones de suavizado (blur-generated models), y que contrastan la hipótesis de si los parámetros que marcan la separabilidad de estos modelos son o no cero utilizando máxima verosimilitud. Otros son procedimientos generales, como el de Mitchell et al. (2005, 2006), que adaptan un test de cocientes de verosimilitudes basado en medidas repetidas multivariantes al contexto espacio-temporal utilizando submuestreo. Otro procedimiento de contraste es el propuesto por Fuentes (2003, 2005a), que utiliza un procedimiento no paramétrico basado en la representación espectral del proceso espacio-temporal.

**Ejemplo 4.3** *Un sencillo ejemplo de modelo separable estacionario es el modelo doblemente exponencial, que viene dado por*

$$C(\mathbf{h}, u) = \sigma^2 \exp(-\|\mathbf{h}\|/a_e - |u|/a_t), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (4.9)$$

siendo  $a_e$  y  $a_t$  dos parámetros positivos de escala espacial y temporal, respectivamente, y  $\sigma^2$  la varianza a priori del proceso estocástico.

La Figura 4.3 muestra dos casos particulares de este modelo, el primer ejemplo obtenido con  $\sigma^2 = 10$ ,  $a_e = 2$  y  $a_t = 2$ , y el segundo generado con  $\sigma^2 = 10$ ,  $a_e = 1$  y  $a_t = 2$ . Observemos que si consideramos  $1/a_t = -\log(\rho)$ , esta reparametrización permite escribir (4.9) como

$$C(\mathbf{h}, u) = \sigma^2 \exp(-\|\mathbf{h}\|/a_e) \exp(|u| \log(\rho)) = \sigma^2 \rho^{|u|} \exp(-\|\mathbf{h}\|/a_e),$$

que es el producto de un modelo temporal de correlación de una serie temporal AR(1) y un modelo espacial exponencial.

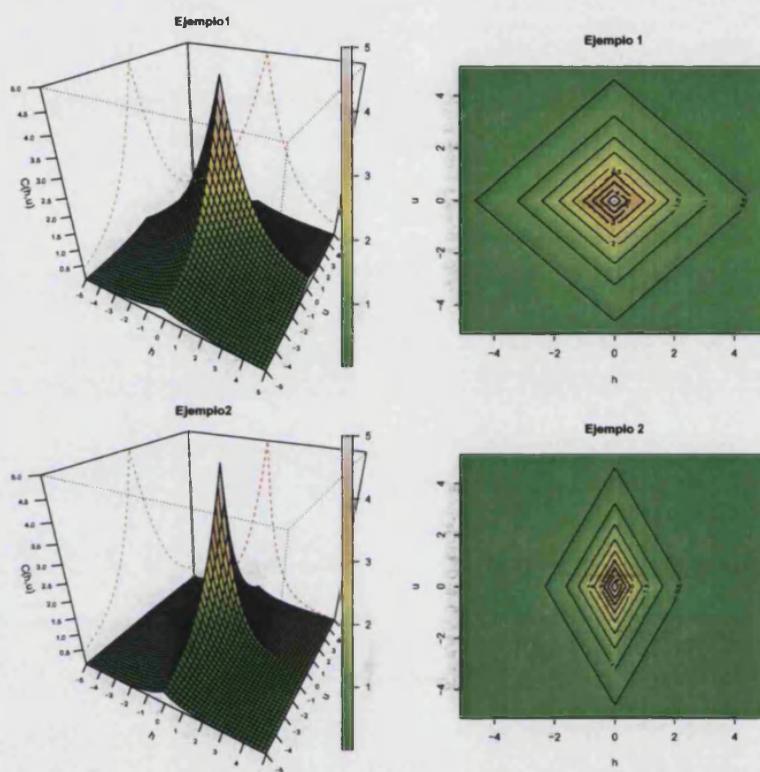


Figura 4.3: Diferentes representaciones del modelo producto (4.9).

## 4.5. Modelos propuestos por Cressie y Huang

En esta sección vamos a estudiar los modelos generados mediante el método descrito en Cressie y Huang (1999). Este procedimiento permite construir modelos paramétricos de funciones de covarianza espacio-temporal estacionarias no separables, y por tanto incluir la posible interacción espacio-temporal del proceso. Se basa fundamentalmente en representaciones espectrales de la estructura de covarianza espacio-temporal teórica, generalizando los resultados para procesos meramente espaciales de Matérn (1960). En esencia, Matérn construye un conjunto de familias paramétricas de modelos para la función de covarianza de procesos espaciales (que incluye la familia que lleva su propio nombre) mediante inversión directa de densidades espectrales. Cressie y Huang utilizan este mismo procedimiento, e incluso en ocasiones las mismas fórmulas, para construir familias paramétricas que modelan la función de covarianza de los procesos espacio-temporales. Aunque esta técnica es muy potente, está restringida a un conjunto de funciones de covarianza relativamente pequeño que son integrables y para las que tenemos una solución analítica de la integral de Fourier correspondiente sobre  $\mathcal{R}^d$ .

Sea  $C(\mathbf{h}, u)$  la función de covarianza espacio-temporal de un proceso estacionario. Si suponemos que  $C$  es continua y que su función de distribución espectral posee una densidad

espectral  $g(\omega, \tau) \geq 0$ , entonces por el teorema de Bochner

$$C(\mathbf{h}, u) = \int \int e^{i(\mathbf{h}'\omega + u\tau)} g(\omega, \tau) d\omega d\tau. \quad (4.10)$$

Si además  $C(\cdot, \cdot)$  es integrable, entonces podemos escribir  $g$  como una transformada de Fourier escalar en  $\tau$

$$g(\omega, \tau) = \frac{1}{(2\pi)^{d+1}} \int \int e^{-i(\mathbf{h}'\omega + u\tau)} C(\mathbf{h}, u) d\mathbf{h} du = \frac{1}{2\pi} \int e^{-iu\tau} h(\omega, u) du, \quad (4.11)$$

donde

$$h(\omega, u) = \frac{1}{(2\pi)^d} \int e^{-i\mathbf{h}'\omega} C(\mathbf{h}, u) d\mathbf{h} = \int e^{iu\tau} g(\omega, \tau) d\tau. \quad (4.12)$$

De (4.10) y (4.12) se deduce que podemos expresar la función de covarianza como

$$C(\mathbf{h}, u) = \int e^{i\mathbf{h}'\omega} h(\omega, u) d\omega. \quad (4.13)$$

Supongamos que  $h(\omega, u)$  puede descomponerse como un producto

$$h(\omega, u) = k(\omega)\rho(\omega, u), \quad (4.14)$$

donde  $k(\omega)$  es la densidad espectral de un proceso meramente espacial y  $\rho(\omega, \cdot)$  una función de autocorrelación temporal, que suponemos cumplen las condiciones siguientes:

(C1) Para cada  $\omega \in \mathfrak{R}^d$ ,  $\rho(\omega, \cdot)$  es una función de autocorrelación continua,  $\int \rho(\omega, u) du < \infty$  y  $k(\omega) > 0$ .

(C2)  $\int k(\omega) d\omega < \infty$ .

Observemos que esta descomposición (4.14) es siempre posible, pues basta con definir

$$\rho(\omega, u) = \frac{h(\omega, u)}{\int g(\omega, \tau) d\tau}, \quad k(\omega) = \int g(\omega, \tau) d\tau.$$

Bajo estas condiciones,  $C(\mathbf{h}, u)$  puede expresarse como

$$C(\mathbf{h}, u) = \int e^{i\mathbf{h}'\omega} k(\omega)\rho(\omega, u) d\omega \quad (4.15)$$

y es una función de covarianza válida del proceso espacio-temporal estacionario.

Este procedimiento permite obtener modelos de covarianza espacio-temporal válidos simplemente seleccionando dos funciones  $\rho(\omega, u)$  y  $k(\omega)$  satisfaciendo (C1) y (C2) y que permitan el desarrollo analítico de la integral (4.15), siendo esta última condición una restricción meramente práctica. Las funciones de covarianza construidas son en general no separables, aunque el modelo producto es un caso particular de la expresión (4.15) considerando  $\rho(\omega, u)$  como una función independiente de  $\omega$ . En su artículo, Cressie y Huang proponen siete casos particulares obtenidos con este procedimiento, aunque posteriormente Gneiting (2002) prueba que los ejemplos 5 y 6 no son modelos válidos al no cumplirse las condiciones necesarias. Vamos a describir brevemente los cinco modelos restantes.

**Ejemplo 4.4** Consideremos las funciones

$$\rho(\omega, u) = \exp\left(-\frac{\|\omega\|^2 u^2}{4}\right) \exp(-\delta u^2), \quad k(\omega) = \exp\left(-\frac{c_0 \|\omega\|^2}{4}\right),$$

con  $\delta > 0$  y  $c_0 > 0$ , que satisfacen las condiciones (C1) y (C2). Por (4.15), tenemos que

$$C(\mathbf{h}, u) = \int \exp(i\mathbf{h}^t \omega) \exp\left(-\frac{c_0 \|\omega\|^2}{4}\right) \exp\left(-\frac{\|\omega\|^2 u^2}{4}\right) \exp(-\delta u^2) d\omega,$$

es una función de covarianza espacio-temporal continua en  $\mathfrak{R}^d \times \mathfrak{R}$ . Operando en esta expresión y aplicando la transformada de Fourier adecuada, se obtiene que

$$C(\mathbf{h}, u) = (2\sqrt{\pi})^d \frac{1}{(u^2 + c_0)^{d/2}} \exp\left(-\frac{\|\mathbf{h}\|^2}{u^2 + c_0}\right) \exp(-\delta u^2).$$

Observemos que la condición  $\delta > 0$  es necesaria para asegurar que se cumple (C1) cuando  $\omega = 0$ . La función resultante tomando límites cuando  $\delta \rightarrow 0$  también será una función de covarianza espacio-temporal válida, por ser límite de una sucesión de funciones de covarianza espacio-temporales válidas, y vendrá expresada por

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a^2 u^2 + 1)^{d/2}} \exp\left(-\frac{b^2 \|\mathbf{h}\|^2}{a^2 u^2 + 1}\right), \quad (4.16)$$

que será una familia paramétrica de funciones de covarianza espacio-temporales no separables, con vector de parámetros  $\theta = (a, b, \sigma^2)'$ , siendo  $a \geq 0$  un parámetro de escala temporal,  $b \geq 0$  un parámetro de escala espacial, y  $\sigma^2 = C(\mathbf{0}, 0) > 0$ . De la relación existente entre una función de covarianza estacionaria y su semivariograma asociado dada por (3.6), se tiene que el semivariograma correspondiente a este modelo vendrá dado por

$$\gamma(\mathbf{h}, u) = \sigma^2 \left(1 - \frac{1}{(a^2 u^2 + 1)^{d/2}} \exp\left(-\frac{b^2 \|\mathbf{h}\|^2}{a^2 u^2 + 1}\right)\right).$$

**Ejemplo 4.5** Consideremos las funciones

$$\rho(\omega, u) = \exp\left(-\frac{\|\omega\|^2 |u|}{4}\right) \exp(-\delta u^2), \quad k(\omega) = \exp\left(-\frac{c_0 \|\omega\|^2}{4}\right),$$

con  $\delta > 0$  y  $c_0 > 0$ , que cumplen (C1) y (C2) y darán lugar por tanto al modelo de covarianza espacio-temporal continuo

$$C(\mathbf{h}, u) \propto \frac{1}{(|u| + c_0)^{d/2}} \exp\left(-\frac{\|\mathbf{h}\|^2}{|u| + c_0}\right) \exp(-\delta u^2).$$

En este caso, cuando  $\delta \rightarrow 0$  tenemos la familia de funciones de covarianza espacio-temporales

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|u| + 1)^{d/2}} \exp\left(-\frac{b^2 \|\mathbf{h}\|^2}{a|u| + 1}\right), \quad (4.17)$$

donde  $a \geq 0$  es un parámetro de escala temporal,  $b \geq 0$  es un parámetro de escala espacial y  $\sigma^2 = C(\mathbf{0}, 0) > 0$ .

**Ejemplo 4.6** En este caso consideramos

$$\rho(\omega, u) = \exp(-\|\omega\|u^2) \exp(-\delta u^2), \quad k(\omega) = \exp(-c_0\|\omega\|),$$

con  $\delta > 0$  y  $c_0 > 0$ , lo que da lugar al modelo de covarianza espacio-temporal continuo

$$C(\mathbf{h}, u) \propto \frac{1}{(u^2 + c_0)^d} \left(1 + \frac{\|\mathbf{h}\|^2}{(u^2 + c_0)^2}\right)^{-(d+1)/2} \exp(-\delta u^2).$$

En este caso cuando  $\delta \rightarrow 0$  tenemos la familia de funciones de covarianza espacio-temporales

$$C(\mathbf{h}, u) = \frac{\sigma^2(a^2u^2 + 1)}{((a^2u^2 + 1)^2 + b^2\|\mathbf{h}\|^2)^{(d+1)/2}}, \quad (4.18)$$

donde  $a \geq 0$  es un parámetro de escala temporal,  $b \geq 0$  es un parámetro de escala espacial y  $\sigma^2 = C(0, 0) > 0$ .

**Ejemplo 4.7** Para construir este modelo definimos

$$\rho(\omega, u) = \exp(-\|\omega\||u|) \exp(-\delta u^2), \quad k(\omega) = \exp(-c_0\|\omega\|),$$

con  $\delta > 0$  y  $c_0 > 0$ , lo que da lugar al modelo de covarianza espacio-temporal continuo

$$C(\mathbf{h}, u) \propto \frac{1}{(|u| + c_0)^d} \left(1 + \frac{\|\mathbf{h}\|^2}{(|u| + c_0)^2}\right)^{-(d+1)/2} \exp(-\delta u^2).$$

Si consideramos  $\delta \rightarrow 0$ , tenemos el modelo de covarianza espacio-temporal

$$C(\mathbf{h}, u) = \frac{\sigma^2(a|u| + 1)}{((a|u| + 1)^2 + b^2\|\mathbf{h}\|^2)^{(d+1)/2}}, \quad (4.19)$$

donde  $a \geq 0$  es un parámetro de escala temporal,  $b \geq 0$  es un parámetro de escala espacial y  $\sigma^2 = C(0, 0) > 0$ .

**Ejemplo 4.8** En este caso consideramos

$$\rho(\omega, u) = (u^2 + 1 + (u^2 + c)\|\omega\|^2)^{-(\nu+d)/2} (1 + c\|\omega\|^2)^{(\nu+d)/2}, \quad k(\omega) = (1 + c\|\omega\|^2)^{-(\nu+d)/2},$$

con  $\delta > 0$ ,  $c > 0$  y  $\nu > 0$ , que satisfacen las condiciones (C1) y (C2) y que por tanto darán lugar a una familia de funciones de covarianza espacio-temporal continuas sobre  $\mathbb{R}^d \times \mathbb{R}$  dada por

$$C(\mathbf{h}, u) \propto \begin{cases} \frac{1}{(u^2+1)^\nu(u^2+c)^{d/2}} \left( \left( \frac{u^2+1}{u^2+c} \right)^{1/2} \|\mathbf{h}\| \right)^\nu K_\nu \left( \left( \frac{u^2+1}{u^2+c} \right)^{1/2} \|\mathbf{h}\| \right), & \text{si } \|\mathbf{h}\| > 0, \\ \frac{1}{(u^2+1)^\nu(u^2+c)^{d/2}}, & \text{si } \|\mathbf{h}\| = 0, \end{cases}$$

donde  $K_\nu$  es la función de Bessel modificada de segunda especie de orden  $\nu$ . Por tanto, la función

$$C(\mathbf{h}, u) = \begin{cases} \frac{\sigma^2(2c^{d/2})}{(a^2u^2 + 1)^\nu(a^2u^2 + c)^{d/2}\Gamma(\nu)} \left( \frac{b}{2} \left( \frac{a^2u^2 + 1}{a^2u^2 + c} \right)^{1/2} \|\mathbf{h}\| \right)^\nu \\ \quad \times K_\nu \left( b \left( \frac{a^2u^2 + 1}{a^2u^2 + c} \right)^{1/2} \|\mathbf{h}\| \right), & \text{si } \|\mathbf{h}\| > 0, \\ \frac{\sigma^2 c^{d/2}}{(a^2u^2 + 1)^\nu(a^2u^2 + c)^{d/2}}, & \text{si } \|\mathbf{h}\| = 0, \end{cases} \quad (4.20)$$

será una familia de funciones de covarianza espacio-temporales en general no separables y que dependen del vector de parámetros  $\theta = (a, b, c, \nu, \sigma^2)'$ , donde  $a \geq 0$  es un parámetro de escala temporal,  $b \geq 0$  un parámetro de escala espacial,  $c > 0$  es un parámetro de interacción espacio-temporal,  $\nu > 0$  un parámetro de suavidad, y  $\sigma^2 = C(\mathbf{0}, 0) > 0$ . Para  $c = 1$  tenemos una función de covarianza separable.

La Figura 4.4 muestra diferentes representaciones de la familia de covarianzas (4.20) para un conjunto de posibles valores de los parámetros  $\theta = (a, b, c, \nu, \sigma^2)'$  que la definen, en la que podemos ver el efecto de la variación de dichos parámetros. El primer ejemplo representa la función de covarianza obtenida para  $\theta = (1, 1, 0.5, 0.5, 5)'$ , el segundo el obtenido para  $\theta = (1, 1, 1.5, 1, 5)'$ , el tercero correspondiente a  $\theta = (1, 5, 1.5, 0.5, 5)'$  y el cuarto obtenido con  $\theta = (0.5, 1, 1.5, 0.5, 5)'$ . Observemos que para este modelo, la función de covarianza meramente espacial asociada  $C(\mathbf{h}, 0)$  pertenece al modelo de Matérn (2.32), y aparece representada con un trazo discontinuo de color rojo, mientras que en verde se representa la covarianza  $C(\mathbf{0}, u)$  correspondiente.

Si en el modelo (4.20) consideramos  $\nu = 1/2$ , obtendremos una familia de covarianzas espacio-temporales que depende de cuatro parámetros y viene dada por

$$C(\mathbf{h}, u) = \frac{\sigma^2 c^{d/2}}{(a^2u^2 + 1)^{1/2}(a^2u^2 + c)^{d/2}} \exp \left( -b \left( \frac{a^2u^2 + 1}{a^2u^2 + c} \right)^{1/2} \|\mathbf{h}\| \right).$$

Todas las funciones de covarianzas espacio-temporales que hemos visto en esta sección son estacionarias y continuas. Podemos lograr una discontinuidad en el origen añadiendo un efecto pepita espacial, temporal o espacio-temporal como vimos en (3.20). Además, todos los ejemplos son espacialmente isotrópicos, pues dependen de la distancia espacial  $\mathbf{h}$  únicamente a través su módulo  $\|\mathbf{h}\|$ . No obstante, esta restricción puede relajarse introduciendo una anisotropía espacial geométrica sustituyendo en los modelos anteriores  $\|\mathbf{h}\|$  por  $\|A\mathbf{h}\|$ , siendo  $A$  una matriz no singular.

## 4.6. Modelo suma-producto

De Cesare et al. (2001a) introducen este nuevo modelo de covarianza espacio-temporal estacionario, que supone una generalización del modelo producto y que se obtiene com-

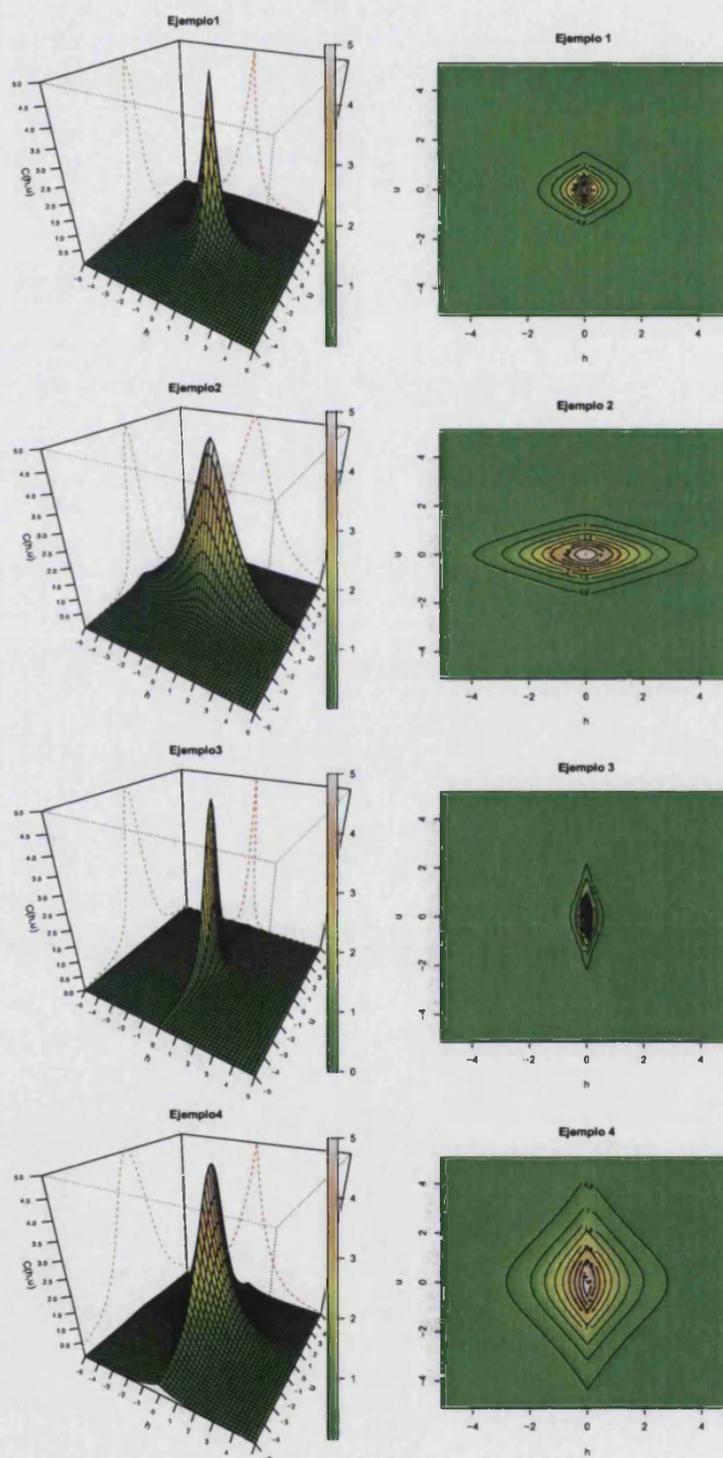


Figura 4.4: Diferentes representaciones del modelo de Cressie y Huang (4.20).

binando sumas y productos de las covarianzas meramente espacial y temporal del proceso. Permite obtener funciones de covarianza que, en general, no corresponden al uso de ninguna métrica en el espacio-tiempo, son no separables y no integrables (es decir, que no pueden obtenerse mediante una representación del modelo de Cressie y Huang). Esta nueva colección de modelos de covarianza válidos para procesos espacio-temporales estacionarios tiene una ventaja importante, la sencillez con que se ajustan a los datos, ya que para ello emplearemos las mismas técnicas usadas en la modelización espacial de variogramas.

El modelo suma-producto viene dado por la siguiente expresión

$$C(\mathbf{h}, u) = k_1 C_e(\mathbf{h}) C_t(u) + k_2 C_e(\mathbf{h}) + k_3 C_t(u), \quad (4.21)$$

siendo  $C_e$  un modelo de covarianza espacial en  $D \subset \mathfrak{R}^d$  y  $C_t$  un modelo de covarianza temporal en  $T \subset \mathfrak{R}^+$ . De (4.21) se observa que tanto el modelo producto como el modelo lineal son casos particulares del modelo suma-producto, considerando  $k_2 = k_3 = 0$  y  $k_1 = 0$ , respectivamente. La condición  $k_1 > 0$ ,  $k_2 \geq 0$ ,  $k_3 \geq 0$  nos asegura la validez del modelo propuesto, pues en ese caso la función (4.21) es definida positiva.

Por la relación existente entre el variograma y la función de covarianza (3.6), tenemos que

$$k_1 C_e(\mathbf{0}) C_t(0) + k_2 C_e(\mathbf{0}) + k_3 C_t(0) = C(\mathbf{0}, 0), \quad (4.22)$$

por lo que es sencillo obtener la expresión del semivariograma asociado a este modelo, que viene dada por

$$\gamma(\mathbf{h}, u) = [k_2 + k_1 C_t(0)] \gamma_e(\mathbf{h}) + [k_3 + k_1 C_e(\mathbf{0})] \gamma_t(u) - k_1 \gamma_e(\mathbf{h}) \gamma_t(u), \quad (4.23)$$

donde  $\gamma_e(\cdot)$  y  $\gamma_t(\cdot)$  son los semivariogramas asociados a las covarianzas  $C_e(\cdot)$  y  $C_t(\cdot)$ , respectivamente, y donde  $C(\mathbf{0}, 0)$  es la meseta de  $\gamma$ ,  $C_e(\mathbf{0})$  es la meseta de  $\gamma_e$  y  $C_t(0)$  es la meseta de  $\gamma_t$ . Por definición, sabemos que  $\gamma(\mathbf{0}, 0) = \gamma_e(\mathbf{0}) = \gamma_t(0) = 0$ .

De (4.23) tenemos que los semivariogramas marginales son en este caso

$$\gamma(\mathbf{h}, 0) = [k_2 + k_1 C_t(0)] \gamma_e(\mathbf{h}), \quad \gamma(\mathbf{0}, u) = [k_3 + k_1 C_e(\mathbf{0})] \gamma_t(u),$$

es decir, son iguales a los semivariogramas meramente espacial y temporal, respectivamente, excepto por unas constantes de proporcionalidad.

Para facilitar el ajuste de este modelo de covarianza espacio-temporal, De Cesare et al. (2001a) establecen una restricción sobre los coeficientes  $k_1$ ,  $k_2$  y  $k_3$  al suponer que

$$k_2 + k_1 C_t(0) = k_3 + k_1 C_e(\mathbf{0}) = 1. \quad (4.24)$$

De (4.22) y (4.24) tenemos que

$$k_1 = \frac{C_e(\mathbf{0}) + C_t(0) - C(\mathbf{0}, 0)}{C_e(\mathbf{0}) C_t(0)}, \quad k_2 = \frac{C(\mathbf{0}, 0) - C_t(0)}{C_e(\mathbf{0})}, \quad k_3 = \frac{C(\mathbf{0}, 0) - C_e(\mathbf{0})}{C_t(0)}. \quad (4.25)$$

Por lo tanto, al modelizar el semivariograma espacial y el temporal tendremos que tener presente que las mesetas de todos ellos deberán hacer positivos los numeradores de las expresiones anteriores.

De Cesare et al. (2001b) muestran una aplicación de este modelo suma-producto en el estudio de la concentración de dióxido de nitrógeno en Milán a partir de medidas horarias obtenidas en 48 estaciones de control. De Cesare et al. (2002) proporciona algunas modificaciones de la librería GSLIB para FORTRAN77 que permiten la modelización de determinados variogramas espacio-temporales. En concreto, se incluye el código que permite el ajuste del modelo producto (4.6) y del modelo suma-producto (4.21), además de una modificación del kriging que permite incorporar dichos modelos espacio-temporales.

De Iaco et al. (2001) generalizan el modelo anterior, suavizando las restricciones (4.24) impuestas a los coeficientes del modelo de covarianza que no son necesarias para asegurar su validez, y que imponen una especie de simetría en el impacto de la componente de la variación espacial y temporal que no tiene porqué ocurrir. Esto permitirá obtener una familia de funciones de covarianza suma-producto más general en la que, tras la modelización de las componentes espacial y temporal, la modelización espacio-temporal dependerá de un único parámetro extra. Se establece también la relación entre este nuevo parámetro y los valores meseta de los variogramas espacial y temporal, así como el rango de valores que aseguran la validez del modelo resultante.

Aplicando la relación existente entre la covarianza de un proceso y su semivariograma asociado dada por (3.6), se tiene que

$$\gamma(\mathbf{h}, u) = C(\mathbf{0}, 0) - C(\mathbf{h}, u), \quad \gamma(\mathbf{h}, 0) = C(\mathbf{0}, 0) - C(\mathbf{h}, 0), \quad \gamma(\mathbf{0}, u) = C(\mathbf{0}, 0) - C(\mathbf{0}, u).$$

El siguiente teorema analiza el comportamiento asintótico de estos tres semivariogramas, calculando sus respectivas mesetas. En De Iaco et al. (2001) puede consultarse su demostración, que se basa en las relaciones anteriores.

**Teorema 4.1** *Sea  $Z(s, t)$  un proceso espacio-temporal estacionario de segundo orden. Supongamos que su función de covarianza tiene la expresión de un modelo suma-producto, y que además es una función continua en el espacio-tiempo. Entonces se tiene que*

$$\begin{aligned} \lim_{\mathbf{h} \rightarrow \text{inf}} \lim_{u \rightarrow \text{inf}} \gamma_{et}(\mathbf{h}, u) &= C_{et}(\mathbf{0}, 0), \\ \lim_{\mathbf{h} \rightarrow \text{inf}} \gamma_{et}(\mathbf{h}, 0) &= k_e C_e(\mathbf{0}), \\ \lim_{u \rightarrow \text{inf}} \gamma_{et}(\mathbf{0}, u) &= k_t C_t(0). \end{aligned}$$

Observemos que podemos expresar el modelo (4.23) como

$$\begin{aligned} \gamma(\mathbf{h}, u) &= k_e \gamma_e(\mathbf{h}) + k_t \gamma_t(u) - k_1 \gamma_e(\mathbf{h}) \gamma_t(u) \\ &= \gamma(\mathbf{h}, 0) + \gamma(\mathbf{0}, u) - k_1 \frac{\gamma(\mathbf{h}, 0)}{k_s} \frac{\gamma(\mathbf{0}, u)}{k_t} \\ &= \gamma(\mathbf{h}, 0) + \gamma(\mathbf{0}, u) - k \gamma(\mathbf{h}, 0) \gamma(\mathbf{0}, u), \end{aligned} \quad (4.26)$$

siendo

$$k = \frac{k_1}{k_e k_t} = \frac{k_e C_e(\mathbf{0}) + k_t C_t(0) - C(\mathbf{0}, 0)}{k_e C_e(\mathbf{0}) k_t C_t(0)}. \quad (4.27)$$

El siguiente teorema establece como condición necesaria y suficiente de la validez del modelo (4.26) que el parámetro  $k$  sea un valor positivo y menor que la inversa del máximo de las mesetas de los semivariogramas  $\gamma(\mathbf{h}, 0)$  y  $\gamma(\mathbf{0}, u)$ .

**Teorema 4.2** Sea  $Z(s, t)$  un proceso espacio-temporal estacionario de segundo orden. Supongamos que su función de covarianza es continua y tiene la expresión de un modelo suma-producto. Entonces  $k_1 > 0$ ,  $k_2 \geq 0$  y  $k_3 \geq 0$  si y sólo si el coeficiente  $k$  definido en (4.27) satisface la desigualdad siguiente

$$0 \leq k \leq \frac{1}{\max\{k_e C_e(0), k_t C_t(0)\}}. \quad (4.28)$$

Observemos que este modelo puede generar funciones de covarianza no integrables, lo que no era posible con el modelo de Cressie y Huang que hemos visto en la Sección 4.5. Se puede probar fácilmente que si  $k_2 \neq 0$  y  $k_3 \neq 0$ , entonces la función de covarianza  $C(\mathbf{h}, \mathbf{u})$  obtenida en (4.21) no es integrable, independientemente de la elección de los modelos  $C_e(\mathbf{h})$  y  $C_t(u)$ . De Iaco et al. (2002a) utilizan este modelo suma-producto generalizado para estimar la contaminación total del aire (TAP) en el distrito de Milán (Italia) durante un conjunto de días que se consideran de riesgo durante 1999. Este valor TAP es un indicador combinado de la contaminación del tráfico a partir de datos de óxido nítrico (NO), dióxido de nitrógeno (NO<sub>2</sub>) y monóxido de carbono (CO). Los autores indican que una de las ventajas de este modelo es que permite la modelización del proceso espacio-temporal a partir únicamente de las funciones de covarianza (o los semivariogramas) marginales. No obstante, esto supone una restricción en su aplicabilidad, pues no nos permitirá introducir una determinada interacción entre el espacio y el tiempo en el fenómeno analizado.

**Ejemplo 4.9** Podemos obtener un ejemplo de este modelo suma-producto considerando una función de covarianza espacial definida por la correlación del modelo de Matérn  $C_e(\mathbf{h}) = (2^{1-\nu}/\Gamma(\nu))(b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|)$  y una función de covarianza temporal definida por la función de autocorrelación de un proceso AR(1) de la forma  $C_t(u) = \alpha^{|u|}$ . Con estas funciones de covarianza, tenemos el modelo

$$C(\mathbf{h}, \mathbf{u}) = k_1 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|) \alpha^{|\mathbf{u}|} + k_2 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|) + k_3 \alpha^{|\mathbf{u}|}, \quad (4.29)$$

donde  $k_1 > 0$ ,  $k_2 \geq 0$  y  $k_3 \geq 0$  son los pesos de la combinación lineal que define el modelo de covarianza,  $\alpha \in (0, 1]$  es un parámetro de suavidad temporal,  $b > 0$  es un parámetro de escala espacial y  $\nu > 0$  un parámetro de suavidad espacial.

La Figura 4.5 muestra diferentes representaciones de esta familia de covarianzas (4.29) para un conjunto de posibles valores de los parámetros que la definen. En concreto, si  $\theta = (k_1, k_2, k_3, \alpha, b, \nu)'$  es el vector de parámetros que definen este modelo (4.29), se han representado las funciones de covarianza correspondientes a  $\theta = (2, 1.5, 1.5, 0.4, 0.3, 1)'$  para el primer ejemplo,  $\theta = (2, 1.5, 1.5, 0.8, 0.3, 1)'$  para el segundo,  $\theta = (2, 1.5, 1.5, 0.4, 1.1, 0.5)'$  para el tercero y  $\theta = (4, 0.5, 0.5, 0.4, 0.3, 1)'$  para el cuarto.

## 4.7. Modelos propuestos por Gneiting

Gneiting (2002) propone un nuevo procedimiento para la construcción de modelos de funciones de covarianza espacio-temporal estacionarias que en general serán no separables.

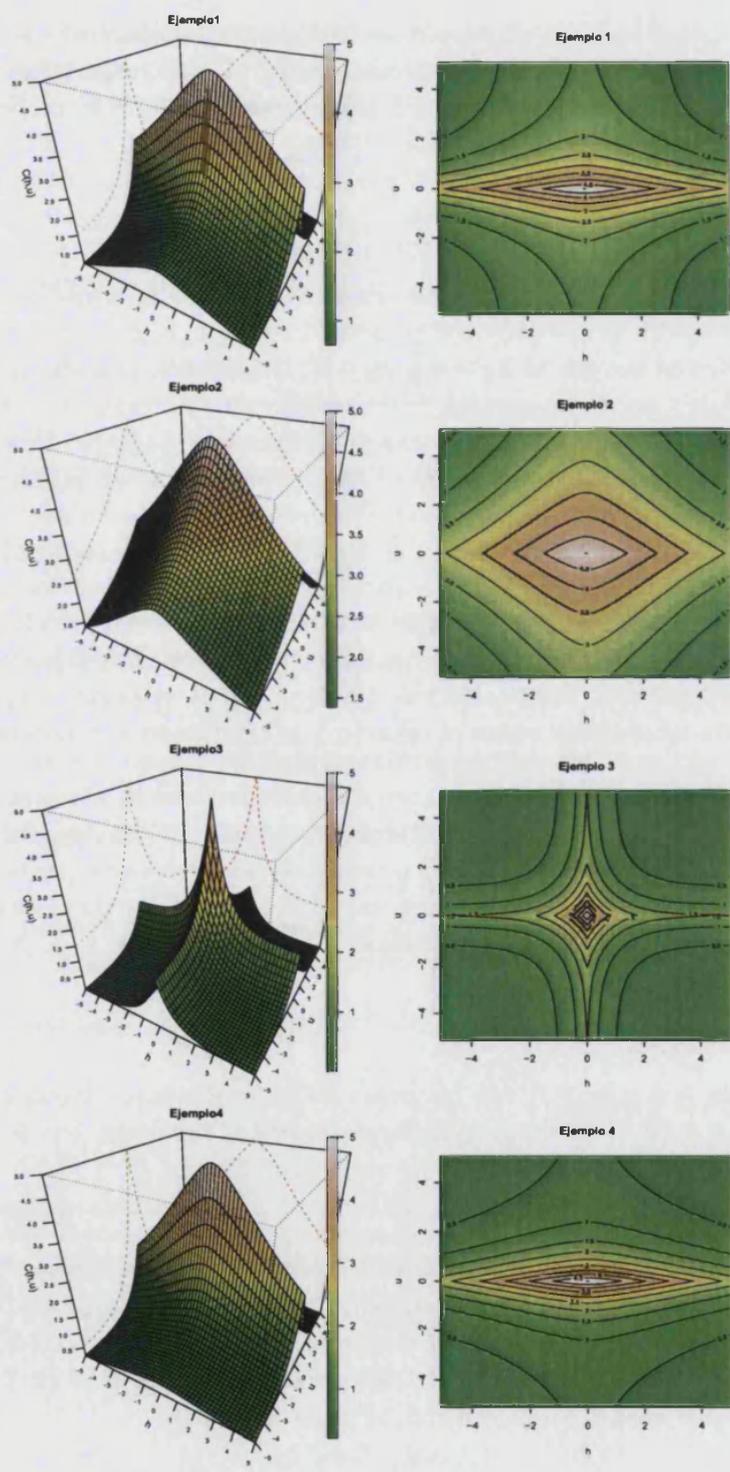


Figura 4.5: Diferentes representaciones del modelo suma-producto (4.29).

Función	Parámetros
$\varphi(x) = \exp(-cx^\gamma)$	$c > 0, 0 < \gamma \leq 1$
$\varphi(x) = (2^{\nu-1}\Gamma(\nu))^{-1} (cx^{1/2})^\nu K_\nu(cx^{1/2})$	$c > 0, \nu > 0$
$\varphi(x) = (1 + cx^\gamma)^{-\nu}$	$c > 0, 0 < \gamma \leq 1, \nu > 0$
$\varphi(x) = 2^\nu (\exp(cx^{1/2}) + \exp(-cx^{1/2}))^{-\nu}$	$c > 0, \nu > 0$

Tabla 4.1: Ejemplos de funciones completamente monótonas.

Esta técnica no precisa del uso de transformadas de Fourier como ocurría con el método de Huang y Cressie que hemos visto en la Sección 4.5, sino que construye familias paramétricas de funciones de covarianza espacio-temporal directamente en el dominio espacio-tiempo.

En la definición de esta clase de modelos intervienen dos tipos de funciones muy relacionadas con la estadística espacial, las funciones completamente monótonas y las funciones de Bernstein. Veamos brevemente la definición y principales propiedades de estos tipos de funciones.

**Definición 4.1** Decimos que una función continua no negativa  $\varphi(x)$ , con  $x \geq 0$ , es una función completamente monótona si existen sus derivadas de cualquier orden  $\varphi^{(n)}$  y además

$$(-1)^n \varphi^{(n)}(x) \geq 0, \quad x > 0, n \in \mathbb{N}.$$

Las funciones de covarianza isotrópicas y las funciones completamente monótonas están estrechamente relacionadas, ya que se cumple que una función continua e isotrópica  $C(\mathbf{h}) = \varphi(\|\mathbf{h}\|^2)$  definida sobre  $\mathbb{R}^d$  es una función de covarianza espacial para cualquier dimensión  $d$  si y sólo si  $\varphi(x)$ ,  $x \geq 0$ , es completamente monótona.

El teorema de Bernstein (1928) establece que, dada una función  $\varphi(x)$  completamente monótona, existe una medida finita y no negativa en  $\mathbb{R}^+$  de forma que

$$\varphi(x) = \int_0^{+\infty} e^{-rx} dF(r), \quad x > 0.$$

Por tanto, existe una correspondencia entre las funciones completamente monótonas y las transformadas de Laplace de medidas finitas y no negativas sobre  $\mathbb{R}^+$ .

En Miller y Samko (2001) podemos encontrar ejemplos y propiedades de este tipo de funciones. La Tabla 4.1 muestra algunos ejemplos de funciones completamente monótonas, donde  $K_\nu$  representa la función de Bessel modificada de segunda especie de orden  $\nu$ . Por ejemplo, la segunda entrada de esta tabla tiene asociada una función de covarianza isotrópica dada por

$$C(\mathbf{h}) = \varphi(\|\mathbf{h}\|^2) = \sigma^2 (2^{\nu-1}\Gamma(\nu))^{-1} (c\|\mathbf{h}\|)^\nu K_\nu(c\|\mathbf{h}\|)$$

que corresponde al modelo de funciones de covarianzas de Matérn visto en (2.32).

**Definición 4.2** Decimos que una función continua no negativa  $\psi(x)$ , con  $x \geq 0$ , es una función de Bernstein si es la integral de una función completamente monótona  $\varphi(r)$ , esto

Función	Parámetros
$\psi(x) = (ax^\alpha + 1)^\beta$	$a > 0, 0 < \alpha \leq 1, 0 \leq \beta \leq 1$
$\psi(x) = \ln(ax^\alpha + 1)$	$a > 0, 0 < \alpha \leq 1$
$\psi(x) = \ln(ax^\alpha + b) / \ln(b)$	$a > 0, b > 1, 0 < \alpha \leq 1$
$\psi(x) = (ax^\alpha + b) / (b(ax^\alpha + 1))$	$a > 0, 0 < b \leq 1, 0 < \alpha \leq 1$

Tabla 4.2: Ejemplos de funciones de Bernstein.

es, si

$$\psi(x) = \int_0^x \varphi(r)dr = \int_0^\infty (1 - e^{-rx})r^{-1}dF(r), \quad x > 0$$

donde  $F(r)$ ,  $r > 0$ , es una función no decreciente y  $\int_0^1 r^{-1}dF(r) < \infty$ .

Las funciones de Bernstein tendrán por tanto derivada completamente monótona. Se puede comprobar que una función continua e isotrópica,  $\gamma(\mathbf{h}) = \psi(\|\mathbf{h}\|^2)$ , definida sobre  $\mathfrak{R}^d$  es un semivariograma para cualquier dimensión  $d$  si y sólo si  $\psi(x)$ ,  $x \geq 0$ , es una función de Bernstein con  $\psi(0) = 0$ . La Tabla 4.2 muestran algunos ejemplos de funciones de este tipo.

Como veremos a continuación, el procedimiento propuesto por Gneiting (2002) construye funciones de covarianza espacio-temporales utilizando las funciones anteriores en base al siguiente resultado.

**Teorema 4.3** Sea  $\varphi(x)$ , con  $x \geq 0$ , una función completamente monótona y  $\psi(t)$ , con  $t \geq 0$ , una función de Bernstein, entonces la función

$$C(\mathbf{h}, u) = \frac{\sigma^2}{\psi(|u|^2)^{d/2}} \varphi\left(\frac{\|\mathbf{h}\|^2}{\psi(|u|^2)}\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (4.30)$$

es una función de covarianza espacio-temporal.

La demostración de este teorema puede consultarse en Gneiting (2002). Las funciones de covarianza generadas por (4.30) serán completamente simétricas.

El procedimiento de construcción que se deriva del Teorema 4.3 es muy general y flexible. Por ejemplo, todos los modelos de Cressie y Huang (1999) pueden expresarse como (4.30), a excepción del modelo (4.8) para el caso en que  $c < 1$ . Aunque la función de covarianza obtenida sea en general no separable, las funciones  $\varphi(x)$  y  $\psi(t)$  pueden asociarse, respectivamente, con la estructura espacial y temporal de los datos. Veamos a continuación los dos ejemplos que Gneiting analiza en este trabajo.

**Ejemplo 4.10** Consideremos las funciones  $\varphi(x) = \exp(-cx^\gamma)$ , con  $x \geq 0$ , y  $\psi(t) = (at^\alpha + 1)^\beta$ , con  $t \geq 0$ . Se puede comprobar que, para  $c > 0$  y  $0 < \gamma \leq 1$ ,  $\varphi(x)$  es una función completamente monótona y que, para  $a > 0$ ,  $0 < \alpha \leq 1$  y  $0 \leq \beta \leq 1$ ,  $\psi(t)$  es una función positiva con derivada completamente monótona. Luego por el Teorema 4.3, la función

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\beta d/2}} \exp\left(-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

es un modelo de covarianza espacio-temporal válido sobre  $\mathbb{R}^d \times \mathbb{R}$ . Su producto con la función de covarianza meramente temporal  $(a|u|^{2/\alpha} + 1)^{-\delta}$  dará lugar al siguiente modelo válido de función de covarianza espacio-temporal

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\delta + \beta d/2}} \exp\left(-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (4.31)$$

donde  $c > 0$  y  $a > 0$  son dos parámetros positivos de escala espacial y temporal, respectivamente,  $\gamma$  y  $\alpha$  son dos parámetros de suavidad espacial y temporal que toman valores en  $]0, 1]$ ,  $\beta$  es un parámetro de interacción espacio-temporal que toma valores en  $[0, 1]$ ,  $\delta \geq 0$ , y  $\sigma^2$  la varianza del proceso espacio-temporal. La función de covarianza meramente espacial  $C(\mathbf{h}, 0)$  asociada a este modelo pertenece a la familia de covarianzas de potencia exponencial, mientras que la función meramente temporal  $C(\mathbf{0}, u)$  pertenece a la familia de covarianzas de Cauchy. Para  $\beta = 0$  obtenemos un modelo de covarianza espacio-temporal separable, mientras que para  $\beta = 1$  obtenemos el caso más extremo de modelo no separable de esta familia. Observemos que para  $\alpha = 1$ ,  $\beta = 1$ ,  $\delta = 0$ ,  $a = d^2$ ,  $c = b^2$  y  $\gamma = 1$  obtenemos el modelo (4.16) de la Sección 4.5. Por otro lado, para  $\alpha = 1/2$ ,  $\beta = 1$ ,  $\delta = 0$ ,  $c = b^2$  y  $\gamma = 1$  obtenemos el modelo (4.17) de la Sección 4.5.

**Ejemplo 4.11** Consideremos las funciones  $\varphi(x) = (2^{\nu-1}\Gamma(\nu))^{-1} (cx^{1/2})^\nu K_\nu(cx^{1/2})$ , con  $x \geq 0$ , y  $\psi(t) = (at^\alpha + 1)^\beta$ , con  $t \geq 0$ . Se puede comprobar que, para  $c > 0$  y  $\nu > 0$ ,  $\varphi(x)$  es una función completamente monótona y que, para  $a > 0$ ,  $0 < \alpha \leq 1$  y  $0 \leq \beta \leq 1$ ,  $\psi(t)$  es una función positiva con derivada completamente monótona. Luego por el Teorema 4.3, la función

$$C(\mathbf{h}, u) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)(a|u|^{2\alpha} + 1)^{\beta d/2}} \left(\frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right)^\nu K_\nu\left(\frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right),$$

es un modelo de covarianza espacio-temporal válido sobre  $\mathbb{R}^d \times \mathbb{R}$ . Su producto con la función de covarianza meramente temporal  $(a|u|^{2/\alpha} + 1)^{-\delta}$  dará lugar al siguiente modelo válido de función de covarianza espacio-temporal

$$C(\mathbf{h}, u) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)(a|u|^{2\alpha} + 1)^{\delta + \beta d/2}} \left(\frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right)^\nu K_\nu\left(\frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right), \quad (4.32)$$

que depende de un vector de parámetros  $\theta = (a, c, \alpha, \beta, \nu, \delta, \sigma^2)'$ , donde  $c > 0$  y  $a > 0$  son dos parámetros positivos de escala espacial y temporal, respectivamente,  $\nu > 0$  es un parámetro de suavizado espacial,  $\alpha \in ]0, 1]$  es un parámetro de suavizado temporal,  $\beta \in [0, 1]$  es un parámetro relacionado con la interacción espacio-temporal,  $\delta \geq 0$  y  $\sigma^2 > 0$  la varianza del proceso espacio-temporal. Como antes, para  $\beta = 0$  tenemos una función de covarianza espacio-temporal separable, mientras que para  $\beta = 1$  obtenemos el representante menos separable de esta familia de covarianzas.

La Figura 4.6 muestra diferentes representaciones de esta familia de covarianzas (4.32) para un conjunto de posibles valores de los parámetros  $\theta = (a, c, \alpha, \beta, \nu, \delta, \sigma^2)'$  que la

definen, en la que podemos ver el efecto de la variación de cada uno de los parámetros anteriores. El primer ejemplo representa la función de covarianza obtenida para  $\theta = (0.5, 0.3, 0.9, 0, 1, 0.5, 5)'$ , el segundo el obtenido para  $\theta = (0.5, 0.3, 0.9, 0.9, 1, 0.5, 5)'$ , el tercero correspondiente a  $\theta = (0.5, 1.5, 0.9, 0, 1, 0.5, 5)'$  y el cuarto obtenido con  $\theta = (0.5, 0.3, 0.9, 0, 0.5, 0.5, 5)'$ . Observemos que para este modelo, la función de covarianza meramente espacial asociada  $C(\mathbf{h}, 0)$  pertenece al modelo de Matérn (2.32), y aparece representada con un trazo discontinuo de color rojo, mientras que en verde se representa la covarianza  $C(\mathbf{0}, u)$  correspondiente, que viene dada por  $\sigma^2/(a|u|^{2\alpha} + 1)^{\delta+\beta d/2}$ ,  $u \in \mathfrak{R}$ .

Para  $\nu = 1/2$ , la función de covarianza espacio-temporal (4.32) se reduce a

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\delta+\beta d/2}} \exp\left(-\frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right),$$

que coincide con (4.31) para  $\gamma = 1/2$ , mientras que para  $\nu = 3/2$  tenemos

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\delta+\beta d/2}} \left(1 + \frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right) \exp\left(-\frac{c\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}}\right).$$

Se puede comprobar que tomando  $\varphi(x) = (1 + bu)^{-(d+1)/2}$  y  $\psi(t) = (at + 1)^d$ , y tomando  $\varphi(x) = (1 + bu)^{-(d+1)/2}$  y  $\psi(t) = (at^{1/2} + 1)^d$ , obtenemos respectivamente los modelos (4.18) y (4.19) de Cressie y Huang que hemos visto en la Sección 4.5.

## 4.8. Modelos producto y suma-producto integrado

De Iaco et al. (2002b) utilizan la segunda propiedad de estabilidad de las funciones de covarianza (3.14) aplicada sobre los modelos producto y suma-producto para construir nuevos modelos válidos estacionarios de funciones de covarianza y de semivariograma, y que serán mixturas particulares de funciones de covarianza espaciales y temporales. Aplicando estas propiedades, es sencillo demostrar el siguiente resultado:

**Teorema 4.4** *Sea  $\mu(a)$  una medida acotada no negativa en  $U \subseteq \mathfrak{R}$ , y sean  $C_e(\mathbf{h}|a)$  y  $C_t(u|a)$  dos funciones de covarianza sobre  $D \subset \mathfrak{R}^d$  y  $T \subset \mathfrak{R}$ , respectivamente, para cada  $a \in V \subseteq U$ .*

(a) *Si  $C_e(\mathbf{h}|a) \cdot C_t(u|a)$  es integrable respecto a la medida  $\mu(a)$  sobre  $V$  para cada  $\mathbf{h} \in \mathfrak{R}^d$  y  $u \in \mathfrak{R}$ , entonces dado  $k > 0$ , la función*

$$C(\mathbf{h}, u) = \int_V k C_e(\mathbf{h}|a) C_t(u|a) d\mu(a), \quad (4.33)$$

*es una función de covarianza sobre  $D \times T$ .*

(b) *Además, si  $k_1 C_e(\mathbf{h}|a) C_t(u|a) + k_2 C_e(\mathbf{h}|a) + k_3 C_t(u|a)$  es integrable con respecto a la medida  $\mu(a)$  sobre  $V$  para cada  $\mathbf{h} \in \mathfrak{R}^d$  y  $u \in \mathfrak{R}$ , con  $k_1 > 0$ ,  $k_2 \geq 0$  y  $k_3 \geq 0$ , entonces la función*

$$C(\mathbf{h}, u) = \int_V [k_1 C_e(\mathbf{h}|a) C_t(u|a) + k_2 C_e(\mathbf{h}|a) + k_3 C_t(u|a)] d\mu(a), \quad (4.34)$$

*es una función de covarianza sobre  $D \times T$ .*

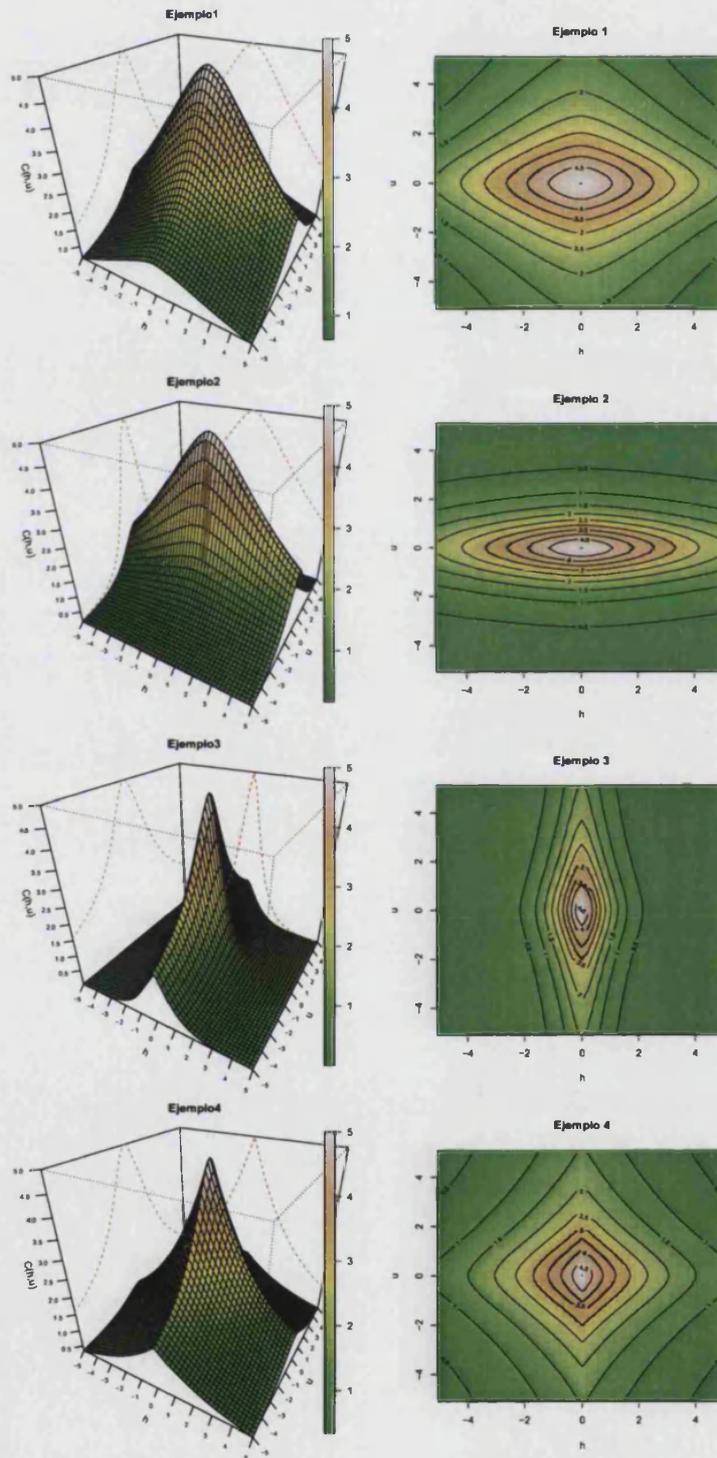


Figura 4.6: Diferentes representaciones del modelo de Gneiting (4.32).

Con estos modelos podemos generar funciones de covarianza espacio-temporales no separables y no integrables. De la relación (3.5) entre el semivariograma y la función de covarianza, podemos expresar el resultado anterior en términos del semivariograma del proceso como sigue

$$\gamma(\mathbf{h}, u) = \int_V k [C_t(0|a)\gamma_s(\mathbf{h}|a) + C_e(\mathbf{0}|a)\gamma_t(u|a) - \gamma_s(\mathbf{h}|a)\gamma_t(u|a)] d\mu(a),$$

y

$$\gamma(\mathbf{h}, u) = \int_V [(k_2 + k_1 C_t(0|a))\gamma_s(\mathbf{h}|a) + (k_3 + k_1 C_e(\mathbf{0}|a))\gamma_t(u|a) - k_1 \gamma_s(\mathbf{h}|a)\gamma_t(u|a)] d\mu(a),$$

donde  $\gamma_s(\mathbf{h}|a)$  y  $\gamma_t(u|a)$  son dos modelos válidos de semivariograma espacial y temporal, respectivamente, para cualquier  $a \in V$ , siendo  $C_e(\mathbf{0}|a)$  y  $C_t(0|a)$  sus correspondientes mesetas.

Veamos a continuación algunos ejemplos de estos modelos. En todos ellos supondremos que la medida de probabilidad  $\mu$  está generada por una función absolutamente continua  $\Phi$  (su función de distribución), por lo que existirá una función  $\phi$  tal que  $d\Phi(a) = \phi(a)da$  casi por todas partes (su función de densidad).

**Ejemplo 4.12** Consideremos la función de densidad

$$\phi(a, n, \beta) = \frac{\beta^{n+1}}{\Gamma(n+1)} a^n \exp(-\beta a), \quad n \geq 0, \beta > 0,$$

que es la asociada a la distribución Gamma de parámetros  $n+1$  y  $1/\beta$ . Consideremos:

$$C_s(\mathbf{h}|a, b, \alpha) = \exp\left(\frac{-a\|\mathbf{h}\|^\alpha}{b}\right), \quad 1 \leq \alpha \leq 2, a > 0, b > 0,$$

$$C_t(u|a, c, \delta) = \exp\left(\frac{-au^\delta}{c}\right), \quad 1 \leq \delta \leq 2, c > 0,$$

donde  $C_s(\mathbf{h}|a, b, \alpha)$  y  $C_t(u|a, c, \delta)$  son dos modelos de covarianza válidos espacial y temporal, respectivamente, para cualquier  $a \in V = [0, \infty[$ . En este caso, aplicando el primer apartado del Teorema 4.4

$$\begin{aligned} C(\mathbf{h}, u) &= \int_V k \exp\left(\frac{-a\|\mathbf{h}\|^\alpha}{b}\right) \exp\left(\frac{-au^\delta}{c}\right) \frac{\beta^{n+1}}{\Gamma(n+1)} a^n \exp(-\beta a) da \\ &= \frac{k\beta^{n+1}}{\Gamma(n+1)} \int_V a^n \exp\left(-a\left(\frac{\|\mathbf{h}\|^\alpha}{b} + \frac{u^\delta}{c} + \beta\right)\right) da = \frac{k\beta^{n+1}}{\left(\frac{\|\mathbf{h}\|^\alpha}{b} + \frac{u^\delta}{c} + \beta\right)^{n+1}}, \end{aligned}$$

es una función de covarianza espacio-temporal válida que depende de  $\theta = (b, c, n, k, \alpha, \beta, \delta)'$ , mientras que aplicando el segundo apartado se tiene que

$$\begin{aligned} C(\mathbf{h}, u) &= \int_V \left[ k_1 \exp\left(\frac{-a\|\mathbf{h}\|^\alpha}{b}\right) \exp\left(\frac{-au^\delta}{c}\right) + k_2 \exp\left(\frac{-a\|\mathbf{h}\|^\alpha}{b}\right) + k_3 \exp\left(\frac{-au^\delta}{c}\right) \right] \\ &\times \frac{\beta^{n+1}}{\Gamma(n+1)} a^n \exp(-\beta a) da = k_1 \frac{\beta^{n+1}}{\left(\frac{\|\mathbf{h}\|^\alpha}{b} + \frac{u^\delta}{c} + \beta\right)^{n+1}} + k_2 \frac{\beta^{n+1}}{\left(\frac{\|\mathbf{h}\|^\alpha}{b} + \beta\right)^{n+1}} + k_3 \frac{\beta^{n+1}}{\left(\frac{u^\delta}{c} + \beta\right)^{n+1}}, \end{aligned}$$

es también una función de covarianza espacio-temporal válida que depende del vector de parámetros  $\theta = (b, c, n, k_1, k_2, k_3, \alpha, \beta, \delta)'$ .

**Ejemplo 4.13** Consideremos la función  $\phi(a, \beta) = \beta \exp(-a\beta)$ , con  $\beta > 0$ , que es la función de densidad de la distribución exponencial de parámetro  $\beta$ . Consideremos:

$$C_s(\mathbf{h}|a, \omega) = \cos(a(\omega\|\mathbf{h}\|)) \quad , \quad a > 0, \omega \in \mathfrak{R},$$

$$C_t(u|a, c, \delta) = \exp\left(\frac{-au^\delta}{c}\right) \quad , \quad 1 \leq \delta \leq 2, c > 0,$$

donde  $C_s(\mathbf{h}|a, b, \alpha)$  y  $C_t(u|a, c, \delta)$  son dos modelos de covarianza válidos espacial y temporal, respectivamente, para cualquier  $a \in V = [0, \infty[$ , por lo que por (4.33) la función

$$C(\mathbf{h}, u) = \int_V k \exp\left(\frac{-au^\delta}{c}\right) \cos(a(\omega\|\mathbf{h}\|)) \beta \exp(-\beta a) da$$

$$= k\beta \int_V \exp\left(-a\left(\frac{u^\delta}{c} + \beta\right)\right) \cos(a(\omega\|\mathbf{h}\|)) da = \frac{k\beta\left(\frac{u^\delta}{c} + \beta\right)}{\left(\frac{u^\delta}{c} + \beta\right)^2 + (\omega\|\mathbf{h}\|)^2},$$

es una función de covarianza espacio-temporal válida que depende de  $\theta = (c, k, \omega, \beta, \delta)'$ , mientras que por (4.34) se tiene que

$$C(\mathbf{h}, u) = \int_V \left[ k_1 \exp\left(\frac{-au^\delta}{c}\right) \cos(a(\omega\|\mathbf{h}\|)) + k_2 \cos(a(\omega\|\mathbf{h}\|)) + k_3 \exp\left(\frac{-au^\delta}{c}\right) \right]$$

$$\beta \exp(-a\beta) da = k_1 \frac{\beta\left(\frac{u^\delta}{c} + \beta\right)}{\left(\frac{u^\delta}{c} + \beta\right)^2 + (\omega\|\mathbf{h}\|)^2} + k_2 \frac{\beta^2}{\beta^2 + (\omega\|\mathbf{h}\|)^2} + k_3 \frac{\beta}{\frac{u^\delta}{c} + \beta},$$

es una función de covarianza espacio-temporal válida que depende del vector de parámetros  $\theta = (c, k_1, k_2, k_3, \omega, \beta, \delta)'$ .

Además de los ejemplos anteriores, De Iaco et al. (2002b) proponen otros modelos seleccionando otras funciones de densidad y otras funciones de covarianza espaciales y temporales. También muestran el procedimiento para, a partir de unos datos observados, seleccionar los modelos de variograma o de función de covarianza a utilizar, la función de densidad adecuada y ajustar los parámetros del modelo a los datos observados. En Myers et al. (2002) podemos encontrar un resumen de todos los modelos de covarianza que hemos visto hasta el momento. De Iaco et al. (2003) ofrecen una extensión del uso de los modelos suma-producto, suma-producto generalizado y suma-producto integrado para fenómenos naturales donde el proceso de interés sea multivariante, y en concreto la generalización al caso espacio-temporal del Modelo de Corregionalización Lineal utilizado en el análisis geoestadístico multivariante espacial.

En Ma (2003a) podemos encontrar la siguiente generalización del modelo producto integrado:

**Teorema 4.5** Sea  $d_0 \in \mathbb{N}$  y  $\mu(\omega)$  una medida acotada no negativa en  $\mathbb{R}_+^{d_0}$ . Sea  $C_e(\mathbf{h}|\omega)$  una función de covarianza sobre  $\mathbb{R}^d$  para todo  $\omega \in \mathbb{R}_+^{d_0}$  y una función medible de  $\omega \in \mathbb{R}_+^{d_0}$  para todo  $\mathbf{h} \in \mathbb{R}^d$ . Sea  $C_t(u|\omega)$  una función de covarianza sobre  $\mathcal{T}$  para todo  $\omega \in \mathbb{R}_+^{d_0}$  y una función medible de  $\omega \in \mathbb{R}_+^{d_0}$  para todo  $u \in \mathcal{T}$ . Si la integral

$$C(\mathbf{h}, u) = \int_{\mathbb{R}_+^{d_0}} C_e(\mathbf{h}|\omega)C_t(u|\omega)d\mu(\omega), \quad (4.35)$$

existe, para todo  $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathcal{T}$ , entonces (4.35) es una función de covarianza espacio-temporal estacionaria.

La covarianza (4.35) puede verse como una mixtura de covarianzas separables, dado que para cada  $\omega \in \mathbb{R}_+^{d_0}$ , el integrando de (4.35) es simplemente el producto de la covarianza meramente espacial  $C_e(\mathbf{s}|\omega)$  y una covarianza meramente temporal  $C_t(t|\omega)$ .

A continuación se citan algunos casos particulares del Teorema 4.5 de especial interés para la construcción de funciones de covarianza espacio-temporales estacionarias:

- Sea  $\mu(\omega)$  una medida acotada no negativa sobre  $\mathbb{R}_+^d$ . Sea  $C_t(u|\omega)$  una función de covarianza temporal estacionaria definida sobre  $\mathcal{T}$  para todo  $\omega \in \mathbb{R}_+^d$ , que es una función medible de  $\omega \in \mathbb{R}_+^d$  para todo  $u \in \mathcal{T}$ . Entonces por el Teorema 4.5,

$$C(\mathbf{h}, u) = \int_{\mathbb{R}_+^d} \cos(\omega'\mathbf{h})C_t(u|\omega)d\mu(\omega), \quad (4.36)$$

es una función de covarianza espacio-temporal estacionaria sobre  $\mathbb{R}^d \times \mathcal{T}$ .

- Sea  $\mu(\omega)$  una medida acotada no negativa sobre  $\mathbb{R}_+^d$ . Sea  $\gamma_t(u|\omega)$  un semivariograma temporal intrínsecamente estacionario sobre  $\mathcal{T}$  para todo  $\omega \in \mathbb{R}_+^d$ , que es una función medible de  $\omega \in \mathbb{R}_+^d$  para todo  $u \in \mathcal{T}$ . Entonces por el Teorema 4.5,

$$C(\mathbf{h}, u) = \int_{\mathbb{R}_+^d} \cos(\omega'\mathbf{h}) \exp(-\gamma_t(u|\omega))d\mu(\omega), \quad (4.37)$$

es una función de covarianza espacio-temporal estacionaria sobre  $\mathbb{R}^d \times \mathcal{T}$ .

- Sea  $\mu(\omega)$  una medida acotada no negativa sobre  $\mathbb{R}_+$ . Sea  $C_e(\mathbf{h}|\omega)$  una función de covarianza espacial estacionaria definida sobre  $\mathbb{R}^d$  para todo  $\omega \geq 0$ , que es una función medible de  $\omega \in \mathbb{R}_+$  para todo  $\mathbf{h} \in \mathbb{R}^d$ . Entonces por el Teorema 4.5,

$$C(\mathbf{h}, u) = \int_0^{+\infty} \cos(\omega t)C_e(\mathbf{h}|\omega)d\mu(\omega), \quad (4.38)$$

es una función de covarianza espacio-temporal estacionaria sobre  $\mathbb{R}^d \times \mathcal{T}$ .

- Sea  $\mu(\omega)$  una medida acotada no negativa sobre  $\mathbb{R}_+$ . Sea  $\gamma_e(\mathbf{h}|\omega)$  un semivariograma temporal intrínsecamente estacionario sobre  $\mathbb{R}^d$  para todo  $\omega \geq 0$ , que es una función medible de  $\omega \in \mathbb{R}_+$  para todo  $\mathbf{h} \in \mathbb{R}^d$ . Entonces por el Teorema 4.5,

$$C(\mathbf{h}, u) = \int_0^{+\infty} \cos(\omega t) \exp(-\gamma_e(\mathbf{h}|\omega))d\mu(\omega), \quad (4.39)$$

es una función de covarianza espacio-temporal estacionaria sobre  $\mathbb{R}^d \times T$ .

Ma (2003a) utiliza estas propiedades para construir modelos de funciones de covarianza espacio-temporales estacionarias con determinadas marginales espaciales. En concreto, genera funciones con marginales espaciales gaussianas, del tipo  $(1 + \gamma_e(\mathbf{h}))^{-1/2}$  o de la familia de Matérn, entre otras. A continuación vamos a describir a modo de ejemplo la construcción de funciones de covarianza espacio-temporales estacionarias con marginales espaciales de Matérn.

**Ejemplo 4.14** Sean  $\gamma_e(\mathbf{h})$  y  $\gamma_t(u)$  dos variogramas intrínsecamente estacionarios sobre  $\mathbb{R}^d$  y  $T$ , respectivamente. Consideremos la función

$$C(\mathbf{h}, u) = \left( \frac{\gamma_e(\mathbf{h})}{1 + \gamma_t(u)} \right)^{\nu/2} K_\nu \left( (\gamma_e(\mathbf{h})(1 + \gamma_t(u)))^{1/2} \right)$$

donde  $K_\nu$  es la función de Bessel modificada de segunda especie de orden  $\nu$ . Se puede comprobar que

$$C(\mathbf{h}, u) = 2^{\nu-1} \Gamma(\nu) \int_0^{+\infty} \exp\left(-\frac{\gamma_e(\mathbf{h})}{4\omega}\right) \exp(-\gamma_t(u)) d\mu(\omega),$$

donde  $\mu(\omega)$  es la función de distribución Gamma con densidad  $(1/\Gamma(\nu))\omega^{\nu-1} \exp(-\omega)$ ,  $\omega \in \mathbb{R}_+$ . Luego por el Teorema 4.5, esta mixtura de covarianzas separables será una función de covarianza espacio-temporal estacionaria sobre  $\mathbb{R}^d \times T$ . Su marginal temporal viene dada por  $C(\mathbf{0}, u) = 2^{\nu-1} \Gamma(\nu)(1 + \gamma_t(u))^{-\nu}$ , y su marginal espacial por  $C(\mathbf{h}, 0) = (\gamma_e(\mathbf{h}))^{\nu/2} K_\nu((\gamma_e(\mathbf{h}))^{1/2})$  (que se reduce al modelo de Matérn para  $\gamma_e(\mathbf{h}) = \alpha^2 \|\mathbf{h}\|^2$ ).

**Ejemplo 4.15** Si en (4.36) consideramos

$$C_t(u|\omega) = (\gamma_t(u) + 1 + (\gamma_t(u) + c)\|\omega\|^2)^{-\nu-d/2} (1 + c\|\omega\|^2)^{\nu+d/2},$$

$$d\mu(\omega) = (1 + c\|\omega\|^2)^{-\nu-d/2} d\omega,$$

entonces obtenemos otro modelo de funciones de covarianza espacio-temporales estacionarias con marginales espaciales pertenecientes a la clase de Matérn dada por

$$C(\mathbf{h}, u) = \frac{1}{(\gamma_t(u) + 1)^\nu (\gamma_t(u) + c)^{d/2}} \left( \frac{(\gamma_t(u) + 1)^{1/2}}{\gamma_t(u) + c} \|\mathbf{h}\| \right)^\nu K_\nu \left( \frac{(\gamma_t(u) + 1)^{1/2}}{\gamma_t(u) + c} \|\mathbf{h}\| \right) \tag{4.40}$$

con  $\nu$  y  $c$  dos constantes positivas. El modelo (4.20) de Cressie y Huang (1999), que hemos visto en la Sección 4.5, es un caso particular de (4.40) considerando  $\gamma_t(u) = u^2$ ,  $u \in \mathbb{R}$ .

## 4.9. Modelos de mixturas propuestos por Ma

Ma (2002, 2003b) ofrece un método muy simple basado en diferentes mixturas de funciones de covarianza puramente espaciales y temporales, que permite construir nuevas familias de modelos de covarianza espacio-temporales estacionarias y es válido tanto para tiempos discretos como continuos.

### 4.9.1. Generación mediante mixturas escalares

Sea  $C_e(\mathbf{h})$  una función de covarianza espacial estacionaria sobre  $\mathfrak{R}^d$  y  $C_t(u)$  una función de covarianza temporal estacionaria sobre  $\mathcal{T}$ . Es sencillo comprobar que, para cualesquiera  $a \in \mathfrak{R}$  y  $b \in \mathcal{T}$ , las funciones  $C_e(a\mathbf{h})$  y  $C_t(bu)$  son covarianzas estacionarias válidas sobre  $\mathfrak{R}^d$  y  $\mathcal{T}$ , respectivamente. Por tanto, la función generada por

$$C(\mathbf{h}, u) = \int_{\mathfrak{R} \times \mathcal{T}} C_e(a\mathbf{h})C_t(bu)d\mu(a, b), \quad (4.41)$$

es una función de covarianza espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ , donde  $\mu(a, b)$  es una medida finita no negativa sobre  $\mathfrak{R} \times \mathcal{T}$ . Veamos a continuación un sencillo ejemplo de este tipo de funciones de covarianza espacio-temporales estacionarias.

**Ejemplo 4.16** Sea  $(U, V)$  un vector aleatorio con distribución Bernoulli bivariada, y sean  $p_{ij} = \text{Prob}(U = i, V = j)$ , para  $i, j = 0, 1$ . Entonces por (4.41), la función

$$C(\mathbf{h}, u) = C_e(\mathbf{h})C_t(u)p_{11} + C_e(\mathbf{h})C_t(0)p_{10} + C_e(\mathbf{0})C_t(u)p_{01} + C_e(\mathbf{0})C_t(0)p_{00},$$

es una función de covarianza espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ . Se puede observar que este modelo está muy relacionado con el modelo suma-producto que vimos en la Sección 4.6.

En la Sección 4.14.2 veremos la generalización de este método para la generación de funciones de covarianza espacio-temporales no estacionarias.

### 4.9.2. Generación mediante mixturas de potencias positivas

Sea  $\rho_e(\mathbf{h})$  una función de correlación espacial estacionaria sobre  $\mathfrak{R}^d$  y  $\rho_t(u)$  una función de correlación temporal estacionaria sobre  $\mathcal{T}$ . Dado  $k \in \mathbb{N}$ ,  $\rho_e^k(\mathbf{h})$  y  $\rho_t^k(u)$  son también dos funciones de correlación estacionarias, la primera espacial sobre  $\mathfrak{R}^d$  y la segunda temporal sobre  $\mathcal{T}$ . Sea

$$\rho(\mathbf{h}, u) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \rho_e^i(\mathbf{h})\rho_t^j(u)p_{ij}, \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathcal{T}, \quad (4.42)$$

con  $\{p_{ij}, (i, j) \in \mathbb{N}^2\}$  constantes no negativas que cumplen que  $\sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} p_{ij} = 1$  (para que  $\rho(\mathbf{0}, 0) = 1$ ). Entonces es fácil comprobar que (4.42) es una función de correlación estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

Esta expresión (4.42) nos ofrece un método sencillo para generar funciones de correlación estacionarias a partir de una correlación estacionaria espacial, una correlación estacionaria temporal y una función de probabilidad discreta  $\{p_{ij}, (i, j) \in \mathbb{N}^2\}$  de un vector bivariado  $(U, V)$  de variables aleatorias no negativas.

Sea  $g(z_1, z_2) = E(z_1^U z_2^V)$ , con  $-1 \leq z_1, z_2 \leq 1$ , la función generatriz de probabilidad del vector aleatorio  $(U, V)$ , que será siempre absolutamente convergente en el rectángulo unidad  $[-1, 1]^2$ . Podemos entonces expresar (4.42) como

$$\rho(\mathbf{h}, u) = g(\rho_e(\mathbf{h}), \rho_t(u)), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathcal{T}. \quad (4.43)$$

Observemos que al considerar  $\{p_{ij}, (i, j) \in \mathbb{N}^2\}$ , la distribución de probabilidad discreta que tiene toda su masa confinada a la recta  $i = j$ , esto es,

$$p_{ij} = \begin{cases} p_i, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}$$

con  $\{p_i, i \in \mathbb{N}\}$  una distribución de probabilidad discreta, la función de correlación (4.42) quedaría

$$\rho(\mathbf{h}, \mathbf{u}) = \sum_{k=0}^{+\infty} (\rho_e(\mathbf{h})\rho_t(\mathbf{u}))^k p_k, \quad (\mathbf{h}, \mathbf{u}) \in \mathfrak{R}^d \times \mathcal{T}. \quad (4.44)$$

Veamos a continuación alguno de los ejemplos propuestos en Ma (2002) obtenidos a partir de mixturas de potencias positivas:

**Ejemplo 4.17** Sea  $g(z_1, z_2)$  la función generatriz de probabilidad asociada a la distribución de Poisson bivariada

$$g(z_1, z_2) = \exp(\theta_1(z_1 - 1) + \theta_2(z_2 - 1) + \theta_{12}(z_1 z_2 - 1)), \quad (z_1, z_2) \in [-1, 1]^2$$

donde  $\theta_1, \theta_2, \theta_{12}$  son constantes no negativas, con  $\theta_{12} \leq \min(\theta_1, \theta_2)$ . Entonces por (4.43) la función

$$\rho(\mathbf{h}, \mathbf{u}) = g(\rho_e(\mathbf{h}), \rho_t(\mathbf{u})) = \exp(\theta_1(\rho_e(\mathbf{h}) - 1) + \theta_2(\rho_t(\mathbf{u}) - 1) + \theta_{12}(\rho_e(\mathbf{h})\rho_t(\mathbf{u}) - 1))$$

es una función de correlación estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ , con  $\rho_e(\mathbf{h})$  y  $\rho_t(\mathbf{u})$  dos funciones de correlación estacionarias espacial y temporal, respectivamente.

**Ejemplo 4.18** Sea  $g(z)$  la función generatriz de probabilidad asociada a la distribución binomial negativa

$$g(z) = \left( \frac{\theta}{1 - (1 - \theta)z} \right)^\alpha,$$

con  $0 < \theta < 1$  y  $\alpha > 0$ . Entonces por (4.44) la función

$$\rho(\mathbf{h}, \mathbf{u}) = g(\rho_e(\mathbf{h})\rho_t(\mathbf{u})) = \left( \frac{\theta}{1 - (1 - \theta)\rho_e(\mathbf{h})\rho_t(\mathbf{u})} \right)^\alpha$$

es una función de correlación estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ , con  $\rho_e(\mathbf{h})$  y  $\rho_t(\mathbf{u})$  dos funciones de correlación estacionarias espacial y temporal, respectivamente.

En principio podríamos plantearnos la extensión natural de (4.42), pasando de una función de distribución de probabilidad discreta  $\{p_{ij}, (i, j) \in \mathbb{N}^2\}$  a una distribución de probabilidad continua sobre  $\mathfrak{R}_+^2$ . Esto es, considerar

$$\rho(\mathbf{h}, \mathbf{u}) = \int_0^{+\infty} \int_0^{+\infty} \rho_e^u(\mathbf{h})\rho_t^v(\mathbf{u})p(u, v)dudv, \quad (\mathbf{h}, \mathbf{u}) \in \mathfrak{R}^d \times \mathcal{T}, \quad (4.45)$$

con  $p(u, v)$  la función de densidad conjunta de un vector aleatorio  $(U, V)$  bivariado no negativo. No obstante, esta extensión presenta diferentes problemas de definición:

- Si  $\rho_e(\mathbf{h})$  o  $\rho_t(u)$  son negativos, entonces la función de correlación  $\rho(\mathbf{h}, u)$  no estaría bien definida, ya que tomaría valores complejos.
- En general,  $\rho_e^u(\mathbf{h})$  y  $\rho_t^v(u)$  pueden no ser una función de correlación espacial y temporal, respectivamente, para todo  $u > 0$  o  $v > 0$ . Por ejemplo,  $\rho_t(u) = \cos^2(u)$  es una función de correlación sobre  $\mathfrak{R}$ , pero en cambio para  $v = 1/2$ , la potencia  $\rho_t^{1/2}(u) = |\cos(u)|$  no es una función de correlación temporal.
- La expresión (4.45) puede expresarse como

$$\rho(\mathbf{h}, u) = \mathcal{L}(-\log(\rho_e(\mathbf{h})) - \log(\rho_t(u))), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathcal{T}, \quad (4.46)$$

siendo  $\mathcal{L}(\theta_1, \theta_2) = E[e^{-\theta_1 U - \theta_2 V}]$  la transformada de Laplace del vector aleatorio  $(U, V)$  bivariado no negativo. Al contrario que ocurría con la función generatriz de probabilidad de un vector aleatorio bivariado discreto no negativo, la transformada de Laplace de un vector aleatorio bivariado continuo no negativo no siempre toma valores finitos.

En Ma (2003b) se demuestra la validez de este resultado para  $\rho_e(\mathbf{h}) = \exp(-\gamma_e(\mathbf{h}))$  y  $\rho_t(u) = \exp(-\gamma_t(u))$ , con  $\gamma_e(\mathbf{h})$  un semivariograma estacionario espacial sobre  $\mathfrak{R}^d$  y  $\gamma_t(u)$  un semivariograma estacionario temporal sobre  $\mathcal{T}$ .

**Teorema 4.6** Sea  $\mathcal{L}(\theta_1, \theta_2)$  la transformada de Laplace de un vector aleatorio bivariado no negativo  $(U, V)$ . Sea  $\gamma_e(\mathbf{h})$  un semivariograma estacionario espacial sobre  $\mathfrak{R}^d$  y  $\gamma_t(u)$  un semivariograma estacionario temporal sobre  $\mathcal{T}$ . Entonces

$$\rho(\mathbf{h}, u) = \mathcal{L}(\gamma_e(\mathbf{h}), \gamma_t(u)) \quad (4.47)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

**Ejemplo 4.19** Sea

$$\mathcal{L}(\theta_1, \theta_2) = \frac{p_1 p_2 e^{-(\theta_1 + \theta_2)}}{1 - p_0 e^{-(\theta_1 + \theta_2)}} \frac{e^{-\theta_1} + e^{-\theta_2} - (1 + p_0)e^{-(\theta_1 + \theta_2)}}{(1 - (1 - p_1)e^{-\theta_1})(1 - (1 - p_2)e^{-\theta_2})}$$

la transformada de Laplace asociada a la distribución geométrica bivariada

$$P(\mathcal{X}_1 = x_1, \mathcal{X}_2 = x_2) = \begin{cases} p_1 p_2 p_0^{x_1 - 1} (p_0 + p_1)^{x_2 - x_1 - 1}, & \text{si } x_1 < x_2 \\ p_1 p_2 p_0^{x_2 - 1} (p_0 + p_2)^{x_1 - x_2 - 1}, & \text{si } x_2 < x_1 \end{cases}$$

para  $x_1, x_2 \in \mathbb{N}$ , con  $0 \leq p_0, p_1, p_2 \leq 1$  y  $p_0 + p_1 + p_2 = 1$ . Entonces la función

$$\rho(\mathbf{h}, u) = \frac{p_1 p_2 e^{-\gamma_e(\mathbf{h}) - \gamma_t(u)}}{1 - p_0 e^{-\gamma_e(\mathbf{h}) - \gamma_t(u)}} \frac{e^{-\gamma_e(\mathbf{h})} + e^{-\gamma_t(u)} - (1 + p_0)e^{-\gamma_e(\mathbf{h}) - \gamma_t(u)}}{(1 - (1 - p_1)e^{-\gamma_e(\mathbf{h})})(1 - (1 - p_2)e^{-\gamma_t(u)})}$$

dada por (4.47) será una función de correlación espacio-temporal sobre  $\mathfrak{R}^d \times \mathcal{T}$ , con  $\gamma_e(\mathbf{h})$  un semivariograma estacionario espacial sobre  $\mathfrak{R}^d$  y  $\gamma_t(u)$  un semivariograma estacionario temporal sobre  $\mathcal{T}$ .

En la Sección 4.14.2 veremos bajo que condiciones será posible la generalización de este método para la generación de funciones de covarianza espacio-temporales no estacionarias.

### 4.10. Modelos generados mediante combinaciones lineales propuestos por Ma

En Ma (2005a,b), el autor ofrece diferentes modelos de covarianza y de variograma espacio-temporales estacionarios generados mediante combinaciones lineales de funciones de covarianza y variogramas espacio-temporales. En la Sección 3.3 hemos visto que, tanto las funciones de covarianza como las funciones de correlación o los semivariogramas, formaban un cono convexo en  $\mathbb{R}^d \times T$ . Por tanto, dadas dos constantes no negativas  $a, b$  y dos funciones de covarianza espacio-temporales  $C_1((s_1, t_1), (s_2, t_2))$  y  $C_2((s_1, t_1), (s_2, t_2))$  sobre  $\mathbb{R}^d \times T$ , entonces  $aC_1((s_1, t_1), (s_2, t_2)) + bC_2((s_1, t_1), (s_2, t_2))$  es también una función de covarianza sobre  $\mathbb{R}^d \times T$ , y lo mismo ocurre con los semivariogramas o las funciones de correlación.

No obstante, estas combinaciones convexas, aunque aseguran la validez del modelo resultante, no permiten valores negativos u oscilaciones entre valores positivos y negativos, algo que es necesario para la modelización de determinados procesos naturales. Por esta razón vamos a plantear combinaciones lineales generales de la forma

$$C((s_1, t_1), (s_2, t_2)) = aC_1((s_1, t_1), (s_2, t_2)) + bC_2((s_1, t_1), (s_2, t_2)), \quad (4.48)$$

donde  $a, b$  serán dos constantes cualesquiera, sujetas únicamente a la restricción de que  $Var(Z(s, t)) = a\sigma_1^2 + b\sigma_2^2 \geq 0$ , con  $\sigma_1^2$  y  $\sigma_2^2$  las varianzas a priori de las funciones de covarianza  $C_1$  y  $C_2$ . Vamos a suponer el caso no trivial en que  $Var(Z(s, t)) > 0$ . Si definimos  $a' = a\sigma_1^2$  y  $b' = b\sigma_2^2$ , con  $a' + b' > 0$ , podemos escribir (4.48) en función de las funciones de correlación como

$$C((s_1, t_1), (s_2, t_2)) = a'\rho_1((s_1, t_1), (s_2, t_2)) + b'\rho_2((s_1, t_1), (s_2, t_2)).$$

Si dividimos esta expresión por  $Var(Z(s, t)) = a' + b' > 0$ , tenemos que

$$\rho((s_1, t_1), (s_2, t_2)) = \theta\rho_1((s_1, t_1), (s_2, t_2)) + (1 - \theta)\rho_2((s_1, t_1), (s_2, t_2)), \quad (4.49)$$

donde  $\theta = a'/(a' + b')$  será una constante cualquiera. Observemos que la validez del modelo (4.49) sobre  $\mathbb{R}^d \times T$  estará garantizada para  $\theta \in [0, 1]$ , al dar lugar a una combinación convexa. En Ma (2005a) el autor analiza si es posible aumentar este rango de variación de  $\theta$  en diferentes situaciones concretas bajo la condición de estacionariedad, lo que daría lugar a combinaciones con valores negativos con las ventajas prácticas que hemos comentado.

El siguiente teorema muestra un procedimiento para generar funciones de correlación espacio-temporales mediante combinaciones lineales de la forma (4.49) con una marginal temporal que tiene un decrecimiento potencial.

**Teorema 4.7** Sea  $d \in \mathbb{N}$  cualquiera y  $\varphi$  una función completamente monótona con  $\varphi(0) = 1$ , entonces

$$\rho(h, u) = \theta(1 + \alpha_1|u|)^{-d/2}\varphi\left(\frac{\|h\|^2}{1 + \alpha_1|u|}\right) + (1 - \theta)(1 + \alpha_2|u|)^{-d/2}\varphi\left(\frac{\|h\|^2}{1 + \alpha_2|u|}\right) \quad (4.50)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathfrak{R}$  si y sólo si

$$\theta \in \left[ \left(1 - \frac{\alpha_2}{\alpha_1}\right)^{-1}, \left(1 - \frac{\alpha_1}{\alpha_2}\right)^{-1} \right] \quad (4.51)$$

Asimismo

$$\rho(\mathbf{h}, u) = \theta(1 + \alpha_1 u^2)^{-d/2} \varphi\left(\frac{\|\mathbf{h}\|^2}{1 + \alpha_1 u^2}\right) + (1 - \theta)(1 + \alpha_2 u^2)^{-d/2} \varphi\left(\frac{\|\mathbf{h}\|^2}{1 + \alpha_2 u^2}\right) \quad (4.52)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathfrak{R}$  si y sólo si

$$\theta \in \left[ \left(1 - \left(\frac{\alpha_2}{\alpha_1}\right)^{1/2}\right)^{-1}, 1 \right]. \quad (4.53)$$

Nótese que tanto (4.50) como (4.52) son funciones de correlación espacio-temporales sobre  $\mathfrak{R}^d \times \mathfrak{R}$  bajo las condiciones (4.51) y (4.53), respectivamente, con independencia del valor de  $d$ , ya que las restricciones para  $\theta$  no dependen de la dimensión espacial.

Las marginales temporales de (4.50) y (4.52) son, respectivamente,  $\rho(\mathbf{0}, u) = \theta(1 + \alpha_1 |u|)^{-d/2} + (1 - \theta)(1 + \alpha_2 |u|)^{-d/2}$  y  $\rho(\mathbf{0}, u) = \theta(1 + \alpha_1 u^2)^{-d/2} + (1 - \theta)(1 + \alpha_2 u^2)^{-d/2}$ , mostrando ambas un decrecimiento potencial. La marginal espacial en ambos casos coincide, y viene dada por  $\rho(\mathbf{h}, 0) = \varphi(\|\mathbf{h}\|^2)$ , la cual es siempre no negativa para todo  $\mathbf{h} \in \mathfrak{R}^d$ . No obstante, tomará valores negativos cuando  $\theta$  sea igual al valor inferior del intervalo (4.51) y (4.53), por lo que (4.50) y (4.52) tomarán valores negativos en esos casos.

**Ejemplo 4.20** Consideremos la función  $\varphi(x) = (1 + x)^{-\beta}$ , con  $x \geq 0$  y  $\beta > 0$ , que corresponde a la tercera entrada de la Tabla 4.1 para  $c = 1$  y  $\gamma = 1$ . Por el Teorema 4.7, la función

$$\rho(\mathbf{h}, u) = \theta(1 + \alpha_1 |u|)^{-d/2} \left(1 + \frac{\|\mathbf{h}\|^2}{1 + \alpha_1 |u|}\right)^{-\beta} + (1 - \theta)(1 + \alpha_2 |u|)^{-d/2} \left(1 + \frac{\|\mathbf{h}\|^2}{1 + \alpha_2 |u|}\right)^{-\beta}$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathfrak{R}$  para  $\theta$  verificando (4.51).

También se ofrece otro teorema para obtener funciones de correlación cuya marginal espacial decaiga siguiendo una ley de potencia.

**Teorema 4.8** Sea  $d \in \mathbb{N}$  cualquiera y  $\varphi$  una función completamente monótona con  $\varphi(0) = 1$ , entonces

$$\rho(\mathbf{h}, u) = \theta(1 + \alpha_1 \|\mathbf{h}\|)^{-1/2} \varphi\left(\frac{|u|^2}{1 + \alpha_1 \|\mathbf{h}\|}\right) + (1 - \theta)(1 + \alpha_2 \|\mathbf{h}\|)^{-1/2} \varphi\left(\frac{|u|^2}{1 + \alpha_2 \|\mathbf{h}\|}\right) \quad (4.54)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathfrak{R}$  si y sólo si

$$\theta \in \left[ \left(1 - \left(\frac{\alpha_2}{\alpha_1}\right)^d\right)^{-1}, \left(1 - \frac{\alpha_1}{\alpha_2}\right)^{-1} \right]. \quad (4.55)$$

Asimismo

$$\rho(\mathbf{h}, u) = \theta(1 + \alpha_1 \|\mathbf{h}\|^2)^{-1/2} \varphi\left(\frac{u^2}{1 + \alpha_1 \|\mathbf{h}\|^2}\right) + (1 - \theta)(1 + \alpha_2 \|\mathbf{h}\|^2)^{-1/2} \varphi\left(\frac{u^2}{1 + \alpha_2 \|\mathbf{h}\|^2}\right) \quad (4.56)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathbb{R}^d \times \mathbb{R}$  si y sólo si

$$\theta \in \left[ \left(1 - \left(\frac{\alpha_2}{\alpha_1}\right)^{d/2}\right)^{-1}, 1 \right]. \quad (4.57)$$

En este caso, las condiciones de validez de (4.54) y (4.56) dadas por (4.55) y (4.57), respectivamente, dependen del valor de  $d$ , a menos que  $\theta$  tome valores entre 0 y  $(\alpha_2 - \alpha_1)/\alpha_2$ , en cuyo caso son válidos para cualquier valor  $d$ .

La marginal temporal en ambos casos coincide, y viene dada por  $\rho(\mathbf{h}, 0) = \varphi(u^2)$ . Las marginales espaciales de (4.54) y (4.56) decaen al aumentar la distancia, en ambos casos, según una ley potencial. Como antes, cuando  $\theta$  tome los valores dados por los límites inferiores de su rango de validez, las funciones resultantes podrán tomar valores negativos.

**Ejemplo 4.21** Consideremos la función completamente monótona  $\varphi(x) = \exp(-x^\beta)$ , con  $x \geq 0$  y  $0 < \beta \leq 1$ , que corresponde a la primera entrada de la Tabla 4.1 para  $c = 1$ . Por el Teorema 4.8, la función

$$\rho(\mathbf{h}, u) = \theta(1 + \alpha_1 \|\mathbf{h}\|)^{-1/2} \exp\left(-\frac{|u|^{2\beta}}{(1 + \alpha_1 \|\mathbf{h}\|)^\beta}\right) + (1 - \theta)(1 + \alpha_2 \|\mathbf{h}\|)^{-1/2} \exp\left(-\frac{|u|^{2\beta}}{(1 + \alpha_2 \|\mathbf{h}\|)^\beta}\right)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathbb{R}^d \times \mathbb{R}$  para  $\theta$  verificando (4.55). (4.55).

Ma (2005a) analiza el caso particular de combinaciones lineales de modelos de covarianza espacio-temporales separables, esto es, del tipo

$$\rho((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \theta \rho_{11}(\mathbf{s}_1, \mathbf{s}_2) \theta \rho_{11}(t_1, t_2) + (1 - \theta) \rho_{21}(\mathbf{s}_1, \mathbf{s}_2) \theta \rho_{22}(t_1, t_2), \quad (4.58)$$

con  $\theta$  una constante cualquiera no necesariamente restringida al intervalo  $[0, 1]$ . En concreto, estudia las combinaciones (4.58) para el caso particular en el que se consideran correlaciones espaciales y temporales estacionarias del modelo de Matérn que hemos analizado en la Sección 2.4.9.

Consideremos la función

$$\rho(\mathbf{h}, u) = \theta(\alpha_1 \|\mathbf{h}\|)^{\nu_1} K_{\nu_1}(\alpha_1 \|\mathbf{h}\|) (\beta_1 |u|)^{\nu_2} K_{\nu_2}(\beta_1 |u|) + (1 - \theta)(\alpha_2 \|\mathbf{h}\|)^{\nu_1} K_{\nu_1}(\alpha_2 \|\mathbf{h}\|) (\beta_2 |u|)^{\nu_2} K_{\nu_2}(\beta_2 |u|), \quad (4.59)$$

donde  $\alpha_1, \alpha_2, \beta_1, \beta_2, \nu_1, \nu_2$  son constantes positivas, con  $\alpha_1 \leq \alpha_2$  y  $\beta_1 \leq \beta_2$ , y  $K_\nu$  es la función de Bessel modificada de segunda especie de orden  $\nu$ . En Ma (2005a) se demuestra

que (4.59) es una función de correlación espacio-temporal sobre  $\mathfrak{R}^d \times \mathfrak{R}$  si y sólo si

$$\theta \in \left[ \left( 1 - \left( \frac{\alpha_2}{\alpha_1} \right)^d \frac{\beta_2}{\beta_1} \right)^{-1}, \left( 1 - \left( \frac{\alpha_1}{\alpha_2} \right)^{2\nu_1} \left( \frac{\beta_1}{\beta_2} \right)^{2\nu_2} \right)^{-1} \right], \quad (4.60)$$

donde dicho intervalo se considera como  $\mathfrak{R}$  cuando  $\alpha_1 = \alpha_2$  y  $\beta_1 = \beta_2$ . Tenemos que, mientras la cota inferior de  $\theta$  en (4.60) no depende de la dimensión  $d$  del espacio, la cota superior decrece cuando  $d$  aumenta y tiende a cero cuanto  $d$  tiende a infinito. Observar que si  $\alpha_1 = \alpha_2$  o  $\beta_1 = \beta_2$ , entonces el modelo (4.59) es un modelo espacio-temporal separable.

Para el caso exponencial ( $\nu_1 = \nu_2 = 1/2$ ) tendríamos que la función

$$\rho(\mathbf{h}u) = \theta \exp(-\alpha_1 \|\mathbf{h}\|) \exp(-\beta_1 |u|) + (1 - \theta) \exp(-\alpha_2 \|\mathbf{h}\|) \exp(-\beta_2 |u|), \quad (4.61)$$

donde  $\alpha_1, \alpha_2, \beta_1, \beta_2$  constantes positivas, con  $\alpha_1 \leq \alpha_2$  y  $\beta_1 \leq \beta_2$ , será una función de correlación espacio-temporal sobre  $\mathfrak{R}^d \times \mathfrak{R}$  si y sólo si

$$\theta \in \left[ \left( 1 - \left( \frac{\alpha_2}{\alpha_1} \right)^d \frac{\beta_2}{\beta_1} \right)^{-1}, \left( 1 - \left( \frac{\alpha_1}{\alpha_2} \right) \left( \frac{\beta_1}{\beta_2} \right) \right)^{-1} \right]. \quad (4.62)$$

A partir de estos resultados, Ma obtiene un caso especial de combinación lineal de las funciones obtenidas por el Teorema 4.6.

**Teorema 4.9** Sea  $\mathcal{L}(\theta_1, \theta_2)$  la transformada de Laplace de un vector aleatorio bivariado no negativo  $(U, V)$ . Entonces

$$\rho(\mathbf{h}, u) = \theta \mathcal{L}(\alpha_1 \|\mathbf{h}\|, \beta_1 |u|) + (1 - \theta) \mathcal{L}(\alpha_2 \|\mathbf{h}\|, \beta_2 |u|) \quad (4.63)$$

es una función de correlación espacio-temporal estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

#### 4.11. Modelos propuestos por Stein

Stein (2005a) muestra un procedimiento que funciona en el dominio espectral, y que permite generar una familia de densidades espectrales que dan como resultado funciones de covarianza espacio-temporales estacionarias en general no separables y con un comportamiento analítico flexible. Estas densidades extienden las del modelo de Matérn.

Las funciones de covarianza separables generalmente implican que pequeños cambios en las localizaciones de las observaciones pueden dar lugar a grandes cambios en las relaciones entre determinadas combinaciones lineales de las observaciones. La fuente de esta discontinuidad puede explicarse por el hecho de que las covarianzas así definidas no presentan más suavidad lejos del origen que en el origen. Stein muestra en sus trabajos que modelos de covarianza espacio-temporales con mayor suavidad lejos del origen dan lugar a predictores lineales óptimos con mejores propiedades. En este artículo, el autor repasa los diferentes modelos introducidos con anterioridad, evaluando las características que presentan lejos del origen y la verificación de esta propiedad.

Stein propone un nuevo modelo de función de covarianza espacio-temporal estacionaria que sí cumple la propiedad de ser más suave lejos del origen que en el origen. Estas covarianzas serán generadas a través de la función de densidad espectral

$$f(\omega, \tau) = (c_1 (a_1^2 + \|\omega\|^2)^{\alpha_1} + c_2 (a_2^2 + |\tau|^2)^{\alpha_2})^{-\nu}, \quad (4.64)$$

con  $c_1$  y  $c_2$  dos constantes positivas,  $a_1^2 + a_2^2 > 0$ ,  $\alpha_1$  y  $\alpha_2$  dos enteros positivos y  $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$ . La función de covarianza asociada a esta función de densidad espectral vendría dada por

$$C(\mathbf{h}, \mathbf{u}) = \int \int e^{i(\omega' \mathbf{h} + \tau u)} f(\omega, \tau) d\omega d\tau. \quad (4.65)$$

Aplicando uno de los resultados demostrados en el artículo anteriormente citado, se verifica que bajo estas condiciones la función de covarianza espacio-temporal dada por (4.65) posee todas las derivadas parciales de todos los órdenes para  $(\mathbf{h}, u) \neq (0, 0)$ . Además, depende únicamente de  $\|\mathbf{h}\|$  y de  $|u|$ . También presenta una propiedad práctica de gran utilidad, y es que podemos controlar el grado de suavidad espacial y el grado de suavidad temporal de forma independiente. Desafortunadamente, se dispone en muy limitados casos de expresiones explícitas para las funciones de covarianza (4.65) generadas mediante este procedimiento.

**Ejemplo 4.22** En el caso en que  $\alpha_1 = \alpha_2 = 1$ , podemos considerar sin pérdida de generalidad que  $a_2 = 0$ . Sea  $\beta = (c_1/c_2)^{1/2}$ , entonces puede verificarse que la covarianza (4.65) es proporcional a  $\mathcal{M}_{\nu-(d_1+d_2)/2}(a_1(\|\omega\|^2 + \beta^2|\tau|^2)^{1/2})$ , siendo  $\mathcal{M}_{\nu}(r) = r^{\nu} \mathcal{K}_{\nu}(r)$  y  $\mathcal{K}_{\nu}(r)$  la función de Bessel modificada de orden  $\nu$ . Por tanto, esta covarianza será proporcional al modelo de Matérn que analizamos en la Sección 2.4.9, que era una función de covarianza isotrópica para cualquier dimensión.

**Ejemplo 4.23** Si  $\alpha_1 \neq \alpha_2$ , con  $\alpha_1, \alpha_2, \nu$  enteros, en ocasiones es posible encontrar expresiones explícitas para la función de covarianza asociada. Por ejemplo, si consideramos  $d_1 = 3$  (dimensión espacial),  $d_2 = 1$  (dimensión temporal),  $\alpha_1 = 2, \alpha_2 = 1, a_2 = 0$  y  $\nu = 2 > 1$  (condición necesaria para asegurar la integrabilidad de  $f$ ), tenemos una función de densidad espectral de la forma  $f(\omega, \tau) = (c(a^2 + \|\omega\|^2)^2 + |\tau|^2)^{-2}$  cuya función de covarianza asociada viene dada por

$$\begin{aligned} C(\mathbf{h}, u) = & \frac{\pi^2}{16c^6} e^{a\|\mathbf{h}\|} \operatorname{erfc} \left( ca|u|^{1/2} + \frac{\|\mathbf{h}\|}{2c|u|^{1/2}} \right) \left( \frac{1}{a^3} - \frac{\|\mathbf{h}\|}{a^2} + \frac{4c^4 u^2}{\|\mathbf{h}\|} \right) + \\ & \frac{\pi^2}{16c^6} e^{-a\|\mathbf{h}\|} \operatorname{erfc} \left( ca|u|^{1/2} - \frac{\|\mathbf{h}\|}{2c|u|^{1/2}} \right) \left( \frac{1}{a^3} + \frac{\|\mathbf{h}\|}{a^2} - \frac{4c^4 u^2}{\|\mathbf{h}\|} \right) + \\ & \frac{\pi^{3/2}|u|^{1/2}}{4c^5 a^2} \exp \left( -c^2 a^2 |u| - \frac{\|\mathbf{h}\|^2}{4c^2 |u|} \right), \end{aligned}$$

siendo  $\operatorname{erfc}$  la función de error complementaria (Abramowitz y Stegun, 1965), y donde la función de covarianza en  $\mathbf{h} = \mathbf{0}$  o  $u = 0$  se define por continuidad, esto es,

$$C(\mathbf{h}, 0) = \frac{\pi^2}{8c^6 a^3} e^{-a\|\mathbf{h}\|} (1 + a\|\mathbf{h}\|),$$

$$C(\mathbf{0}, u) = \frac{\pi^2}{8c^6} \left( \frac{1}{a^3} + 4c^2|u|^2 \right) \operatorname{erfc}(ca|u|^{1/2}) + \frac{\pi^{3/2}e^{-c^2a^2|u|}}{c^6} \left( \frac{c|u|^{1/2}}{4a^2} - \frac{c^3|u|^{3/2}}{2} \right).$$

Este modelo es infinitamente diferenciable en cualquier punto distinto del origen, para  $u = 0$  únicamente es una vez diferenciable en cualquier dirección espacial y en cambio para  $\mathbf{h} = \mathbf{0}$  el proceso no es diferenciable en el tiempo.

#### 4.12. Modelo suma de productos generalizado

Sean  $\{C_{ei} : i = 1, 2, \dots, n\}$  y  $\{C_{ti} : i = 1, 2, \dots, n\}$  dos conjuntos de funciones de covarianza estacionarias, continuas e integrables espaciales y temporales, respectivamente, con  $n \in \mathbb{N}$ . Consideremos la función

$$C(\mathbf{h}, u) = \sum_{i=1}^n k_i C_{ei}(\mathbf{h}) C_{ti}(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \quad (4.66)$$

Evidentemente, por las propiedades de las funciones de covarianza que hemos visto en el Capítulo 3, la función (4.66) será una función de covarianza espacio-temporal válida sobre  $\mathbb{R}^d \times \mathbb{R}$  siempre que se obtenga mediante una combinación lineal convexa, esto es, que las constantes  $k_1, \dots, k_n$  que la definen sean no negativas. El objetivo que nos planteamos en este capítulo es analizar la posibilidad de definir funciones con esta misma estructura, que es intuitivamente sencilla y de fácil implementación, pero donde las constantes de la combinación lineal que la definen no estén sujetas a la restricción de ser no negativas sino que puedan tomar valores negativos. Si logramos este objetivo, tendremos funciones de covarianza que podrán tomar valores negativos, o valores que oscilen entre positivos y negativos. Como ya comentamos en la Introducción, esto tiene importantes implicaciones prácticas muy útiles para la modelización de determinados procesos naturales.

En la Sección 4.10 hemos visto una primera aproximación a este problema, donde se establecen un conjunto de condiciones necesarias y suficientes para el caso en que  $n = 2$  y las funciones de covarianza espaciales y temporales tienen una misma estructura y pertenecen al modelo de Matérn (o al exponencial, que es un caso particular del anterior). Vamos en esta sección a generalizar estos resultados cubriendo un espectro mucho mayor de posibilidades.

A continuación se presenta una condición necesaria y una condición suficiente para la validez de este modelo (4.66) bajo el caso general en que las constantes  $k_1, \dots, k_n$  sean valores reales cualesquiera.

**Teorema 4.10** Sean  $\{C_{ei} : i = 1, 2, \dots, n\}$  y  $\{C_{ti} : i = 1, 2, \dots, n\}$  dos conjuntos de  $n$  funciones de covarianza integrables y continuas espaciales y temporales, respectivamente, con  $n \in \mathbb{N}$ . Consideremos la función

$$C(\mathbf{h}, u) = \sum_{i=1}^n k_i C_{ei}(\mathbf{h}) C_{ti}(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (4.67)$$

con  $k_1, \dots, k_n \in \mathbb{R}$ . Sean  $f_{ei}$  y  $f_{ti}$ , con  $i = 1, 2, \dots, n$ , las transformadas de Fourier de las covarianzas  $C_{ei}$  and  $C_{ti}$ , respectivamente. Supongamos que al menos un par de

estas funciones  $(f_{ei}, f_{ti})$  no son nulas, y que denotaremos sin pérdida de generalidad como  $(f_{en}, f_{tn})$ . Sean

$$m_{ti} = \inf_{\tau \in \mathfrak{R}} \frac{f_{ti}(\tau)}{f_{tn}(\tau)}, \quad M_{ti} = \sup_{\tau \in \mathfrak{R}} \frac{f_{ti}(\tau)}{f_{tn}(\tau)}, \quad m_{ei} = \inf_{\omega \in \mathfrak{R}^d} \frac{f_{ei}(\omega)}{f_{en}(\omega)}, \quad M_{ei} = \sup_{\omega \in \mathfrak{R}^d} \frac{f_{ei}(\omega)}{f_{en}(\omega)}.$$

(i) Si (4.67) es una función de covarianza espacio-temporal, entonces

$$k_n \geq - \sum_{i=1}^{n-1} k_i [M_{ei} M_{ti} \mathbf{1}_{\{k_i \geq 0\}} + m_{ei} m_{ti} \mathbf{1}_{\{k_i < 0\}}].$$

(ii) Si

$$k_n \geq - \sum_{i=1}^{n-1} k_i [m_{ei} m_{ti} \mathbf{1}_{\{k_i \geq 0\}} + M_{ei} M_{ti} \mathbf{1}_{\{k_i < 0\}}],$$

entonces la función (4.67) es una función de covarianza espacio-temporal.

**Demostración.-**

(i) Dada una función  $C(\mathbf{h}, u)$  continua definida sobre  $\mathfrak{R}^d \times \mathfrak{R}$  con función de densidad espectral  $f(\omega, \tau)$ , entonces hemos visto que como consecuencia del teorema de Bochner es equivalente que  $C(\mathbf{h}, u)$  sea una función de covarianza espacio-temporal a que  $f(\omega, \tau) \geq 0$ , para todo  $(\omega, \tau) \in \mathfrak{R}^d \times \mathfrak{R}$ .

En este caso, la función de densidad espectral asociada a la función  $C(\mathbf{h}, u)$  definida por (4.67) viene dada por

$$f(\omega, \tau) = \sum_{i=1}^n k_i f_{ei}(\omega) f_{ti}(\tau), \quad (\omega, \tau) \in \mathfrak{R}^d \times \mathfrak{R}.$$

Luego por hipótesis

$$0 \leq f(\omega, \tau) = \sum_{i=1}^n k_i f_{ei}(\omega) f_{ti}(\tau) = f_{en}(\omega) f_{tn}(\tau) \left( \sum_{i=1}^{n-1} k_i \frac{f_{ei}(\omega)}{f_{en}(\omega)} \frac{f_{ti}(\tau)}{f_{tn}(\tau)} + k_n \right).$$

Como  $C_{en}(\mathbf{h})$  y  $C_{tn}(u)$  son dos funciones de covarianza sobre  $\mathfrak{R}^d$  y  $\mathfrak{R}$ , respectivamente, entonces se tendrá que  $f_{en}(\omega) \geq 0, \forall \omega \in \mathfrak{R}^d$  y  $f_{tn}(\tau) \geq 0, \forall \tau \in \mathfrak{R}$ . Por tanto,

$$0 \leq \sum_{i=1}^{n-1} k_i \frac{f_{ei}(\omega)}{f_{en}(\omega)} \frac{f_{ti}(\tau)}{f_{tn}(\tau)} + k_n = \sum_{\substack{i=1 \\ k_i \geq 0}}^{n-1} k_i \frac{f_{ei}(\omega)}{f_{en}(\omega)} \frac{f_{ti}(\tau)}{f_{tn}(\tau)} + \sum_{\substack{i=1 \\ k_i < 0}}^{n-1} k_i \frac{f_{ei}(\omega)}{f_{en}(\omega)} \frac{f_{ti}(\tau)}{f_{tn}(\tau)} + k_n.$$

Por definición,

$$0 \leq m_{ei} \leq \frac{f_{ei}(\omega)}{f_{en}(\omega)} \leq M_{ei}, \quad 0 \leq m_{ti} \leq \frac{f_{ti}(\tau)}{f_{tn}(\tau)} \leq M_{ti}, \quad \forall (\omega, \tau) \in \mathfrak{R}^d \times \mathfrak{R}.$$

Luego

$$0 \leq m_{ei}m_{ti} \leq \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} \leq M_{ei}M_{ti}, \forall (\omega, \tau) \in \mathbb{R}^d \times \mathbb{R}. \quad (4.68)$$

Por tanto

$$\begin{aligned} 0 \leq \sum_{\substack{i=1 \\ k_i \geq 0}}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + \sum_{\substack{i=1 \\ k_i < 0}}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + k_n &\leq \sum_{\substack{i=1 \\ k_i \geq 0}}^{n-1} k_i M_{ei}M_{ti} + \sum_{\substack{i=1 \\ k_i < 0}}^{n-1} k_i m_{ei}m_{ti} + k_n \\ \Rightarrow k_n &\geq - \sum_{i=1}^{n-1} k_i [M_{ei}M_{ti} \mathbf{1}_{\{k_i \geq 0\}} + m_{ei}m_{ti} \mathbf{1}_{\{k_i < 0\}}]. \end{aligned}$$

(ii) Supongamos que  $k_n \geq - \sum_{i=1}^{n-1} k_i [m_{ei}m_{ti} \mathbf{1}_{\{k_i \geq 0\}} + M_{ei}M_{ti} \mathbf{1}_{\{k_i < 0\}}]$ , entonces

$$k_n + \sum_{i=1}^{n-1} k_i [m_{ei}m_{ti} \mathbf{1}_{\{k_i \geq 0\}} + M_{ei}M_{ti} \mathbf{1}_{\{k_i < 0\}}] \geq 0. \quad (4.69)$$

Vamos a demostrar que la función  $C(h, u)$  definida por (4.67) es una función de covarianza espacio-temporal sobre  $\mathbb{R}^d \times \mathbb{R}$ , o lo que es lo mismo, que su función de densidad espectral asociada

$$f(\omega, \tau) = \sum_{i=1}^n k_i f_{ei}(\omega) f_{ti}(\tau) = f_{en}(\omega) f_{tn}(\tau) \left( \sum_{i=1}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + k_n \right) \geq 0,$$

para todo  $(\omega, \tau) \in \mathbb{R}^d \times \mathbb{R}$ . Como  $f_{en}(\omega) \geq 0$  y  $f_{tn}(\tau) \geq 0$  por ser las funciones de densidad espectral asociadas a dos funciones de covarianza, entonces tendremos que probar que

$$\sum_{i=1}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + k_n = \sum_{\substack{i=1 \\ k_i \geq 0}}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + \sum_{\substack{i=1 \\ k_i < 0}}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + k_n \geq 0.$$

Por (4.68) se tendrá que

$$\begin{aligned} \sum_{i=1}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + k_n &= \sum_{\substack{i=1 \\ k_i \geq 0}}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + \sum_{\substack{i=1 \\ k_i < 0}}^{n-1} k_i \frac{f_{ei}(\omega) f_{ti}(\tau)}{f_{en}(\omega) f_{tn}(\tau)} + k_n \\ &\geq \sum_{\substack{i=1 \\ k_i \geq 0}}^{n-1} k_i m_{ei}m_{ti} + \sum_{\substack{i=1 \\ k_i < 0}}^{n-1} k_i M_{ei}M_{ti} + k_n, \end{aligned}$$

que será no negativo por (4.69).

A continuación vamos a centrarnos en el caso  $n = 2$ , es decir, aquellos modelos generados como combinación lineal de dos modelos de covarianza espacio-temporal separables.

**Teorema 4.11** Sean  $C_{e1}$  y  $C_{e2}$  dos funciones de covarianza espaciales definidas sobre  $\mathbb{R}^d$ , y sean  $C_{t1}$  y  $C_{t2}$  dos funciones de covarianza temporales definidas sobre  $\mathbb{R}$ , todas ellas integrables y continuas. Consideremos la función

$$C(\mathbf{h}, u) = k_1 C_{e1}(\mathbf{h}) C_{t2}(u) + k_2 C_{e2}(\mathbf{h}) C_{t1}(u) \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (4.70)$$

con  $k_1, k_2 \in \mathbb{R}$ . Sean  $f_{e1}, f_{e2}, f_{t1}$  y  $f_{t2}$  las transformadas de Fourier de las covarianzas  $C_{e1}, C_{e2}, C_{t1}$  y  $C_{t2}$ , respectivamente, donde supondremos que  $f_{e2}$  y  $f_{t2}$  no son funciones nulas. Sean

$$m_{t1} = \inf_{\tau \in \mathbb{R}} \frac{f_{t1}(\tau)}{f_{t2}(\tau)}, \quad M_{t1} = \sup_{\tau \in \mathbb{R}} \frac{f_{t1}(\tau)}{f_{t2}(\tau)}, \quad m_{e1} = \inf_{\mathbf{w} \in \mathbb{R}^d} \frac{f_{e1}(\mathbf{w})}{f_{e2}(\mathbf{w})}, \quad M_{e1} = \sup_{\mathbf{w} \in \mathbb{R}^d} \frac{f_{e1}(\mathbf{w})}{f_{e2}(\mathbf{w})}.$$

Entonces (4.70) es una función de covarianza espacio-temporal si y sólo si

$$k_2 \geq -k_1 [m_{e1} m_{t1} \mathbf{1}_{\{k_1 \geq 0\}} + M_{e1} M_{t1} \mathbf{1}_{\{k_1 < 0\}}].$$

**Demostración.-** Supongamos que (4.70) es una función de covarianza espacio-temporal. Su función de densidad espectral asociada será no negativa, esto es,

$$0 \leq f(\omega, \tau) = k_1 f_{e1}(\omega) f_{t1}(\tau) + k_2 f_{e2}(\omega) f_{t2}(\tau) = f_{e2}(\omega) f_{t2}(\tau) \left( k_1 \frac{f_{e1}(\omega) f_{t1}(\tau)}{f_{e2}(\omega) f_{t2}(\tau)} + k_2 \right).$$

Como  $C_{e2}$  y  $C_{t2}$  son dos funciones de covarianza, sus densidades espectrales asociadas  $f_{e1}$  y  $f_{e2}$  serán no negativas, por lo que

$$0 \leq k_1 \frac{f_{e1}(\omega) f_{t1}(\tau)}{f_{e2}(\omega) f_{t2}(\tau)} + k_2 \Rightarrow k_2 \geq -k_1 \frac{f_{e1}(\omega) f_{t1}(\tau)}{f_{e2}(\omega) f_{t2}(\tau)}, \quad \forall (\omega, \tau) \in \mathbb{R}^d \times \mathbb{R}.$$

Si  $k_1 \geq 0$ , tenemos que

$$-k_2 \leq k_1 \frac{f_{e1}(\omega) f_{t1}(\tau)}{f_{e2}(\omega) f_{t2}(\tau)}, \quad \forall (\omega, \tau) \in \mathbb{R}^d \times \mathbb{R}.$$

En particular, esta desigualdad se mantendrá para los ínfimos de ambos cocientes, por lo que  $-k_2 \leq k_1 m_e m_t$ , o lo que es lo mismo,  $k_2 \geq -k_1 m_e m_t$ .

Si  $k_1 < 0$ , tenemos que

$$k_2 \geq -k_1 \frac{f_{e1}(\omega) f_{t1}(\tau)}{f_{e2}(\omega) f_{t2}(\tau)}, \quad \forall (\omega, \tau) \in \mathbb{R}^d \times \mathbb{R}.$$

En particular, esta desigualdad se mantendrá para los supremos de ambos cocientes, por lo que  $k_2 \geq -k_1 M_e M_t$ .

Por tanto, se cumplirá que

$$k_2 \geq -k_1 [m_{e1} m_{t1} \mathbf{1}_{\{k_1 \geq 0\}} + M_{e1} M_{t1} \mathbf{1}_{\{k_1 < 0\}}].$$

El recíproco es inmediato, como consecuencia del apartado (ii) del Teorema 4.10.

Vamos ahora a considerar sumas ponderadas de dos funciones de covarianza espacio-temporales separables de forma que los coeficientes que definen la combinación lineal sumen 1. Bajo esta reparametrización, el modelo que consideraremos en adelante será

$$C(\mathbf{h}, u) = \theta C_{s1}(\mathbf{h})C_{t1}(u) + (1 - \theta)C_{s2}(\mathbf{h})C_{t2}(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

con  $\theta \in \mathbb{R}$ , siendo  $C_{e1}$  y  $C_{e2}$  dos funciones de covarianza espaciales definidas sobre  $\mathbb{R}^d$ , y  $C_{t1}$  y  $C_{t2}$  dos funciones de covarianza temporales definidas sobre  $\mathbb{R}$ , todas ellas integrables y continuas.

**Corolario 4.1** Sean  $C_{e1}$  y  $C_{e2}$  dos funciones de covarianza espaciales definidas sobre  $\mathbb{R}^d$ , y sean  $C_{t1}$  y  $C_{t2}$  dos funciones de covarianza temporales definidas sobre  $\mathbb{R}$ , todas ellas integrables y continuas. Consideremos la función

$$C(\mathbf{h}, u) = \theta C_{e1}(\mathbf{h})C_{t2}(u) + (1 - \theta)C_{e2}(\mathbf{h})C_{t2}(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \quad (4.71)$$

con  $\theta \in \mathbb{R}$ . Sean  $f_{e1}$ ,  $f_{e2}$ ,  $f_{t1}$  y  $f_{t2}$  las transformadas de Fourier de las covarianzas  $C_{e1}$ ,  $C_{e2}$ ,  $C_{t1}$  y  $C_{t2}$ , respectivamente, donde supondremos que  $f_{e2}$  y  $f_{t2}$  no son funciones nulas. Sean

$$m_t = \inf_{\tau \in \mathbb{R}} \frac{f_{t1}(\tau)}{f_{t2}(\tau)}, \quad M_t = \sup_{\tau \in \mathbb{R}} \frac{f_{t1}(\tau)}{f_{t2}(\tau)}, \quad m_e = \inf_{\mathbf{w} \in \mathbb{R}^d} \frac{f_{e1}(\mathbf{w})}{f_{e2}(\mathbf{w})}, \quad M_e = \sup_{\mathbf{w} \in \mathbb{R}^d} \frac{f_{e1}(\mathbf{w})}{f_{e2}(\mathbf{w})}.$$

Entonces (4.71) es una función de covarianza espacio-temporal si y sólo si

$$[1 - \max(1, M_e M_t)]^{-1} \leq \theta \leq [1 - \min(1, m_e m_t)]^{-1}, \quad (4.72)$$

donde  $0^{-1} = -\infty$  y  $(-\infty)^{-1} = 0$  a la izquierda de la desigualdad, y  $0^{-1} = +\infty$  a la derecha de la misma.

**Demostración.-** Consideremos  $k_1 = \theta$  y  $k_2 = 1 - \theta$ . Por el Teorema 4.11, (4.71) es una función de covarianza espacio-temporal si y sólo si

$$1 - \theta \geq -\theta [m_e m_t \mathbf{1}_{\{\theta \geq 0\}} + M_e M_t \mathbf{1}_{\{\theta < 0\}}]. \quad (4.73)$$

Veamos que esta expresión es equivalente a la condición (4.72) del enunciado del teorema.

o Supongamos que  $M_e M_t \geq m_e m_t \geq 1$  y que  $\theta \geq 0$ . Si (4.71) es una función de covarianza espacio-temporal, aplicando (4.73) tenemos que

$$1 - \theta \geq -\theta m_e m_t \geq -\theta M_e M_t \Rightarrow 1 \geq \theta(1 - M_e M_t) \Rightarrow (1 - M_e M_t)^{-1} \leq \theta \leq +\infty,$$

por lo que se cumple la condición (4.72).

Recíprocamente, si se cumple la condición (4.72), entonces  $(1 - M_e M_t)^{-1} \leq \theta \leq +\infty$ . Luego  $\theta \geq (1 - M_e M_t)^{-1} \geq (1 - m_e m_t)^{-1}$ , ya que  $1 - M_e M_t \leq 1 - m_e m_t \leq 0$ . Por tanto, se cumple la condición (4.73), por lo que (4.71) será una función de covarianza espacio-temporal.

o Supongamos que  $M_e M_t \geq m_e m_t \geq 1$  y que  $\theta < 0$ . En ese caso es inmediata la equivalencia de ambas expresiones, ya que

$$1 - \theta \geq -\theta M_e M_t \Leftrightarrow 1 \geq \theta(1 - M_e M_t) \Leftrightarrow (1 - M_e M_t)^{-1} \leq \theta \leq +\infty.$$

o Supongamos que  $M_e M_t \geq 1 \geq m_e m_t \geq 0$  y que  $\theta \geq 0$ . Si (4.71) es una función de covarianza espacio-temporal, aplicando (4.73) tenemos que  $1 - \theta \geq -\theta m_e m_t$ , donde  $-\theta m_e m_t \geq -\theta M_e M_t$ . Luego

$$\begin{aligned} 1 - \theta &\geq -\theta m_e m_t \Rightarrow 1 \geq \theta(1 - m_e m_t) \Rightarrow (1 - m_e m_t)^{-1} \geq \theta, \\ 1 - \theta &\geq -\theta M_e M_t \Rightarrow 1 \geq \theta(1 - M_e M_t) \Rightarrow (1 - M_e M_t)^{-1} \leq \theta, \end{aligned}$$

y por tanto se cumple la condición (4.72).

Recíprocamente, si se cumple la condición (4.72), entonces  $(1 - M_e M_t)^{-1} \leq \theta \leq (1 - m_e m_t)^{-1}$ . Luego  $\theta \leq (1 - m_e m_t)^{-1} \Rightarrow 1 - \theta \geq -\theta m_e m_t$ . Por tanto, se cumple la condición (4.73), y (4.71) será una función de covarianza espacio-temporal.

o Supongamos que  $M_e M_t \geq 1 \geq m_e m_t \geq 0$  y que  $\theta < 0$ . Si (4.71) es una función de covarianza espacio-temporal, aplicando (4.73) tenemos que  $1 - \theta \geq -\theta M_e M_t$ , donde  $-\theta M_e M_t \geq -\theta m_e m_t$ . Luego

$$\begin{aligned} 1 - \theta &\geq -\theta m_e m_t \Rightarrow 1 \geq \theta(1 - m_e m_t) \Rightarrow (1 - m_e m_t)^{-1} \geq \theta, \\ 1 - \theta &\geq -\theta M_e M_t \Rightarrow 1 \geq \theta(1 - M_e M_t) \Rightarrow (1 - M_e M_t)^{-1} \leq \theta, \end{aligned}$$

y por tanto se cumple la condición (4.72).

Recíprocamente, si se cumple la condición (4.72), entonces  $(1 - M_e M_t)^{-1} \leq \theta \leq (1 - m_e m_t)^{-1}$ . Luego  $\theta \geq (1 - M_e M_t)^{-1} \Rightarrow 1 - \theta \geq -\theta M_e M_t$ . Por tanto, se cumple la condición (4.73), y (4.71) será una función de covarianza espacio-temporal.

o Supongamos que  $1 \geq M_e M_t \geq m_e m_t \geq 0$  y que  $\theta \geq 0$ . En ese caso es inmediata la equivalencia de ambas expresiones, ya que

$$1 - \theta \geq -\theta m_e m_t \Leftrightarrow 1 \geq \theta(1 - m_e m_t) \Leftrightarrow (1 - m_e m_t)^{-1} \geq \theta \geq -\infty.$$

o Supongamos que  $1 \geq M_e M_t \geq m_e m_t \geq 0$  y que  $\theta < 0$ . Si (4.71) es una función de covarianza espacio-temporal, aplicando (4.73) tenemos que

$$1 - \theta \geq -\theta M_e M_t \geq -\theta m_e m_t \Rightarrow 1 \geq \theta(1 - m_e m_t) \Rightarrow (1 - m_e m_t)^{-1} \geq \theta \leq +\infty,$$

por lo que se cumple la condición (4.72).

Recíprocamente, si se cumple la condición (4.72), entonces  $-\infty \leq \theta \leq (1 - m_e m_t)^{-1}$ . Luego  $\theta \leq (1 - m_e m_t)^{-1} \leq (1 - M_e M_t)^{-1}$ , ya que  $0 \leq 1 - M_e M_t \leq 1 - m_e m_t \leq 1$ . Por tanto, se cumple la condición (4.73), por lo que (4.71) será una función de covarianza espacio-temporal.

Obsérvese que este resultado permite a su vez definir funciones de covarianza puramente espaciales mediante combinación lineal (no necesariamente convexa) de funciones de covarianza puramente espaciales, como se muestra a continuación, y cuya demostración se obtiene fácilmente aplicando todos los resultados anteriores al caso meramente espacial.

**Corolario 4.2** Sean  $C_{e1}$  y  $C_{e2}$  dos funciones de covarianza espaciales definidas sobre  $\mathbb{R}^d$  integrables y continuas. Consideremos la función

$$C(\mathbf{h}) = \theta C_{e1}(\mathbf{h}) + (1 - \theta)C_{e2}(\mathbf{h}), \quad \mathbf{h} \in \mathbb{R}^d, \quad (4.74)$$

con  $\theta \in \mathbb{R}$ . Sean  $f_{e1}$  y  $f_{e2}$  las transformadas de Fourier asociadas a las covarianzas  $C_{e1}$  y  $C_{e2}$ , respectivamente, donde supondremos que  $f_{e2}$  no es una función nula. Sean

$$m = \inf_{\boldsymbol{\omega} \in \mathbb{R}^d} \frac{f_{e1}(\boldsymbol{\omega})}{f_{e2}(\boldsymbol{\omega})}, \quad M = \sup_{\boldsymbol{\omega} \in \mathbb{R}^d} \frac{f_{e1}(\boldsymbol{\omega})}{f_{e2}(\boldsymbol{\omega})}.$$

Entonces (4.74) es una función de covarianza espacial si y sólo si

$$[1 - \text{máx}(1, M)]^{-1} \leq \theta \leq [1 - \text{mín}(1, m)]^{-1}, \quad (4.75)$$

donde  $0^{-1} = -\infty$  y  $(-\infty)^{-1} = 0$  a la izquierda de la desigualdad, y  $0^{-1} = +\infty$  a la derecha de la misma.

A continuación vamos a calcular las cotas que definen el rango de variación posible de  $\theta$  que permite la generación de funciones de covarianza espacio-temporales válidas del tipo (4.71) bajo determinados modelos. En particular se muestran los casos en los que las covarianzas consideradas pertenezcan al modelo gaussiano o al modelo de Matérn, para las que su función de densidad espectral tiene una expresión analítica.

**Teorema 4.12** Sean  $G1(\mathbf{r}|\sigma_{g1}^2, \alpha_{g1})$  y  $G2(\mathbf{r}|\sigma_{g2}^2, \alpha_{g2})$  dos funciones de covarianzas pertenecientes al modelo gaussiano sobre  $\mathbb{R}^k$ . Sean  $f_{G1}(\boldsymbol{\omega})$  y  $f_{G2}(\boldsymbol{\omega})$  sus funciones de densidad espectrales asociadas, y sean

$$m_{G1, G2} = \inf_{\boldsymbol{\omega} \in \mathbb{R}^k} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})}, \quad M_{G1, G2} = \sup_{\boldsymbol{\omega} \in \mathbb{R}^k} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})},$$

(i) Si  $\alpha_{g1} < \alpha_{g2}$ , entonces

$$m_{G1, G2} = 0, \quad M_{G1, G2} = \frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left( \frac{\alpha_{g2}}{\alpha_{g1}} \right)^{k/2}.$$

(ii) Si  $\alpha_{g1} = \alpha_{g2}$ , entonces

$$m = \frac{\sigma_{g1}^2}{\sigma_{g2}^2}, \quad M = \frac{\sigma_{g1}^2}{\sigma_{g2}^2}.$$

(iii) Si  $\alpha_{g1} > \alpha_{g2}$ , entonces

$$m = \frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left( \frac{\alpha_{g2}}{\alpha_{g1}} \right)^{k/2}, \quad M = +\infty.$$

**Demostración.-** Como hemos visto en la Sección 2.4.5, la función de covarianza asociada al modelo gaussiano venía dada por

$$G(\mathbf{r}|\sigma_g^2, \alpha_g) = \sigma_g^2 \exp(-\alpha_g \|\mathbf{r}\|^2), \quad \mathbf{r} \in \mathfrak{R}^k,$$

donde  $\sigma_g^2 > 0$  es la varianza del proceso y  $\alpha_g$  es un parámetro de escala de forma que  $\alpha_g^{-1} > 0$  indica la velocidad con la que el proceso decae. La densidad espectral asociada a este modelo tiene la misma forma, y viene dada por

$$f_G(\boldsymbol{\omega}) = \sigma_g^2 \pi^{k/2} \alpha_g^{-k/2} \exp(-\|\boldsymbol{\omega}\|^2/4\alpha_g), \quad \boldsymbol{\omega} \in \mathfrak{R}^k.$$

Luego dadas dos funciones de covarianza pertenecientes a este modelo, el cociente de sus funciones de densidades espectrales vendrá dado por

$$\begin{aligned} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} &= \frac{\sigma_{g1}^2 \pi^{k/2} \alpha_{g1}^{-k/2} \exp(-\|\boldsymbol{\omega}\|^2/4\alpha_{g1})}{\sigma_{g2}^2 \pi^{k/2} \alpha_{g2}^{-k/2} \exp(-\|\boldsymbol{\omega}\|^2/4\alpha_{g2})} \\ &= \frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left(\frac{\alpha_{g1}}{\alpha_{g2}}\right)^{-k/2} \exp\left(-\frac{1}{4} \left(\frac{1}{\alpha_{g1}} - \frac{1}{\alpha_{g2}}\right) \|\boldsymbol{\omega}\|^2\right) = h(\|\boldsymbol{\omega}\|^2), \end{aligned}$$

donde  $h(x) = a \exp(-bx)$ , con  $x \geq 0$ , es una función continua de una variable definida en  $[0, +\infty[$  cuyo comportamiento depende del signo de  $b$ , siendo

$$a = \frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left(\frac{\alpha_{g1}}{\alpha_{g2}}\right)^{-k/2}, \quad b = \frac{1}{4} \left(\frac{1}{\alpha_{g1}} - \frac{1}{\alpha_{g2}}\right).$$

(i) Si  $\alpha_{g1} < \alpha_{g2}$ , entonces  $b > 0$ , por lo que la función  $h(x)$  será decreciente. En ese caso su supremo coincidirá con su máximo, que se alcanzará en  $x = \|\boldsymbol{\omega}\|^2 = 0$  y valdrá  $h(0) = a$ , mientras que su ínfimo se alcanzará en el límite y valdrá  $\lim_{x \rightarrow +\infty} h(x) = 0$ . Por tanto,

$$m = \inf_{\boldsymbol{\omega} \in \mathfrak{R}^k} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} = 0, \quad M = \sup_{\boldsymbol{\omega} \in \mathfrak{R}^k} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} = \frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left(\frac{\alpha_{g2}}{\alpha_{g1}}\right)^{k/2}.$$

(ii) Si  $\alpha_{g1} = \alpha_{g2}$ , entonces  $b = 0$ , con lo que la función  $h(x)$  será constante e igual a  $a$  en todo su dominio. Luego

$$m = \inf_{\boldsymbol{\omega} \in \mathfrak{R}^d} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} = \frac{\sigma_{g1}^2}{\sigma_{g2}^2}, \quad M = \sup_{\boldsymbol{\omega} \in \mathfrak{R}^d} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} = \frac{\sigma_{g1}^2}{\sigma_{g2}^2}.$$

(iii) Si  $\alpha_{g1} > \alpha_{g2}$ , entonces  $b < 0$ , con lo que la función  $h(x)$  será creciente. Su ínfimo coincidirá con su mínimo, que se alcanzará en  $x = \|\boldsymbol{\omega}\|^2 = 0$  y valdrá  $h(0) = a$ , mientras que su supremo se alcanzará en el límite y será  $\lim_{x \rightarrow +\infty} h(x) = +\infty$ . Luego en ese caso

$$m = \inf_{\boldsymbol{\omega} \in \mathfrak{R}^k} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} = \frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left(\frac{\alpha_{g2}}{\alpha_{g1}}\right)^{k/2}, \quad M = \sup_{\boldsymbol{\omega} \in \mathfrak{R}^k} \frac{f_{G1}(\boldsymbol{\omega})}{f_{G2}(\boldsymbol{\omega})} = +\infty.$$

**Teorema 4.13** Sean  $M1(\mathbf{r}|\sigma_{m1}^2, \alpha_{m1}, \nu_{m1})$  y  $M2(\mathbf{r}|\sigma_{m2}^2, \alpha_{m2}, \nu_{m2})$  dos funciones de covarianzas pertenecientes al modelo de Matérn sobre  $\mathfrak{R}^k$ . Sean  $f_{M1}(\omega)$  y  $f_{M2}(\omega)$  sus funciones de densidad espectrales asociadas, y sean

$$m_{M1,M2} = \inf_{\omega \in \mathfrak{R}^k} \frac{f_{M1}(\omega)}{f_{M2}(\omega)}, \quad M_{M1,M2} = \sup_{\omega \in \mathfrak{R}^k} \frac{f_{M1}(\omega)}{f_{M2}(\omega)},$$

(i) Si  $\nu_{m1} = \nu_{m2}$  y  $\alpha_{m1} = \alpha_{m2}$ , entonces

$$m_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2}, \quad M_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2}.$$

(ii) Si  $\nu_{m1} = \nu_{m2}$  y  $\alpha_{m1} > \alpha_{m2}$ , entonces

$$m_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k, \quad M_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m1}}{\alpha_{m2}} \right)^{2\nu_{m1}}.$$

(iii) Si  $\nu_{m1} = \nu_{m2}$  y  $\alpha_{m1} < \alpha_{m2}$ , entonces

$$m_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m1}}{\alpha_{m2}} \right)^{2\nu_{m1}}, \quad M_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k.$$

(iv) Si  $\nu_{m1} > \nu_{m2}$  y  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} \geq \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces

$$m_{M1,M2} = 0, \quad M_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1} - \nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k.$$

(v) Si  $\nu_{m1} > \nu_{m2}$  y  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} < \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces

$$m_{M1,M2} = 0, \\ M_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left( \frac{\alpha_{m2}^2 - \alpha_{m1}^2}{2(\nu_{m2} - \nu_{m1})} \right)^{\nu_{m2} - \nu_{m1}} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}}.$$

(vi) Si  $\nu_{m1} < \nu_{m2}$  y  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} \geq \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces

$$m_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left( \frac{\alpha_{m2}^2 - \alpha_{m1}^2}{2(\nu_{m2} - \nu_{m1})} \right)^{\nu_{m2} - \nu_{m1}} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}}, \\ M_{M1,M2} = +\infty.$$

(vii) Si  $\nu_{m1} < \nu_{m2}$  y  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} < \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces

$$m_{M1,M2} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1} - \nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k, \\ M_{M1,M2} = +\infty.$$

**Demostración.-** Como hemos visto en la Sección 2.4.9, la función de covarianza asociada al modelo de Matérn viene dada por

$$M(\mathbf{r}|\sigma_m^2, \alpha_m, \nu_m) = \sigma_m^2 (\alpha_m \|\mathbf{r}\|)^{\nu_m} \mathcal{K}_{\nu_m}(\alpha_m \|\mathbf{r}\|), \quad \mathbf{r} \in \mathfrak{R}^k,$$

donde  $\sigma_m^2 > 0$  es la varianza del proceso,  $\alpha_m$  es un parámetro de escala de forma que  $\alpha_m^{-1} > 0$  indica la velocidad con la que el proceso decae y  $\nu_m$  mide el grado de suavidad del proceso asociado. La densidad espectral asociada a este modelo viene dada por

$$f_M(\omega) = \sigma_m^2 2^{\nu_m-1} \pi^{-k/2} \Gamma\left(\nu_m + \frac{k}{2}\right) \alpha_m^{2\nu_m} (\alpha_m^2 + \|\omega\|^2)^{-\nu_m - \frac{k}{2}}, \quad \omega \in \mathfrak{R}^k.$$

Luego dadas dos funciones de covarianza pertenecientes a este modelo, el cociente de sus funciones de densidades espectrales vendrá dado por

$$\begin{aligned} \frac{f_{M1}(\omega)}{f_{M2}(\omega)} &= \frac{\sigma_{m1}^2 2^{\nu_{m1}-1} \pi^{-k/2} \Gamma\left(\nu_{m1} + \frac{k}{2}\right) \alpha_{m1}^{2\nu_{m1}} (\alpha_{m1}^2 + \|\omega\|^2)^{-\nu_{m1} - \frac{k}{2}}}{\sigma_{m2}^2 2^{\nu_{m2}-1} \pi^{-k/2} \Gamma\left(\nu_{m2} + \frac{k}{2}\right) \alpha_{m2}^{2\nu_{m2}} (\alpha_{m2}^2 + \|\omega\|^2)^{-\nu_{m2} - \frac{k}{2}}} \\ &= \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1}-\nu_{m2}} \cdot \frac{\Gamma\left(\nu_{m1} + \frac{k}{2}\right)}{\Gamma\left(\nu_{m2} + \frac{k}{2}\right)} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\alpha_{m2}^2 + \|\omega\|^2)^{\nu_{m2} + \frac{k}{2}}}{(\alpha_{m1}^2 + \|\omega\|^2)^{\nu_{m1} + \frac{k}{2}}} = h(\|\omega\|^2), \end{aligned}$$

donde  $h(x) = a(b+x)^e/(c+x)^f$ ,  $x \geq 0$ , es una función continua de una variable definida sobre  $[0, +\infty[$ , siendo

$$a = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1}-\nu_{m2}} \cdot \frac{\Gamma\left(\nu_{m1} + \frac{k}{2}\right)}{\Gamma\left(\nu_{m2} + \frac{k}{2}\right)} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}}, \quad b = \alpha_{m2}^2, \quad c = \alpha_{m1}^2, \quad e = \nu_{m2} + \frac{k}{2}, \quad f = \nu_{m1} + \frac{k}{2}.$$

El supremo e ínfimo de la función  $h(x)$  se alcanzará en sus extremos (en  $x = 0$  o en su límite cuando  $x \rightarrow +\infty$ ) o en sus puntos singulares (que anulen su primera derivada). Su derivada viene dada por

$$h(x) = a \frac{(b+x)^e}{(c+x)^f} \Rightarrow h'(x) = a \frac{(b+x)^{e-1}}{(c+x)^{f+1}} ((e-f)x + (ec-fb)),$$

que tendrá un punto singular en  $[0, +\infty[$  si y sólo si  $e-f \neq 0$ , esto es, si y sólo si  $\nu_{m1} \neq \nu_{m2}$ .

(i-iii) Si  $\nu_{m1} = \nu_{m2}$ , entonces  $e-f = 0$ , por lo que la función  $h(x)$  alcanzará su supremo e ínfimo en sus extremos, que tendremos que evaluar.

$$h(0) = a \left(\frac{b}{c}\right)^e, \quad \lim_{x \rightarrow +\infty} h(x) = \lim_{x \rightarrow +\infty} a \left(\frac{b+x}{c+x}\right)^e = a.$$

Luego el ínfimo y supremo de la función  $h(x)$  dependen del cociente  $b/c$ .

(i) Si  $\alpha_{m1} = \alpha_{m2}$ , entonces  $b = c$ , con lo que la función  $h(x) = a$  en todo su dominio. En este caso

$$m = \inf_{\omega \in \mathfrak{R}^k} \frac{f_{M1}(\omega)}{f_{M2}(\omega)} = a = \frac{\sigma_{m1}^2}{\sigma_{m2}^2}, \quad M = \sup_{\omega \in \mathfrak{R}^k} \frac{f_{G1}(\omega)}{f_{G2}(\omega)} = a = \frac{\sigma_{m1}^2}{\sigma_{m2}^2}.$$

(ii) Si  $\alpha_{m1} > \alpha_{m2}$ , entonces  $b < c$ , con lo que  $b/c < 1$ . En este caso

$$m = \inf_{\omega \in \mathbb{R}^k} \frac{f_{M1}(\omega)}{f_{M2}(\omega)} = a \left(\frac{b}{c}\right)^e = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left(\frac{\alpha_{m1}}{\alpha_{m2}}\right)^{2\nu_{m1}} \left(\frac{\alpha_{m2}}{\alpha_{m1}}\right)^{2\nu_{m1}+k} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left(\frac{\alpha_{m2}}{\alpha_{m1}}\right)^k,$$

$$M = \sup_{\omega \in \mathbb{R}^k} \frac{f_{G1}(\omega)}{f_{G2}(\omega)} = a = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left(\frac{\alpha_{m1}}{\alpha_{m2}}\right)^{2\nu_{m1}}.$$

(iii) Si  $\alpha_{m1} < \alpha_{m2}$ , entonces  $b > c$ , con lo que  $b/c > 1$ . En este caso

$$m = \inf_{\omega \in \mathbb{R}^k} \frac{f_{M1}(\omega)}{f_{M2}(\omega)} = a = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left(\frac{\alpha_{m1}}{\alpha_{m2}}\right)^{2\nu_{m1}},$$

$$M = \sup_{\omega \in \mathbb{R}^k} \frac{f_{G1}(\omega)}{f_{G2}(\omega)} = a \left(\frac{b}{c}\right)^e = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left(\frac{\alpha_{m2}}{\alpha_{m1}}\right)^k.$$

(iv-vii) Si  $\nu_{m1} \neq \nu_{m2}$ , entonces la función  $h(x)$  tiene un punto singular en  $x^* = (fb - ec)/(e - f)$ . En este punto, la función  $h(x)$  toma el valor

$$h(x^*) = a \frac{(b+x^*)^e}{(c+x^*)^f} = a \left(\frac{b-c}{e-f}\right)^{e-f} \frac{e^e}{f^f}.$$

Luego el ínfimo y el supremo de la función  $h(x)$  se alcanzará en este punto o en sus dos extremos

$$h(0) = a \frac{b^e}{c^f}, \quad \lim_{x \rightarrow +\infty} h(x) = \lim_{x \rightarrow +\infty} a \frac{(b+x)^e}{(c+x)^f}.$$

(iv-v) Si  $\nu_{m1} > \nu_{m2}$ , entonces  $f > e$ , por lo que

$$\lim_{x \rightarrow +\infty} h(x) = \lim_{x \rightarrow +\infty} a \frac{(b+x)^e}{(c+x)^f} = 0.$$

Luego en este caso  $m = 0$ . Veamos ahora donde se alcanza el supremo, si en  $h(0)$  o en  $h(x^*)$ . Para ello vamos a calcular el cociente entre ambas cantidades.

$$\frac{h(x^*)}{h(0)} = \frac{a \left(\frac{b-c}{e-f}\right)^{e-f} \frac{e^e}{f^f}}{a \frac{b^e}{c^f}} = \left(\frac{b-c}{e-f}\right)^{e-f} \left(\frac{e}{b}\right)^e \left(\frac{c}{f}\right)^f.$$

(iv) Si  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} \geq \frac{\nu_{m2}+k/2}{\nu_{m1}+k/2}$ , entonces  $c/f \leq b/e$ , por lo que

$$\frac{h(x^*)}{h(0)} = \left(\frac{b-c}{e-f}\right)^{e-f} \left(\frac{e}{b}\right)^e \left(\frac{c}{f}\right)^f \leq \left(\frac{b-c}{e-f}\right)^{e-f} \left(\frac{e}{b}\right)^e \left(\frac{b}{e}\right)^f = \left(\frac{e(b-c)}{b(e-f)}\right)^{e-f} \leq 1.$$

Luego en este caso  $h(0) \geq h(x^*)$ , por lo que el supremo se alcanzará en  $h(0)$ , y valdrá

$$M = a \frac{b^e}{c^f} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1}-\nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \frac{\alpha_{m1}^{2\nu_{m1}} \alpha_{m2}^{2\nu_{m2}+k}}{\alpha_{m2}^{2\nu_{m2}} \alpha_{m1}^{2\nu_{m1}+k}}$$

$$= \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1}-\nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left(\frac{\alpha_{m2}}{\alpha_{m1}}\right)^k.$$

(v) Si  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} < \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces  $c/f > b/e$ , por lo que

$$\frac{h(x^*)}{h(0)} = \left(\frac{b-c}{e-f}\right)^{e-f} \left(\frac{e}{b}\right)^e \left(\frac{c}{f}\right)^f \geq \left(\frac{e(b-c)}{b(e-f)}\right)^{e-f} \geq 1.$$

Luego en este caso  $h(0) \leq h(x^*)$ , por lo que el supremo se alcanzará en  $h(x^*)$ , y valdrá

$$\begin{aligned} M &= a \left(\frac{b-c}{e-f}\right)^{e-f} \frac{e^e}{f^f} = \\ &= \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1} - \nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \left(\frac{\alpha_{m2}^2 - \alpha_{m1}^2}{\nu_{m2} - \nu_{m1}}\right)^{\nu_{m2} - \nu_{m1}} \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}} \\ &= \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left(\frac{\alpha_{m2}^2 - \alpha_{m1}^2}{2(\nu_{m2} - \nu_{m1})}\right)^{\nu_{m2} - \nu_{m1}} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}}. \end{aligned}$$

(vi-vii) Si  $\nu_{m1} < \nu_{m2}$ , entonces  $f < e$ , por lo que

$$\lim_{x \rightarrow +\infty} h(x) = \lim_{x \rightarrow +\infty} a \frac{(b+x)^e}{(c+x)^f} = +\infty.$$

Luego en este caso el supremo es  $M = +\infty$ . Veamos ahora si el ínfimo se alcanza en  $h(0)$  o en  $h(x^*)$ , para lo que calcularemos el cociente de ambas cantidades.

$$\frac{h(x^*)}{h(0)} = \left(\frac{b-c}{e-f}\right)^{e-f} \left(\frac{e}{b}\right)^e \left(\frac{c}{f}\right)^f.$$

(vi) Si  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} \geq \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces procediendo como antes se llega a que

$$m = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left(\frac{\alpha_{m2}^2 - \alpha_{m1}^2}{2(\nu_{m2} - \nu_{m1})}\right)^{\nu_{m2} - \nu_{m1}} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}}.$$

(vii) Si  $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} < \frac{\nu_{m2} + k/2}{\nu_{m1} + k/2}$ , entonces siguiendo el razonamiento anterior llegamos a que

$$m = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1} - \nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{d}{2})} \cdot \left(\frac{\alpha_{m2}}{\alpha_{m1}}\right)^k.$$

**Teorema 4.14** Sea  $G(r|\sigma_g^2, \alpha_g)$  una función de covarianza gaussiana y  $M(r|\sigma_m^2, \alpha_m, \nu_m)$  una función de covarianza perteneciente al modelo de Matérn, ambas sobre  $\mathbb{R}^k$ . Sean  $f_G(\omega)$  y  $f_M(\omega)$  sus funciones de densidad espectrales asociadas, y sean

$$m_{G,M} = \inf_{\omega \in \mathbb{R}^k} \frac{f_G(\omega)}{f_M(\omega)}, \quad M_{G,M} = \sup_{\omega \in \mathbb{R}^k} \frac{f_G(\omega)}{f_M(\omega)},$$

(i) Si  $\frac{\alpha_m^2}{\nu_m + \frac{k}{2}} > 4\alpha_g$ , entonces

$$m_{G,M} = 0,$$

$$M_{G,M} = \frac{\sigma_g^2}{\sigma_m^2} \cdot 2^{k+1} \cdot \pi^k \left( \frac{2\alpha_g}{\alpha_m^2} \right)^{\nu_m} \frac{(\nu_m + \frac{k}{2})^{\nu_m + \frac{k}{2}}}{\Gamma(\nu_m + \frac{k}{2})} \exp\left( \frac{\alpha_m^2}{4\alpha_g} - \left( \nu_m + \frac{k}{2} \right) \right).$$

(ii) Si  $\frac{\alpha_m^2}{\nu_m + \frac{k}{2}} \leq 4\alpha_g$ , entonces

$$m_{G,M} = 0,$$

$$M_{G,M} = \frac{\sigma_g^2}{\sigma_m^2} \cdot \frac{2^{1-\nu_m}}{\Gamma(\nu_m + \frac{k}{2})} \cdot \left( \frac{\pi^2 \alpha_m^2}{\alpha_g} \right)^{k/2}.$$

**Demostración.-** En ese caso, el cociente de sus funciones de densidades espectrales vendrá dado por

$$\frac{f_G(\omega)}{f_M(\omega)} = \frac{\sigma_g^2 \pi^{k/2} \alpha_g^{-k/2} \exp(-\|\omega\|^2 / 4\alpha_g)}{\sigma_m^2 2^{\nu_m-1} \pi^{-k/2} \Gamma(\nu_m + \frac{k}{2}) \alpha_m^{2\nu_m} (\alpha_m^2 + \|\omega\|^2)^{-\nu_m - \frac{k}{2}}}$$

$$= \frac{\sigma_g^2}{\sigma_m^2} \cdot \pi^k \cdot \frac{2^{1-\nu_m}}{\Gamma(\nu_m + \frac{k}{2})} \cdot \frac{\alpha_g^{-k/2}}{\alpha_m^{2\nu_m}} \cdot \exp\left(-\frac{1}{4\alpha_g} \|\omega\|^2\right) (\alpha_m^2 + \|\omega\|^2)^{\nu_m + \frac{k}{2}} = h(\|\omega\|^2),$$

donde  $h(x) = a \exp(-bx)(c+x)^e$ ,  $x \geq 0$ , que es una función continua de una variable definida sobre  $[0, +\infty[$ , siendo

$$a = \frac{\sigma_g^2}{\sigma_m^2} \pi^k \cdot \frac{2^{1-\nu_m}}{\Gamma(\nu_m + \frac{k}{2})} \cdot \frac{\alpha_g^{-k/2}}{\alpha_m^{2\nu_m}}, \quad b = \frac{1}{4\alpha_g}, \quad c = \alpha_m^2, \quad e = \nu_m + \frac{k}{2}.$$

El supremo e ínfimo de la función  $h(x)$  se alcanzará en sus extremos (en  $x = 0$  o en su límite cuando  $x \rightarrow +\infty$ ) o en sus puntos singulares (que anulen su primera derivada). En este caso

$$\lim_{x \rightarrow +\infty} h(x) = 0,$$

por lo que el ínfimo del cociente anterior será  $m = 0$ . Vamos ahora a buscar el supremo, que se alcanzará en  $h(0)$  o en sus puntos singulares. En este caso,  $h'(x) = a(c+x)^{e-1} \exp(-bx)(-b(c+x) + e)$ , que se anula en  $x^* = (e/b) - c$ , siendo el valor de la función en este punto es  $h(x^*) = a \exp(bc - e)(e/b)^e$ . Por tanto, tenemos que comparar los valores de  $h(0)$  y  $h(x^*)$  y buscar el mayor. El cociente de ambas cantidades viene definido por

$$\frac{h(x^*)}{h(0)} = \frac{a \exp(bc - e)(e/b)^e}{a c^e} = \exp(bc - e) \left( \frac{e}{bc} \right)^e. \quad (4.76)$$

(i) Supongamos que  $\frac{\alpha_m^2}{\nu_m + \frac{k}{2}} > 4\alpha_g$ , entonces  $bc > e$ , con lo que el cociente (4.76) es siempre mayor o igual a 1. Por tanto,  $h(x^*) \geq h(0)$ , por lo que

$$\begin{aligned} M_{G,M} &= a \exp(bc - e)(e/b)^e \\ &= \frac{\sigma_g^2}{\sigma_m^2} \cdot 2^{k+1} \cdot \pi^k \left( \frac{2\alpha_g}{\alpha_m^2} \right)^{\nu_m} \frac{(\nu_m + \frac{k}{2})^{\nu_m + \frac{k}{2}}}{\Gamma(\nu_m + \frac{k}{2})} \exp\left( \frac{\alpha_m^2}{4\alpha_g} - \left( \nu_m + \frac{k}{2} \right) \right). \end{aligned}$$

(ii) Supongamos ahora que  $\frac{\alpha_m^2}{\nu_m + \frac{k}{2}} \leq 4\alpha_g$ , entonces  $bc \leq e$ , con lo que el cociente (4.76) será siempre menor o igual a 1. En ese caso,  $h(x^*) \leq h(0)$ , por lo que

$$M_{G,M} = ac^e = \frac{\sigma_g^2}{\sigma_m^2} \cdot \frac{2^{1-\nu_m}}{\Gamma(\nu_m + \frac{k}{2})} \cdot \left( \frac{\pi^2 \alpha_m^2}{\alpha_g} \right)^{k/2}.$$

La Tabla 4.3 resume los valores de las constantes obtenidas en los Teoremas 4.12, 4.13 y 4.14. Permite, por tanto, obtener el rango de variación de la constante  $\theta$  que da lugar a funciones de covarianza espacio-temporales del modelo suma de productos generalizado (4.71). En concreto, a partir de la forma de las funciones de covarianza espaciales  $C_{e1}$  y  $C_{e2}$  consideradas en el modelo (ambas gaussianas, ambas de Matérn o una de cada modelo), esta tabla ofrece los valores  $M_e$  y  $m_e$  correspondientes. Procediendo de forma análoga para la parte temporal, obtendremos los valores  $M_t$  y  $m_t$  correspondientes. A partir de estas cuatro constantes, obtendremos el rango de variación de  $\theta$  utilizando la expresión (4.72). La primera parte de la Tabla 4.3 muestra los valores de  $m_e$  y  $M_e$  en el caso en que las covarianzas espaciales de la expresión (4.71) sean ambas gaussianas (o equivalentemente, los valores de  $m_t$  y  $M_t$  en el caso en que las covarianzas temporales sean gaussianas), en función de los parámetros de los modelos considerados. Del mismo modo, la segunda parte muestra estas mismas constantes en el caso en que ambas covarianzas espaciales (o temporales) pertenezcan al modelo de Matérn. Por último, la tercera parte de la tabla ofrece los valores correspondientes al caso en que la primera covarianza espacial de (4.71) sea gaussiana y la segunda sea de Matérn (y lo mismo para la parte temporal). Los valores correspondientes al caso opuesto en que la primera covarianza espacial de (4.71) sea de Matérn y la segunda gaussiana se obtienen de la tercera parte de la tabla, tomando  $m_e$  y  $M_e$  la inversa de los valores  $M_e$  y  $m_e$  correspondientes, respectivamente. Esta tabla nos permite obtener las constantes correspondientes al modelo (4.71) cuando se consideran covarianzas espaciales o temporales pertenecientes al modelo exponencial, por ser este un caso particular del modelo de Matérn con  $\nu = 1/2$ . Como hemos dicho, esta table permite tanto el cálculo de  $M_e$  y  $m_e$  correspondientes a las covarianzas espaciales como  $M_t$  y  $m_t$  correspondientes a las temporales, siendo  $k$  en cada caso la dimensión del espacio asociado ( $d$  en el caso espacial y 1 en el caso temporal). A continuación se muestra un ejemplo que muestra la aplicación de esta Tabla 4.3 para un caso práctico.

**Ejemplo 4.24** *Vamos a describir a continuación un ejemplo de este modelo suma de productos generalizado (4.71) para  $n = 2$  definido a partir de dos modelos de covarianza espaciales de Matérn con un mismo grado de suavización  $\nu$  y una misma varianza  $a$*

Gaussiana <sub>1</sub> / Gaussiana <sub>2</sub>		
Parámetros	$m_{G_1, G_2}$	$M_{G_1, G_2}$
$\alpha_{g1} < \alpha_{g2}$	0	$\frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left( \frac{\alpha_{g2}}{\alpha_{g1}} \right)^{k/2}$
$\alpha_{g1} = \alpha_{g2}$	$\frac{\sigma_{g1}^2}{\sigma_{g2}^2}$	$\frac{\sigma_{g1}^2}{\sigma_{g2}^2}$
$\alpha_{g1} > \alpha_{g2}$	$\frac{\sigma_{g1}^2}{\sigma_{g2}^2} \left( \frac{\alpha_{g2}}{\alpha_{g1}} \right)^{k/2}$	$+\infty$
Matérn <sub>1</sub> / Matérn <sub>2</sub>		
Parámetros	$m_{M_1, M_2}$	$M_{M_1, M_2}$
$\nu_{m1} = \nu_{m2}$ y $\alpha_{m1} \geq \alpha_{m2}$	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k$	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m1}}{\alpha_{m2}} \right)^{2\nu_1}$
$\nu_{m1} = \nu_{m2}$ y $\alpha_{m1} < \alpha_{m2}$	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m1}}{\alpha_{m2}} \right)^{2\nu_1}$	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k$
$\nu_{m1} > \nu_{m2}$ y $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} \geq \frac{\nu_{m2} + \frac{k}{2}}{\nu_{m1} + \frac{k}{2}}$	0	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1} - \nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{k}{2})} \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k$
$\nu_{m1} > \nu_{m2}$ y $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} < \frac{\nu_{m2} + \frac{k}{2}}{\nu_{m1} + \frac{k}{2}}$	0	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{k}{2})} \cdot \left( \frac{\alpha_{m2}^2 - \alpha_{m1}^2}{2(\nu_{m2} - \nu_{m1})} \right)^{\nu_{m2} - \nu_{m1}} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}}$
$\nu_{m1} < \nu_{m2}$ y $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} \geq \frac{\nu_{m2} + \frac{k}{2}}{\nu_{m1} + \frac{k}{2}}$	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{k}{2})} \cdot \left( \frac{\alpha_{m2}^2 - \alpha_{m1}^2}{2(\nu_{m2} - \nu_{m1})} \right)^{\nu_{m2} - \nu_{m1}} \cdot \frac{\alpha_{m1}^{2\nu_{m1}}}{\alpha_{m2}^{2\nu_{m2}}} \cdot \frac{(\nu_{m2} + \frac{k}{2})^{\nu_{m2} + \frac{k}{2}}}{(\nu_{m1} + \frac{k}{2})^{\nu_{m1} + \frac{k}{2}}}$	$+\infty$
$\nu_{m1} < \nu_{m2}$ y $\frac{\alpha_{m2}^2}{\alpha_{m1}^2} < \frac{\nu_{m2} + \frac{k}{2}}{\nu_{m1} + \frac{k}{2}}$	$\frac{\sigma_{m1}^2}{\sigma_{m2}^2} \cdot 2^{\nu_{m1} - \nu_{m2}} \cdot \frac{\Gamma(\nu_{m1} + \frac{k}{2})}{\Gamma(\nu_{m2} + \frac{k}{2})} \cdot \left( \frac{\alpha_{m2}}{\alpha_{m1}} \right)^k$	$+\infty$
Gaussiana / Matérn		
Parámetros	$m_{G, M}$	$M_{G, M}$
$\frac{\alpha_m^2}{\nu_m + \frac{k}{2}} > 4\alpha_g$	0	$\frac{\sigma_g^2}{\sigma_m^2} \cdot \frac{2^{1-\nu_m}}{\Gamma(\nu_m + \frac{k}{2})} \cdot \left( \frac{\pi^2 \alpha_m^2}{\alpha_g} \right)^{k/2}$
$\frac{\alpha_m^2}{\nu_m + \frac{k}{2}} \leq 4\alpha_g$	0	$\frac{\sigma_g^2}{\sigma_m^2} \cdot \frac{2^{1-\nu_m}}{\Gamma(\nu_m + \frac{k}{2})} \cdot \left( \frac{\pi^2 \alpha_m^2}{\alpha_g} \right)^{k/2}$

Tabla 4.3: Valores de  $m_t$ ,  $M_t$ ,  $m_e$  y  $M_e$  para  $n = 2$  en el caso en que se han combinado funciones de covarianza gaussianas y de Matérn en función de sus parámetros.

priori  $\sigma_e^2$ , y dos modelos de covarianza temporales exponenciales (que sabemos son casos particulares del modelo de Matérn para  $\nu = 1/2$ ) con una misma varianza a priori  $\sigma_t^2$ . El modelo considerado será

$$C(\mathbf{h}, u) = \theta C_{e1}(\mathbf{h})C_{t1}(u) + (1 - \theta)C_{e2}(\mathbf{h})C_{t2}(u), \quad (\mathbf{h}, u) \in \mathfrak{R}^d \times \mathfrak{R}, \quad (4.77)$$

con

$$\begin{aligned} C_{e1}(\mathbf{h}) &= M(\mathbf{h}|\sigma_e^2, \beta_1, \nu) = \sigma_e^2(2^{1-\nu}/\Gamma(\nu))(\beta_1\|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\beta_1\|\mathbf{h}\|), \\ C_{e2}(\mathbf{h}) &= M(\mathbf{h}|\sigma_e^2, \beta_2, \nu) = \sigma_e^2(2^{1-\nu}/\Gamma(\nu))(\beta_2\|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\beta_2\|\mathbf{h}\|), \\ C_{t1}(u) &= M(\mathbf{h}|\sigma_t^2, \alpha_1, \nu = 1/2) = \sigma_t^2 \exp(-\alpha_1|u|), \\ C_{t2}(u) &= M(\mathbf{h}|\sigma_t^2, \alpha_2, \nu = 1/2) = \sigma_t^2 \exp(-\alpha_2|u|). \end{aligned}$$

Por ejemplo, de la Tabla 4.3 se tiene que, para  $\beta_1 \geq \beta_2$  y  $\alpha_1 \geq \alpha_2$ ,

$$m_e = \left(\frac{\beta_2}{\beta_1}\right)^d, \quad M_e = \left(\frac{\beta_1}{\beta_2}\right)^{2\nu}, \quad m_t = \frac{\alpha_2}{\alpha_1}, \quad M_t = \frac{\alpha_1}{\alpha_2},$$

Luego

$$\begin{aligned} \text{máx}(1, M_e M_t) &= \text{máx}\left(1, \left(\frac{\beta_1}{\beta_2}\right)^{2\nu} \frac{\alpha_1}{\alpha_2}\right) = \left(\frac{\beta_1}{\beta_2}\right)^{2\nu} \frac{\alpha_1}{\alpha_2} \\ \text{mín}(1, m_e m_t) &= \text{mín}\left(1, \left(\frac{\beta_2}{\beta_1}\right)^d \frac{\alpha_2}{\alpha_1}\right) = \left(\frac{\beta_2}{\beta_1}\right)^d \frac{\alpha_2}{\alpha_1} \end{aligned}$$

Por tanto, en este caso la función (4.77) será un modelo de covarianza espacio-temporal si y sólo si

$$\left(1 - \left(\frac{\beta_1}{\beta_2}\right)^{2\nu} \frac{\alpha_1}{\alpha_2}\right)^{-1} \leq \theta \leq \left(1 - \left(\frac{\beta_2}{\beta_1}\right)^d \frac{\alpha_2}{\alpha_1}\right)^{-1}$$

Del mismo modo, se tiene que:

- Si  $\beta_1 \geq \beta_2$  y  $\alpha_1 < \alpha_2$ , entonces la función (4.77) es un modelo de covarianza espacio-temporal si y sólo si

$$\left(1 - \left(\frac{\beta_2}{\beta_1}\right)^d \frac{\alpha_2}{\alpha_1}\right)^{-1} \leq \theta \leq \left(1 - \left(\frac{\beta_1}{\beta_2}\right)^{2\nu} \frac{\alpha_1}{\alpha_2}\right)^{-1} \quad (4.78)$$

- Si  $\beta_1 < \beta_2$  y  $\alpha_1 \geq \alpha_2$ , entonces la función (4.77) es un modelo de covarianza espacio-temporal si y sólo si

$$\left(1 - \left(\frac{\beta_2}{\beta_1}\right)^d \frac{\alpha_1}{\alpha_2}\right)^{-1} \leq \theta \leq \left(1 - \left(\frac{\beta_1}{\beta_2}\right)^{2\nu} \frac{\alpha_2}{\alpha_1}\right)^{-1}$$

- Si  $\beta_1 < \beta_2$  y  $\alpha_1 < \alpha_2$ , entonces la función (4.77) es un modelo de covarianza espacio-temporal si y sólo si

$$\left(1 - \left(\frac{\beta_2}{\beta_1}\right)^d \frac{\alpha_2}{\alpha_1}\right)^{-1} \leq \theta \leq \left(1 - \left(\frac{\beta_1}{\beta_2}\right)^{2\nu} \frac{\alpha_1}{\alpha_2}\right)^{-1}$$

Como hemos comentado anteriormente, una de las principales ventajas de este modelo suma de productos generalizado es que permite covarianzas negativas o con oscilaciones entre valores positivos y negativos, lo que es de especial interés en la modelización de determinados fenómenos naturales. Por ejemplo, si en (4.77) consideramos  $\alpha_1 = 0.5 < \alpha_2 = 1$ ,  $\beta_1 = \beta_2 = 0.5$ ,  $\nu = 0.3$  y  $\sigma^2 = \sigma_e^2 \sigma_t^2 = 5$ , entonces por (4.78) este modelo es válido siempre que  $-1 \leq \theta \leq 2$ . Si consideramos  $\theta = -1$ , podemos ver a la izquierda de la Figura 4.7 cómo la función de covarianza resultante puede tomar valores negativos. A la derecha de la Figura 4.7 podemos ver la representación de la marginal temporal  $C(0, u)$  correspondiente a este modelo y su comportamiento respecto al cero.

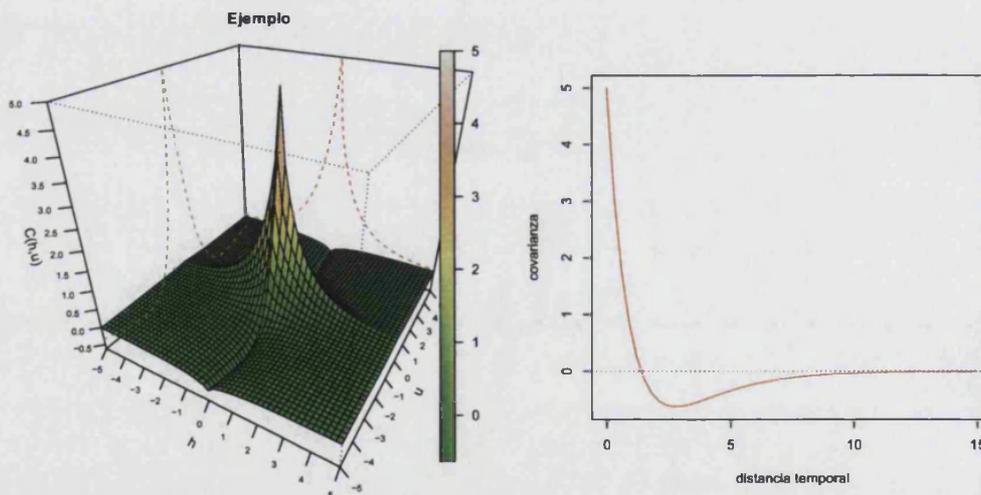


Figura 4.7: Ejemplo del modelo suma de productos generalizado (4.77) con covarianzas negativas.

La Figura 4.8 muestra diferentes representaciones de esta familia de covarianzas (4.77) para un conjunto de posibles valores de los parámetros que la definen. En concreto, sean  $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \nu, \theta, \sigma_e^2, \sigma_t^2)'$  los parámetros que definen este modelo (4.77), se han representado las funciones de covarianza correspondientes a  $\alpha_1 = 0.5$ ,  $\alpha_2 = 1.5$ ,  $\beta_1 = 0.3$ ,  $\beta_2 = 0.6$ ,  $\nu = 0.3$  en el ejemplo 1,  $\alpha_1 = 0.5$ ,  $\alpha_2 = 1.5$ ,  $\beta_1 = 1.5$ ,  $\beta_2 = 0.6$ ,  $\nu = 0.3$  en el ejemplo 2,  $\alpha_1 = 0.5$ ,  $\alpha_2 = 1.5$ ,  $\beta_1 = 0.3$ ,  $\beta_2 = 1.2$ ,  $\nu = 0.3$  en el ejemplo 3 y  $\alpha_1 = 0.5$ ,  $\alpha_2 = 1.5$ ,  $\beta_1 = 0.3$ ,  $\beta_2 = 1.2$ ,  $\nu = 0.8$  en el ejemplo 4, donde en todos los casos se ha considerado  $\theta = 0.5$  y  $\sigma^2 = \sigma_e^2 \sigma_t^2 = 5$ .

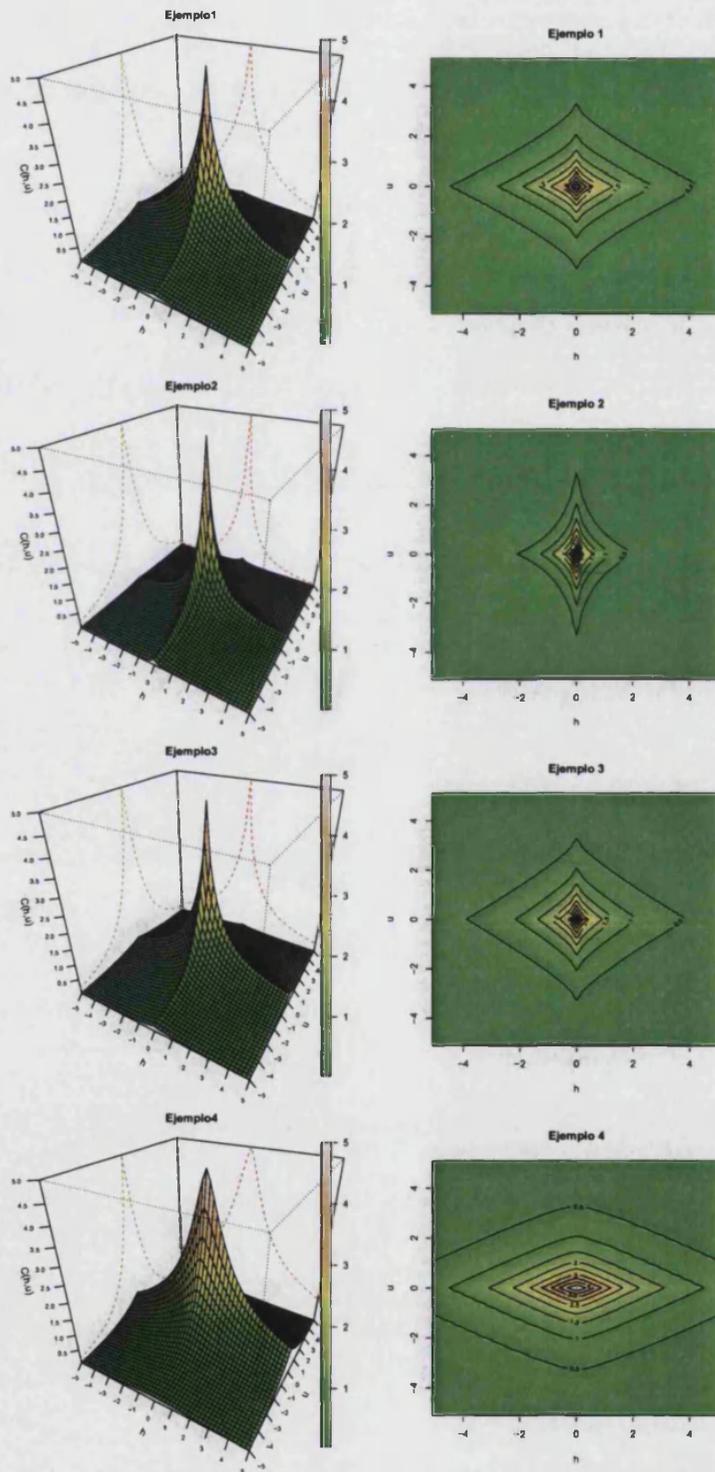


Figura 4.8: Diferentes representaciones del modelo suma de productos generalizado (4.77).

### 4.13. Modelos que no satisfacen la simetría completa

Como ya hemos comentado en el Capítulo 3, existen determinados procesos espacio-temporales que por su propia naturaleza no son completamente simétricos. Por ejemplo, pueden existir determinadas direcciones espaciales donde el proceso se manifieste de forma diferente que en el resto, por el efecto de vientos predominantes, corrientes oceánicas, etc. Las covarianzas que hemos visto en las secciones anteriores son completamente simétricas, por lo que no permitirán la modelización de estas situaciones.

Una forma de tratar estos problemas es generar funciones de covarianza definidas mediante (3.16), esto es, de la forma

$$C(\mathbf{h}, \mathbf{u}) = E_{\mathcal{V}}[C_e(\mathbf{h} + \mathcal{V}\mathbf{u})] \quad (4.79)$$

donde  $\mathcal{V}$  es un vector de velocidad aleatorio y  $C_e(\mathbf{h})$  es una función de covarianza espacial estacionaria. Un ejemplo de utilización de este tipo de funciones de covarianza podemos encontrarlo en Cox e Isham (1988). Es sencillo verificar la validez de las covarianzas definidas por (4.79), ya que

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n a_i a_j C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E_{\mathcal{V}}[C_e((\mathbf{s}_i - \mathbf{s}_j) + \mathcal{V}(t_i - t_j))] = \\ E_{\mathcal{V}} \left[ \sum_{i=1}^n \sum_{j=1}^n a_i a_j C_e((\mathbf{s}_i + \mathcal{V}t_i) - (\mathbf{s}_j + \mathcal{V}t_j)) \right] &= E_{\mathcal{V}} \left[ \sum_{i=1}^n \sum_{j=1}^n a_i a_j C_e(\mathbf{r}_i - \mathbf{r}_j) \right] \geq 0, \end{aligned}$$

por ser la esperanza de una cantidad positiva, al ser  $C_e(\mathbf{h})$  una función de covarianza espacial sobre  $\mathbb{R}^d$ .

Para la especificación del vector de velocidad aleatorio existen diferentes elecciones razonables desde el punto de vista físico. El caso más sencillo es seleccionar  $\mathcal{V} = \mathbf{v}$  como el vector medio de los valores observados. Otra posibilidad es asignar al vector aleatorio  $\mathcal{V}$  una distribución empírica de vectores de velocidad obtenida a partir de un conjunto de registros históricos. También puede actualizarse la distribución de  $\mathcal{V}$  de forma dinámica de acuerdo con el estado actual del fenómeno, lo que dará lugar a estructuras de covarianza no estacionarias como las que aparecen en Huang y Hsu (2004), entre otros.

En la literatura podemos encontrar diferentes funciones de covarianza espacio-temporales estacionarias que no son completamente simétricas obtenidas a través de ecuaciones de difusión o de ecuaciones estocásticas de derivadas parciales. Pueden citarse Jones y Zhang (1997), Christakos (2000), Brown et al. (2000), Kolovos et al. (2004) o Stein (2005a), entre otros.

### 4.14. Modelos no estacionarios

Como ya hemos comentado a lo largo del Capítulo 3, la admisión de la hipótesis de estacionariedad no será conveniente en la modelización de determinados fenómenos naturales. No obstante, la literatura referida a modelos no estacionarios no es tan amplia como la

de modelos estacionarios, aunque está en paulatino crecimiento durante los últimos años. En esta sección vamos a resumir algunas de las principales estrategias utilizadas en la literatura para la construcción de funciones de covarianza no estacionarias. En concreto, analizaremos la generalización del procedimiento de construcción de funciones espaciales no estacionarias propuesto por Fuentes y Smith (2001) al caso espacio-temporal, utilizando funciones de covarianza que serán sumas ponderadas de covarianzas localmente estacionarias. También analizaremos los modelos no estacionarios propuestos en (Ma, 2002, 2003c). Por último, describiremos las funciones de covarianza no estacionarias construidas por Porcu y Mateu (2007).

#### 4.14.1. Mixtura de procesos estacionarios localmente ortogonales

En Chen et al. (2006) podemos encontrar una generalización al caso espacio-temporal del procedimiento de generación de funciones de covarianza espaciales no estacionarias mediante una mixtura de procesos estacionarios localmente ortogonales (Fuentes y Smith, 2001; Fuentes, 2002a,b).

Sea  $Z(\mathbf{s}, t)$  un proceso espacio-temporal de media cero no estacionario, con  $(\mathbf{s}, t) \in D \times T$ , siendo  $D \subseteq \mathbb{R}^d$  y  $T \subseteq \mathbb{R}$ . Vamos a descomponer el proceso  $Z(\mathbf{s}, t)$  como una mixtura de procesos espacio-temporales estacionarios ortogonales

$$Z(\mathbf{s}, t) = \sum_{i=1}^k K(\mathbf{s} - \mathbf{s}_i, t - t_i) Z_i(\mathbf{s}, t) \quad (4.80)$$

donde  $\mathbf{s}_i$  es el centroide de cada una de las regiones  $S_i$ , siendo  $\{S_1, \dots, S_n\}$  una partición del dominio espacio-temporal de interés  $D \times T$ ,  $Z_i(\mathbf{s}, t)$  es un proceso espacio-temporal localmente estacionario en  $S_i$  que explica la estructura espacio-temporal del proceso general en dicha región, y  $K(\mathbf{s} - \mathbf{s}_i, t - t_i)$  es una función de ponderación sobre el espacio y el tiempo que da más importancia al proceso  $Z_i(\mathbf{s}, t)$  cuando  $\mathbf{s}$  está cerca de  $\mathbf{s}_i$  y menos importancia cuando  $\mathbf{s}$  está lejos de  $\mathbf{s}_i$ .

Si el proceso espacio-temporal  $Z(\mathbf{s}, t)$  es estacionario en el tiempo pero no en el espacio, podemos escribir (4.80) como

$$Z(\mathbf{s}, t) = \sum_{i=1}^k K(\mathbf{s} - \mathbf{s}_i) Z_i(\mathbf{s}, t)$$

donde  $Z_i(\mathbf{s}, t)$  es un proceso espacio-temporal localmente estacionario dentro de la región  $D_i \times T$  y  $K(\mathbf{s} - \mathbf{s}_i)$  una función de ponderación centrada en  $\mathbf{s}_i$ , siendo  $\{D_1 \times T, \dots, D_k \times T\}$  una partición del dominio espacio-temporal  $D \times T$  y  $\mathbf{s}_i$  el centroide de  $D_i$ . Del mismo modo, si el proceso espacio-temporal  $Z(\mathbf{s}, t)$  es estacionario en el espacio pero no en el tiempo, podemos escribir (4.80) como

$$Z(\mathbf{s}, t) = \sum_{i=1}^k K(t - t_i) Z_i(\mathbf{s}, t)$$

donde cada elemento de la expresión se define de forma análoga al caso anterior.

#### 4.14.2. Modelos no estacionarios propuestos por Ma

En este apartado vamos a estudiar las principales técnicas de generación de modelos no estacionarios propuestas por Ma. En primer lugar vamos a analizar el procedimiento descrito en Ma (2002), que permite la generación de funciones de covarianza no estacionarias utilizando mixturas escalares y mixturas de potencias positivas de funciones de covarianza, generalizando los resultados que hemos visto en las secciones 4.9.1 y 4.9.2. También estudiaremos la técnica descrita en Ma (2003c), que permite la generación de funciones de covarianza y semivariogramas espacio-temporales no estacionarios mediante combinaciones lineales de funciones de covarianza espacio-temporales estacionarias y variogramas intrínsecamente estacionarios.

##### Mixturas escalares

Vamos a generalizar el procedimiento de mixturas escalares que hemos descrito en la Sección 4.9.1 para la generación de funciones de covarianza no estacionarias. Sea  $C_e(\mathbf{s}_1, \mathbf{s}_2)$  una función de covarianza espacial sobre  $\mathfrak{R}^d$  y  $C_t(t_1, t_2)$  una función de covarianza temporal sobre  $\mathcal{T}$ , tenemos que las funciones  $C_e(u\mathbf{s}_1, u\mathbf{s}_2)$  y  $C_t(vt_1, vt_2)$  son covarianzas válidas sobre  $\mathfrak{R}^d$  y  $\mathcal{T}$ , respectivamente, para cualesquiera  $u \in \mathfrak{R}$  y  $v \in \mathcal{T}$ . Luego la función generada por

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \int_{\mathfrak{R} \times \mathcal{T}} C_e(\mathbf{s}_1 u, \mathbf{s}_2 u) C_t(t_1 v, t_2 v) d\mu(u, v), \quad (4.81)$$

será una función de covarianza espacio-temporal sobre  $\mathfrak{R}^d \times \mathcal{T}$  en general no estacionaria, donde  $\mu(u, v)$  es una medida finita no negativa sobre  $\mathfrak{R} \times \mathcal{T}$ .

Veamos cómo construir un proceso espacio-temporal que tenga esta función de covarianza. Sean  $\{Z_e(\mathbf{s}), \mathbf{s} \in \mathfrak{R}^d\}$  un proceso espacial con función de covarianza  $C_e(\mathbf{s}_1, \mathbf{s}_2)$  y  $\{Z_t(t), t \in \mathcal{T}\}$  un proceso temporal con función de covarianza temporal  $C_t(t_1, t_2)$ . Supongamos que ambos procesos son independientes. Sea  $(U, V)$  un vector aleatorio con función de distribución  $\mu(u, v)$ , que sea independiente de  $\{Z_e(\mathbf{s}), Z_t(t), \mathbf{s} \in \mathfrak{R}^d, t \in \mathcal{T}\}$ , entonces el proceso espacio-temporal  $\{Z(\mathbf{s}, t), \mathbf{s} \in \mathfrak{R}^d, t \in \mathcal{T}\}$ , con  $Z(\mathbf{s}, t) = Z_e(\mathbf{s}U)Z_t(tV)$ , tiene a (4.81) como función de covarianza. Si las variables aleatorias  $U$  y  $V$  son independientes, entonces la función de distribución  $\mu(u, v)$  es separable y la función (4.81) se reduce a una función de covarianza separable.

Se muestran a continuación algunos de los ejemplos que aparecen en Ma (2002), y que cubren tanto el caso en que la componente temporal toma valores discretos como continuos.

**Ejemplo 4.25** Sea  $(U, V)$  un vector aleatorio bivariado no negativo con función de distribución  $\mu(u, v)$ . Consideremos la función de covarianza espacial estacionaria

$$C_e(\mathbf{s}_1, \mathbf{s}_2) = \exp(-\theta h), \quad \mathbf{s}_1, \mathbf{s}_2 \in \mathfrak{R}^d$$

con  $h = \|\mathbf{s}_1 - \mathbf{s}_2\|$  y  $\theta > 0$ , y la función de covarianza temporal no estacionaria

$$C_t(t_1, t_2) = \exp(-\gamma|t_1 - t_2|) - \exp(-\gamma t_1 - \gamma t_2), \quad t_1, t_2 \in \mathfrak{R}_+$$

con  $\gamma > 0$ . Por (4.81), la función

$$C((s_1, t_1), (s_2, t_2)) = \int_{\mathbb{R}_+^2} \exp(-\theta hu) (\exp(-\gamma|t_1 - t_2|v) - \exp(-\gamma t_1 v - \gamma t_2 v)) d\mu(u, v),$$

será una función de covarianza espacio-temporal espacialmente estacionaria. Sea  $\mathcal{L}$  la transformada de Laplace de la función de distribución  $\mu(u, v)$ , entonces podemos expresar la función anterior como

$$C((s_1, t_1), (s_2, t_2)) = \mathcal{L}(\theta h, \gamma|t_1 - t_2|) - \mathcal{L}(\theta h, \gamma(t_1 + t_2)), \quad (4.82)$$

lo que nos permitirá generar diferentes familias paramétricas de funciones de covarianza seleccionando diferentes valores de los parámetros y diferentes transformadas de Laplace.

Por ejemplo, si seleccionamos la transformada de Laplace de la distribución gamma bivariada de Cheriyan-Ramabhardran

$$\mathcal{L}(z_1, z_2) = (1 + z_1)^{-\alpha_1} (1 + z_2)^{-\alpha_2} (1 + z_1 + z_2)^{-\alpha_{12}},$$

con  $\alpha_1 \geq 0$ ,  $\alpha_2 \geq 0$  y  $\alpha_{12} \geq 0$ , entonces por (4.82) la función

$$C((s_1, t_1), (s_2, t_2)) = (1 + \theta h)^{-\alpha_1} ((1 + \gamma|t_1 - t_2|)^{-\alpha_2} (1 + \theta h + \gamma|t_1 - t_2|)^{-\alpha_{12}} - (1 + \gamma|t_1 - t_2|)^{-\alpha_2} (1 + \theta h + \gamma|t_1 + t_2|)^{-\alpha_{12}}),$$

será una función de covarianza espacio-temporal sobre  $\mathbb{R}^d \times \mathbb{R}_+$ .

**Ejemplo 4.26** Sea  $(U, V)$  un vector aleatorio bivariado no negativo con función de distribución  $\mu(u, v)$ . Consideremos la función de covarianza espacial estacionaria

$$C_e(s_1, s_2) = \exp(-\theta h), \quad s_1, s_2 \in \mathbb{R}^d$$

con  $h = \|s_1 - s_2\|$  y  $\theta > 0$ , y la función de covarianza temporal no estacionaria

$$C_t(t_1, t_2) = \exp(-\theta_1 t_1 - \theta_2 t_2 - \theta_{12} \max\{t_1, t_2\}), \quad t_1, t_2 \in \mathbb{R}_+$$

con  $\theta_1, \theta_2$  y  $\theta_{12}$  constantes positivas. Por (4.81), la función

$$C((s_1, t_1), (s_2, t_2)) = \int_{\mathbb{R}_+^2} \exp(-\theta hu) (\exp(-\theta_1 t_1 + \theta_2 t_2 + \theta_{12} \max\{t_1, t_2\})v) d\mu(u, v),$$

será una función de covarianza espacio-temporal espacialmente estacionaria. Sea  $\mathcal{L}$  la transformada de Laplace de la función de distribución  $\mu(u, v)$ , entonces podemos expresar esta integral como

$$C((s_1, t_1), (s_2, t_2)) = \mathcal{L}(\theta h, \theta_1 t_1 + \theta_2 t_2 + \theta_{12} \max\{t_1, t_2\}), \quad (4.83)$$

lo que nos permitirá generar diferentes familias paramétricas de funciones de covarianza seleccionando diferentes transformadas de Laplace y diferentes parámetros  $\theta_i$ .

### Mixturas de potencias positivas

Respecto a la generación de funciones de covarianza no estacionarias mediante el procedimiento de mixturas de potencias positivas que hemos visto en la Sección 4.3.2, la extensión de los modelos (4.42) y (4.44) al caso no estacionario será posible únicamente si las funciones de covarianza espacial y temporal no estacionarias consideradas están acotadas (esto es, si  $|C_e(\mathbf{s}_1, \mathbf{s}_2)| < c_e$  y  $|C_t(t_1, t_2)| < c_t$ ). En ese caso, las funciones

$$\rho((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} \left( \frac{C_e(\mathbf{s}_1, \mathbf{s}_2)}{c_e} \right)^i \left( \frac{C_t(t_1, t_2)}{c_t} \right)^j p_{ij}, \quad (4.84)$$

$$\rho((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \sum_{k=0}^{+\infty} \left( \frac{C_e(\mathbf{s}_1, \mathbf{s}_2) C_t(t_1, t_2)}{c_e c_t} \right)^k p_k, \quad (4.85)$$

serán funciones de correlación espacio-temporales sobre  $\mathfrak{R}^d \times \mathcal{T}$ , con  $\{p_{ij}, (i, j) \in \mathbb{N}^2\}$  y  $\{p_k, k \in \mathbb{N}\}$  distribuciones de probabilidad. Si las funciones de covarianza  $C_e$  y  $C_t$  no están acotadas, entonces las funciones (4.84) y (4.85) no tienen porque estar bien definidas. No obstante, siempre será posible definir cualquier mixtura finita de potencias positivas de funciones de covarianza espaciales y temporales no estacionarias.

### Combinación lineal de funciones de covarianza y variogramas

En la Sección 3.3 hemos visto que si  $\gamma(\mathbf{s}, t)$  es un semivariograma intrínsecamente estacionario sobre  $\mathfrak{R}^d \times \mathcal{T}$ , entonces la función

$$\omega((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \gamma(\mathbf{s}_1, t_1) + \gamma(\mathbf{s}_2, t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2)$$

es una función de covarianza sobre  $\mathfrak{R}^d \times \mathcal{T}$ . Vamos a considerar además las funciones:

$$\phi((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = 2(\gamma(\mathbf{s}_1, t_1) + \gamma(\mathbf{s}_2, t_2)) - \gamma(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2), \quad (4.86)$$

$$\kappa((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \gamma(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2). \quad (4.87)$$

Nótese que

$$\omega((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \frac{1}{2}\phi((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) + \frac{1}{2}\kappa((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)).$$

En particular, si  $\phi((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = 0$ , entonces  $\gamma(\mathbf{s}, t)$  es una forma cuadrática y  $\omega((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \frac{1}{2}\kappa((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  una función de covarianza sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

En Ma (2003c) se demuestra que si  $\gamma(\mathbf{s}, t)$  es un semivariograma intrínsecamente estacionario, entonces la función  $\phi((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  definida por (4.86) es una función de covarianza no negativa sobre  $\mathfrak{R}^d \times \mathcal{T}$ , con variograma asociado  $\frac{1}{2}\phi((\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2), (\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2))$ . Como consecuencia, dada una función de covarianza espacio-temporal estacionaria  $C(\mathbf{h}, u)$  sobre  $\mathfrak{R}^d \times \mathcal{T}$ , entonces la función

$$C(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2) + C(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2) - 2(C(\mathbf{s}_1, t_1) + C(\mathbf{s}_2, t_2) - C(\mathbf{0}, 0)), \quad (4.88)$$

es una función de covarianza espacio-temporal no negativa y en general no estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

También se demuestra que si  $\gamma(\mathbf{s}, t)$  es un semivariograma intrínsecamente estacionario, entonces la función  $\kappa((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2))$  definida por (4.87) es una función de covarianza sobre  $\mathfrak{R}^d \times \mathcal{T}$ . Como consecuencia, dada una función de covarianza espacio-temporal estacionaria  $C(\mathbf{h}, u)$  sobre  $\mathfrak{R}^d \times \mathcal{T}$ , entonces la función

$$C(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2) - C(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2), \quad (4.89)$$

es una función de covarianza espacio-temporal en general no estacionaria sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

A partir de estos resultados podemos construir modelos de covarianza no estacionarios utilizando semivariogramas intrínsecamente estacionarios puramente espaciales y temporales. Veamos a continuación uno de los ejemplos propuestos en Ma (2003c).

**Ejemplo 4.27** Sean  $\gamma_e(\mathbf{h})$  un semivariograma espacial intrínsecamente estacionario sobre  $\mathfrak{R}^d$ , entonces hemos visto en la Sección 3.3 que  $\gamma_e(\mathbf{h} - \boldsymbol{\theta}t)$ , para  $\boldsymbol{\theta} \in \mathfrak{R}^d$ , es también un semivariograma intrínsecamente estacionario, por lo que  $\exp(-\gamma_e(\mathbf{h} - \boldsymbol{\theta}t))$  será una función de covarianza espacio-temporal estacionaria. Luego por los resultados anteriores, las funciones

$$\begin{aligned} & 2(\gamma_e(\mathbf{s}_1 - t_1\boldsymbol{\theta}) + \gamma_e(\mathbf{s}_2 - t_2\boldsymbol{\theta})) - \gamma_e(\mathbf{s}_1 + \mathbf{s}_2 - (t_1 + t_2)\boldsymbol{\theta}) - \gamma_e(\mathbf{s}_1 - \mathbf{s}_2 - (t_1 - t_2)\boldsymbol{\theta}), \\ & \gamma_e(\mathbf{s}_1 + \mathbf{s}_2 - (t_1 + t_2)\boldsymbol{\theta}) - \gamma_e(\mathbf{s}_1 - \mathbf{s}_2 - (t_1 - t_2)\boldsymbol{\theta}), \\ & \exp(-\gamma_e(\mathbf{s}_1 - \mathbf{s}_2 - (t_1 - t_2)\boldsymbol{\theta})) - \exp(-\gamma_e(\mathbf{s}_1 + \mathbf{s}_2 - (t_1 + t_2)\boldsymbol{\theta})), \end{aligned}$$

son funciones de covarianza espacio-temporales no estacionarias sobre  $\mathfrak{R}^d \times \mathcal{T}$ .

#### 4.14.3. Modelos no estacionarios propuestos por Porcu y Mateu

Porcu y Mateu (2007) muestran un procedimiento que permite la construcción de modelos no estacionarios utilizando funciones completamente monótonas.

Sea  $\gamma(\mathbf{h}, u)$  un semivariograma intrínsecamente estacionario con  $\gamma(\mathbf{0}, 0) = 0$ , y sea  $\varphi(t)$ ,  $t > 0$ , una función completamente monótona con  $\varphi(0) = 1$ . Entonces la función

$$C((\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2)) = \varphi(\psi(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)), \quad (4.90)$$

con

$$\psi(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = \frac{1}{2}(\gamma(2\mathbf{s}_1, 2t_1) + \gamma(2\mathbf{s}_2, 2t_2)) - (\gamma(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2)), \quad (4.91)$$

o

$$\psi(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = 1 + \frac{1}{2}(\gamma(2\mathbf{s}_1, 2t_1) + \gamma(2\mathbf{s}_2, 2t_2)) - \left( \frac{1 + \gamma(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2)}{1 + \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2)} \right), \quad (4.92)$$

es una función de covarianza no estacionaria válida sobre  $\mathfrak{R}^d \times \mathfrak{R}$ . Además, es fácil comprobar que las funciones

$$\begin{aligned}C((s_1, t_1), (s_2, t_2)) &= \varphi \left( \frac{\psi(s_1, s_2, t_1, t_2)}{1 + \psi(s_1, s_2, t_1, t_2)} \right), \\C((s_1, t_1), (s_2, t_2)) &= \varphi \left( 1 - \frac{1 - \psi(s_1, s_2, t_1, t_2)}{1 + \psi(s_1, s_2, t_1, t_2)} \right), \\C((s_1, t_1), (s_2, t_2)) &= \varphi \left( (1 + \psi(s_1, s_2, t_1, t_2))^\rho - 1 \right),\end{aligned}$$

siendo  $\psi$  la función dada por (4.91) o (4.92) y  $\rho \in [0, 1[$ , generan también funciones de covarianza no estacionarias válidas sobre  $\mathfrak{R}^d \times \mathfrak{R}$ .

## Capítulo 5

# Comparación de modelos mediante simulación

### 5.1. Introducción

Como ya hemos comentado en los capítulos anteriores, existe una creciente demanda de modelos estadísticos que permitan analizar fenómenos naturales que se desarrollan tanto en el espacio como en el tiempo. Esto hace que la búsqueda y análisis de modelos válidos para la función de covarianza de procesos espacio-temporales sea una materia de gran interés en los últimos años con una bibliografía que crece rápidamente.

En el Capítulo 3 hemos introducido el problema de la modelización espacio-temporal y citado algunas de las principales referencias de la literatura aparecidas en los últimos años. También se han descrito las principales propiedades de los procesos espacio-temporales y la extensión a este contexto de las diferentes técnicas de predicción utilizadas en la geoestadística espacial clásica. En el Capítulo 4 se han analizado con profundidad algunos de los principales modelos utilizados para la función de covarianza a lo largo de los últimos años en la literatura de la modelización espacio-temporal.

En la modelización práctica, el investigador tiene que hacer frente a la importante decisión de seleccionar aquel modelo espacio-temporal que se ajusta mejor a los datos empíricos observados. Con el objetivo de ayudar a los investigadores en esta tarea, se presenta en este capítulo un exhaustivo estudio de simulación que compara la potencia predictiva de algunos de los principales modelos de función de covarianza utilizados en la modelización estadística espacio-temporal. La batería de modelos utilizados en este análisis cubre tanto funciones de covarianza separables como no separables, así como diferentes grados de dependencia espacial, temporal y espacio-temporal. También se analiza el coste computacional de las diferentes alternativas propuestas, ya que este es un elemento muy a tener en cuenta cuando nos enfrentamos con la modelización de datos reales, como veremos en el capítulo siguiente.

En particular, en este análisis comparativo hemos seleccionado cinco de los modelos válidos para la función de covarianza de procesos espacio-temporales presentados en la literatura. Una primera parte de este análisis se recoge en un trabajo ya publicado (Huang

et al., 2007) que se reproduce en el Capítulo 6. El primero de los modelos considerados es el modelo separable obtenido al aplicar un determinado filtro de Kalman espacio-temporal descrito en Huang y Cressie (1996), y que hemos estudiado en la Sección 3.5.1. El segundo de los modelos analizados es uno de los modelos no separables que se introducen en Cressie y Huang (1999), que hemos estudiado en la Sección 4.5. El tercero pertenece a la familia de modelos no separables presentados en Gneiting (2002), y que hemos estudiado en la Sección 4.7. El cuarto de los modelos considerados es también un modelo no separable basado en una estructura de suma-producto que se estudia en De Cesare et al. (2001a), y que hemos analizado en la Sección 4.6. Por último, el quinto de los modelos considerados es el modelo suma de productos generalizado, que da lugar a una estructura no separable tal y como hemos definido en la Sección 4.12. El conjunto de modelos considerados cubren las clases de funciones de covarianza espacio-temporal más utilizadas en la modelización práctica.

Como veremos a lo largo de este capítulo, el ajuste de los parámetros de los diferentes modelos considerados se ha llevado a cabo mediante el método de máxima verosimilitud, que presentamos en la Sección 3.6. Suponiendo normalidad en el proceso, este procedimiento de ajuste es más eficaz que los procedimientos basados en el método de los momentos, tan utilizados en la práctica geoestadística. También se han comparado diferentes estrategias automáticas de selección de modelos para analizar su comportamiento sobre los procesos simulados. En concreto, se han implementado dos de los criterios de selección de modelos más utilizados, el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC), que tratan de buscar un compromiso entre la bondad del ajuste del modelo respecto a los datos y la complejidad del modelo seleccionado. Hasta el momento, no se ha conseguido analizar completamente el comportamiento asintótico de estos criterios de selección para los modelos geoestadísticos, por lo que es de gran interés investigar si es posible trasladar al contexto de la modelización espacio-temporal los resultados tradicionales de consistencia del BIC y de eficiencia asintótica del AIC en la selección de modelos de regresión y de series temporales, aunque se trata de un problema de gran complejidad. En este capítulo se analiza el comportamiento de ambos criterios sobre los experimentos simulados, obteniendo resultados prometedores para la selección de modelos espacio-temporales.

La Sección 5.2 introduce la forma general de los procesos espacio-temporales considerados. Se describen los diferentes modelos de covarianza espacio-temporal, los parámetros involucrados y las condiciones que deben cumplir para asegurar la validez del modelo final. La Sección 5.3 describe las combinaciones paramétricas consideradas para cada uno de los modelos anteriores, junto con las características generales del procedimiento de simulación utilizado. La Sección 5.4 compara los resultados obtenidos en cada una de las simulaciones con los diferentes modelos considerados, basándonos en determinados estadísticos de resumen. También analizaremos las estimaciones de los parámetros obtenidas con cada uno de los modelos, comparando los modelos obtenidos con el que ha generado los datos en cada caso. Por último, la Sección 5.5 resume las principales conclusiones extraídas del trabajo de simulación presentado en este capítulo.

## 5.2. Modelos espacio-temporales considerados

Consideremos un proceso estocástico espacio-temporal de valores reales  $\{Y(\mathbf{s}, t), \mathbf{s} \in D, t \in \mathcal{T}\}$  que toma valores sobre el dominio espacial  $D \subseteq \mathbb{R}^2$  y sobre el dominio temporal  $\mathcal{T} \subseteq \mathbb{Z}$  o  $\mathbb{R}$ . Como hemos visto en el Capítulo 3, bajo la condición de regularidad del proceso podemos definir la función de covarianza y el semivariograma del proceso como

$$\begin{aligned} C(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) &= \text{Cov}[Y(\mathbf{s}_1, t_1), Y(\mathbf{s}_2, t_2)], \\ \gamma(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) &= \frac{1}{2} \text{Var}[Y(\mathbf{s}_1, t_1) - Y(\mathbf{s}_2, t_1)], \end{aligned}$$

para todo  $(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2) \in D \times \mathcal{T}$ . Ambas funciones están relacionadas entre sí mediante la expresión (3.5), según la cual

$$2\gamma(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = C(\mathbf{s}_1, \mathbf{s}_1, t_1, t_1) + C(\mathbf{s}_2, \mathbf{s}_2, t_2, t_2) - 2C(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2).$$

Hemos visto que el proceso  $Y(\mathbf{s}, t)$  es estacionario de segundo orden (o simplemente estacionario) si la media es constante y la covarianza entre dos localizaciones  $C(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$  depende únicamente de la distancia espacial  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  y temporal  $u = t_1 - t_2$  existente entre ellas. En ese caso, denotábamos la covarianza entre dos localizaciones espacio-temporales como  $C(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = C(\mathbf{h}, u)$  y recibía el nombre de función de covarianza estacionaria.

Del mismo modo, decíamos que el proceso espacio-temporal  $Y(\mathbf{s}, t)$  es intrínsecamente estacionario si la media es constante y el variograma entre dos localizaciones  $2\gamma(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$  depende únicamente de la distancia espacial  $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$  y temporal  $u = t_1 - t_2$  existente entre ellas. La estacionariedad intrínseca es más general que la estacionariedad de segundo orden, ya que existen procesos que son intrínsecamente estacionarios pero no estacionarios de segundo orden. Por tanto, si  $Y(\mathbf{s}, t)$  es estacionario, entonces será intrínsecamente estacionario.

Vamos a considerar a lo largo de todo el capítulo que el proceso de interés  $\{Y(\mathbf{s}, t), \mathbf{s} \in D, t \in \mathcal{T}\}$ , con  $D \subseteq \mathbb{R}^2$  y  $\mathcal{T} \subseteq \mathbb{Z}$ , es un proceso espacio-temporal Gaussiano estacionario con media cero (es decir, sin tendencia) y con función de covarianza espacio-temporal  $C(\mathbf{h}, u)$ . Vamos a suponer que en lugar de observar directamente el proceso de interés  $Y(\mathbf{s}, t)$  observamos una versión  $Z(\mathbf{s}, t)$  afectada por un determinado error de medida, de acuerdo con la siguiente ecuación

$$Z(\mathbf{s}, t) = Y(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t),$$

donde  $\varepsilon(\cdot, \cdot)$  es un proceso de ruido blanco con media cero y varianza  $\sigma_\varepsilon^2$ .

Supongamos que hemos observado el valor del proceso sobre un conjunto de  $n$  localizaciones espaciales para  $T$  instantes temporales, obteniendo un conjunto de  $T$  vectores  $\mathbf{Z}(t) = (Z(\mathbf{s}_1, t), \dots, Z(\mathbf{s}_n, t))'$ , para  $t = 1, \dots, T$ . En la Sección 3.7 hemos visto que el predictor lineal óptimo de  $Y(\mathbf{s}, T)$  basado en los valores  $\mathbf{Z}(1), \dots, \mathbf{Z}(T)$  podía escribirse como

$$\hat{Y}(\mathbf{s}, T) = \text{Cov}(Y(\mathbf{s}, T), \mathbf{Z}) \text{Var}(\mathbf{Z})^{-1} \mathbf{Z}, \quad (5.1)$$

siendo  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')$ , y recibe el nombre de kriging simple. El error de predicción asociado viene dado por

$$\text{Var}(Y(\mathbf{s}, T)) - \text{Cov}(Y(\mathbf{s}, T), \mathbf{Z}) \text{Var}(\mathbf{Z})^{-1} \text{Cov}(\mathbf{Z}, Y(\mathbf{s}, T)). \quad (5.2)$$

Como hemos comentado en la Introducción, vamos a considerar para la simulación cinco modelos de covarianza distintos. Todos ellos han sido descritos con detalle en su forma general en el Capítulo 4, aunque a continuación vamos a introducir de forma resumida los casos particulares aquí considerados, con énfasis particular en la estructura de la función de verosimilitud.

### 5.2.1. Modelo espacio-temporal dinámico (Huang y Cressie, 1996)

El primero de los modelos considerados es el modelo presentado en Huang y Cressie (1996) basado en el filtro de Kalman espacio-temporal que hemos estudiado en la Sección 3.5.1, y que considera una estructura de covarianza separable obtenida a partir de una determinada función de covarianza espacial y una función de covarianza temporal con estructura autorregresiva de orden  $p$  (AR( $p$ )). En este caso, vamos a considerar una función de covarianza espacial para  $Y(\cdot, t)$  perteneciente al modelo de Matérn (2.33) que viene dada por

$$C_{\phi}(\mathbf{h}) = \sigma_Y^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^{\nu} K_{\nu}(b\|\mathbf{h}\|),$$

donde  $\phi \equiv (b, \nu, \sigma_Y^2)^t$ , con  $\sigma_Y^2 = \text{Var}(Y(\mathbf{s}, t))$ ,  $b > 0$  es un parámetro de escala y  $\nu > 0$  un parámetro de suavizado. La función de covarianza temporal para  $Y(\mathbf{s}, \cdot)$  vendrá dada por la función de autocorrelación de un modelo AR(1), por lo que la función de covarianza espacio-temporal de  $Y(\cdot, \cdot)$  será

$$C_Y(\mathbf{h}, u | \phi, \alpha) = \alpha^{|u|} C_{\phi}(\mathbf{h}), \quad (5.3)$$

que es una función de covarianza espacio-temporal separable.

Vamos a suponer que el proceso espacio-temporal  $Y(\mathbf{s}, t)$  evoluciona en el tiempo según la ecuación

$$Y(\mathbf{s}, t) = \alpha Y(\mathbf{s}, t-1) + \xi(\mathbf{s}, t), \quad (\mathbf{s}, t) \in D \times T,$$

siendo  $\xi(\cdot, \cdot)$  un proceso de error Gaussiano estacionario espacialmente e independiente temporalmente.

El predictor óptimo de  $Y(\mathbf{s}, t)$  basado en los datos  $\mathbf{Z}(1), \dots, \mathbf{Z}(t)$  viene dado por

$$\hat{Y}(\mathbf{s}, t) \equiv E(Y(\mathbf{s}, t) | \mathbf{Z}(1), \dots, \mathbf{Z}(t)).$$

La covarianza del error de predicción de  $Y(\mathbf{s}, t)$  e  $Y(\mathbf{s}', t)$  basada en los datos  $\mathbf{Z}(1), \dots, \mathbf{Z}(t)$  es

$$\begin{aligned} P_t(\mathbf{s}, \mathbf{s}') &\equiv \text{Cov}(\hat{Y}(\mathbf{s}, t) - Y(\mathbf{s}, t), \hat{Y}(\mathbf{s}', t) - Y(\mathbf{s}', t)) \\ &= \text{Cov}(Y(\mathbf{s}, t), Y(\mathbf{s}', t) | \mathbf{Z}(1), \dots, \mathbf{Z}(t)). \end{aligned}$$

En la sección 3.5.1 se muestra que  $\hat{Y}(s, t)$  and  $P_t(s, s')$ , para cualesquiera  $s, s' \in D$ , puede calcularse de forma recursiva utilizando el algoritmo del filtro de Kalman

$$\begin{aligned}\hat{Y}(s, t) &= \alpha \hat{Y}(s, t-1) + Q_t(s)' F_t^{-1} (Z(t) - \alpha \hat{Y}(t-1)), \\ P_t(s, s') &= P_{t|t-1}(s, s') + Q_t(s)' F_t^{-1} Q_t(s'),\end{aligned}$$

para  $t = 1, \dots, T$ , donde

$$\begin{aligned}\hat{Y}(t-1) &= (\hat{Y}(s_1, t-1), \dots, \hat{Y}(s_n, t-1))', \\ P_{t|t-1}(s, s') &= \alpha^2 P_{t-1}(s, s') + Cov(\xi(s, t), \xi(s', t)), \\ Q_t(s) &= (P_{t|t-1}(s_1, s), \dots, P_{t|t-1}(s_n, s))', \\ F_t &= (P_{t|t-1}(s_i, s_j))_{n \times n} + \sigma_\epsilon^2 I_n.\end{aligned}$$

El algoritmo del filtro de Kalman se inicializa con  $\hat{Y}(s, 0) = 0$  and  $P_{1|0}(s, s') = Cov(Y(s, t), Y(s', t))$ .

La distribución conjunta de  $Z$  puede escribirse como

$$P(Z) = P(Z(1)) \prod_{t=2}^T P(Z(t) | Z(1), \dots, Z(t-1)),$$

donde

$$Z(t) | Z(1), \dots, Z(t-1) \sim \mathcal{N}(\alpha \hat{Y}(t-1), F_t).$$

Por tanto, la función de log-verosimilitud de  $\theta = (\alpha, b, \nu, \sigma_Y^2, \sigma_\epsilon^2)'$  basada en  $Z = (Z(1)', \dots, Z(T)')'$  vendrá dada por

$$L(\theta | Z) = \log(2\pi)^{-nT/2} - \frac{1}{2} \left\{ \sum_{t=1}^T \left( \log |F_t| + (Z(t) - \alpha \hat{Y}(t-1))' F_t^{-1} (Z(t) - \alpha \hat{Y}(t-1)) \right) \right\}.$$

Observar que  $\{F_t\}$  y  $\{\hat{Y}(t)\}$  dependen del vector de parámetros  $\theta$  y pueden calcularse de forma recursiva mediante el filtro de Kalman. Por tanto, mientras que en Huang y Cressie (1996) se utilizaba el método de los momentos para obtener una estimación de  $\theta$ , nosotros utilizaremos un procedimiento más eficiente basado en la log-verosimilitud  $L(\theta | Z)$  que puede calcularse mediante el filtro de Kalman.

### 5.2.2. Modelo espacio-temporal no separable (Cressie y Huang, 1999)

El segundo de los modelos considerados en este estudio comparativo es el modelo (4.20), que era uno de los modelos no separables introducidos por Cressie y Huang (1999) que hemos estudiado en la Sección 4.5.

En concreto, este modelo de función de covarianza espacio-temporal estacionaria y no separable es un modelo paramétrico que depende de cinco parámetros y que para  $d = 2$  viene dado por

$$C(\mathbf{h}, u) = \begin{cases} \frac{\sigma_Y^2 \beta}{(a^2 u^2 + 1)^\nu (a^2 u^2 + \beta)} & , \text{ si } \|\mathbf{h}\| = 0, \\ \frac{2\sigma_Y^2 \beta}{(a^2 u^2 + 1)^\nu (a^2 u^2 + \beta) \Gamma(\nu)} \left( \frac{b}{2} \left( \frac{a^2 u^2 + 1}{a^2 u^2 + \beta} \right)^{1/2} \|\mathbf{h}\| \right)^\nu & (5.4) \\ \times K_\nu \left( b \left( \frac{a^2 u^2 + 1}{a^2 u^2 + \beta} \right)^{1/2} \|\mathbf{h}\| \right) & , \text{ si } \|\mathbf{h}\| > 0, \end{cases}$$

siendo  $K_\nu$  la función de Bessel modificada de segunda especie de orden  $\nu$ ,  $a \geq 0$  un parámetro de escala temporal,  $b \geq 0$  un parámetro de escala espacial,  $\beta > 0$  un parámetro de interacción espacio-temporal,  $\nu > 0$  un parámetro de suavizado, y  $\sigma_Y^2 = C(\mathbf{0}, 0) > 0$ . La marginal espacial  $C(\mathbf{h}, 0)$  de este modelo (5.4) es una función de covarianza perteneciente al modelo de Matérn que hemos visto en (2.32). Para  $\beta = 1$  obtenemos una función de covarianza separable dada por

$$C(\mathbf{h}, u) = \left( \sigma_Y^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|) \right) \left( \frac{1}{(a^2 u^2 + 1)^{\nu+1}} \right),$$

que es el producto de una función de covarianza espacial de Matérn y una función de correlación temporal.

Bajo este modelo (5.4) para la función de covarianza espacio-temporal, el proceso resultante dependerá de un vector de seis parámetros  $\theta = (a, b, \beta, \nu, \sigma_Y^2, \sigma_\varepsilon^2)'$ , los cinco asociados a la función de covarianza y la varianza  $\sigma_\varepsilon^2 \geq 0$  asociada al error de medida con el que el proceso es observado. La función de log-verosimilitud de  $\theta$  basada en los datos observados  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')$  puede escribirse como

$$L(\theta|\mathbf{Z}) = \log(2\pi)^{-nT/2} - \frac{1}{2}(\log|\Sigma| + \mathbf{Z}^t \Sigma^{-1} \mathbf{Z}), \quad (5.5)$$

siendo  $\Sigma = \text{Var}(\mathbf{Z})$  la matriz cuyo elemento  $(i, j)$ -ésimo viene dado por

$$C(\mathbf{s}_{i^*} - \mathbf{s}_{j^*}, t_i - t_j | \theta) + \sigma_\varepsilon^2 \delta_{ij},$$

con  $i^* = i - n[(i-1)/n]$ ,  $j^* = j - n[(j-1)/n]$ ,  $t_i = [(i-1)/n]$ ,  $t_j = [(j-1)/n]$ , siendo  $[x]$  el mayor entero menor o igual a  $x$ , y  $\delta_{ij}$  la función  $\delta$  de Kronecker.

### 5.2.3. Modelo espacio-temporal no separable (Gneiting, 2002)

El tercero de los modelos considerados pertenece a la familia de modelos no separables de Gneiting (2002), y que hemos estudiado en la Sección 4.7.

En concreto, se ha seleccionado el modelo (4.32), que es un modelo de función de covarianza espacio-temporal estacionario y no separable cuya expresión para  $d = 2$  viene dada por

$$C(\mathbf{h}, u) = \begin{cases} \frac{\sigma_Y^2}{(a|u|^{2\alpha} + 1)^{\delta+\beta}} & , \text{ si } \|\mathbf{h}\| = 0, \\ \frac{\sigma_Y^2}{2^{\nu-1}\Gamma(\nu)(a|u|^{2\alpha} + 1)^{\delta+\beta}} \left( \frac{b\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right)^\nu & (5.6) \\ \times K_\nu \left( \frac{b\|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right) & , \text{ si } \|\mathbf{h}\| > 0, \end{cases}$$

donde  $a > 0$  es un parámetro de escala temporal,  $b > 0$  es un parámetro de escala espacial,  $\alpha \in (0, 1]$  es un parámetro de suavizado temporal,  $\nu > 0$  es un parámetro de suavizado espacial,  $\beta \in [0, 1]$ ,  $\delta \geq 0$  y  $\sigma_Y^2 > 0$ . La marginal espacial  $C(\mathbf{h}, 0)$  asociada a este modelo pertenece a la clase de Matérn (2.33). Para  $\beta = 0$  tenemos una función de covarianza separable dada por

$$C(\mathbf{h}, u) = \left( \sigma_Y^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|) \right) \left( \frac{1}{(a|u|^{2\alpha} + 1)^\delta} \right),$$

que es el producto de una función de covarianza espacial de Matérn y una función de correlación temporal.

Bajo este modelo (5.6), el proceso resultante dependerá de un vector de ocho parámetros  $\theta = (a, b, \alpha, \beta, \nu, \delta, \sigma_Y^2, \sigma_\epsilon^2)'$ , siendo  $\sigma_\epsilon^2 \geq 0$  la varianza del error de medida asociado al proceso observado. La función de log-verosimilitud de  $\theta$  basada en los datos observados  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')$  viene definida por (5.5).

#### 5.2.4. Modelo suma-producto (De Cesare et al., 2001a)

El cuarto de los modelos utilizados en este análisis comparativo mediante simulación corresponde a una modelo no separable basado en la estructura suma-producto introducida por De Cesare et al. (2001a), y que hemos analizado en la Sección 4.6.

En concreto se ha considerado el modelo (4.29) generado a partir de una función de correlación espacial de Matérn  $C_e(\mathbf{h}) = (2^{1-\nu}/\Gamma(\nu))(b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|)$  y una función de autocorrelación de un proceso  $AR(1)$  dada por  $C_t(u) = \alpha^{|u|}$ . Por tanto, el modelo considerado vendría dado por

$$C(\mathbf{h}, u) = k_1 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|) \alpha^{|u|} + k_2 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|) + k_3 \alpha^{|u|},$$

donde  $k_1 > 0$ ,  $k_2 \geq 0$  y  $k_3 \geq 0$  son los pesos de la combinación lineal que define el modelo de covarianza,  $\alpha \in (0, 1]$  es un parámetro de suavización temporal,  $b > 0$  es un parámetro de escala espacial y  $\nu > 0$  un parámetro de suavización espacial. Si denotamos

por  $\sigma_Y^2 = C(\mathbf{0}, 0)$  la varianza a priori del proceso y definimos  $k_i = m_i/\sigma_Y^2$ , para  $i = 1, 2, 3$ , el modelo anterior puede escribirse como

$$C(\mathbf{h}, u) = \sigma_Y^2 (k_1 C_e(\mathbf{h}) C_t(u) + k_2 C_e(\mathbf{h}) + (1 - k_1 - k_2) C_t(u)). \quad (5.7)$$

Bajo este modelo para la función de covarianza espacio-temporal, el proceso resultante dependerá de un vector de siete parámetros  $\theta = (k_1, k_2, \alpha, b, \nu, \sigma_Y^2, \sigma_\varepsilon^2)'$ , donde  $k_1$  y  $k_2$  son los parámetros que definen la combinación lineal que define el modelo de covarianza,  $\alpha \in (0, 1]$  es un parámetro de suavizado temporal,  $b > 0$  es un parámetro de escala espacial,  $\nu > 0$  es un parámetro de suavizado espacial,  $\sigma_Y^2 > 0$  es la varianza a priori del proceso y  $\sigma_\varepsilon^2 \geq 0$  la varianza asociada al error de medida del proceso observado. La función de log-verosimilitud de  $\theta$  basada en los datos observados  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')$  viene definida por (5.5).

### 5.2.5. Modelo suma de productos generalizado

El último de los modelos considerados en este análisis comparativo es el modelo no separable con una estructura de suma de productos que hemos presentado en la Sección 4.12.

En concreto, hemos considerado el modelo (4.77) definido por

$$C(\mathbf{h}, u) = \theta C_{e1}(\mathbf{h}) C_{t1}(u) + (1 - \theta) C_{e2}(\mathbf{h}) C_{t2}(u), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R} \quad (5.8)$$

donde las covarianzas espaciales que lo definen pertenecen ambas al modelo de Matérn con un mismo grado de suavización  $\nu$  y una misma varianza a priori  $\sigma_e^2$ , mientras que las covarianzas temporales pertenecen ambas al modelo exponencial (que es un caso particular del modelo de Matérn para  $\nu = 1/2$ ) con una misma varianza a priori  $\sigma_t^2$ . Luego

$$C_{e1}(\mathbf{h}) = \sigma_e^2 (2^{1-\nu}/\Gamma(\nu)) (\beta_1 \|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\beta_1 \|\mathbf{h}\|), \quad C_{t1}(u) = \sigma_t^2 \exp(-\alpha_1 |u|),$$

$$C_{e2}(\mathbf{h}) = \sigma_e^2 (2^{1-\nu}/\Gamma(\nu)) (\beta_2 \|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\beta_2 \|\mathbf{h}\|), \quad C_{t2}(u) = \sigma_t^2 \exp(-\alpha_2 |u|),$$

donde el parámetro  $\theta$  toma valores en el intervalo definido por (4.72), que depende de los parámetros de los modelos considerados y que hemos descrito en el Ejemplo 4.24.

Tenemos que  $C(\mathbf{0}, 0) = \theta \sigma_e^2 \sigma_t^2 + (1 - \theta) \sigma_e^2 \sigma_t^2 = \sigma_e^2 \sigma_t^2$ . Si denotamos por  $\sigma_Y^2 = C(\mathbf{0}, 0) = \sigma_e^2 \sigma_t^2$  la varianza a priori del proceso, entonces el modelo (5.8) puede escribirse como

$$C(\mathbf{h}, u) = \sigma_Y^2 [\theta (2^{1-\nu}/\Gamma(\nu)) (\beta_1 \|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\beta_1 \|\mathbf{h}\|) \exp(-\alpha_1 |u|) + (1 - \theta) (2^{1-\nu}/\Gamma(\nu)) (\beta_2 \|\mathbf{h}\|)^\nu \mathcal{K}_\nu(\beta_2 \|\mathbf{h}\|) \exp(-\alpha_2 |u|)] \quad (5.9)$$

donde  $\alpha_1 > 0$  y  $\alpha_2 > 0$  son dos parámetros de suavizado temporal,  $\beta_1 > 0$  y  $\beta_2 > 0$  dos parámetros de escala espacial,  $\nu > 0$  un parámetro de suavizado espacial,  $\theta$  el peso en la combinación lineal que define el modelo de covarianza y que tomará valores en un intervalo que depende de los valores del resto de parámetros, y  $\sigma_Y^2 > 0$  la varianza a priori. Bajo este modelo para la función de covarianza espacio-temporal, el proceso resultante dependerá de un vector de ocho parámetros  $\theta = (\alpha_1, \alpha_2, \beta_1, \beta_2, \nu, \theta, \sigma_Y^2, \sigma_\varepsilon^2)'$ , siendo  $\sigma_\varepsilon^2 \geq 0$  la varianza asociada al error de medida del proceso observado. La función de log-verosimilitud de  $\theta$  basada en los datos observados  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')$  viene definida por (5.5).

### 5.3. Estudio de simulación

Como hemos comentado en la Introducción, nuestro objetivo es examinar la versatilidad de los cinco modelos espacio-temporales descritos en la Sección 5.2 al aplicarlos sobre un conjunto suficientemente amplio de simulaciones. Estas simulaciones han sido generadas para muy diversos escenarios, incluyendo tanto funciones de covarianza separables como no separables, diferentes intensidades de dependencia espacial, temporal y espacio-temporal, diferentes grados de suavizado o diferentes niveles de ruido.

Con el fin de considerar la mayor cantidad de diferentes escenarios posibles, para cada uno de los modelos de covarianza se han considerado 16 combinaciones distintas de los parámetros que los definen, por lo que tendremos 16 funciones de covarianza distintas pertenecientes a cada uno de los 5 modelos considerados. Por tanto, tendremos en total 90 escenarios de análisis distintos, uno para cada una de las funciones de covarianza consideradas (que vendrá determinada por el modelo y la combinación paramétrica correspondiente).

#### Parámetros considerados

Se han considerado 16 combinaciones de parámetros diferentes para cada uno de los modelos incluidos en el estudio, tratando de cubrir un espectro de escenarios lo más amplio posible. A continuación se detallan cada una de las combinaciones consideradas:

- Modelo I:  $\alpha \in \{0.2, 0.5\}$ ,  $b \in \{0.4, 0.8\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 \in \{1, 5\}$ .
- Modelo II:  $a = 1$ ,  $b \in \{1, 5\}$ ,  $\beta \in \{0.5, 1, 5\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 \in \{1, 5\}$ .
- Modelo III:  $a = 0.5$ ,  $b \in \{0.3, 1.5\}$ ,  $\alpha \in \{0.1, 0.9\}$ ,  $\beta \in \{0, 0.9\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\delta = 0.5$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 = 1$ .
- Modelo IV:  $(k_1, k_2) \in \{(0.8, 0.1), (0.4, 0.3)\}$ ,  $\alpha \in \{0.4, 0.8\}$ ,  $b \in \{0.3, 1.1\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 = 1$ .
- Modelo V:  $(\alpha_1, \alpha_2) \in \{(0.5, 1.5), (1.5, 0.5)\}$ ,  $\beta_1 \in \{0.3, 1.5\}$ ,  $\beta_2 \in \{0.6, 1.2\}$ ,  $\nu \in \{0.3, 0.8\}$ ,  $\theta = 0.5$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 = 1$ .

Como hemos visto en la Sección 5.2, el Modelo I está definido por una función de covarianza separable obtenida como producto de una función de correlación temporal del modelo AR(1) y una función de covarianza espacial del modelo de Matérn. El modelo resultante depende de cuatro parámetros: un parámetro de dependencia temporal  $\alpha \in (0, 1)$ , un parámetro de escala espacial  $b \geq 0$ , un parámetro de suavidad espacial  $\nu > 0$  y la varianza a priori del proceso  $\sigma^2 = C(0, 0) > 0$ . En este análisis hemos considerado  $\alpha = 0,2$  (0,5) que corresponde a una estructura de dependencia temporal pequeña (grande),  $b = 0,4$  (0,8) que corresponde a un rango espacial grande (pequeño), y  $\nu$  que es un parámetro de suavidad espacial para el que  $\nu = 1$  dará lugar a un proceso más suave que  $\nu = 0,5$ .

Por otro lado, el Modelo II es un modelo paramétrico que depende de cinco parámetros: un parámetro de escala temporal  $a \geq 0$ , un parámetro de escala espacial  $b \geq 0$ , un

parámetro de interacción espacio-temporal  $\beta > 0$  (donde  $\beta = 1$  corresponde a una función de covarianza separable), un parámetro de suavidad espacial  $\nu > 0$  y la varianza a priori del proceso  $\sigma^2 = C(\mathbf{0}, \mathbf{0}) > 0$ . En este análisis hemos considerado  $a = 1$ ,  $b = 1$  (5) que corresponde a un rango espacial grande (pequeño),  $\beta = 0,5$  (1,5) que corresponde a diferentes interacciones espacio-temporales, y  $\nu = 1$  (0,5) que dará lugar a un proceso espacialmente más suave (menos suave). En la Figura 4.4 podemos ver cuatro de las combinaciones paramétricas aquí consideradas.

Respecto al Modelo III, hemos visto que es un modelo de función de covarianza estacionaria que depende de siete parámetros: un parámetro de escala espacial  $a > 0$ , un parámetro de escala temporal  $b > 0$ , un parámetro de suavizado temporal  $\alpha \in (0, 1]$ , un parámetro de suavizado espacial  $\nu > 0$ , un parámetro de interacción espacio-temporal  $\beta \in [0, 1]$  (para  $\beta = 0$  obtenemos una función de covarianza separable), un parámetro  $\delta \geq 0$  y la varianza a priori del proceso  $\sigma_Y^2 > 0$ . En nuestras simulaciones hemos considerado  $b = 0,3$  (1,5) que corresponde a un rango espacial grande (pequeño),  $\alpha = 0,3$  (0,9) que corresponde a una estructura de dependencia temporal grande (pequeña),  $\beta = 0$  (0,9) que corresponde a un modelo separable sin interacción espacio-temporal (con gran interacción espacio-temporal), y  $\nu = 1$  (0,5) que dará lugar a un proceso espacialmente más suave (menos suave). En la Figura 4.6 podemos ver cuatro de las combinaciones paramétricas aquí consideradas.

El Modelo IV es un modelo de función de covarianza estacionaria que depende de seis parámetros: dos parámetros  $k_1$  y  $k_2$  que definen la combinación lineal que genera el modelo de covarianza, un parámetro de suavizado temporal  $\alpha \in (0, 1]$ , un parámetro de escala espacial  $b > 0$ , un parámetro de suavizado espacial  $\nu > 0$  y la varianza a priori del proceso  $\sigma_Y^2 > 0$ . En nuestras simulaciones hemos considerado diferentes valores de  $(k_1, k_2)$  que corresponden a diferentes pesos para las componentes de la mixtura,  $b = 0,3$  (1,5) que corresponde a un rango espacial grande (pequeño),  $\alpha = 0,4$  (0,8) que corresponde a una estructura de dependencia temporal pequeña (grande), y  $\nu = 1$  (0,5) que dará lugar a un proceso espacialmente más suave (menos suave). En la Figura 4.5 podemos ver cuatro de las combinaciones paramétricas aquí consideradas.

Por último, el Modelo V es un modelo paramétrico que depende de siete parámetros: dos parámetros de suavizado temporal  $\alpha_1 > 0$  y  $\alpha_2 > 0$ , dos parámetros de escala espacial  $\beta_1 > 0$  y  $\beta_2 > 0$ , un parámetro de suavizado espacial  $\nu > 0$ , el peso en la combinación lineal que define el modelo de covarianza  $\theta$  y que tomará valores en un intervalo que depende de los valores del resto de parámetros, y la varianza a priori  $\sigma_Y^2 > 0$ . En nuestras simulaciones hemos considerado diferentes valores de  $(\alpha_1, \alpha_2)$  que corresponden a diferentes suavizados temporales en ambas componentes,  $\beta_1 = 0,3$  (1,5) que corresponde a un rango espacial de la primera componente grande (pequeño),  $\beta_2 = 0,6$  (1,2) que corresponde a un rango espacial de la segunda componente grande (pequeño), y  $\nu = 0,3$  (0,8) que dará lugar a un proceso espacialmente menos suave (más suave). En la Figura 4.8 podemos ver cuatro de las combinaciones paramétricas aquí consideradas.

### Simulación de datos

Para este estudio, se han considerado  $N = 100$  localizaciones espaciales formado un retículo de  $10 \times 10$  puntos  $\{\mathbf{s}_i = (x_j, y_k) = (j - 0,5, k - 0,5)/10, \text{ con } j, k = 1, \dots, 10\}$  distribuidos de forma regular dentro del cuadrado unidad  $[0, 1] \times [0, 1]$ . Para cada una de estas localizaciones se han generado datos para  $T = 10$  instantes temporales, por lo que en total cada una de las simulaciones analizadas estará formada por  $N \times T = 1000$  localizaciones espacio-temporales.

Para cada modelo y combinación de parámetros, podemos generar simulaciones de observaciones del proceso  $Y(\mathbf{s}, t)$  sobre el anterior conjunto de localizaciones espacio-temporales mediante el método de simulación directa. Para ello construiremos la matriz de varianzas-covarianzas  $\Sigma$ , cuyo elemento  $(i, j)$ -ésimo viene dado por  $C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)$ , donde  $C(\mathbf{h}, u)$  es la función de covarianza determinada por el modelo y la combinación de parámetros seleccionados. Entonces, dado  $\mathbf{X}$  un vector de  $N \times T$  variables aleatorias independientes gaussianas con media cero y varianza uno, se tiene que  $\mathbf{Y} = \mathbf{X}\Sigma^{1/2}$  es un vector multivariante gaussiano con media cero y matriz de varianzas-covarianzas  $\Sigma$ . Este procedimiento permite obtener simulaciones gaussianas de media cero con una estructura de dependencia espacio-temporal determinada de una manera sencilla y con costes computacionales aceptables debido a que las dimensiones con las que trabajamos no son excesivamente elevadas. En Schlather (1999) pueden consultarse otros procedimientos de simulación.

Vamos a suponer que el proceso de interés  $Y(\mathbf{s}, t)$  no es observado directamente, sino que está afectado por un proceso de ruido blanco aditivo, con lo que el proceso realmente observado sería

$$Z(\mathbf{s}, t) = Y(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \quad (5.10)$$

donde  $\varepsilon(\mathbf{s}, t)$  es un proceso gaussiano de variables independientes de media cero y varianza  $\sigma_\varepsilon^2 > 0$ .

### Ajuste y comparación de modelos

Con el fin de ajustar los cinco modelos a los datos simulados con el procedimiento anterior, se ha seleccionado una muestra aleatoria de  $n = 20$  localizaciones espaciales  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  de las 100 consideradas inicialmente en el cuadrado unidad. Estas 20 localizaciones se seguirán durante los  $T = 10$  tiempos, por lo que en total consideraremos 200 datos. Para estas localizaciones espacio-temporales generaremos los valores  $Z(\mathbf{s}_k, t_k)$  correspondientes, con  $k = 1, \dots, 200$ , mediante la ecuación (5.10).

Para cada una de las simulaciones generadas por el procedimiento anterior, se han ajustado los cinco modelos descritos en la Sección 5.2. Estos ajustes se han llevado a cabo mediante el procedimiento de máxima verosimilitud, que es más eficiente bajo el supuesto de normalidad que el estimador de los momentos utilizado tradicionalmente. Como vimos en la Sección 3.6.2, el estimador máximo verosímil se basa en calcular mediante algún procedimiento de optimización el valor  $\hat{\theta}$  que minimiza la función definida como menos dos veces el logaritmo de la función de verosimilitud del modelo correspondiente, esto es,

consiste en resolver

$$\hat{\theta} = \arg \min_{\theta} -2L(\theta|Z).$$

Para llevar a cabo estas optimizaciones, se ha utilizado en este trabajo de simulación el procedimiento propuesto por Byrd et al. (1995), que es un método quasi-Newton que permite la inclusión de cotas inferiores y superiores de la variación de los parámetros, necesario para muchos de los modelos de covarianza aquí considerados.

Como hemos comentado en la Sección 5.1, uno de los objetivos de este trabajo de simulación es, además de analizar la flexibilidad de los modelos de función de covarianza espacio-temporal considerados, estudiar la eficacia de las herramientas de selección de modelos basados en la minimización de determinados criterios de información, como el criterio de información de Akaike (*AIC*) o el criterio de información bayesiano (*BIC*). Como hemos visto en la Sección 3.6.4, estos criterios se definían como

$$AIC = -2 \log L(\hat{\theta}|Z) + 2p, \quad BIC = -2 \log L(\hat{\theta}|Z) + \log(n \times T)p,$$

siendo  $p$  el número de parámetros del vector  $\theta$ .

Para comparar la capacidad predictiva de cada uno de los modelos anteriores, hemos calculado las predicciones obtenidas mediante kriging simple y las varianzas del error de predicción correspondientes. Estos valores vienen dados por (5.1) y (5.2), esto es,

$$\hat{Y}(s, T) = Cov(Y(s, T), Z)Var(Z)^{-1}Z,$$

$$Var(Y(s, T)) - Cov(Y(s, T), Z)Var(Z)^{-1}Cov(Z, Y(s, T)).$$

A partir de los resultados obtenidos para cada uno de los métodos se ha calculado el error de predicción cuadrático medio (*EPCM*) para el último de los tiempos observados, el tiempo 10, que vendrá dado por

$$EPCM = \frac{1}{100} \sum_{i=1}^{10} \sum_{j=1}^{10} \left( \hat{Y}(s(i, j), 10) - Y(s(i, j), 10) \right)^2. \quad (5.11)$$

En cada una de las simulaciones, además del valor del *EPCM* asociado al ajuste de los cinco modelos de covarianza espacio-temporal de la Sección 5.2, se ha definido el asociado al Modelo *AIC* y el asociado al Modelo *BIC* como el *EPCM* de aquel modelo cuyo ajuste tenga un menor *AIC* y *BIC*, respectivamente. Esto nos permitirá comparar los errores de predicción de los cinco modelos considerados, y el que se obtendría si se incluyese un criterio de selección de modelo automático basado en el valor del *AIC* o *BIC* de los modelos ajustados. Además, para cada simulación también se ha calculado también el *EPCM* asociado al modelo definido por la función de covarianza generadora de los datos, que llamaremos Modelo *T* (True). En resumen, para cada simulación tendremos los cinco errores correspondientes al ajuste de los cinco modelos de covarianza considerados, los dos errores asociados a los criterios *AIC* y *BIC*, y el error asociado al modelo real *T* que ha generado los datos.

Hay que tener en cuenta que, incluso aunque el proceso de optimización se inicie con unos parámetros iniciales cercanos a los óptimos, podríamos quedar atrapados en

un óptimo local en lugar del óptimo global, dando lugar a valores *EPCM* inusualmente elevados. Con el fin de eliminar la posible influencia de estos valores anómalos, se han eliminado aquellas simulaciones en las que el valor del *EPCM* asociado a alguno de los cinco modelos ajustados fuese 10 veces mayor al valor mínimo observado para estos cinco modelos. Finalmente se han seleccionado 100 simulaciones para cada uno de estos escenarios que cumpliesen este criterio.

## 5.4. Resultados

Los resultados, expresados mediante medias y errores típicos de los valores del *EPCM* obtenidos a partir de las 100 réplicas simuladas para las 16 combinaciones de parámetros de cada uno de los 5 modelos, se resumen en las Figuras 5.1, 5.2, 5.3, 5.4 y 5.5.

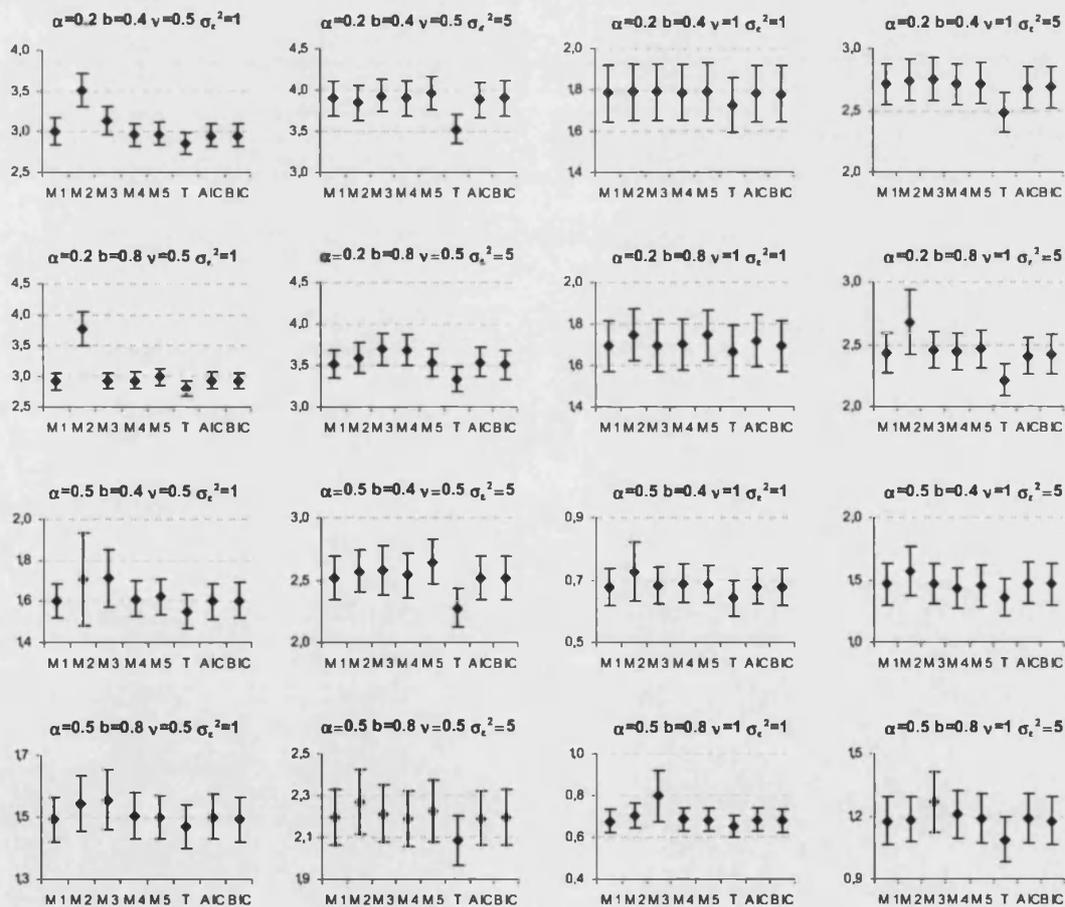


Figura 5.1: Intervalo de confianza al 95 % para la media de los valores del *EPCM* basado en las 100 réplicas de datos generados con el Modelo I para las 16 combinaciones de parámetros utilizadas.

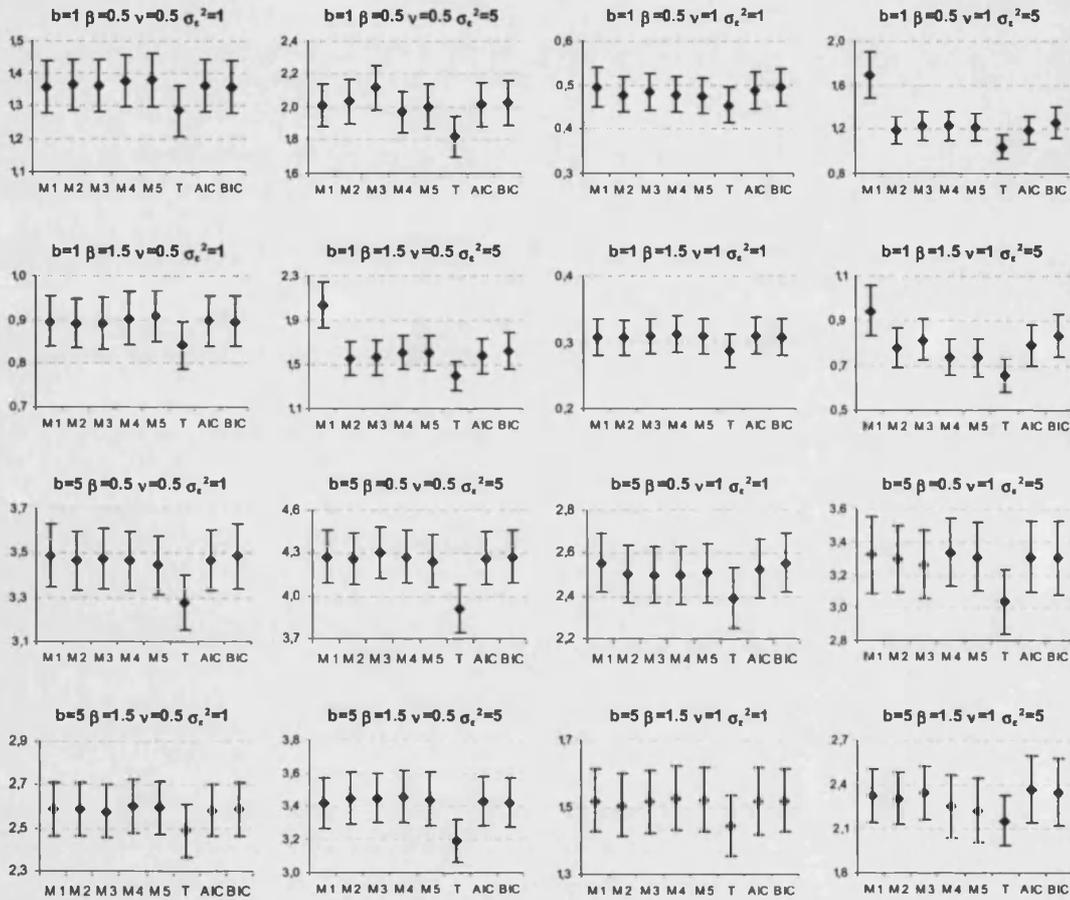


Figura 5.2: Intervalo de confianza al 95 % para la media de los valores del *EPCM* basado en las 100 réplicas de datos generados con el Modelo II para las 16 combinaciones de parámetros utilizadas.

La Tabla 5.1 resume las gráficas anteriores al recoger las medias, medianas y desviaciones típicas del *EPCM* de todas simulaciones de un mismo modelo, independientemente de la combinación de parámetros que la generó. Es decir, se han considerado conjuntamente los valores del *EPCM* en las  $1600 = 100 \times 16$  simulaciones generadas con cada modelo, cuando se predice a partir de las modelizaciones llevadas a cabo con los 5 modelos en estudio, incluido el que las generó. La tabla incluye tres columnas más, la encabezada por *T* que se refiere al verdadero modelo que utiliza exactamente los valores de los parámetros con los que se simularon los datos. Para la obtención de las estadísticos descriptivos del *EPCM* asociados a las columnas encabezadas por *AIC* y *BIC*, respectivamente, se han utilizado en cada simulación los modelos con menor valor para los criterios *AIC* y *BIC*, lo que puede implicar cambio de modelo para cada criterio y para cada simulación.

Las Figuras 5.6, 5.7, 5.8, 5.9 y 5.10 muestran, a modo de ejemplo, las funciones de

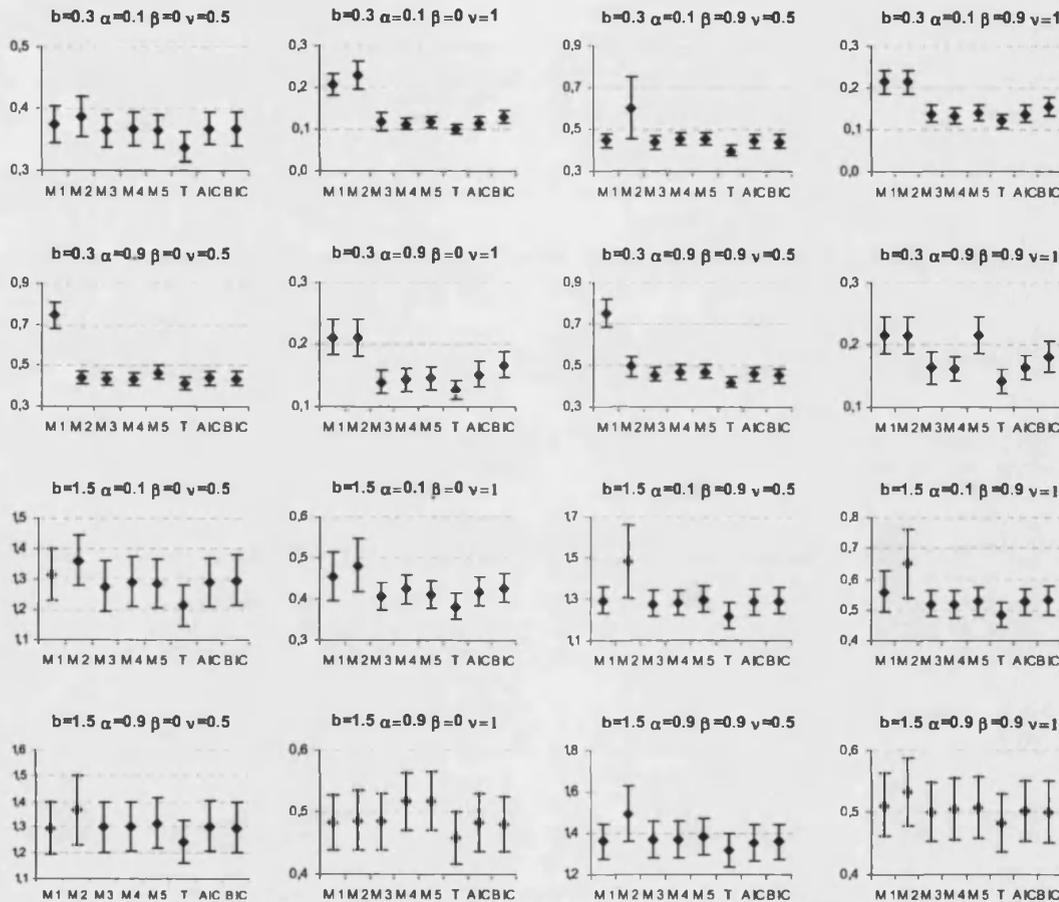


Figura 5.3: Intervalo de confianza al 95 % para la media de los valores del *EPCM* basado en las 100 réplicas de datos generados con el Modelo III para las 16 combinaciones de parámetros utilizadas.

covarianza espacio-temporal que han generado los datos para cada uno de los cinco modelos con una de las 16 combinaciones de parámetros consideradas. En ellas podemos ver, tanto la forma de la covarianza generadora de los datos como la forma de las covarianzas ajustadas para cada uno de los cinco modelos. En concreto, se representa la covarianza de los cinco modelos que tiene como parámetros la mediana de los parámetros ajustados para las 100 simulaciones consideradas.

La primera de las gráficas de la Figura 5.6 muestra la forma de la función de covarianza del Modelo I para una de las combinaciones de parámetros consideradas, la correspondiente a los valores  $\alpha=0.5$ ,  $b=0.8$ ,  $\nu=0.5$  y  $\sigma_Y^2=5$ . Como ya hemos comentado en la Sección 5.3, con esta función de covarianza se han generado 100 simulaciones, ajustando para los datos de cada una de ellas los cinco modelos de covarianza incluidos en este estudio de simulación. En esta Figura 5.6 se muestra la función de covarianza de cada

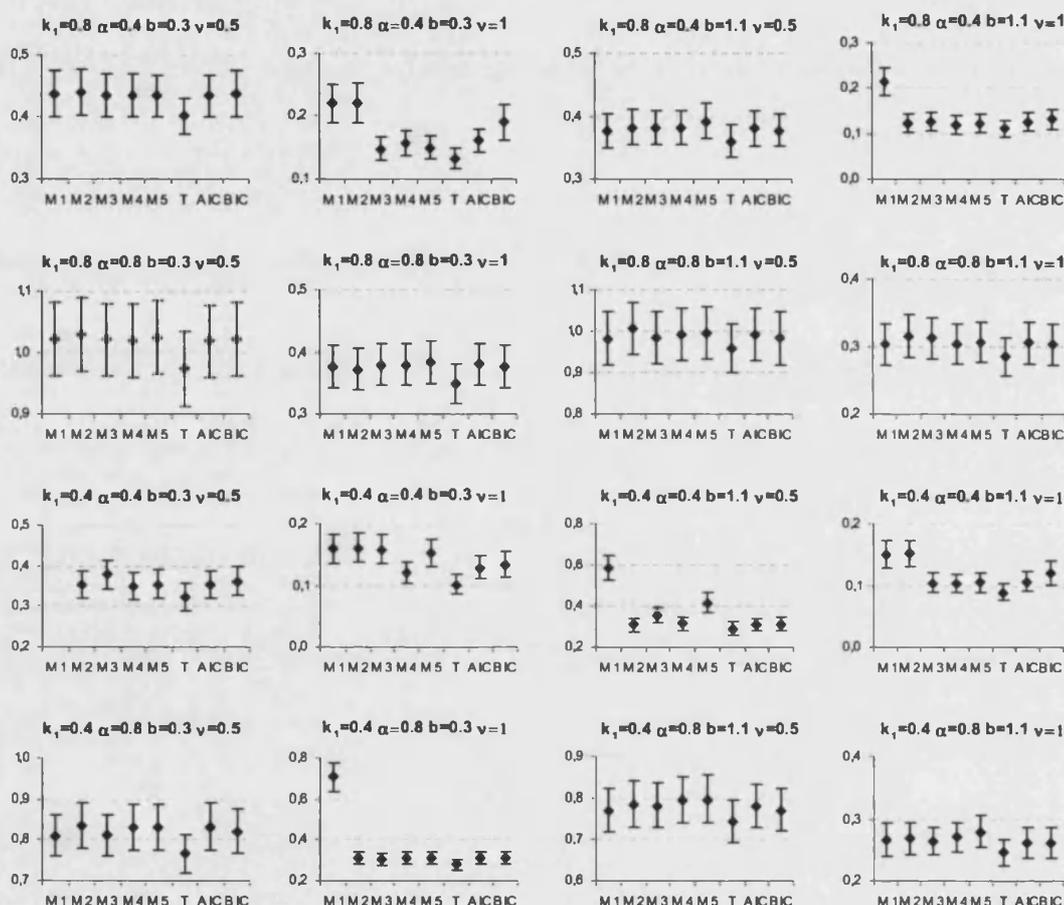


Figura 5.4: Intervalo de confianza al 95 % para la media de los valores del  $EPCM$  basado en las 100 réplicas de datos generados con el Modelo IV para las 16 combinaciones de parámetros utilizadas.

uno de estos modelos con parámetros la mediana de los parámetros ajustados para las 100 simulaciones. Estas medianas de los parámetros ajustados son  $\alpha=0.4799$ ,  $b=0.8112$ ,  $\nu=0.5527$  y  $\sigma_Y^2=1.7960$  para el Modelo I;  $a=0.5881$ ,  $b=12.3275$ ,  $\beta=31.0825$ ,  $\nu=0.5574$  y  $\sigma_Y^2=4.3061$  para el Modelo II;  $a=0.0134$ ,  $b=1.8840$ ,  $\alpha=0.5185$ ,  $\beta=0.0002$ ,  $\nu=0.4979$ ,  $\delta=15.5960$  y  $\sigma_Y^2=5.1496$  para el Modelo III;  $k_1=0.9341$ ,  $k_2=0.0659$ ,  $\alpha=0.7612$ ,  $b=2.9143$ ,  $\nu=0.5855$  y  $\sigma_Y^2=4.0459$  para el Modelo IV;  $\alpha_1=0.1911$ ,  $\alpha_2=19.3065$ ,  $\beta_1=2.7665$ ,  $\beta_2=0.8361$ ,  $\nu=0.6416$ ,  $\theta=0.9881$  y  $\sigma_Y^2=4.8699$  para el Modelo V. Como vemos, para esta combinación de parámetros, el Modelo III ha captado perfectamente el carácter separable de la estructura de covarianza generadora de los datos ( $\beta \approx 0$ ). Lo mismo ocurre con el Modelo IV ( $k_3 \approx 0$ ) y con el modelo V ( $\theta \approx 1$ ). No obstante, en todos los modelos se ha estimado una dependencia temporal mayor que la que ha generado los datos.

En la Figura 5.7 se muestra esto mismos gráficos para una de las combinaciones gene-

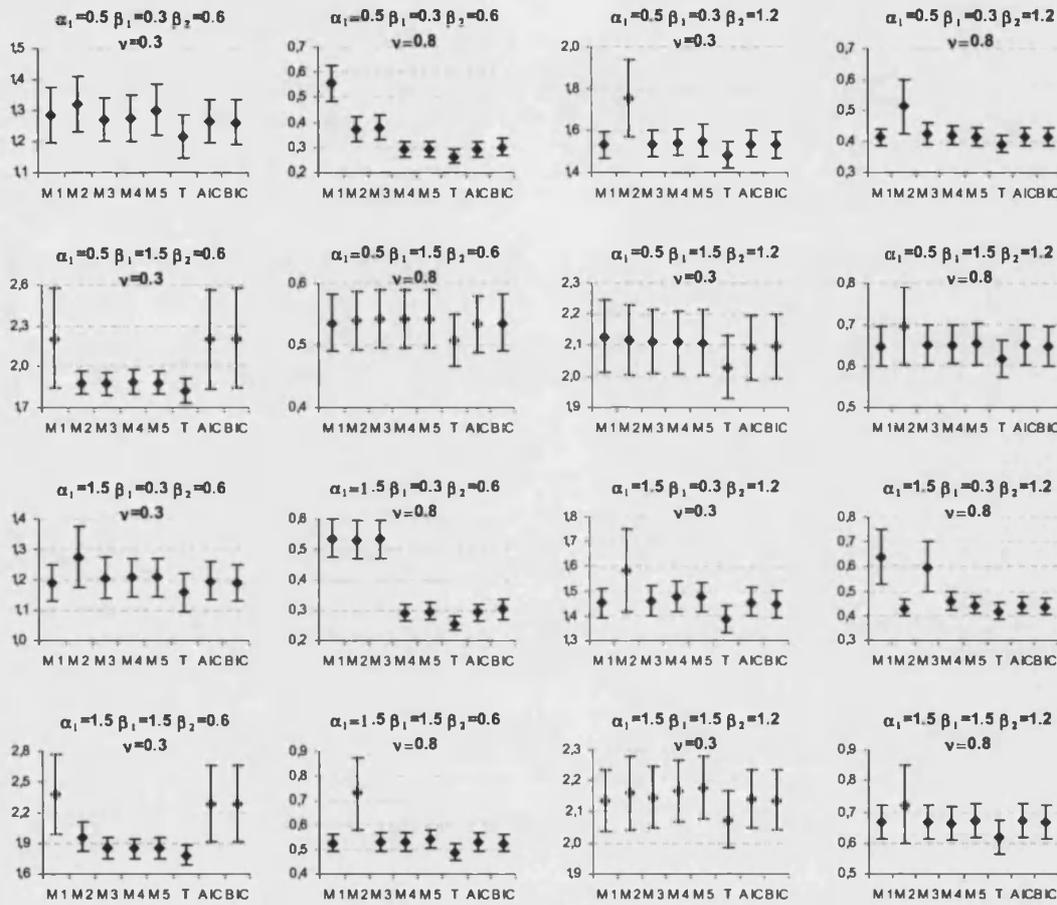


Figura 5.5: Intervalo de confianza al 95 % para la media de los valores del *EPCM* basado en las 100 réplicas de datos generados con el Modelo V para las 16 combinaciones de parámetros utilizadas.

radoras de datos correspondientes al Modelo II, en concreto para  $a=1$ ,  $b=5$ ,  $\beta=1.5$ ,  $\nu=0.5$  y  $\sigma_Y^2=5$ . En este caso, las medianas de los parámetros ajustados para las simulaciones generadas con la combinación anterior son  $\alpha=0.4080$ ,  $b=0.2397$ ,  $\nu=0.4652$  y  $\sigma_Y^2=4.4142$  para el Modelo I;  $a=0.7526$ ,  $b=6.0003$ ,  $\beta=1.3472$ ,  $\nu=0.5738$  y  $\sigma_Y^2=5.0946$  para el Modelo II;  $a=0.1252$ ,  $b=4.2405$ ,  $\alpha=0.9633$ ,  $\beta=0.0001$ ,  $\nu=0.4804$ ,  $\delta=6.4818$  y  $\sigma_Y^2=5.4383$  para el Modelo III;  $k_1=0.9998$ ,  $k_2=0.0001$ ,  $\alpha=0.3803$ ,  $b=5.5659$ ,  $\nu=0.5136$  y  $\sigma_Y^2=5.0170$  para el Modelo IV;  $\alpha_1=1.4449$ ,  $\alpha_2=0.7407$ ,  $\beta_1=8.1323$ ,  $\beta_2=4.3284$ ,  $\nu=0.5922$ ,  $\theta=0.0001$  y  $\sigma_Y^2=5.4547$  para el Modelo V. Como vemos, para el ajuste de esta combinación del Modelo II no separable ( $\beta=1.5$ ) los modelos III, IV y V han utilizado combinaciones separables.

La Figura 5.8 muestra la forma de la función de covarianza del Modelo III para una de las 16 combinaciones consideradas, en concreto para  $a=0.5$ ,  $b=1.5$ ,  $\alpha=0.1$ ,  $\beta=0.9$ ,  $\nu=0.5$ ,  $\delta=0.5$  y  $\sigma^2=5$ . También aparece la forma de las covarianzas para la mediana de los

Teórico	EPCM	M1	M2	M3	M4	M5	T	AIC	BIC
Mod1	Media	2,111	2,248	2,167	2,124	2,137	1,983	2,109	2,105
	Mediana	2,016	2,094	2,048	2,023	2,031	1,888	2,000	2,009
	D.Típ.	0,677	0,853	0,749	0,671	0,681	0,608	0,673	0,669
Mod2	Media	2,077	1,999	2,013	2,002	1,996	1,856	2,007	2,019
	Mediana	1,956	1,884	1,909	1,899	1,886	1,763	1,891	1,908
	D.Típ.	0,677	0,610	0,614	0,623	0,629	0,563	0,632	0,648
Mod3	Media	0,653	0,667	0,587	0,594	0,603	0,554	0,591	0,596
	Mediana	0,597	0,582	0,539	0,548	0,560	0,513	0,549	0,552
	D.Típ.	0,275	0,381	0,226	0,224	0,227	0,201	0,222	0,228
Mod4	Media	0,499	0,442	0,434	0,431	0,441	0,401	0,430	0,433
	Mediana	0,447	0,401	0,396	0,394	0,402	0,368	0,394	0,392
	D.Típ.	0,222	0,187	0,177	0,176	0,185	0,163	0,176	0,183
Mod5	Media	1,177	1,162	1,112	1,086	1,089	1,033	1,126	1,125
	Mediana	1,062	1,059	1,050	1,041	1,038	0,982	1,037	1,034
	D.Típ.	0,541	0,525	0,345	0,307	0,315	0,292	0,477	0,479
Total	Media	1,303	1,304	1,263	1,247	1,253	1,165	1,253	1,256
	Mediana	1,216	1,204	1,189	1,181	1,183	1,103	1,174	1,179
	D.Típ.	0,478	0,511	0,422	0,400	0,407	0,366	0,436	0,441

Tabla 5.1: Media del EPCM para todas las simulaciones y todas las combinaciones de cada uno de los modelos.

parámetros ajustados en cada modelo. Estas medianas son  $\alpha=0.6764$ ,  $b=0.4576$ ,  $\nu=0.7613$  y  $\sigma^2=2.5806$  para el Modelo I;  $a=0.6937$ ,  $b=2.1874$ ,  $\beta=6.3380$ ,  $\nu=0.5918$  y  $\sigma^2=4.4725$  para el Modelo II;  $a=0.8724$ ,  $b=1.8802$ ,  $\alpha=0.1063$ ,  $\beta=0.0556$ ,  $\nu=0.5453$ ,  $\delta=0.1451$  y  $\sigma^2=4.7657$  para el Modelo III;  $k_1=0.5926$ ,  $k_2=0.4074$ ,  $\alpha=0.1322$ ,  $b=2.8255$ ,  $\nu=0.7819$  y  $\sigma^2=3.9280$  para el Modelo IV;  $\alpha_1=0.0292$ ,  $\alpha_2=10.8202$ ,  $\beta_1=2.1312$ ,  $\beta_2=2.8147$ ,  $\nu=0.6884$ ,  $\theta=0.5658$  y  $\sigma^2=4.8426$  para el Modelo V. En este caso se ha generado una función con mucha interacción espacio-temporal ( $\beta \approx 1$ ), que ni el Modelo I ni el Modelo II parecen haber ajustado correctamente. Se observa que la estimación del parámetro de interacción espacio-temporal del propio Modelo III no ha sido muy satisfactoria para esta combinación.

La Figura 5.9 muestra la forma de la función de covarianza del Modelo IV para  $k_1=0.8$ ,  $k_2=0.1$ ,  $\alpha=0.4$ ,  $b=1.1$ ,  $\nu=0.5$  y  $\sigma_Y^2=5$ . También muestra las covarianzas obtenidas con la mediana de los parámetros ajustados en las 100 simulaciones obtenidas con esta función de covarianza. Estas medianas, para esta combinación de parámetros, son  $\alpha=0.4706$ ,  $b=0.8879$ ,  $\nu=0.4533$  y  $\sigma_Y^2=3.6439$  para el Modelo I;  $a=0.6082$ ,  $b=0.7076$ ,  $\beta=0.5440$ ,  $\nu=0.4856$  y  $\sigma_Y^2=4.7516$  para el Modelo II;  $a=0.5797$ ,  $b=1.0121$ ,  $\alpha=0.5738$ ,  $\beta=0.0001$ ,  $\nu=0.4348$ ,  $\delta=1.4239$  y  $\sigma_Y^2=5.1242$  para el Modelo III;  $k_1=0.8315$ ,  $k_2=0.0799$ ,  $\alpha=0.3622$ ,  $b=2.3748$ ,  $\nu=0.6489$  y  $\sigma_Y^2=3.6801$  para el Modelo IV;  $\alpha_1=0.3184$ ,  $\alpha_2=3.5763$ ,  $\beta_1=1.2609$ ,  $\beta_2=1.3775$ ,  $\nu=0.5695$ ,  $\theta=0.6333$  y  $\sigma_Y^2=4.7004$  para el Modelo V. Como vemos, parece que todos los modelos han generado covarianzas similares a la que ha originado los datos, que era próxima a una covarianza separable ( $k_2 = k_3 = 0,1 \approx 0$ ) (de hecho, incluso el Modelo III ha utilizado una covarianza separable para su modelización).

La última de las figuras de esta serie, la Figura 5.10, muestra la forma de la función de

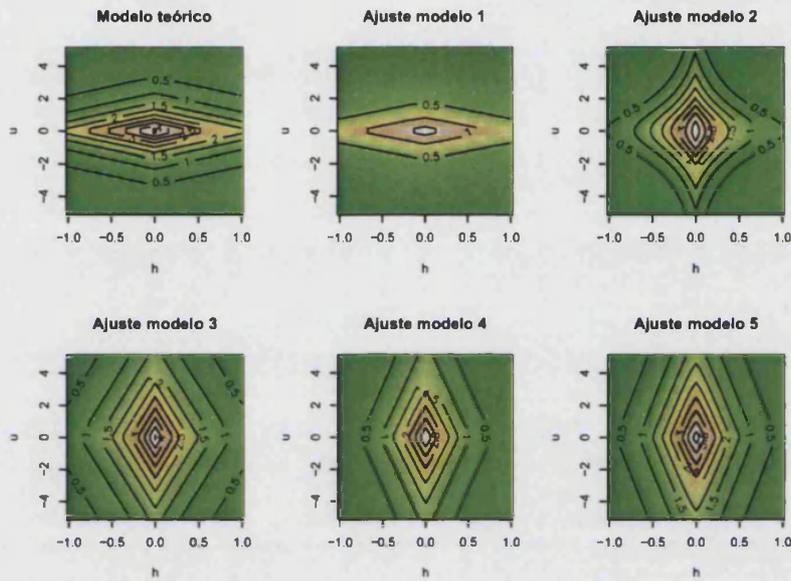


Figura 5.6: Ejemplo de función de covarianza teórica del Modelo I y funciones ajustadas con los cinco modelos considerados.

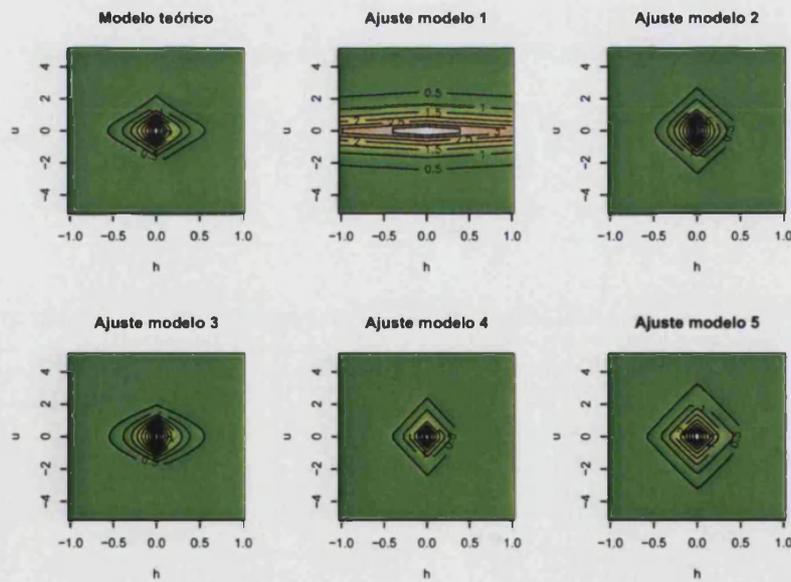


Figura 5.7: Ejemplo de función de covarianza teórica del Modelo II y funciones ajustadas con los cinco modelos considerados.

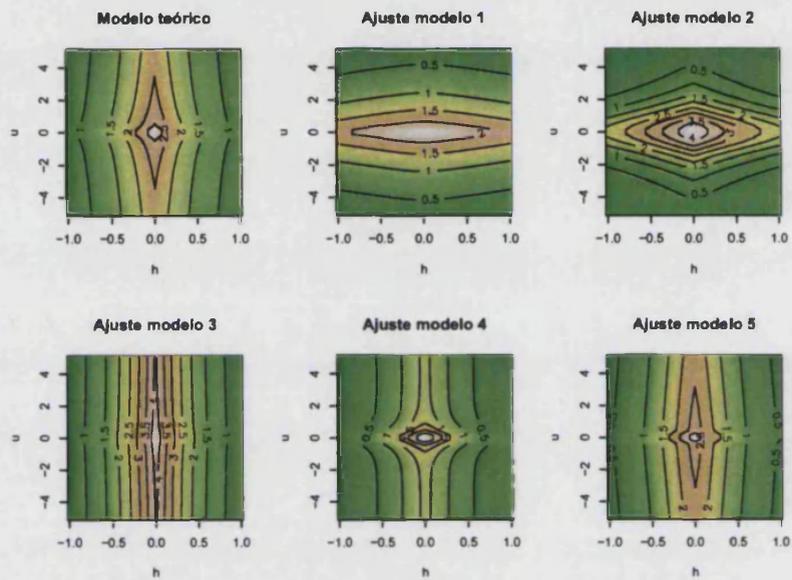


Figura 5.8: Ejemplo de función de covarianza teórica del Modelo III y funciones ajustadas con los cinco modelos considerados.

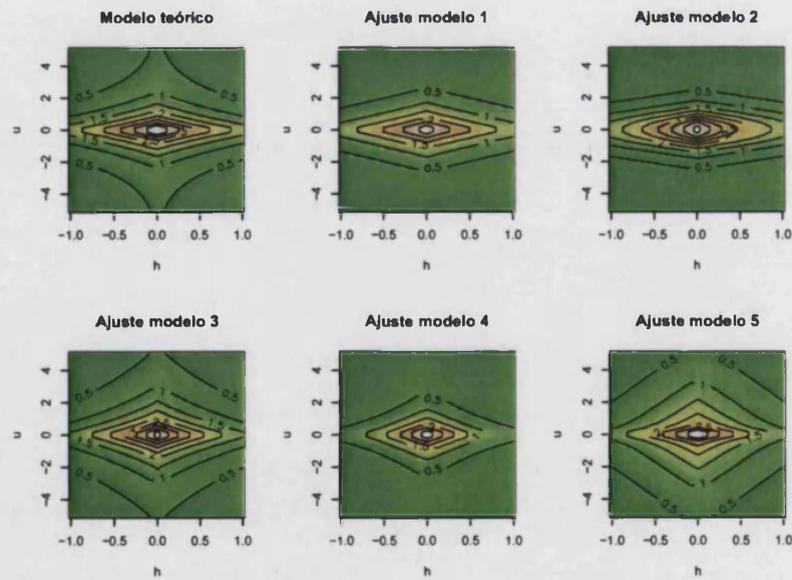


Figura 5.9: Ejemplo de función de covarianza teórica del Modelo IV y funciones ajustadas con los cinco modelos considerados.

covarianza del modelo (5.9) para  $\alpha_1=0.5$ ,  $\alpha_2=1.5$ ,  $\beta_1=1.5$ ,  $\beta_2=1.2$ ,  $\nu=0.3$ ,  $\theta=0.5$  y  $\sigma_Y^2=5$ . También muestra las covarianzas obtenidas con la mediana de los parámetros ajustados en las 100 simulaciones obtenidas con esta función de covarianza. Estas medianas, para esta combinación de parámetros, son  $\alpha=0.4313$ ,  $b=0.6237$ ,  $\nu=0.3273$  y  $\sigma_Y^2=4.0669$  para el Modelo I;  $a=1.0954$ ,  $b=2.6664$ ,  $\beta=3.4549$ ,  $\nu=0.3668$  y  $\sigma_Y^2=4.5995$  para el Modelo II;  $a=0.1344$ ,  $b=1.5843$ ,  $\alpha=0.6021$ ,  $\beta=0.0022$ ,  $\nu=0.3187$ ,  $\delta=5.8077$  y  $\sigma_Y^2=5.1302$  para el Modelo III;  $k_1=0.9764$ ,  $k_2=0.0120$ ,  $\alpha=0.3944$ ,  $b=2.6240$ ,  $\nu=0.3836$  y  $\sigma_Y^2=4.1567$  para el Modelo IV;  $\alpha_1=0.5188$ ,  $\alpha_2=1.5709$ ,  $\beta_1=2.2367$ ,  $\beta_2=1.6608$ ,  $\nu=0.3972$ ,  $\theta=0.5293$  y  $\sigma_Y^2=4.9790$  para el Modelo V. Como vemos, excepto el Modelo I, el resto de modelos han ajustado correctamente el comportamiento de la función teórica.

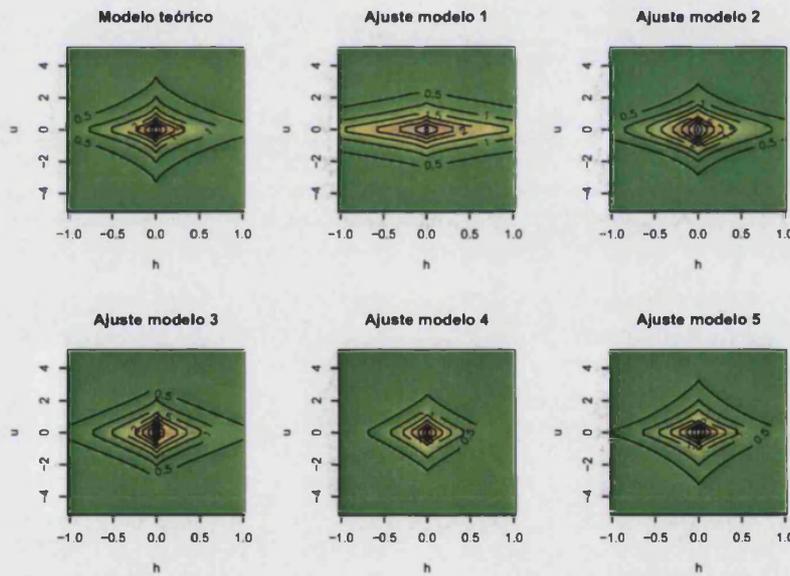


Figura 5.10: Ejemplo de función de covarianza teórica del Modelo V y funciones ajustadas con los cinco modelos considerados.

## 5.5. Conclusiones

La modelización de procesos espacio-temporales he tenido un gran auge en los últimos años, sin embargo es muy poco lo que se ha hecho acerca de las consecuencias de utilizar distintos tipos de modelos. Con el objetivo de aportar algún resultado a este problema, en un primer trabajo (Huang et al., 2007) presentamos un estudio de simulación que comparaba la habilidad predictiva de cuatro clases distintas de modelos espacio-temporales, utilizando los valores AIC y BIC como criterios de selección.

En este capítulo hemos completado aquel estudio añadiendo un quinto modelo, el modelo suma de productos que hemos introducido en la Sección 4.12, utilizando la combinación de parámetros detallada en la Sección 5.3.



Desde el punto de vista de modelizar en la práctica un conjunto de observaciones provenientes de un fenómeno espacio-temporal, sugerimos como primera acción algún tipo de análisis exploratorio de los datos, con el fin de comprobar su estacionariedad espacial y temporal. Conseguida ésta, si fuera necesario mediante la eliminación de la tendencia a gran escala, caso de estar presente, la siguiente etapa consiste en contrastar la separabilidad de la función de covarianza espacio-temporal. Para ello puede bastar una simple inspección visual de la covarianza empírica o, más formalmente, se puede utilizar alguno de los test citados en la Sección 4.4. Si la separabilidad es aceptable, el Modelo I es una buena elección, probando diferentes modelos de covarianzas espaciales y temporales. Si un modelo separable no fuera apropiado, conviene recurrir a alguno de los cuatro restantes, utilizando los criterios AIC o BIC para elegir el modelo final.

De acuerdo con los resultados obtenidos tanto en el trabajo original (Huang et al., 2007) como en la extensión desarrollada en este capítulo, podemos destacar, a modo de conclusión, los siguientes puntos:

1. La intensidad del suavizado y el nivel de ruido  $\sigma_\varepsilon^2$  del proceso espacio-temporal juegan un papel crucial en la predicción. En particular, el ajuste de un modelo teórico presentará menos dificultades en procesos muy suavizados y/o con poco nivel de ruido.
2. Una estructura separable puede proporcionar buenas predicciones incluso para datos generados con estructuras no separables. Ello es debido a la flexibilidad intrínseca a aquel tipo de estructuras. Por ejemplo, la combinación de una covarianza espacial del tipo Matérn y un modelo ARMA para la componente temporal puede dar buenos resultados.
3. Como era de esperar, y los gráficos y tablas anteriores confirman, la predicción basada en el verdadero modelo se comporta siempre mejor que la basada en un modelo ajustado.
4. Sin embargo, por término medio, las predicciones basadas en el modelo ajustado con la clase correcta son mejores que las obtenidas con los otros modelos. La excepción a esta afirmación serían los Modelos 2 y 5, para los cuales los *EPCM* medios obtenidos con los Modelos 5 y 4, respectivamente, son ligeramente mejores (véase la Tabla 5.1).
5. Por lo respecta a los valores de AIC y BIC, son prácticamente iguales en todos los modelos, con una ligera ventaja para el BIC en los Modelos 1 y 5, y para el AIC en los restantes. Sus *EPCM*, en todo caso, no están alejados de los obtenidos para los verdaderos modelos, lo que resulta prometedor a la hora de utilizar estos estadísticos como criterio de selección. Hemos de insistir, no obstante, en lo poco que sabemos sobre su comportamiento y sus propiedades asintóticas en el contexto espacial, un aspecto que requiere la atención de futuras investigaciones. futuro.
6. El proceso espacial subyacente raramente se conoce en la práctica. Así, es conveniente probar varios modelos después de llevar a cabo un análisis exploratorio de los datos,

eligiendo en función de algunas de las medidas de bondad sugeridas o utilizando como criterio el AIC o el BIC.

7. En términos de tiempo de computación y de capacidad de manejo de grandes conjunto de datos, un modelo separable y flexible, tal como el Modelo I, sería la opción. Sin embargo, si se sospecha que la no separabilidad está presente, el Modelo 2 ofrece una estructura no separable, flexible y de fácil cómputo. Pero si el tiempo de cálculo no es problema porque, como ha sido nuestro caso, se tiene acceso a un máquina de cálculo intensivo, modelos más complejos, como el 4 y el 5, ofrecen mejores resultados y se adaptan a una mayor diversidad de situaciones.



## Capítulo 6

# Modelización de datos reales

### 6.1. Introducción

La aplicación a problemas reales de las técnicas descritas en los capítulos anteriores se recogen en una colección de artículos que se reproducen a continuación. Los dos primeros se centran en el análisis de diferentes procesos naturales espaciales, mientras que los tres últimos generalizan los análisis a procesos espacio-temporales.

La Sección 6.2 reproduce el primero de estos trabajos, publicado en Martínez et al. (2003). En él se estudia la evolución espacial de la superficie piezométrica del acuífero de Xàbia a partir de medidas de nivel piezométrico, haciendo intervenir como covariables algunas características físico-químicas del agua subterránea.

El segundo de los trabajos, que se reproduce en la Sección 6.3, está aceptado y pendiente de publicación en Martínez et al. (2007). Es un análisis del efecto de la reducción de los puntos muestrales mediante diferentes esquemas de muestreo sobre un determinado conjunto de datos de radiación solar en superficie, al que se le aplican diferentes modelos geoestadísticos que utilizan la altitud del terreno como información complementaria.

La tercera de las aplicaciones que aquí se presenta corresponde al artículo publicado en Huang et al. (2007), que se reproduce en la Sección 6.4. Además del estudio de simulación que se amplía en el Capítulo 5, incluye un análisis de la radiación solar en superficie sobre una determinada región de la Comunidad Valenciana sobre la que se han aplicado diferentes modelos de covarianza espacio-temporal y criterios de selección de modelos, analizando la capacidad predictiva de cada uno de ellos.

La Sección 6.5 reproduce el trabajo publicado en Bodas et al. (2003), en el que se muestra una aplicación de las técnicas espacio-temporales para la modelización y predicción de la radiación solar en superficie, utilizando las mediciones llevadas a cabo por un satélite meteorológico sobre una región de la Comunidad Valenciana durante un periodo de tiempo determinado.

Por último, la Sección 6.6 reproduce el trabajo enviado a Debón et al. (2007), que presenta un novedoso enfoque en el análisis de las tablas de mortalidad dinámicas aplicando técnicas de estadística espacio-temporal, modelizando la estructura de dependencia existente entre las celdas de las tablas analizadas.

El resultado de este proceso es el deterioro progresivo de la calidad del agua del acuífero que se manifiesta por un rápido incremento de la salinidad y que puede reducir considerablemente los recursos disponibles. Esta situación, muy frecuente en el litoral mediterráneo español, es la principal causa del déficit hídrico que afecta en distinto grado a la mayor parte de los acuíferos costeros y que obliga a actuaciones de regeneración (disminución de los bombeos, recarga artificial) o de búsqueda de recursos alternativos (reutilización de aguas residuales, desalinización). El acuífero de Xàbia es un buen ejemplo de esta problemática, en el que la sobreexplotación y la intrusión marina se conocen desde hace décadas (Pulido-Bosch, 1976).

De acuerdo con todo lo anterior, se puede aceptar que existe una relación primaria entre las variaciones de nivel piezométrico y las características físico-químicas del agua subterránea, si bien estas características no sólo dependen de factores hidrodinámicos sino que están fuertemente influenciadas por procesos hidroquímicos modificadores.

El objetivo de este estudio es conocer la evolución espacio-temporal de la superficie piezométrica del acuífero de Xàbia a partir de medidas de nivel piezométrico haciendo intervenir como covariables algunas características físico-químicas del agua subterránea. Disponemos para ello de observaciones de estas variables en un conjunto de pozos repartidos a lo largo de la superficie del acuífero (ver Figura 2). En este tipo de estudios, la modelización matemático-estadística constituye una herramienta adecuada para analizar la evolución del acuífero y proponer las actuaciones adecuadas sobre el mismo. Así, el conocimiento detallado de la superficie piezométrica del acuífero es esencial para determinar posteriores modelos matemáticos de ecuaciones de flujo o transporte. Y son los modelos estadísticos espaciales y espacio-temporales los que nos permitirán predecir adecuadamente los valores de los niveles piezométricos en cualquier lugar de la superficie del acuífero. En cualquier caso, en este trabajo nos restringimos a los modelos estadísticos, provenientes de la rama de la geoestadística, sin considerar modelos matemáticos, de que seguro se beneficiarán de los análisis propuestos.

El presente estudio se estructura de la siguiente forma: la Sección 2 está dedicada a la descripción de los datos: localización del acuífero y de los pozos, variables y tiempos; la Sección 3 contiene una breve exposición teórica de los modelos espacio-temporales; la Sección 4 muestra los resultados de la aplicación de los distintos modelos a los datos del acuífero, y las conclusiones que de ellos y de su comparación se derivan constituyen la última Sección.

## El acuífero de Xàbia

El acuífero cuaternario de Xàbia pertenece al subsistema *Peñón-Montgó-Bernia-Benissa*, que se enmarca en el sistema denominado *Prebético de Valencia y Alicante* (IGME, 1988). Según la nomenclatura del Plan Hidrológico del Júcar (1977), el acuífero pertenece a la Unidad Hidrogeológica nº 47 del Sistema de Explotación Marina Alta (Figura 1). Se localiza en el extremo nororiental de la provincia de Alicante, en la Marina Alta, encajado entre la Sierra de Benitachell, Castell de la Solana y el Cabo San Martín en la zona sur, y el Macizo del Montgó y el Cabo San Antonio en la zona norte. Está formado por materiales detríticos cuaternarios heterogéneos, gravas, arenas y limos de origen fluvial, con granulometrías muy variadas que van desde bloques a conglomerados heterométricos (DPA, 1999). Información detallada acerca de la climatología de la zona en el período de estudio, así como de las características geológicas e hidrogeológicas pueden encontrarse en Blasco (1988).

El acuífero es libre con valores altos de transmisividad, que oscilan entre 750 y 7500 m<sup>2</sup>/día (Pulido-Bosch, 1976) y con cotas piezométricas que, en general, van desde -3 mb.n.m a 30 ms.n.m., si bien han variado con el tiempo (Armayer et al., 2000). Hay que hacer notar que disponemos de algunos pozos que muestran niveles piezométricos de hasta 80 ms.n.m., los cuáles al ser incluidos en nuestro modelo, evidentemente influirán en las predicciones. El gradiente hidráulico presenta una trayectoria preferencial hacia el este, donde se sitúa el mar Mediterráneo y, por lo tanto, el flujo del agua en régimen estacionario seguirá esta dirección. Solamente el aumento de intensidad en los bombeos estacionales variará las direcciones preferentes de flujo (DPA, 1999). La recarga del acuífero se produce a través del agua de lluvia y agua de escorrentía superficial en gran parte canalizada a favor del río Gorgos; también existe cierta alimentación lateral procedente de las formaciones calizas adyacentes del Tossalet y por los retornos de riego. La descarga principal se realiza a través de bombeos y por salidas naturales al mar. Los recursos hídricos se han cifrado en 3 hm<sup>3</sup>/año con un déficit estimado en 1 hm<sup>3</sup>/año, aunque actualmente se considera equilibrado (DPA, 1999; Armayer et al., 2000).

## Pozos, variables y tiempos

El período de observación y toma de datos abarca desde diciembre de 1985 a septiembre de 1987. En un primer conjunto de 19 pozos, A, se tomaron datos en 18 ocasiones, aproximadamente cada 37 días. Los

# Análisis espacio-temporal del acuífero del cuaternario de Xàbia

F. Martínez<sup>(1)</sup>, J. Mateu<sup>(2)</sup>, F. Montes<sup>(1)</sup> e I. Morell<sup>(3)</sup>

(1) Departamento de Estadística e I.O., Universitat de València, Dr Moliner, 50, E-46100 Burjassot, Spain.  
Email: montes@uv.es

(2) Departamento de Matemáticas, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón, Spain.  
Email: mateu@mat.uji.es

(3) Departamento de Ciències Experimentals, Universitat Jaume I, Campus Riu Sec, E-12071 Castellón, Spain.  
Email: morell@exp.uji.es

## RESUMEN

Los modelos matemáticos usuales de flujo o transporte necesitan del conocimiento del parámetro de interés a lo largo de una fina malla, que cubre la superficie del acuífero. Son varios e importantes los motivos que impiden la toma de datos en cada punto donde es necesario, lo que motiva la necesidad de los modelos estadísticos de predicción espacial en este contexto.

En este trabajo se propone una metodología estadística para la predicción espacio-temporal del nivel piezométrico. Los modelos propuestos analizan y separan la tendencia para posteriormente trabajar con el residuo estacionario. Mediante estos modelos analizamos la superficie piezométrica del acuífero de Xàbia.

Palabras clave: características físico-químicas del agua subterránea, kriging, modelos espacio-temporales, nivel piezométrico, REML, validación cruzada

## *Spatio-temporal analysis of the quaternary aquifer of Xàbia*

### ABSTRACT

*The usual flow or mass transport mathematical models require the knowledge of the corresponding parameters of interest through a fine grid covering the whole surface of the study. However, there are several real circumstances that avoid taking measures where needed, which motivates the necessity of statistical models for spatial prediction.*

*We propose a statistical framework for the spatio-temporal prediction of the piezometric level. The proposed models analyze the trend to work with stationary residuals. These techniques are applied to analyze the piezometric surface of the Xàbia aquifer.*

*Key words: cross-validation, kriging, physical and chemical characteristics of groundwater, piezometric level, REML, spatio-temporal models*

## Introducción

Las áreas costeras de regiones semiáridas se suelen caracterizar por un régimen irregular de precipitaciones, normalmente inferiores a 400 mm/año, por la práctica ausencia de recursos hídricos superficiales y por una creciente demanda hídrica derivada de la presión agrícola y turística. Esta demanda se satisface esencialmente mediante el uso intensivo de las aguas subterráneas que conduce, en muchos casos, a situaciones de sobreexplotación, entendida ésta como el régimen de explotación en que los bombeos superan

a los recursos renovables (recarga). La principal consecuencia de esta sobreexplotación es el rebatimiento progresivo de la superficie piezométrica y el consiguiente desequilibrio hidrodinámico entre el agua dulce y el agua de mar que provoca el proceso denominado intrusión marina. El contacto agua dulce - agua salada en el seno del acuífero da lugar a la formación de una zona de mezcla cuya anchura depende de diversos factores (permeabilidad en la franja costera, espesor saturado del acuífero) y cuyo avance tierra adentro está relacionado básicamente con las variaciones piezométricas.

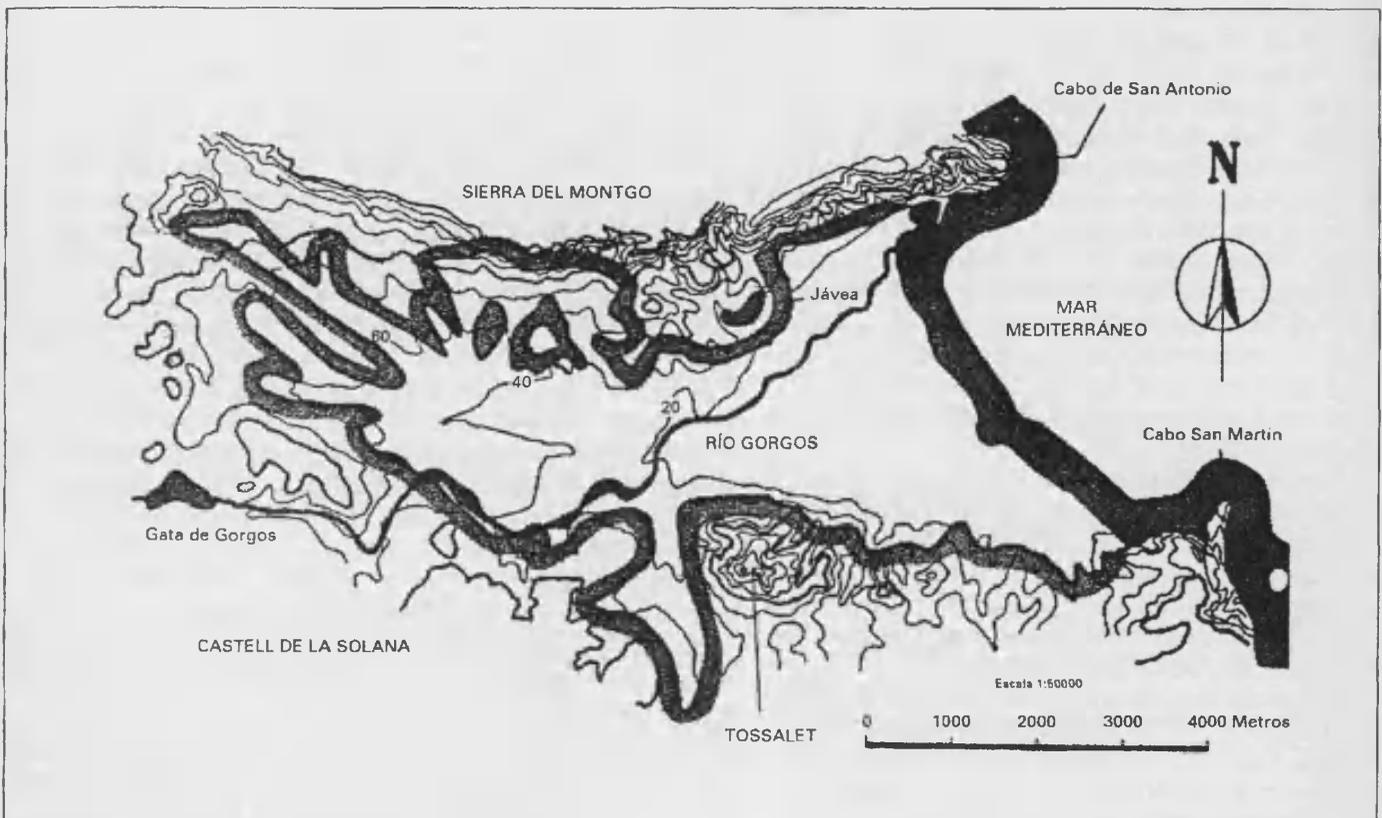


Fig. 1. Localización del acuífero del cuaternario Xàbia  
 Fig. 1. Location map of the quaternary aquifer of Xàbia (Spain)

parámetros medidos fueron las concentraciones de  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{HCO}_3^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Br}^-$ ,  $\text{NO}_3^-$ ,  $\text{B}$ , el pH, la conductividad y el nivel piezométrico. Además se dispone de observaciones para septiembre de 1987 en un conjunto adicional de 124 pozos,  $B$ , si bien en todos ellos, y para ambas fechas, falta la medida del nivel piezométrico. Las localizaciones de los pozos en el acuífero se muestran en la Figura 2. En ella, los 19 pozos con seguimiento temporal aparecen señalados con un punto ( $\bullet$ ) y los 124 restantes con una equis ( $x$ ).

### Modelos espacio-temporales

Consideremos el proceso estocástico  $Z_t(s)$  en la localización  $s$  y el tiempo  $t$ . Hemos observado el valor del proceso en un conjunto puntos  $\{Z_t(s), s \in D, t \in T\}$ , y deseamos predecir el valor del proceso original,  $Z_{t_0}(s_0)$ , en la localización  $s_0 \in D$  y el tiempo  $t_0$ . Suponemos además que los valores de  $Z$  han sido observados sin error de medida.

Asumimos que podemos descomponer los datos como la suma de una *tendencia* (variación determi-

nista a gran escala) y un *error* (variación aleatoria a pequeña escala),

$$Z_t(s) = \mu_t(s) + \delta_t(s) \quad (1)$$

donde  $E(Z_t(s)) = \mu_t(s)$  and  $\delta_t(\bullet)$  es un proceso intrínsecamente estacionario con media cero y variograma  $2\gamma_u(h) = \text{Var}(Z_{t+u}(s+h) - Z_t(s)) = \text{Var}(\delta_{t+u}(s+h) - \delta_t(s))$  que caracteriza la dependencia espacial.

La media del proceso,  $\mu_t(s)$ , puede expresarse como una función lineal de un conjunto de  $k$  regresores o variables explicativas  $x_t(s) = (x_{t,1}(s) \dots x_{t,k}(s))$

$$\mu_t(s) = x_t(s)' \beta \quad (2)$$

donde  $\beta$  es el vector de coeficientes. Sustituyendo en (1) y utilizando notación matricial,

$$z_t = x_t \beta + \delta_t \quad (3)$$

donde  $z_t$  es el vector de las observaciones,  $X_t$  es la matriz del diseño y  $\delta_t$  el vector de los residuos, cuya matriz de varianza-covarianzas denotamos por  $\Sigma$ . Necesitamos estimar las componentes de  $\beta$  y  $\Sigma$ , que

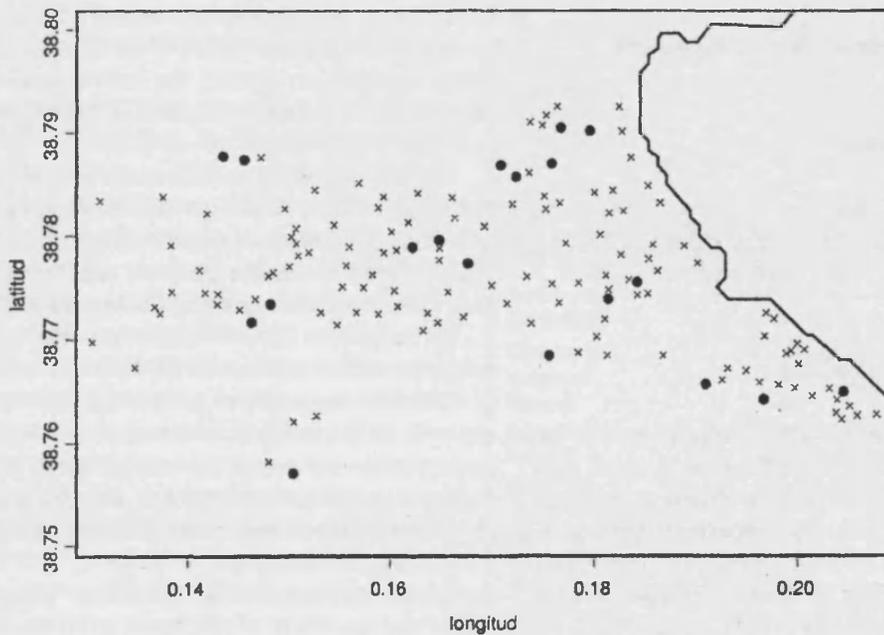


Fig. 2. Localización de los pozos muestreados  
 Fig. 2. Spatial locations of the selected wells

bajo la hipótesis de estacionariedad del proceso dependerá de un vector de parámetros  $\theta$ , es decir,  $\Sigma = \Sigma(\theta)$ .

La estimación de ambos conjuntos de parámetros puede realizarse mediante un procedimiento iterativo y secuencial basado en el GLM (General Linear Model) que consiste en llevar a cabo una primera estimación de  $\beta$  mediante mínimos cuadrados ordinarios, esta estimación se utiliza para estimar los valores de  $\theta$ , con los que se reinicia el proceso y se obtienen, mediante mínimos cuadrados generalizados, nuevas estimaciones de  $\beta$ . El procedimiento se repite hasta que las estimaciones convergen. Este proceso conduce a estimaciones sesgadas de los parámetros. Un método alternativo para soslayar esta dificultad es el llamado REML (Restricted Maximum Likelihood), que permite estimar simultáneamente ambos parámetros (Cressie, 1993).

Las componentes de  $\theta$  son los parámetros de  $2\gamma_u(h)$ , variograma del error aleatorio  $\delta_i$ . Dicho error puede estimarse mediante

$$\delta'_i(s) = Z_i(s) - x_i(s)' \beta \quad (4)$$

El estimador de Cressie y Hawkins (Cressie, 1993) es un estimador robusto para el variograma y tiene por expresión,

$$2\gamma'_u(h) = \left( \frac{1}{|N_u(h)|} \sum_{(s_i, t_i), (s_j, t_j) \in N_u(h)} |\delta'_i(s_i) - \delta'_j(s_j)|^{1/2} \right)^4 / \left( 0,457 + \frac{0,494}{|N_u(h)|} \right) \quad (5)$$

donde  $N_u(h) = \{(s_i, t_i), (s_j, t_j) : s_i - s_j = h, t_i - t_j = u\}$  y  $|N_u(h)|$  es el número de pares distintos en  $N_u(h)$ . Obsérvese que para un variograma isotrópico se verifica,  $2\gamma_{yu}(h) = 2\gamma_{yu}(\|h\|)$ .

Una vez estimados los parámetros podemos obtener predicciones y sus errores en nuevas localizaciones y tiempos,  $\hat{Z}_{i0}(s_0)$ , mediante la suma del valor estimado de la tendencia determinista a gran escala,  $\mu_{i0}(s_0)$ , y el valor estimado del error  $\delta_{i0}(s_0)$ . Los detalles para la obtención de estas estimaciones pueden consultarse en Cressie (1993). Señalemos por último que

este procedimiento puede adaptarse fácilmente al caso puramente espacial eliminado el índice temporal de las anteriores expresiones.

### Modelización de la superficie piezométrica del acuífero de Xàbia

#### Modelo puramente espacial

Aplicaremos en primer lugar un modelo puramente espacial a los niveles piezométricos observados en los 19 pozos del conjunto A en septiembre de 1987.

#### Análisis exploratorio de los datos

El proceso es claramente no estacionario como se deduce de la observación de la Figura 3, que nos muestra cómo el nivel piezométrico crece a medida que nos alejamos de la costa. Para soslayar este problema y conseguir la estacionariedad intrínseca del proceso podemos modelizar la media, lo que hemos llamado *tendencia determinista* en (1), mediante una forma cuadrática de las coordenadas geográficas.

Podemos también incorporar a esta modelización otras covariables, que pueden ser algunas de las componentes del agua o transformaciones de las mismas. Optamos por esta segunda opción y en lugar

de utilizar directamente las otras variables observadas en cada uno de los 19 pozos, hemos recurrido al análisis factorial para reducir el número de covariables sin perder información de interés. Esta técnica se utiliza con frecuencia en estudios hidrogeoquímicos y, además de la reducción de dimensión citada, permite interpretar alguno de los factores en función de fenómenos hidrogeológicos e hidrogeoquímicos de interés.

Hemos llevado a cabo el análisis factorial con todas las observaciones disponibles y con todas las variables químicas excepto la conductividad y el pH, muy correlacionadas con las restantes, y los bromuros, por presentar gran cantidad de datos faltantes.

El resultado del análisis factorial conduce a tres factores que explican el 78 % de la varianza, del que el 49 % corresponde al primer factor, el 18 % al segundo y el 12 % restante al tercero. Las componentes de cada factor se muestran en la Tabla 1 y de ellas se deduce que el primer factor, sin duda el más importante, está relacionado con la intrusión del agua marina en el acuífero.

Propondremos dos modelos para analizar los datos del acuífero: el primero, *modelo 1*, tendrá como covariables solamente las de la forma cuadrática de las coordenadas geográficas con la latitud modificada mediante una traslación negativa de 31°; el segundo, *modelo 2*, contiene además los tres factores como covariables.

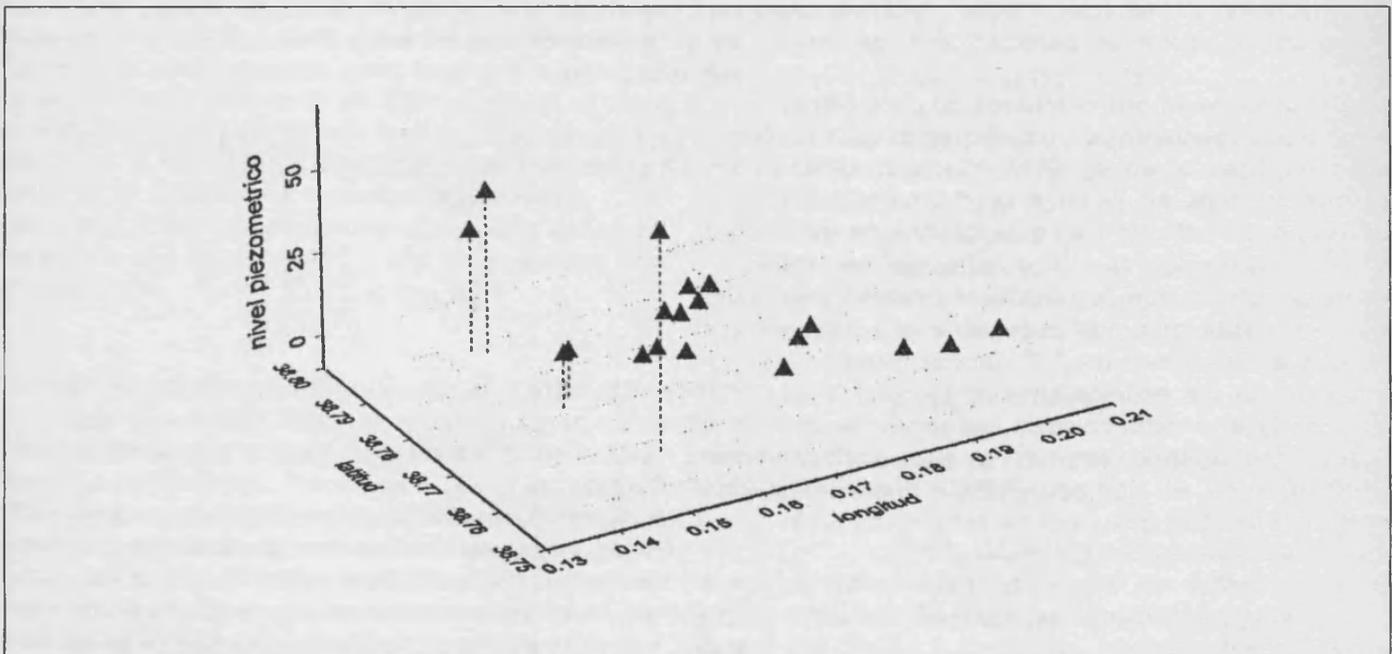


Fig. 3. Gráfico de dispersión 3-D del nivel piezométrico  
Fig. 3. 3-D scatterplot of the piezometric level

**Resultados**

La Tabla 2 resume los resultados para ambos modelos. La primera parte muestra las estimaciones de las componentes del vector  $\beta$  correspondiente a cada modelo. No hay por tanto valores para los tres factores en el *modelo 1*.

La segunda parte de la Tabla 2 recoge las estimaciones de las componentes de  $\theta$ . Se observa que para el *modelo 2* faltan valores. Para comprender los motivos de estas ausencias hemos de remitirnos a la Figura 4 que muestra (gráfica de la derecha) cómo el semivariograma ajustado al error del *modelo 2* es una recta paralela al eje de distancias, lo que significa que no existe prácticamente variación espacial. Al poder suponer constante la varianza, ésta puede estimarse mediante la varianza de los residuos obtenidos después de eliminar la tendencia. El único valor que aparece en la tabla para el *modelo 2* es esta varianza y no el nugget. La Figura 4 muestra también a su izquierda el semivariograma de tipo exponencial ajustado al error del *modelo 1* mediante mínimos cuadrados ponderados, cuyos parámetros son los de la tabla.

La tercera parte de la Tabla 2 proporciona información sobre la bondad de los modelos basada en técnicas de validación cruzada (Huang y Cressie, 1996). La cantidad *CRV1* contrasta la insesgadez de la predicción y debe valer aproximadamente 0; la cantidad *CRV2* depende de la precisión de la desviación típica del error de predicción y su valor debe de estar

	Componentes de los factores		
	1	2	3
Sodio	.8860	-.2770	-.0130
Potasio	.4869	.3851	.2924
Calcio	.7981	-.0609	-.2487
Magnesio	.8930	-.3069	-.0137
Cloruros	.9120	-.3485	-.1257
Bicarbon	.0481	-.1315	.7929
Sulfatos	.6381	.4609	-.0379
Nitratos	.2319	.7366	-.4331
Boro	.5078	.4724	.5533

Tabla 1. Componentes de los tres primeros factores  
*Table 1. Components of the first three factors*

Parámetro	Modelo 1	Modelo 2
Constante	393.8324	369.3819
longitud	-3094.2980	-2837.1140
latitud	-4271.8620	-3039.6380
longitud <sup>2</sup>	7433.8810	6776.5380
latitud <sup>2</sup>	107924.9000	106230.6000
long*lat	-10675.7300	-18034.8200
factor 1	-	3.8120
factor 2	-	-0.2619
factor 3	-	5.5392
nugget	0.0000	11.6677
range	0.0064	-
sill	46.8155	-
CVR1	0.0024	0.0053
CVR2	1.1005	0.8222
CVR3	5.5679	3.3247

Tabla 2. Resumen de los resultados para ambos modelos  
*Table 2. Summary of the results for both models*

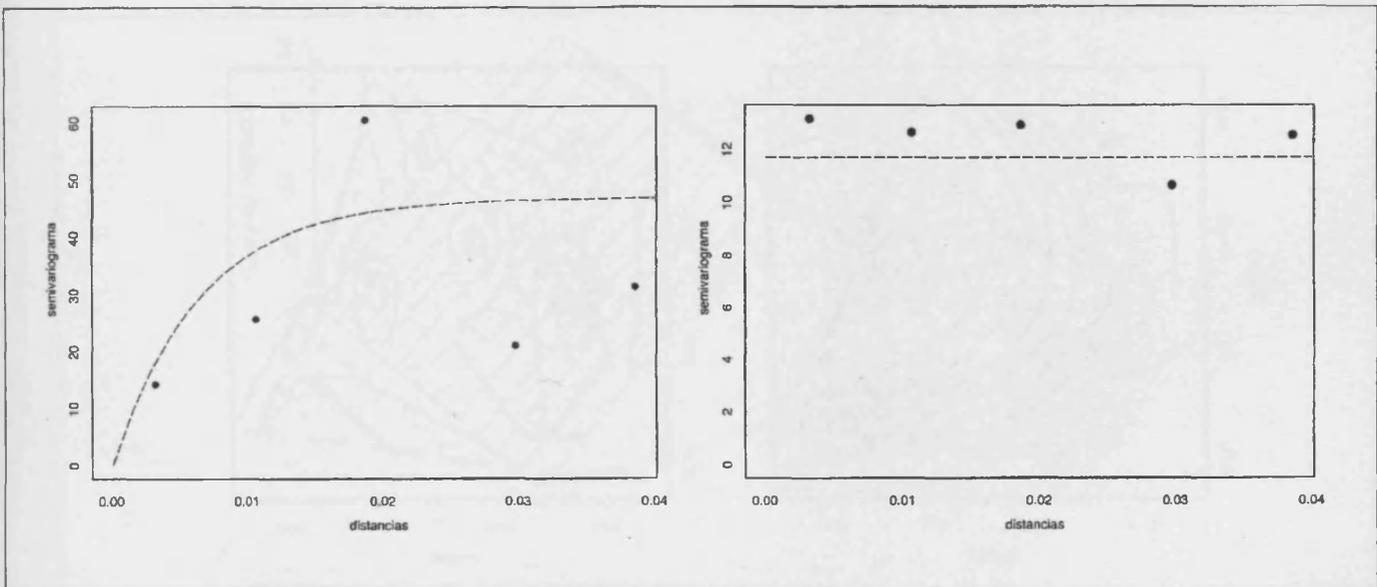


Fig. 4. Semivariograma de los modelos 1 (izquierda) y 2 (derecha)  
*Fig. 4. Semivariogram for models 1 (left) and 2 (right)*

muy próximo a 1; por último, la cantidad *CRV3* es una medida de la bondad de la predicción, concretamente su error estándar.

La Tabla 3 muestra las predicciones del nivel piezométrico obtenidas con ambos modelos para los 19 pozos del conjunto

A. Se recogen también, a efectos de comparación, los valores observados.

La calidad de las predicciones, medida en términos de los errores asociados, puede apreciarse en las curvas de nivel de las Figuras 5 y 6.

**Modelo espacio-temporal**

La variación espacio-temporal del nivel piezométrico de los 19 pozos a lo largo de los 18 tiempos observados se ha estudiado con sendos modelos espacio-temporales, *modelo 3* y *modelo 4* equivalentes a los dos *modelos 1* y *2*, respectivamente. Es decir, las tendencias deterministas a gran escala utilizadas han sido las correspondientes a los *modelos 1* y *2*, que una vez abstraídas dan lugar a sendos errores espacio-temporales a partir de los cuáles hemos obtenido

pozo	nivel p.	Modelo 1		Modelo 2	
		predicción	error	predicción	error
B1	-0.04	-1.4712	6.7514	0.0623	4.5114
C1	0.32	-7.5956	5.4707	-2.4552	4.2025
D1	-0.06	-2.9460	4.9206	-5.2079	3.8225
D2	-0.53	-1.7683	4.2388	1.2769	3.8025
D3	-0.14	0.4868	5.2255	-4.3666	4.2238
F1	-0.40	2.6533	4.5414	1.4482	4.5245
F2	0.21	4.5465	4.7903	9.1489	3.9898
G1	-0.39	5.9330	6.6164	3.4782	4.6162
G2	0.37	1.6439	6.2727	-0.5054	3.9159
G3	-0.10	3.9731	6.0871	1.0899	4.1536
H1	-0.13	-2.8885	5.3100	-0.3972	3.9933
J1	-0.37	-0.6476	4.5952	-3.2564	3.8005
J2	-0.53	2.7423	5.1377	1.0643	3.7438
J4	-0.66	-2.4620	5.5119	-3.8192	3.7180
L1	14.30	15.6160	5.0409	18.0320	4.0353
L2	16.22	20.5712	5.0702	17.3893	4.2605
L4	65.15	58.9140	7.1553	63.0455	4.7695
O1	48.81	33.4911	4.7362	45.2135	4.2995
O2	37.67	48.2473	4.7578	38.4589	4.1924

Tabla 3. Predicción y error de predicción con ambos modelos para los 19 pozos del conjunto A  
 Table 3. Prediction and associated standard error for both models for the 19 selected wells of set A

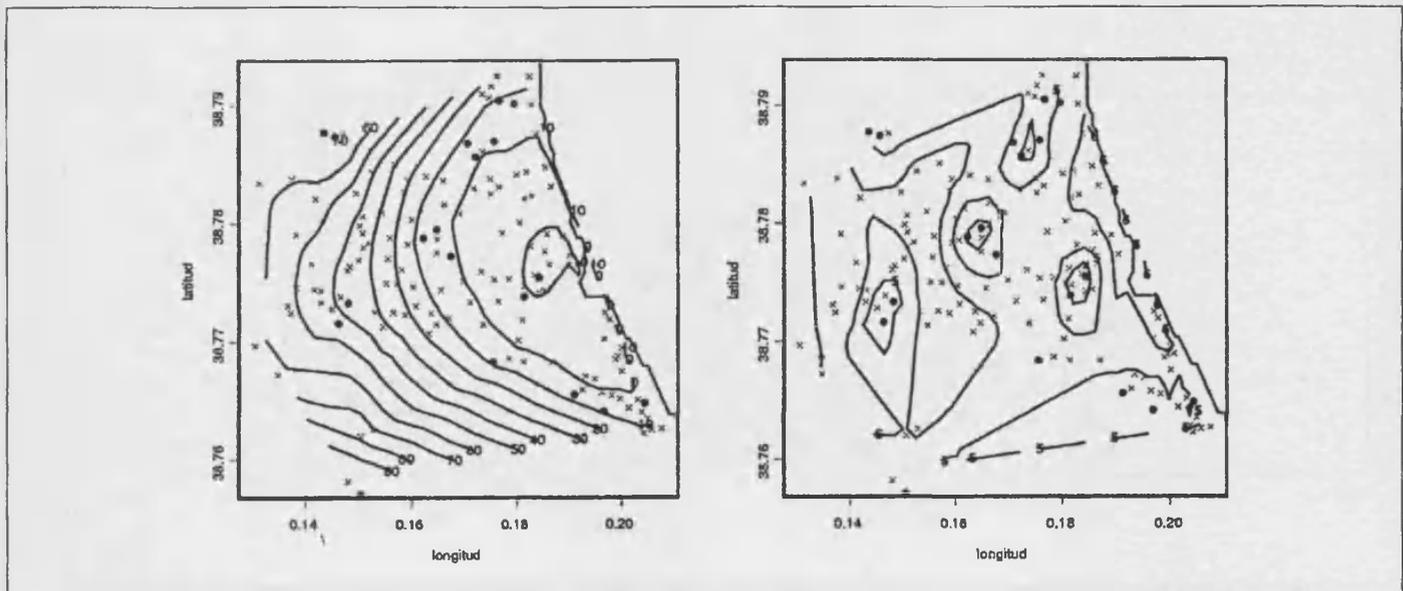


Fig. 5. Predicción (izquierda) y error de predicción (derecha) para el modelo 1 del nivel piezométrico para los 124 pozos del conjunto B  
 Fig. 5. Prediction (left) and associated prediction error (right) of piezometric level for model 1 and the 124 wells of set B

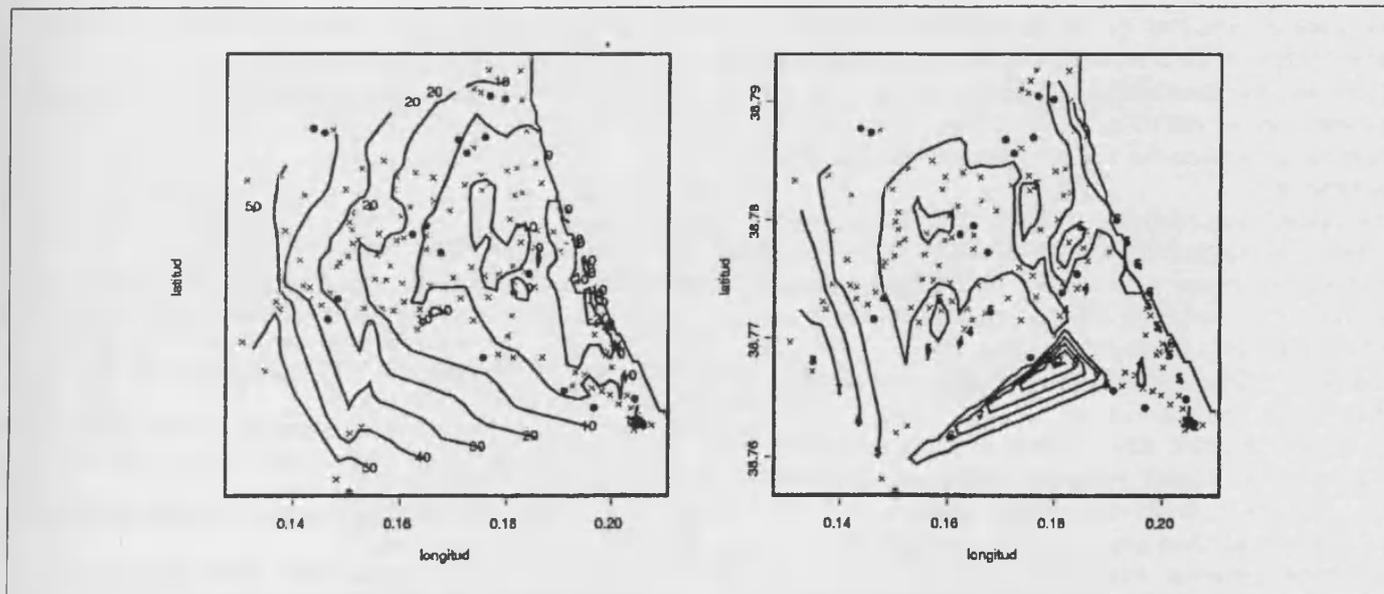


Fig. 6. Predicción (izquierda) y error de predicción (derecha) para el modelo 2 del nivel piezométrico para los 124 pozos del conjunto B  
 Fig. 6. Prediction (left) and associated prediction error (right) of piezometric level for model 2 and the 124 wells of set B

los variogramas correspondientes. En ambos casos, hay carencia de variabilidad temporal, como bien puede apreciarse en la Figura 7 que muestra el semivariograma empírico de los residuos del modelo 3 obtenido según (5). Obsérvese que los variogramas empíricos son paralelos en el tiempo. No ha lugar a estimar los parámetros del semivariograma puesto

que al carecer de componente temporal obtendríamos valores muy parecidos a los de la Tabla 2.

### Conclusiones

En este trabajo se ha presentado una metodología

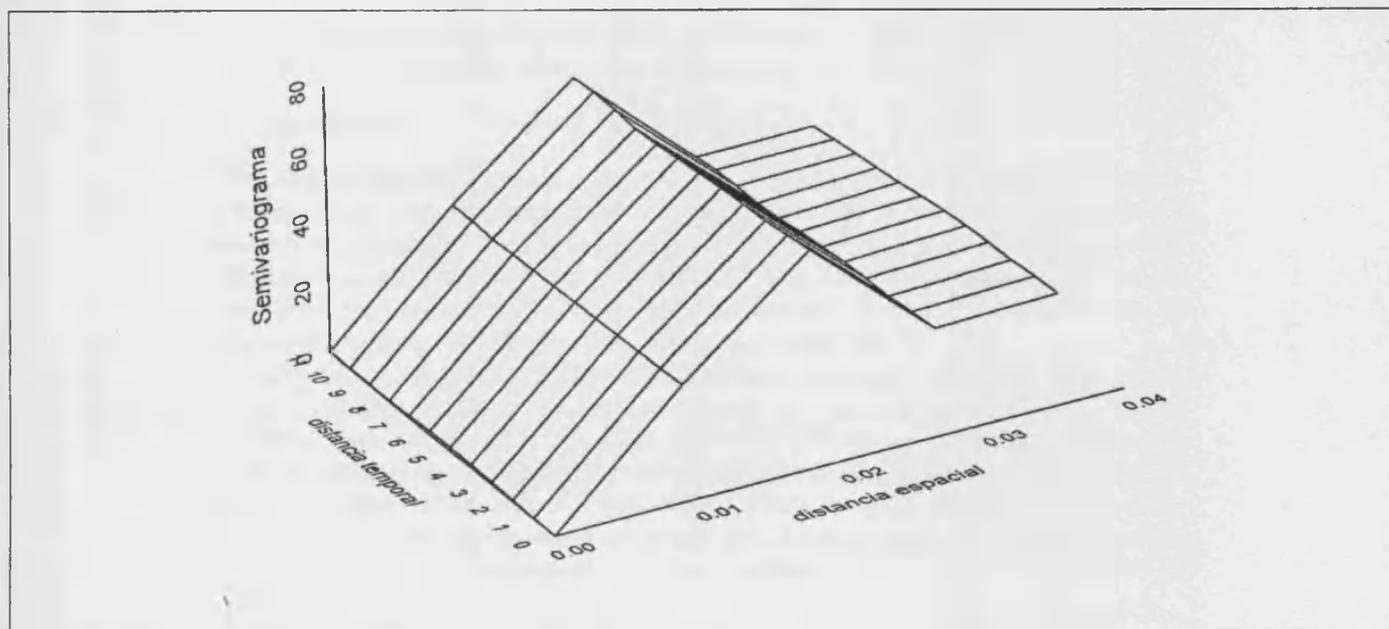


Fig. 7. Semivariograma espacio-temporal  
 Fig. 7. Spatio-temporal semivariogram

válida para el análisis de la variabilidad espacial y espacio-temporal de ciertos parámetros de interés en relación con los acuíferos. Los modelos pueden ser utilizados con el objetivo final de la modelización estadística, o bien como soporte para otros modelos matemáticos.

Los resultados expuestos en la sección anterior permiten concluir que:

1. El *modelo 2* presenta mejores prestaciones que el *modelo 1*. El valor de *CRV3*, error estándar de la predicción, es mucho mejor para aquél, 3.32, que para éste, 5.56, presentando las otras dos cantidades valores semejantes.
2. La mayor bondad del *modelo 2* supone admitir que no existe efecto espacial a pequeña escala y que éste se manifiesta sólo a gran escala, de manera que el nivel piezométrico puede explicarse satisfactoriamente mediante una superficie de regresión espacial que depende de las coordenadas geográficas, a través de una forma cuadrática, y de los tres factores derivados de las características químicas del agua.
3. Los modelos espacio-temporales no han resultado satisfactorios al carecer sus variogramas de variabilidad temporal.

#### Agradecimientos

Los autores quieren agradecer a la Profesora María Pilar Blasco de la Universidad de Alicante su inesti-

mable colaboración al permitirnos utilizar los datos de su tesis doctoral.

Este trabajo ha sido financiado parcialmente con cargo al proyecto BFM-2001-3286 de la DGICYT.

#### Referencias

- Armador, J.L., Murillo, J.M. y Rodríguez, L. 2000. Recarga artificial en el acuífero de Jávea mediante el aprovechamiento de los excedentes hídricos procedentes del río Jalón-Gorgos (Alicante). *V Congreso Geológico de España*. Alicante.
- Blasco, M.P. 1988. *Estudio Hidroquímico del Acuífero del Cuaternario de Jávea (Alicante)*. Tesis Doctoral. Universidad de Alicante.
- Cressie, N. 1993. *Statistics for Spatial Data*, Revised Edition. John Wiley, New York.
- Diputación Provincial de Alicante (DPA). 1999. *Recarga artificial de acuíferos*. Ed. DPA. 157 pp.
- Huang, H.C. y Cressie, N. 1996. Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis*, 22, 159-175.
- Pulido-Bosch, A. 1976. Salinización y recarga artificial en el acuífero de Jávea (Alicante). *Hidrogeología y Recursos Hidráulicos*, 2, 772-787.

Recibido: Mayo 2003

Aceptado: Septiembre 2003

## A comparative analysis of different spatial sampling schemes: Modelling of SSRB data

F. MARTÍNEZ†, J. MATEU‡, F. MONTES\*§, A. BODAS-SALCEDO¶\*\* and E. LÓPEZ-BAEZA¶

†Ajuntament de València, Spain

‡Departament de Matemàtiques, Universitat Jaume I, Spain

§Departament d'Estadística i I.O., Universitat de València, Spain

¶Departament de Termodinàmica, Universitat de València, Spain

(Received 18 July 2005; in final form 2 April 2007)

Low spatial resolution satellite sensors provide information over relatively large targets with typical pixel resolutions of hundreds of km<sup>2</sup>. However, the spatial scales of ground measurements are usually much smaller. Such differences in spatial scales makes the interpretation of comparisons between quantities derived from low resolution sensors and ground measurements particularly difficult. It also highlights the importance of developing appropriate sampling strategies when designing ground campaigns for validation studies of low resolution sensors.

We make use of statistical modelling of high resolution surface shortwave radiation budget (SSRB) data to look into this problem. A spatial model that describes the SSRB over a selected region is proposed, and the impact of different sampling schemes in the performance of the model is analysed. Both systematic and random sampling schemes can efficiently represent the full observations set.

### 1. Introduction

The activity that endeavours to ensure that remote sensing products are highly consistent and reproducible is known as calibration and validation (cal/val). This is an evolving discipline that is becoming increasingly important as more long-term studies on global change are undertaken and new satellite missions are launched. Therefore, well-instrumented benchmark test sites and datasets for the validation of Earth observation satellite products should be supported.

The primary objective of validation is to assess the quality and, as far as possible, to quantify the accuracy of remote sensing data products. Ideally, validation activities seek to compare the data products with more accurate independent measurements of the same quantity over a statistically significant number of samples and wide variety of situations. The problem is that the space and time scales of the *in situ* data and the satellite data are rarely directly comparable. Typically, there are insufficient *in situ* measurements to cover a satellite field of view of several km<sup>2</sup> of

\*Corresponding author. Email: montes@uv.es

\*\* Current affiliation: Meteorological Office, Hadley Centre for Climate Prediction and Research, UK

low spatial resolution sensors (broadband radiometers, L-band radiometers, etc.). Thus, even a perfect remote sensing measurement will be expected to differ from an *in situ* verification measurement because of the inability to match the observations in time or space. *In situ* measurements are invariably taken over small space scales, whilst the time scales may range from instantaneous measurements to accumulated values (e.g. rain gauges). However, a satellite overpass is a snapshot of a few seconds over a large area. For broadband radiometers, a common example is the comparison of a surface flux radiometer with a satellite based estimate for the radiometer field of view. The matching error can be reduced (but not eliminated) by increasing the number of surface observations within the field of view (Li *et al.*, 1995). This error can also be minimized by using very large ensembles of matched data.

The large pixel size of some satellite sensors for climate studies, such as the broadband radiometers GERB (Geostationary Earth Radiation Budget) (Harries *et al.*, 2005), CERES (Clouds and the Earth's Radiant Energy System) (Wielicki *et al.*, 1996), and the L-band radiometer SMOS (Soil Moisture and Ocean Salinity) (Wigneron *et al.*, 2000), introduces a number of scientific issues that make it relevant and even necessary to develop a specific methodology and carry out specific measurements over large extended areas. Dedicated programmes, such as CAGEX (CERES/ARM/GEWEX Experiment) (Charlock and Alberta, 1996) and BSRN (Baseline Surface Radiation Network) (Ohmura *et al.*, 1998), and other long-term sites are considered the best strategy for the validation of satellite products and comparison with models.

Without doubt, the large size of the reference pixel makes it necessary to study the change of scale processes in detail in order to be able to compare measurements with different scales, namely point measurements, aircraft observations, remote sensing images with different spatial resolutions, etc., and establish ground-measurement sampling criteria to get different area averages and validate large scale pixels.

In this respect, remote sensing poses interesting scientific questions that spatial statistics can help to deal with. In particular, spatial statistics addresses the classification of images by supervised and unsupervised procedures and the estimation of spatial structures by variograms, interpolation, simulation, sampling and issues of scale. Some references are (Atkinson *et al.*, 2000; Cressie and Kornak, 2003; Stein *et al.*, 1999). Different spatial statistical techniques have been used to deal with the estimation of areal averages of rainfall at different spatial and time scales, e.g. (Grimes *et al.*, 2003; Lebel and Amani, 1999, and references therein). In this paper, we take advantage of geostatistical tools to look at the problem of comparing measurements on different spatial scales, with the aim of helping in the selection of appropriate sampling schemes in ground validation campaigns.

In this study we make use of surface shortwave radiation budget (SSRB) as derived from high spatial resolution satellite data, and altitude data as ancillary information. These two datasets will be considered our 'ground-truth', which will allow us to analyse the impact of different sampling schemes in the representation of spatially averaged values. This can be considered as an analog of a densely distributed grid of ground sensors in the field of view of a low resolution sensor. A spatial model that describes the SSRB over that region is proposed, and the impact of reducing the number of sampling points by means of different qualitative and quantitative sampling schemes is analysed. We focus on the size of the regular grid provided by the satellite measurements, which is basically a set of 1 km<sup>2</sup> squares. The data provided is an average of the SSRB data over the square. Our concern is how

we can use fewer sampling points while preserving the area associated with each point and, at the same time, keeping the same amount of information as if we had used the whole dataset. We try to use several schemes based on a different number of points and geometrical structures to evaluate the goodness-of-fit of our statistical spatial model. The evaluation will be based on a general evaluation set which is valid for the whole set of samplings. Other approaches have been used for investigating the optimal selection of a sample in spatial context using geostatistical techniques; Pardo (1998), van Groeningen *et al.* (1999), and Zhu and Stein (2006) share the use of the simulated annealing method to optimize the selection and number of location of a sample.

These kinds of studies are crucial for the deployment of ground sensors over extended areas in validation campaigns of low-resolution satellite sensors. Recently, the University of Valencia has set up a robust automatic meteorological station (Valencia Anchor Station) as a reference for calibration and validation activities in low-resolution satellite sensors (López-Baeza *et al.*, 2001). The station also has complementary mobile instrumentation available to produce reliable parameters which are representative of relatively large areas defined around the Anchor Station. Therefore, we consider this study to be a first step that can help in the selection of an objective sampling scheme of the region under study, using a finite number of ground sensors.

The outline of the paper is as follows. Section 2 describes the dataset and deals with aspects of the change of spatial resolution of the altitude database. Section 3 summarizes the methodological development based on spatial statistical models. Section 4 presents the six sampling schemes used in the paper and the six different models used to fit the trend (long-range deterministic variation) of the budget solar radiation. Section 5 analyses the data, comparing the statistical models and the sampling scheme. The paper ends with a conclusions section.

## 2. Region and datasets

The selected geographical region under study is located in the east of the Iberian Peninsula, in the province of Valencia. The region has been selected as a square delimited by geographical coordinates UL= $(-1.42, 39.59)$  and LR= $(-0.93, 39.10)$ , shaded area in figure 1. We have selected this region because it covers the area where the Valencia Anchor Station is located.

The main database used in the analysis are estimates of the clear-sky SSRB over that region for a 10-day period, from 21 to 30 April 1996. The surface shortwave radiation budget (SSRB) is the amount of energy in the solar region of the electromagnetic spectrum ( $0.2\text{--}4.0\ \mu\text{m}$ ) absorbed at the surface. The SSRB is an important quantity in climate studies, as it plays a crucial role in the oceanic poleward heat transport (Gleckler, 2005). The best way of monitoring this data over extended regions is by means of satellites orbiting the Earth. In recent years, several satellite-based datasets have provided a global perspective of the surface radiation budget (Gupta *et al.*, 1999; Zhang *et al.*, 2004). Regional, high spatial resolution description of the SRB is also of interest as indicator of changes and because of its relationship to aridification from well developed vegetation and, as a measurement of the available energy, SRB is a key parameter for the derivation of the surface energy fluxes (Bolle *et al.*, 2006). The SSRB database has been derived from satellite data at the top of the atmosphere and a physical model that relates the shortwave radiation budget at the top of the atmosphere with that at the surface, making use of ancillary information of precipitable water content in the atmospheric column (Li

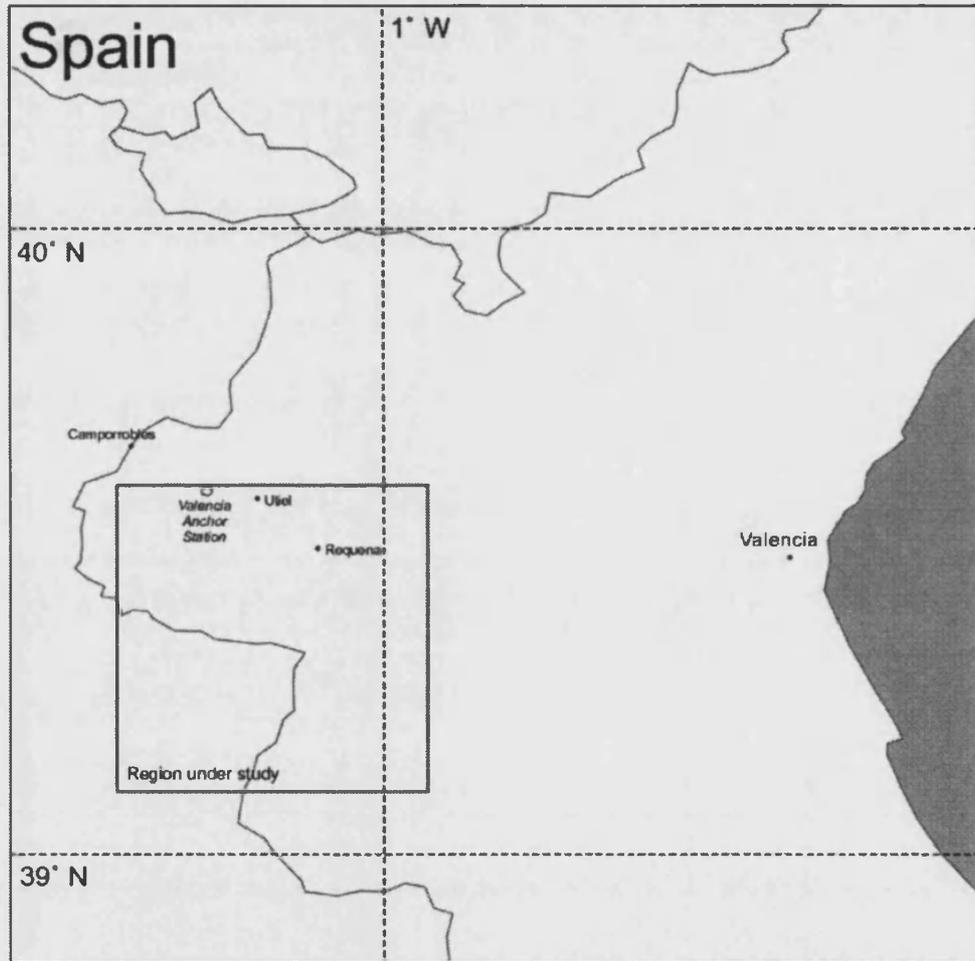


Figure 1. Map of the analysed region.

*et al.*, 1993). For the derivation of the SSRB over the selected study area in the Iberian Peninsula (figure 1), we used the Mediterranean Extended Daily One Km AVHRR DataSet (MEDOKADS) (Koslowsky, 1996). The MEDOKADS database is a high spatial resolution database of 10-day maximum NDVI (Normalized Difference Vegetation Index) composites, obtained from the AVHRR (Advanced Very High Resolution Radiometer) sensors, on board NOAA (National Oceanic and Atmosphere Administration) satellites. The spatial resolution of the SSRB dataset is that of the input satellite data, arranged on a  $0.01^\circ$  lat/long grid, with a pixel size of approximately  $1 \times 1 \text{ km}^2$ . This implies that the region under study is covered by  $50 \times 50$  square cells. Due to the fact that satellite data provide integrated values over each pixel, we will assume that the SSRB is constant within each cell and therefore neglect any subgrid variability. A more detailed description of the derivation of the SSRB dataset can be consulted in Bodas-Salcedo *et al.* (2003). The satellite data will be subsampled to provide both training and validation datasets as described in §4.

The second dataset used is the altitude of the region under study, which will be used as an ancillary input to improve the spatial modelling of the SSRB over that region. This altitude database is defined over rectangular cells of  $0.92 \times 0.70 \text{ km}^2$ ,

which is different to the SSRB resolution. Due to the different spatial resolution of both databases, the first task to be carried out has been the homogenization of both resolutions. There are several methods for adapting the spatial resolution of two grids (Flowerdew and Green, 1992, 1994; Mugglin *et al.*, 2000). Since both datasets have one value in each cell, we have chosen the simplest and most general algorithm, namely, the interpolation of areas. The method uses a weighted mean with the areas of the intersections between cells as weights. Given the different dimensions of both grids, the altitude grid divides every SSRB cell, denoted as  $SSRB_i$ , into four rectangles,  $H_j, j=1, 2, 3, 4$ . If the altitudes of the four cells that intersect with  $SSRB_i$  are  $h_j, j=1, 2, 3, 4$ , the altitude assigned to the cell  $SSRB_i$  is

$$h_i = \sum_{j=1}^4 \frac{|H_j|}{|SSRB_i|} h_j.$$

### 3. Spatial models

#### 3.1 Estimation and prediction

Let us consider the Gaussian stochastic process  $\{Z(\mathbf{s}), \mathbf{s} \in D\}$ , where  $D$  is a particular region of interest. Let us assume that we have  $n$  observations of the process  $z_i = z(\mathbf{s}_i), i=1, 2, \dots, n$ , and that these observations are free of measurement error.

Let us assume that we can decompose the data as the sum of a *trend* (large-scale deterministic variation) and an *error* (small-scale random variation),

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}), \quad (1)$$

where  $E(Z(\mathbf{s})) = \mu(\mathbf{s})$  and  $\delta(\cdot)$  is an intrinsically stationary process with zero-mean and variogram  $2\gamma(\mathbf{h}) = \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = \text{Var}(\delta(\mathbf{s} + \mathbf{h}) - \delta(\mathbf{s}))$ , which characterizes the spatial dependence. Let us express the mean of the process,  $\mu(\mathbf{s})$ , as a linear function of a set of  $k$  regressors or explicative variables  $\mathbf{x}(\mathbf{s}) = (x_1(\mathbf{s}), \dots, x_k(\mathbf{s}))$ ,

$$\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})' \boldsymbol{\beta}, \quad (2)$$

where  $\boldsymbol{\beta}$  is the coefficients vector of this regression model.

Substituting in (1) and using matrix notation,

$$\mathbf{z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}, \quad (3)$$

where  $\mathbf{z}$  is the vector of observations,  $\mathbf{X}$  the design matrix, and  $\boldsymbol{\delta}$  the vector of residuals, whose variance-covariance matrix is denoted as  $\boldsymbol{\Sigma}$ .

We need to estimate the components of  $\boldsymbol{\beta}$  and  $\boldsymbol{\Sigma}$  which, under the stationarity hypothesis of the process, will depend on a vector of parameters  $\boldsymbol{\theta}$ , that is,  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ . The estimate of both sets of parameters,  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ , is carried out using an iterative procedure by means of the restricted maximum likelihood (REML) method, with certain advantages over other common methods like the maximum likelihood or weighted least squares.

Once the previous parameters have been estimated, we can predict the value of the original process,  $Z(\cdot)$ , in a new position  $\mathbf{s}_0 \in D$  using the universal kriging technique. This geostatistical technique provides, in addition to the predicted value, its error. The reader is referred to Cressie (1993) for further technical details.

### 3.2 Variogram model and data spatial independence

In the spatial model described above, the variance–covariance matrix,  $\Sigma$ , is obtained from a parametric model of the variogram,  $2\gamma(h; \theta)$ , which under the isotropy hypothesis only depends on the distance  $h$  and on a vector of parameters  $\theta$ . Once a valid variogram model is selected, it is possible to contrast its adequacy with the observations by means of the test proposed by Maglione and Diblasi (2001). This test is based on the distance between the theoretical model of variogram proposed in null hypothesis ( $H_0$ ), and the experimental variogram obtained by smoothing the empirical variogram directly retrieved from the data. The test statistic is a quadratic form whose distribution, for a Gaussian process and under  $H_0$ , can be approximated by a noncentral  $\chi^2$ .

The same test can be used to test the independence of the data, a necessary condition prior to any spatial analysis of the data using the model explained above. If the observations were spatially independent, any of the classic linear models could be applied. Independent observation is equivalent to assuming a constant variogram (Diblasí and Bowman, 2001).

### 4. Sampling schemes, regression models and validation set

Two basic sampling schemes have been used to extract a subset from the full dataset over the region under study: systematic and random. In addition, five different sample sizes, 50% ( $N=1250$ ), 25% ( $N=625$ ), 12.5% ( $N=312$ ), 6.25% ( $N=156$ ) and 3.125% ( $N=78$ ), have been selected, which leads to a total number of ten sampling schemes. These sample size are imposed by the geometrical structures of systematic sampling. Figure 2 shows the five systematic schemes superimposed on the original grid. Notice that sampling points represent the centre of the cells.

In §3, we expressed the trend,  $\mu(s)$ , as a linear function of a set of  $k$  regressors or explicative variables (2). Different combinations between regressors produce different models with different performances. Comparison among models allows us to select those that perform better. Geographical coordinates and altitude have been used as explicative variables through six different combinations, giving the following models: M1, M3 and M5 are models based on geographical coordinates with constant trend, first order polynomial and second order polynomial, respectively; models M2, M4 and M6 are the analogous models adding altitude.

The comparison of the models and sample schemes is done by means of two types of criteria, one based on the training set, the sample, and another based on a validation set comprising of 78 points also depicted in figure 2. These 78 points were selected through a systematic system in order to cover the whole region of interest while preserving the underlying spatial correlation. Due to the fact that the validation set was selected first, the samples of each scheme were obtained from the complementary set.

It would also be interesting to know how well the various validation networks can simulate the average radiation over the whole domain. In order to do this, the mean SSRB value over the whole region has been estimated as the mean of the 2500 point values over the whole grid, which is an unbiased estimator (Christensen *et al.*, 2001). A fraction of those 2500 values are those of the sample that has been used to build the model, being the rest the predicted values from that model. The mean of the 2500 observations has been used as reference value in order to evaluate the performance of each sampling scheme.

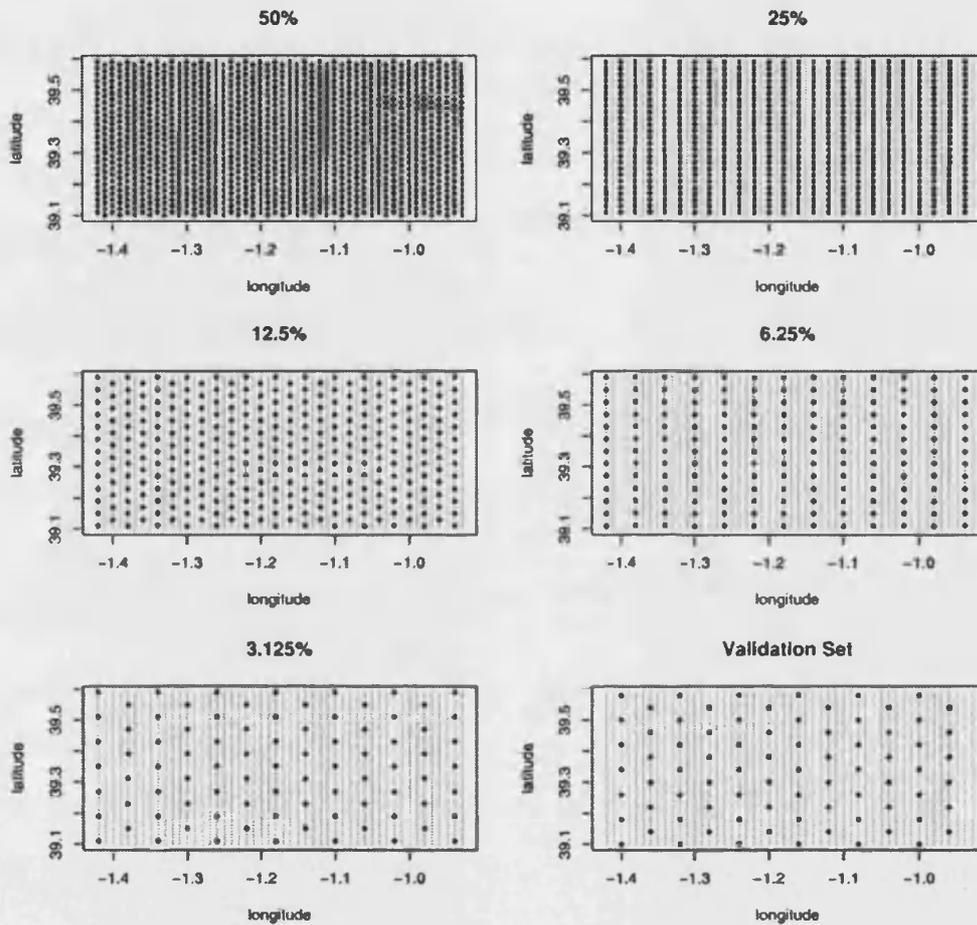


Figure 2. Systematic sampling schemes and validation set.

## 5. Statistical spatial data analysis

### 5.1 Test of spatial independence

The application of a geostatistical model to the analysis of the dataset is justified if the observations are spatially related. This implies that a previous test of independence must be carried out. We applied the test proposed by Dibalasi and Bowman (2001), which rejects in all cases, except for the *R3* scheme, the hypothesis of independence.

### 5.2 Selection of the regression model

The selection of the regression model is done using Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (BIC), and minus twice the value of the log-likelihood ( $-2LL$ ) (Faraway, 2005)

As the REML method estimates the parameters of the regression model and the parameters of the variogram simultaneously, it requires previous modelling of the variogram of the residuals. This analysis was carried out in all cases, after verifying the stationarity and isotropy of the residuals, with the following exponential

variogram:

$$\text{Var}(Z(s_i), Z(s_j)) = 2\gamma(h) = \begin{cases} 0, & h=0; \\ \theta_n - \theta_s \left(1 - \exp\left(-\frac{h}{\theta_r}\right)\right), & h \neq 0 \end{cases} \quad (4)$$

where  $h=|s_i-s_j|$ ,  $\theta_n \geq 0$  is the nugget,  $\theta_s \geq 0$  is the partial sill and  $\theta_r \geq 0$  is the range. The selection of this model of variogram is justified on the basis of comparing the theoretical and the empirical variogram using the test proposed by Maglione and Diblasi (2001). In all cases, the p-values ranged between 0.118 and 0.156.

Table 1 shows the values of  $-2LL$ , AIC and BIC for the six models obtained for some of the ten sampling schemes. It can be observed that model M5, a second order model on longitude and latitude, is the one that behaved better for all sampling schemes. This result led us to reject altitude as the regressor and to model the trend as

$$\mu(s) = \beta_0 + \beta_1 \text{long}(s) + \beta_2 \text{lat}(s) + \beta_3 \text{long}^2(s) + \beta_4 \text{lat}^2(s) + \beta_5 \text{long}(s) \times \text{lat}(s).$$

Figure 3 represents SSRB against longitude and latitude, showing the large-scale trend due to geographical coordinates. An increase in SSRB, as longitude grows, can be seen in the graph. With respect to its relationship to latitude, higher values can be seen in the central zone of the region under study.

The main dependence of the SSRB on geographical location reflects the leading-order relation of clear-sky SSRB with surface albedo. Broadband surface albedo

Table 1. Comparison of the six models' fit via AIC and BIC criteria and minus twice the value of likelihood for the six sampling schemes.

Model	Vario params	Model params	Total params	$-2LL$	AIC	BIC	$-2LL$	AIC	BIC
				<i>Systematic sampling 50%</i>			<i>Random sampling 50%</i>		
M1	3	1	4	12740.4–6374.2	–6384.4		12560.4–6284.2	–6294.4	
M2	3	2	5	12741.9–6375.9	–6388.8		12565.8–6287.9	–6300.7	
M3	3	3	6	12710.0–6361.0	–6376.4		12535.0–6273.5	–6288.9	
M4	3	4	7	12718.3–6366.1	–6384.1		12540.4–6277.2	–6295.2	
M5	3	6	9	12672.7–6345.4	–6368.4		12495.4–6256.7	–6279.8	
M6	3	7	10	12676.9–6348.5	–6374.1		12536.5–6278.2	–6303.9	
				<i>Systematic sampling 12.5%</i>			<i>Random sampling 12.5%</i>		
M1	3	1	4	3283.3–1645.6	–1653.1		3238.7–1623.3	–1630.8	
M2	3	2	5	3288.3–1649.1	–1658.5		3243.5–1626.8	–1636.1	
M3	3	3	6	3258.0–1635.0	–1646.2		3212.8–1612.4	–1623.6	
M4	3	4	7	3264.5–1639.2	–1652.3		3218.2–1616.1	–1629.2	
M5	3	6	9	3214.9–1616.5	–1633.3		3170.3–1594.2	–1611.0	
M6	3	7	10	3220.1–1620.0	–1638.8		3175.7–1597.8	–1616.5	
				<i>Systematic sampling 3.125%</i>			<i>Random sampling 3.125%</i>		
M1	3	1	4	911.2–459.6	–464.5		824.8–416.4	–421.1	
M2	3	2	5	914.3–462.1	–468.3		827.4–418.7	–424.6	
M3	3	3	6	885.1–448.6	–455.9		799.9–405.9	–413.0	
M4	3	4	7	890.6–452.3	–460.9		803.9–408.9	–417.2	
M5	3	6	9	841.7–429.8	–440.8		755.8–386.9	–397.5	
M6	3	7	10	845.5–432.8	–445.0		759.3–389.7	–401.4	

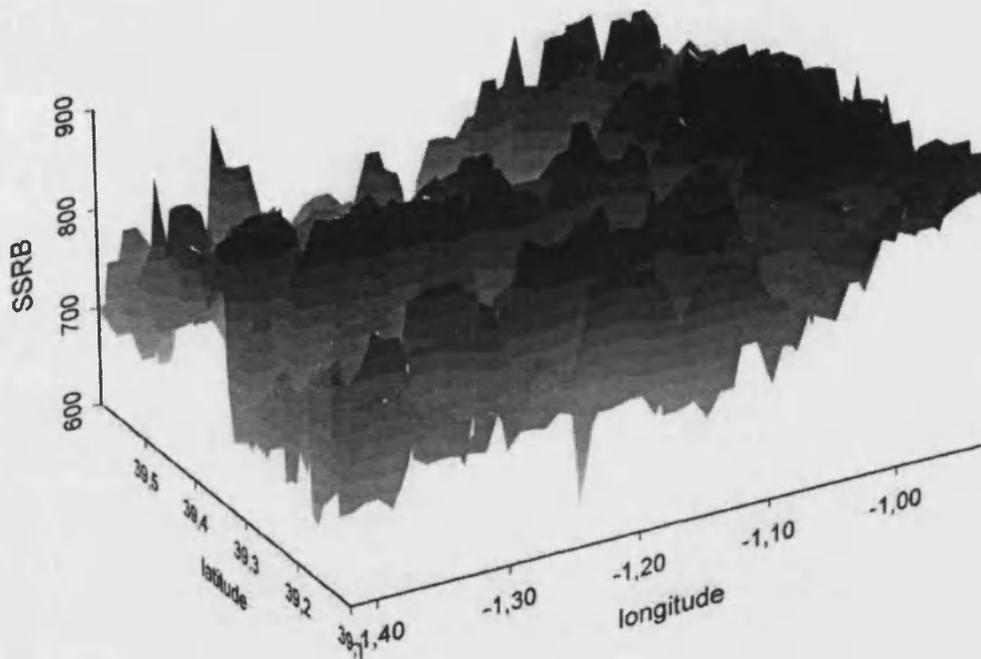


Figure 3. SSRB against geographical coordinates.

over the region under study ranges from  $\sim 0.10$  to  $\sim 0.30$ . For typical values of the incident solar radiation over that region and time of the year, this variation of the surface albedo implies a variation in the SSRB of  $\sim 180 \text{ W m}^{-2}$ . Therefore, the large changes in SSRB with geographic location can be explained by changes in the surface albedo between different geographical locations. However, altitude over that region ranges from 400 to 1000 m, being the impact on the SSRB small, of around  $4 \text{ W m}^{-2}$  for a 100 m change in altitude.

### 5.3 Selection of the sampling scheme: a simulation study

The selection of the best sampling scheme was done by comparing the behaviour of model M5 for the different sampling schemes by means of the value of MSE (mean squared error) over the validation set. In order to do that, we carried out a simulation study by using an ensemble of samples of constant size for each of the ten sampling schemes. This allowed us to extract information on the variability of the MSE in each sampling scheme. Random sampling does not virtually introduce any limitation on the allowed number of repetitions (members of the ensemble), chosen to be 100. That is not the case for the systematic sampling, because each of the sampling schemes shown in figure 2 only allows for a maximum number of realizations that can be obtained from rigid translations of a particular case. In addition to coincidences with previous displacements, any intersection with the validation set was avoided. Table 2 shows a summary of the values of the MSE obtained from this study and the mean SSRB over the whole region in study. As it can be seen, the number of repetitions for the systematic sampling is different for each sample size, decreasing when the sample size increases.

Figure 4 shows the regression curves fitted to the MSE relative mean increments ( $\Delta \text{MSE}$  in table 2), which have been obtained taking  $\text{MSE} = 28.8936$  as the reference

Table 2. Summary of MSE values for validation set obtained with model M5 for  $N$  repetitions of different sampling size. The columns SSRB area refers to the estimation of mean radiation over the whole region in study.

Sampling scheme	Sampling fraction	MSE mean	MSE Std. error	$\Delta$ MSE mean	$\Delta$ MSE Std. error	SSRB area mean	SSRB area Std. error	$N$
Systematic	0.5	29.8114		0.0318		792.5234		1
	0.25	34.8836	1.9149	0.2073	0.0663	792.9562	0.5053	3
	0.125	37.7072	0.6952	0.3050	0.0241	793.0555	0.3503	13
	0.625	42.0204	1.2626	0.4543	0.0437	792.5982	0.8191	15
	0.3125	44.9640	0.5639	0.5562	0.0195	792.8142	0.6675	61
Random	0.5	33.1456	0.1239	0.1472	0.0043	793.1288	0.0656	100
	0.25	36.5198	0.1480	0.2639	0.0051	792.8644	0.1302	100
	0.125	39.8337	0.1909	0.3786	0.0066	792.7223	0.1995	100
	0.625	42.5588	0.2119	0.4730	0.0073	793.0760	0.3361	100
	0.3125	47.1330	0.2873	0.6313	0.0099	792.1999	0.5222	100

value. This is the MSE over the validation set for the model estimated using the whole dataset ( $N=2500-78$ ). These curves have been obtained by means of a weighted least squares (WLS) fitting, using the number of repetitions as the weights. Both curves are logarithmic and have the following expression:

$$\text{Systematic } y = -0.0518 - 0.1761 \ln x \quad R^2 = 0.9929 \quad (5)$$

$$\text{Random } y = 0.0224 - 0.1693 \ln x \quad R^2 = 0.9938. \quad (6)$$

It is concluded from this figure that both sampling schemes behave similarly. The systematic sampling shows better results as the sampling fraction increases.

## 6. Conclusions

Making use of surface shortwave radiation budget (SSRB) as derived from high spatial resolution satellite data, a spatial model that describes the SSRB over a  $\approx 50 \times 50 \text{ km}^2$  region is proposed, and the impact of reducing the sampling points by means of different qualitative and quantitative sampling schemes is analysed. This is a first step that can help in the selection of an objective sampling scheme of the region under study, using a finite number of ground sensors.

Regarding validation activities over that region, the results shown above lead us to conclude that:

- (i) Altitude does not provide relevant information about SSRB and consequently it can be rejected as a regressor when modelling the trend.
- (ii) The complete second order model on longitude and latitude is the one that behaved better for all sampling schemes.
- (iii) The estimated mean values of the SSRB over the whole region, obtained from different sample sizes and sampling schemes, do not differ from the mean value obtained from the 2500 observations in any of the cases analysed. This result somehow reinforces the fact that we can use the various validation networks to estimate the average area over the whole domain. However, it is noted here that, although the estimated mean of the SSRB (see table 2) is not significantly different amongst the various network systems, the standard

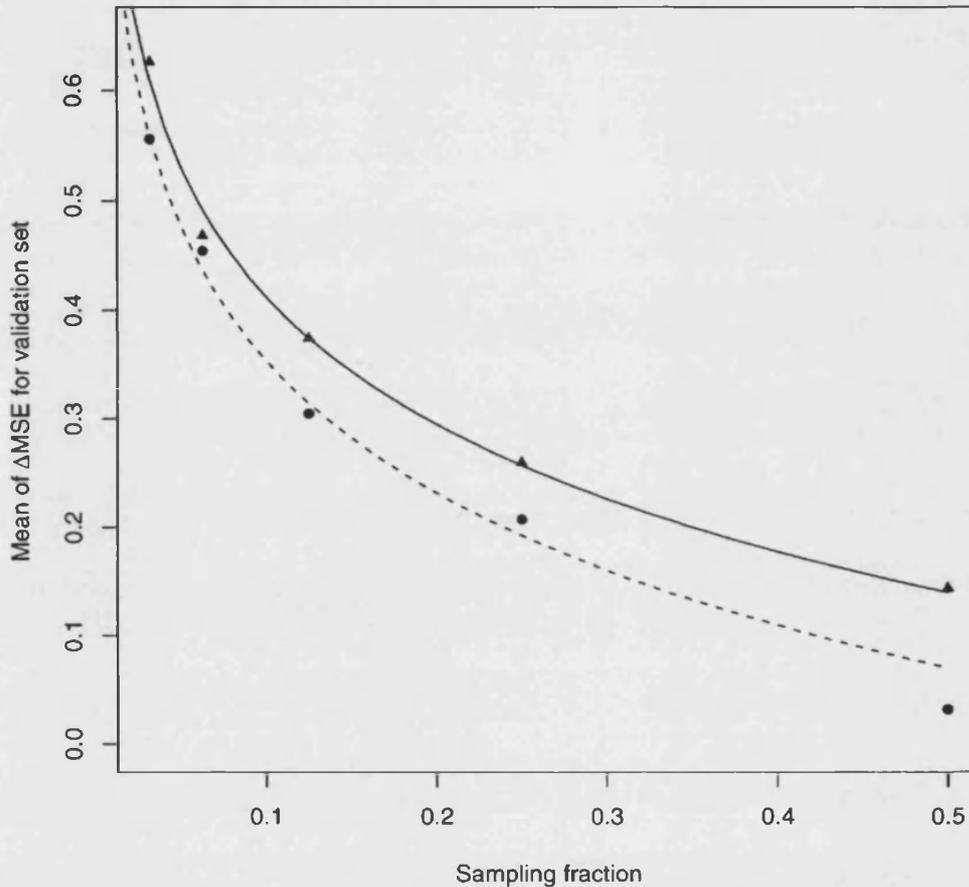


Figure 4. Regression of mean increment of MSE values for validation set over sampling fraction for systematic (●) and random (▲) sampling.

errors make some more real differences, with larger sampling fractions providing lower standard errors.

- (iv) Over the validation set, both sampling schemes show a similar functional dependence of the MSE with the sampling fraction, being this dependence logarithmic. Generally, the systematic sampling behaves better (provides smaller MSE), particularly for large values of the sampling fraction. Consequently, systematic sampling seems to be a better option than random sampling for the deployment of ground instrumentation.
- (v) Figure 4 shows that the difference in MSE between both schemes is small for small values of the sampling fraction. Therefore, if the sampling fraction is constrained to be small, as it is usually the case for ground campaigns, a random sampling may be a better choice if it provides other advantages.
- (vi) The selection of the sampling fraction depends on to competing factors, the cost of the sampling and the MSE. The weight that is applied to each factor may vary for each particular study. In any case, equations like (5) and (6) may be a valuable tool to help in the decision-making process. This is supported by the high  $R^2$  values of the fittings, which provide confidence in the estimation of the MSE from the sampling fraction.

### Acknowledgements

This work has been partly funded by the project BFM-2001-3286 of the DGICYT (Dirección General de Ciencia y Tecnología). The work of Francisco Montes has also been partly funded by the project GRUPOS03/189 of the Generalitat Valenciana.

### References

- ATKINSON, P.M., FOODY, G.M., CURRAN, P.J. and BOYD, D.S., 2000, Assessing the ground data requirements for regional scale remote sensing of tropical forest biophysical properties. *International Journal of Remote Sensing*, **21**, pp. 2571–2887.
- BODAS-SALCEDO, A., LÓPEZ-BAEZA, E., MARTÍNEZ, F., MATEU, J. and MONTES, F., 2003, Spatio-Temporal Modeling and Prediction of Solar Radiation. *J. Geophys. Res.*, **108** (D24).
- BOLLE, H.J. *et al.* (Eds), 2006, *Mediterranean Land-surface Processes Assessed from Space* (Berlin: Springer-Verlag).
- CHARLOCK, T.P. and ALBERTA, T.L., 1996, The CERES/ARM/GEWEX Experiment (CAGEX) for the retrieval of radiative fluxes with satellite data. *Bull. Amer. Meteorol. Soc.*, **77**, pp. 2673–2683.
- CRESSIE, N., 1993, *Statistics for Spatial Data*, Revised Edition (New York: John Wiley).
- CRESSIE, N. and KORNAK, J., 2003, Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statist. Sci.*, **18**, pp. 436–456.
- CHRISTENSEN, O.F., DIGGLE, P.J. and RIBEIRO, P.J., 2001, Analysing positive valued spatial data: the transformed Gaussian model. In *GeoENV III — Geostatistics for Environmental Applications*, P. Monestiez, D. Allard and R. Froidevaux (Eds) (Dordrecht: Kluwer).
- DIBLASI, A. and BOWMAN, A.W., 2001, On the use of the variogram in checking for independence in spatial data. *Biometrics*, **57**, pp. 211–218.
- FARAWAY, J., 2005, *Linear Models with R* (London: Chapman and Hall/CRC).
- FLOWERDEW, R. and GREEN, M., 1992, Developments in areal interpolating methods and GIS. *Annals of Regional Sciences*, **26**, pp. 67–78.
- FLOWERDEW, R. and GREEN, M., 1994, *Spatial Analysis and GIS*, pp. 121–145 (London: Taylor & Francis).
- GLECKLER, P.J., 2005, Surface energy balance errors in AGCMs: Implications for ocean-atmosphere model coupling. *Geophys. Res. Lett.*, **32**, 10.1029/2005GL023061.
- GRIMES, D.I.F., COPPOLA, E., VERDECCHIA, M. and VISCONTI, G., 1993, A neural network approach to real-time rainfall estimation for africa using satellite data. *J. Hydrometeorology*, **4**, pp. 1119–1133.
- GUPTA, S.K., RITCHEY, N.A., WILBER, A.C., WHITLOCK, C.H., GIBSON, G.G. and STACKHOUSE, P.W., 1999, A climatology of surface radiation budget derived from satellite data. *J. Climate*, **12**(8), pp. 2691–2710.
- HARRIES, J.E., *et al.* 2005, The Geostationary Earth Radiation Budget Project. *Bulletin of the American Meteorological Society*, **86**, pp. 945–960.
- KOSLOWSKY, D., 1996, Mehrjaehrige validierte und homogenisierte Reihen des Reflexionsgrades und des Vegetationsindexes von Landoberflaechen aus taeglichen AVHRR-Daten hoher Aufloesung. *Tech. Rep. Neue Folge, Serie A Monographien*, Freie Universitaet Berlin, Meteorologische Abhandlungen.
- LEBEL, T. and AMANI, A., 1999, Rainfall estimation in the Sahel: What is the ground truth? *J. Appl. Meteor.*, **35**, pp. 555–568.
- LI, Z., LEIGHTON, H.G. and CESS, R.D., 1993, Surface net solar radiation estimated from satellite measurements: Comparisons with tower observations. *J. Climate*, **6**, pp. 1764–1772.
- LI, Z., WHITLOCK, C.H. and CHARLOCK, T.P., 1995, Assessment of the global monthly mean surface insolation estimated from satellite measurements using global energy balance archive data. *J. Climate*, **8**, pp. 315–328.

- LÓPEZ-BAEZA, E., VELÁZQUEZ, A., ANTOLÍN, C., BODAS-SALCEDO, A., GIMENO, J.F., SALEH, K., FERRER, F., DOMENECH, C., CASTELL, N. and SÁNCHEZ, M.A., 2001, The Valencia Anchor Station, a Calval reference area for large-scale low spatial resolution remote sensing missions. In *Third Conferencia Internacional sobre Experiencias con Estaciones Meteorológicas Automáticas*, Torremolinos (Málaga), Spain. 19–21 February.
- MAGLIONE, D. and DIBLASI, A., 2001, Exploring a valid model for the variogram of an isotropic spatial process. In *Proceedings of the Annual Conference of the International Association for Mathematical Geology, IAMG2001*, Cancún, Mexico.
- MUGGLIN, A.S., CARLIN, B.P. and GELFAND, A.E., 2000, Fully model-based approaches for spatially misaligned data. *JASA*, **95**, pp. 877–887.
- OHMURA, A., *et al.* 1998, Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bull. Amer. Meteorol. Soc.*, **79**, pp. 2115–2136.
- PARDO-IGÚZQUIZA, E., 1998, Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *J. Hydrology*, **210**, pp. 206–220.
- STEIN, A., VAN DER MEER, F., and GORTE, B.G.H. (Eds), 1999, *Spatial Statistics for Remote Sensing* (Dordrecht: Kluwer Academic).
- VAN GROENINGEN, J.W., SIDERIUS, W. and STEIN, A., 1999, Constrained optimization of soil sampling for minimization of the kriging variance. *Geoderma*, **87**, pp. 239–259.
- WIELICKI, B.A., BARKSTROM, B.R., HARRISON, E.F., LEE III, R.B., SMITH, G.L. and COOPER, J.E., 1996, Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bull. Amer. Meteor. Soc.*, **77**, pp. 853–868.
- WIGNERON, J.P., WALDTEUFEL, P., CHANZY, A., CALVET, J.C. and KERR, Y., 2000, Two-dimensional microwave interferometer retrieval capabilities over land surfaces (SMOS mission). *Remote Sens. Environ.*, **73**, pp. 270–282.
- ZHANG, Y., ROSSOW, W.B., LACIS, A.A., OINAS, V. and MISHCHENKO, M.I., 2004, Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and input data. *J. Geophys. Res.*, **109**(D19,105), 10.1029/2003JD004457.
- ZHU, Z. and STEIN, A., 1999, Spatial sampling design for prediction with estimated parameters. *JABES*, **11**, pp. 24–49.



## Model comparison and selection for stationary space–time models

H.-C. Huang<sup>a</sup>, F. Martinez<sup>b,1</sup>, J. Mateu<sup>c,\*</sup>, F. Montes<sup>b,1</sup>

<sup>a</sup>*Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan*

<sup>b</sup>*Department of Statistics and O.R., Universitat de Valencia, E-46100 Burjassot, Spain*

<sup>c</sup>*Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071 Castellon, Spain*

Received 1 November 2004; received in revised form 1 December 2005; accepted 27 July 2006

Available online 22 August 2006

### Abstract

An intensive simulation study to compare the spatio–temporal prediction performances among various space–time models is presented. The models having separable spatio–temporal covariance functions and nonseparable ones, under various scenarios, are also considered. The computational performance among the various selected models are compared. The issue of how to select an appropriate space–time model by accounting for the tradeoff between goodness-of-fit and model complexity is addressed. Performances of the two commonly used model-selection criteria, Akaike information criterion and Bayesian information criterion are examined. Furthermore, a practical application based on the statistical analysis of surface shortwave radiation budget (SSRB) data is presented.

© 2006 Elsevier B.V. All rights reserved.

**Keywords:** AIC; BIC; Kalman filter; Maximum likelihood; Nonseparable and separable spatio–temporal covariance function; Surface shortwave radiation budget (SSRB)

### 1. Introduction

Many environmental processes, such as groundwater contaminants, air pollution patterns, air–water and air–soil energy exchanges, are characterised by variability and uncertainty that may result from ontologic factors and epistemic limitations. The former are due to the inherent complexity of the natural systems, whereas the latter are associated with incomplete information. Classical statistical approaches often fail to provide a sufficient description of a process evolution across space and time. Spatio–temporal random fields are used to represent correlations between fluctuations in the composite space–time domain. Modelling the effects of fluctuations with suitable covariance functions can improve our ability to characterise and predict space–time variations in various natural systems. Thus, the introduction of models able to integrate spatial and temporal behaviours at different scales constitutes an important topic in probability and statistics in the last decade.

The importance of simultaneously studying spatial and temporal aspects of processes is well-known. Stochastic models describing spatio–temporal processes vary across space and time are essential to the application of statistics to a wide

\* Corresponding author. Tel.: +34 9 64728391; fax: +34 9 64728429.

E-mail addresses: [hhuang@stat.sinica.edu.tw](mailto:hhuang@stat.sinica.edu.tw) (H.-C. Huang), [Francisco.Martinez@uv.es](mailto:Francisco.Martinez@uv.es) (F. Martinez), [mateu@mat.uji.es](mailto:mateu@mat.uji.es) (J. Mateu), [Francisco.Montes@uv.es](mailto:Francisco.Montes@uv.es) (F. Montes).

<sup>1</sup> Research supported by Grant MTM2004-06231 from Spanish Ministry of Science and Education.

range of environmental applications. However, the structural analysis of such processes is more difficult than for spatial or temporal processes. Estimating and modelling the correlation of a space–time process are principal goals in geostatistical analysis. In the last years, several efforts have been made for the construction of valid spatio–temporal covariance structures, making the spatio–temporal modelling an important area that is one of rapid growth at the moment, with various applications in environmental science, geophysical science, biology, epidemiology and others. The reader is referred to the comprehensive reviews of Kyriakidis and Journel (1999) and Kolovos et al. (2004).

Dimitrakopoulos and Lou (1994) proposed some geometrically anisotropic models for spatio–temporal models. Guttorp et al. (1994) proposed some separable covariance structures, obtained with the product of a spatial and a temporal covariance. In a similar context of separability, Rohuani and Hall (1989) proposed the so-called sum-variogram. Other approaches assume that the drift could entirely catch the temporal variability, and consider the spatio–temporal process to be temporally independent (Sampson and Guttorp, 1992). In the nonseparable context, several efforts have been made in order to obtain a class of spatio–temporal covariances which must be as much general as possible. It is worth citing Jones and Zhang (1997), Cressie and Huang (1999), Christakos (2000), Brown et al. (2000), de Iaco et al. (2001), de Cesare et al. (2001a, b), Gneiting (2002), Ma (2003a, b), Stein (1999, 2005) and Fernández-Casal (2003). Most of the contributions regard stationary spatio–temporal covariances assuming isotropy on space and time. Particularly, Cressie and Huang (1999) proposed a spectral approach to obtain spatio–temporal covariances, and Gneiting's (2002) work represent the natural generalisation to this approach, obtained using completely monotone functions and functions whose first derivative is completely monotone. Stein (2005) puts emphasis on the spectral approach. Finally, Fernández-Casal (2003) extends the Shapiro–Botha approach to build flexible variograms in the spatio–temporal context.

Three expressions have been generally used to describe a spatio–temporal random field: a dynamic stochastic representation, a spatio–temporal covariance function, and a spatio–temporal spectral density function (if the spatio–temporal process is stationary).

A *stochastic (partial) differential equation* could be formally used to describe a spatio–temporal random field, although some of the (partial) derivatives in the equation often do not exist in the usual sense. A deterministic physical model often takes the form of an ordinary or partial differential equation. A dynamic or stochastic model may be formally described by a stochastic (partial) differential equation. Whittle (1954) described a spatial random field by using the stochastic partial differential equation driven by white noise. His work was extended by Heine (1955), Whittle (1956, 1962) and Renshaw (1994). Jones and Zhang (1997) considered a separable and nonseparable spatio–temporal random field defined by two different stochastic partial differential equations giving valid correlation functions. See also Christakos and Hristopoulos (1998), and Kolovos et al. (2004) for other models generated from stochastic partial differential equations. There are dynamic space–time models represented using autoregressive time-series models on discrete time points, where Kalman-filter recursions can be applied to obtain optimal predictors. See for example, Huang and Cressie (1996), Mardia et al. (1998), Wikle and Cressie (1999), Stroud et al. (2001), and Huang and Hsu (2004). See also the excellent review on space–time Kalman filtering in Cressie and Wikle (2002).

The *spectral density function* is an important tool to characterise a stationary spatio–temporal random field, just like the covariance function. Mathematically speaking, the spectral density function and the covariance function are closely related as the pair of the Fourier transform. Thus, the spectral density approach is particularly helpful if there is no explicit expression for the covariance function. Stein (2005) considers a parametric class of spectral densities obtaining some desirable features of the class. In particular, it is noted that it may be desirable to use covariance functions that are smooth away from the origin, and that it can be proved that a covariance function is smooth away from the origin by proving its corresponding spectral density has derivatives with certain moments. These results are used to build covariance functions that are infinitely differentiable away from the origin and allow for arbitrary and possibly different degrees of smoothness for the process in space and time.

A separable spatio–temporal covariance model does provide a useful base for deriving nonseparable spatio–temporal covariance models through appropriate mixture procedures, as shown in Cesare et al. (2001a,b) and Ma (2002, 2003a,b). Ma (2005a) analyses the linear combination of two spatial or spatio–temporal covariances (or variograms) which are isotropic in space. Furthermore, in Ma (2005b) linear combinations of spatio–temporal covariances, where the resulting covariance functions are stationary in time but may not be in space, are presented. New spatio–temporal covariances may be obtained by taking partial derivatives of a differentiable spatio–temporal covariance, whenever it exists. Examples can be found in Ma (2005b).

The theory of *spatio-temporal ordinary and generalised random fields* also provides a suitable framework for the statistical analysis and modelling of spatio-temporal data (Angulo et al., 2000; Anh and Leonenko, 2001, 2002).

The different methods recalled here and those shown in Kolovos et al. (2004) give an overall picture of the current developments on the construction of spatio-temporal covariance models. However, there has been little direct work examining the consequences of using various types of models. Additionally, in practice, a tradeoff between mathematical tractability of the analytical expressions and incorporation of more realistic features or physical meanings into the model is often required. This paper focuses on stationary models and our attention in this paper is mostly paid to the covariance function, a widely used measure for space and time interaction as well as dependence. The direct use of the covariance function in data analysis include those for evaluating the simple kriging predictor and for calculating the likelihood function.

In practical modelling, the researcher has to face the important decision of which spatio-temporal model is the one that best fits the empirical data. To the best of our knowledge, actually there is no published paper presenting an intensive simulation study comparing the performance of various such models with the aim of helping the researcher in model selection. Thus, the goal of this paper is to compare the spatio-temporal prediction performances among various space-time models, including those having separable spatio-temporal covariance functions and non-separable ones, under several scenarios. Computational performances among various methods will also be compared. In particular the classes of models used here are: (a) the separable model in Huang and Cressie (1996); (b) the class of models given in Cressie and Huang (1999); (c) the class of nonseparable models presented in Gneiting (2002); and (d) a class of nonseparable models based on a simple product-sum structure, as presented in de Cesare et al. (2001a). These classes of models cover quite a portion of spatio-temporal covariance functions used in the literature.

In the space-time modelling literature, model parameters are traditionally estimated by using some method-of-moments estimation tools. In this paper, a more efficient maximum-likelihood method is used. We also address an important issue, which has somewhat been ignored in the past literature, on how to select an appropriate space-time model by accounting for the tradeoff between goodness-of-fit and model complexity. Performances of two commonly used model-selection criteria, Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978), are examined.

Note that the asymptotic theory of both AIC and BIC for geostatistical models has not been fully studied in the literature, and thus it is of great interest to investigate whether the traditional consistent result of BIC and the asymptotic efficiency result of AIC in regression and time series model selection would yet hold in the framework of space-time modelling. However, this problem is highly nontrivial, because even the maximum-likelihood estimators of some covariance parameters may not be consistent under infill asymptotics (Zhang, 2004), not to mention asymptotic normality. Rather than proving any asymptotic results, our aim here is to remind the readers the importance of model selection for space-time models, a topic that has somewhat been overlooked, and try to highlight throughout our simulation experiments that both AIC and BIC show some promise in space-time model selection.

Finally, but not least, and on a more applied context, we make use of the findings of the simulation study to present a space-time analysis of the surface shortwave radiation budget (SSRB), which is the amount of energy in the solar region of the electromagnetic spectrum (0.2–4.0  $\mu\text{m}$ ) absorbed at the surface. The large pixel size of some satellite sensors for climate studies introduces a number of scientific issues that make relevant, and even necessary, to develop a specific methodology. Without doubt, the large size of the reference pixel makes it necessary to study the change of scale processes in detail in order to be able to compare measurements with different scales (namely point measurements, aircraft observations, remote sensing images with different spatial resolutions), and establish ground-measurement sampling criteria to get different area averages and validate large scale pixels. In this respect, remote sensing poses interesting scientific questions that spatial statistics can help to deal with. In particular, spatial statistics addresses the classification of images by supervised and unsupervised procedures, and the estimation of spatial structures by variograms, interpolation, simulation, sampling and issues of scale. Thus, a space-time analysis can reveal interesting and useful information in, for example, image analysis or remote sensing disciplines.

The plan of the paper is as follows. Section 2 introduces basic concepts for the statistical formulation needed throughout the paper. The various space-time models used in this paper are also described. A comprehensive and intensive simulation study is given in Section 3. Section 4 shows the spatio-temporal analysis of the SSRB data through the models presented in Section 2. The paper ends with some conclusions and discussions.

## 2. Statistical formulation and methodology

### 2.1. Setup

Consider a real-valued spatio-temporal random field  $\{Y(s, t), s \in D, t \in \mathbb{Z}\}$  over the spatial domain  $D \subset \mathbb{R}^2$  and the temporal domain  $\mathbb{Z}$ . Its covariance function,  $C(s_1, s_2, t_1, t_2)$ , and semi-variogram,  $\gamma(s_1, s_2, t_1, t_2)$ , are defined by

$$C(s_1, s_2, t_1, t_2) = E\{[Y(s_1, t_1) - EY(s_1, t_1)][Y(s_2, t_2) - EY(s_2, t_2)]\},$$

$$\gamma(s_1, s_2, t_1, t_2) = \frac{1}{2} \text{Var}\{[Y(s_1, t_1) - EY(s_1, t_1)] - [Y(s_2, t_2) - EY(s_2, t_2)]\},$$

with  $(s_1, t_1), (s_2, t_2) \in D \times \mathbb{Z}$ , and whenever they exist.

Under the assumption that  $\text{var}(Y(s, t)) < \infty$  for all  $(s, t) \in D \times \mathbb{Z}$ , the covariance function and the semi-variogram are well-defined and related through

$$\gamma(s_1, s_2, t_1, t_2) = \frac{1}{2} \{C(s_1, s_1, t_1, t_1) + C(s_2, s_2, t_2, t_2)\} - C(s_1, s_2, t_1, t_2).$$

The random field  $Y(s, t)$  is said to be (weakly, or second-order) stationary in time if  $EY(s, t_1) = EY(s, t_2)$ , and  $C(s_1, s_2, t_1, t_2)$  depends only on  $t_1 - t_2$  for all  $s_1, s_2 \in D$ . It is said to be stationary in space if  $EY(s_1, t) = EY(s_2, t)$ , and  $C(s_1, s_2, t_1, t_2)$  depends only on  $s_1 - s_2$  for all  $t_1, t_2 \in \mathbb{Z}$ . Finally, the random field is said to be stationary in space-time if  $EY(s_1, t_1) = EY(s_2, t_2)$ , and  $C(s_1, s_2, t_1, t_2)$  depends only on  $s_1 - s_2$  and  $t_1 - t_2$ . When the random field is stationary in time (or space), its covariance is denoted by  $C(s_1, s_2, t)$  (or  $C(s, t_1, t_2)$ ). When the random field is stationary in space-time, its covariance is denoted by  $C(s, t)$  and called a stationary covariance function. In the same way, the intrinsic stationarity is characterised in terms of the variogram. Equivalently, a constant-mean random field  $Y(s, t)$  is intrinsically stationary in space-time if the increment process  $Y(s + s_0, t + t_0) - Y(s, t)$  is stationary in space-time for any fixed  $(s_0, t_0) \in D \times \mathbb{Z}$ . The intrinsic stationarity is more general than the second-order stationarity, since they are processes which are intrinsically stationary but not stationary. If  $Y(s, t)$  is stationary, then it is also intrinsically stationary.

In particular, in this paper we consider  $\{Y(s, t); s \in D \subset \mathbb{R}^2, t \in \mathbb{Z}\}$  a spatio-temporal Gaussian process of interest. Suppose that  $Y(\cdot, \cdot)$  is stationary with  $E(Y(s, t)) \equiv 0$  (i.e. there is no trend) and spatio-temporal covariance function

$$C(\mathbf{h}, \mathbf{u}) \equiv \text{Cov}(Y(s, t), Y(s + \mathbf{h}, t + \mathbf{u})); \quad s, s + \mathbf{h} \in D, \quad u, t \in \mathbb{Z},$$

where  $\mathbf{h}$  is the spatial lag and  $\mathbf{u}$  the temporal lag. Also suppose that data  $\mathbf{Z}(t) = (Z(s_1, t), \dots, Z(s_n, t))^t; t = 1, \dots, T$ , are observed at  $n$  locations and  $T$  time points according to the following measurement equation

$$\mathbf{Z}(s, t) = Y(s, t) + \varepsilon(s, t),$$

where  $\varepsilon(\cdot, \cdot)$  is a white-noise process with variance  $\sigma_\varepsilon^2$ .

The optimal linear predictor of  $Y(s, T)$  based on  $\mathbf{Z}(1), \dots, \mathbf{Z}(T)$  can be written as

$$\hat{Y}(s, T) = \text{Cov}(Y(s, T), \mathbf{Z}) \text{Var}(\mathbf{Z})^{-1} \mathbf{Z} \equiv E(Y(s, T) | \mathbf{Z}), \quad (1)$$

where  $\mathbf{Z} \equiv (\mathbf{Z}(1)^t, \dots, \mathbf{Z}(T)^t)^t$ . The prediction method is usually referred to as simple kriging. The corresponding prediction-error variance is

$$\text{Var}(Y(s, T)) - \text{Cov}(Y(s, T), \mathbf{Z}) \text{Var}(\mathbf{Z})^{-1} \text{Cov}(\mathbf{Z}, Y(s, T)). \quad (2)$$

In this paper we selected four spatio-temporal models, covering the situations of separability and nonseparability. Our Model I (Huang and Cressie, 1996) is a typical class of separable model, in which the spatial covariance function is arbitrary, and the temporal covariance function has an  $ARMA(p, q)$  structure. This class of models is quite large, covering many spatio-temporal covariance structures used in the literature. Moreover, inference and prediction can be implemented fast using the Kalman-filter technique. A systematic way to construct classes of nonseparable models can be found in Cressie and Huang (1999). Model II is a class of models given in Cressie and Huang (1999), while Model III (Gneiting, 2002) is a class of nonseparable models that is constructed using an approach extended from Cressie

and Huang (1999). Model IV (de Cesare et al., 2001a) shows another way to build nonseparable models, which has a simple product–sum structure.

These four classes of models cover quite a portion of spatio–temporal covariance functions used in the literature. There are other types of models, which may be very complicated and are, somewhat, ad hoc to the problem in hand and thus are not being considered here.

2.2. Dynamic space–time models (Huang and Cressie, 1996)

Consider a spatio–temporal stationary process,  $Y(s, t)$ , satisfying the relation

$$Y(s, t) = \alpha Y(s, t - k) + \zeta(s, t); \quad s \in D, \quad t \in \mathbb{Z},$$

where  $\zeta(\cdot, \cdot)$  is a spatially stationary and temporally independent Gaussian error process. The spatial covariance function of  $Y(\cdot, t)$  is assumed to belong to the Matérn class (Matérn, 1960)

$$C_\phi(\mathbf{h}) = \sigma_Y^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (b\|\mathbf{h}\|)^\nu K_\nu(b\|\mathbf{h}\|),$$

where  $\phi \equiv (b, \nu, \sigma_Y^2)'$ ,  $\sigma_Y^2 = \text{Var}(Y(s, t))$ ,  $b > 0$  is the scaling parameter, and  $\nu > 0$  is the smoothness parameter. Then, the spatio–temporal covariance function of  $Y(\cdot, \cdot)$  is given by

$$C_Y(\mathbf{h}, u|\phi, \alpha) = \alpha^{|u|} C_\phi(\mathbf{h}),$$

which defines a separable spatio–temporal covariance function.

The optimal predictor of  $Y(s, t)$  based on data  $Z(1), \dots, Z(t)$  is given by

$$\hat{Y}(s, t) \equiv E(Y(s, t)|Z(1), \dots, Z(t)).$$

The prediction-error covariance of  $Y(s, t)$  and  $Y(s', t)$  based on data  $Z(1), \dots, Z(t)$  is given by

$$\begin{aligned} P_t(s, s') &\equiv \text{Cov}(\hat{Y}(s, t) - Y(s, t), \hat{Y}(s', t) - Y(s', t)) \\ &= \text{Cov}(Y(s, t), Y(s', t) | Z(1), \dots, Z(t)). \end{aligned}$$

Huang and Cressie (1996) showed that  $\hat{Y}(s, t)$  and  $P_t(s, s')$ , for any  $s, s' \in D$ , can be computed recursively, using the Kalman-filter algorithm, as follows:

$$\hat{Y}(s, t) = \alpha \hat{Y}(s, t - 1) + Q_t(s)' F_t^{-1} (Z(t) - \alpha \hat{Y}(t - 1)),$$

$$P_t(s, s') = P_{t|t-1}(s, s') + Q_t(s)' F_t^{-1} Q_t(s'),$$

for  $t = 1, \dots, T$ , where

$$\hat{Y}(t - 1) \equiv (\hat{Y}(s_1, t - 1), \dots, \hat{Y}(s_n, t - 1))',$$

$$P_{t|t-1}(s, s') \equiv \alpha^2 P_{t-1}(s, s') + \text{Cov}(\zeta(s, t), \zeta(s', t)),$$

$$Q_t(s) \equiv (P_{t|t-1}(s_1, s), \dots, P_{t|t-1}(s_n, s))',$$

$$F_t \equiv (P_{t|t-1}(s_i, s_j))_{n \times n} + \sigma_\zeta^2 I_n.$$

The Kalman-filter algorithm starts with  $\hat{Y}(s, 0) \equiv 0$  and  $P_{1|0}(s, s') \equiv \text{Cov}(Y(s, 1), Y(s', 1))$ . The joint distribution of  $Z$  can be written as

$$P(Z) = P(Z(1)) \prod_{t=2}^T P(Z(t)|Z(1), \dots, Z(t - 1)),$$

where

$$\mathbf{Z}(t) | \mathbf{Z}(1), \dots, \mathbf{Z}(t-1) \sim \text{Gau}(\alpha \hat{\mathbf{Y}}(t-1), \mathbf{F}_t).$$

Consequently, the log-likelihood function of  $\theta = (a, b, \nu, \sigma_Y^2, \sigma_\epsilon^2)'$  based on  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')'$  is given by

$$L(\theta; \mathbf{Z}) = \log(2\pi)^{-nT/2} - \frac{1}{2} \left\{ \sum_{t=1}^T \left( \log |\mathbf{F}_t| + (\mathbf{Z}(t) - \alpha \hat{\mathbf{Y}}(t-1))' \mathbf{F}_t^{-1} (\mathbf{Z}(t) - \alpha \hat{\mathbf{Y}}(t-1)) \right) \right\}.$$

Note that  $\{\mathbf{F}_t\}$  and  $\{\hat{\mathbf{Y}}(t)\}$  depend on the parameter vector  $\theta$  and can be computed recursively from the Kalman filter. Therefore,  $L(\theta; \mathbf{Z})$  and the ML estimator of  $\theta$  can be computed efficiently using the Kalman filter. In contrast, Huang and Cressie (1996) used a simple but less efficient method-of-moments estimator of  $\theta$ .

### 2.3. Nonseparable space–time models (Cressie and Huang, 1999)

A five-parameter nonseparable spatio–temporal stationary covariance family given by Cressie and Huang (1999) is of the form

$$C(\mathbf{h}, u | \theta) = \begin{cases} \frac{\sigma_Y^2 (2\beta^{d/2})}{(a^2 u^2 + 1)^\nu (a^2 u^2 + \beta)^{d/2} \Gamma(\nu)} \left\{ \frac{b (a^2 u^2 + 1)^{1/2}}{2 (a^2 u^2 + \beta)} \|\mathbf{h}\| \right\}^\nu \\ \quad \times K_\nu \left( b \left( \frac{a^2 u^2 + 1}{a^2 u^2 + \beta} \right)^{1/2} \|\mathbf{h}\| \right) & \text{if } \|\mathbf{h}\| > 0, \\ \frac{\sigma_Y^2 \beta^{d/2}}{(a^2 u^2 + 1)^\nu (a^2 u^2 + \beta)^{d/2}} & \text{if } \|\mathbf{h}\| = 0, \end{cases}$$

where  $\theta \equiv (a, b, \beta, \nu, \sigma_Y^2, \sigma_\epsilon^2)'$ ,  $K_\nu$  is the modified Bessel function of the second kind of order  $\nu$ ,  $a \geq 0$  is the scaling parameter of time,  $b \geq 0$  is the scaling parameter of space,  $\beta > 0$  is a space–time interaction parameter,  $\nu > 0$  is a smoothness parameter, and  $\sigma_Y^2 = C(\mathbf{0}, 0) > 0$ . Notice that a separable covariance function is obtained when  $\beta = 1$ . Also, in this case,  $C(\mathbf{h}, 0 | \theta)$  corresponds to the Matérn class.

The log-likelihood function of  $\theta$  based on  $\mathbf{Z} = (\mathbf{Z}(1)', \dots, \mathbf{Z}(T)')'$  can be written as

$$L(\theta; \mathbf{Z}) = \log(2\pi)^{-nT/2} - \frac{1}{2} \left( \log |\Sigma| + \mathbf{Z}' \Sigma^{-1} \mathbf{Z} \right), \quad (3)$$

where  $\Sigma \equiv \text{Var}(\mathbf{Z})$ . Note that the  $(i, j)$ th entry of  $\Sigma$  is given by

$$C(s_{i^*} - s_{j^*}, t_i - t_j | \theta) + \sigma_\epsilon^2 \delta_{ij},$$

where  $i^* \equiv i - n[(i-1)/n]$ ,  $j^* \equiv j - n[(j-1)/n]$ ,  $t_i \equiv [(i-1)/n]$ ,  $t_j \equiv [(j-1)/n]$ ,  $[x]$  is the largest integer smaller or equal to  $x$ , and  $\delta_{ij}$  is the Kronecker  $\delta$  function.

### 2.4. Nonseparable space–time models (Gneiting, 2002)

A seven-parameter nonseparable spatio–temporal stationary covariance family given by Gneiting (2002) is of the form

$$C(\mathbf{h}, u | \theta) = \begin{cases} \frac{\sigma_Y^2}{2^{\nu-1} \Gamma(\nu) (a|u|^{2\alpha} + 1)^{\delta+\beta}} \left( \frac{b \|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right)^\nu \\ \quad \times K_\nu \left( \frac{b \|\mathbf{h}\|}{(a|u|^{2\alpha} + 1)^{\beta/2}} \right) & \text{if } \|\mathbf{h}\| > 0, \\ \frac{\sigma_Y^2}{(a|u|^{2\alpha} + 1)^{\delta+\beta}} & \text{if } \|\mathbf{h}\| = 0, \end{cases}$$

where  $\theta \equiv (a, b, x, \beta, \nu, \delta, \sigma_Y^2, \sigma_\varepsilon^2)^t$ ,  $a > 0$  and  $b > 0$  are scaling parameters of time and space, respectively,  $\alpha \in (0, 1]$  is the smoothness parameter of time,  $\nu > 0$  is the smoothness parameter of space,  $\beta \in [0, 1]$ ,  $\delta \geq 0$ , and  $\sigma_Y^2 > 0$ . Notice that a separable covariance function is obtained when  $\beta = 0$ , and  $C(\mathbf{h}, 0|\theta)$  corresponds to the Matérn class. The log-likelihood function of  $\theta$  based on  $\mathbf{Z}$  is given in the form of (3).

### 2.5. Product-sum models (de Cesare et al., 2001a)

A seven-parameter nonseparable spatio-temporal stationary covariance family is given by

$$C(\mathbf{h}, u|\theta) = \sigma_Y^2 (k_1 C_s(\mathbf{h}) C_t(u) + k_2 C_s(\mathbf{h}) + (1 - k_1 - k_2) C_t(u)),$$

where  $\theta \equiv (k_1, k_2, b, \alpha, \nu, \sigma_Y^2, \sigma_\varepsilon^2)^t$ ,  $C_s(\mathbf{h}) = (2^{1-\nu}/\Gamma(\nu))(b\|\mathbf{h}\|)^{\nu} K_{\nu}(b\|\mathbf{h}\|)$  is the Matérn spatial correlation class, and  $C_t(u) = \alpha^{|u|}$  is the AR(1) autocorrelation function. The log-likelihood function of  $\theta$  based on  $\mathbf{Z}$  is given in the form of (3).

### 2.6. Parameter estimation and model selection

We used the more efficient maximum likelihood estimation (MLE) rather than the traditional and commonly used method-of-moments. In particular, MLE is based on minimising the minus (log-)likelihood function of the corresponding model,

$$\hat{\theta} = \arg \min_{\theta} -L(\theta; \mathbf{Z}).$$

Rather than entering into a fruitless debate over which approach is better, it is more appropriate to address the question through model selection, which makes the answer problem-specific. Methods of cross-validation, minimising an information criterion, or minimising description length offer alternative ways to discriminate between different models of the same data. Little is known about their properties in the spatial context, a topic of research that needs attention in the future.

Thus, concerning model selection, we focus on both, the Akaike information criterion (AIC) (Akaike, 1973)

$$AIC = -2 \log L(\hat{\theta}; \mathbf{Z}) + 2|\theta|,$$

and the Bayesian Information Criterion (BIC) (Schwarz, 1978)

$$BIC = -2 \log L(\hat{\theta}; \mathbf{Z}) + \log(nT)|\theta|.$$

The properties of AIC and BIC are not well established for space-time models. Recall that, in general, asymptotic efficiency and consistency are viewed as incompatible properties. AIC is the most widely-known representative of the class of asymptotically efficient criteria, and BIC is the most widely-known representative of the class of consistent criteria (McQuarrie and Tsai, 1998).

### 3. Simulation study

In this section, we describe a thorough simulation study examining and comparing the performance of various spatio-temporal models. Simulations were performed under different scenarios, including separable and nonseparable spatio-temporal covariance functions, various intensities of spatial and temporal dependencies, various degrees of smoothness, and various noise levels. Specifically, we considered the four classes of spatio-temporal models presented in Section 2 with the following combinations of model parameters:

- Model I:  $b \in \{0.2, 0.5\}$ ,  $x \in \{0.4, 0.8\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 \in \{1, 5\}$ .
- Model II:  $a = 1$ ,  $b \in \{1, 5\}$ ,  $\beta \in \{0.5, 1, 5\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 \in \{1, 5\}$ .
- Model III:  $a = 0.5$ ,  $b \in \{0.3, 1.5\}$ ,  $\alpha \in \{0.1, 0.9\}$ ,  $\beta \in \{0, 0.9\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\delta = 0.5$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 = 1$ .
- Model IV:  $(k_1, k_2) \in \{(0.8, 0.1), (0.4, 0.3)\}$ ,  $b \in \{0.3, 1.1\}$ ,  $\alpha \in \{0.4, 0.8\}$ ,  $\nu \in \{0.5, 1\}$ ,  $\sigma_Y^2 = 5$ ,  $\sigma_\varepsilon^2 = 1$ .

With the aim of covering as much practical possibilities as possible, the parameter values were chosen as follows. For Model I,  $b = 0.2$  (0.5) corresponds to a large (small) spatial range,  $\alpha = 0.4$  (0.8) corresponds to a small (large) temporal dependence structure, and  $\nu$  is a smoothness parameter in space with  $\nu = 1$  corresponding to a smoother process than  $\nu = 0.5$ . For Model II,  $b = 1$  (5) corresponds to a large (small) spatial range,  $\beta$  is the space–time interaction parameter with  $\beta = 1$  corresponding to a separable spatio–temporal covariance function, and  $\nu$  is a smoothness parameter with  $\nu = 1$  corresponding to a smoother process than  $\nu = 0.5$ . For Model III,  $b = 0.3$  (1.5) corresponds to a large (small) spatial range,  $\alpha = 0.3$  (0.9) corresponds to a small (large) temporal dependence structure,  $\beta$  is the space–time interaction parameter with  $\beta = 0$  corresponding to a separable spatio–temporal covariance function, and  $\nu$  is a smoothness parameter with  $\nu = 1$  corresponding to a smoother process than  $\nu = 0.5$ . For Model IV, different values of  $(k_1, k_2)$  correspond to different weights to the mixture of components,  $b = 0.3$  (1.5) corresponds to a large (small) spatial range,  $\alpha = 0.4$  (0.8) corresponds to a small (large) temporal dependence structure, and  $\nu$  is a smoothness parameter in space with  $\nu = 1$  corresponding to a smoother process than  $\nu = 0.5$ .

We generated spatio–temporal variables  $\{Y(s(i, j), t) : i, j = 1, \dots, 10, t = 1, \dots, 10\}$  on a regular grid of  $(10 \times 10) \times 10$  space–time points, where  $s(i, j) \equiv (i - 0.5, j - 0.5)/10 \in [0, 1] \times [0, 1]$ . Out of the 100 spatial locations, 20 locations were randomly selected. Then, a random sample was taken at these 20 locations for  $t = 1, \dots, 10$  with additive white noise

$$Z(s(i_k, j_k), t) = Y(s(i_k, j_k), t) + \varepsilon(s(i_k, j_k), t); \quad k = 1, \dots, 20, \quad t = 1, \dots, 10,$$

resulting in a dataset of sample size 200, where  $\{\varepsilon(s(i_k, j_k), t)\}$  are Gaussian white-noise variables with variance  $\sigma_\varepsilon^2$ .

For each dataset, the four classes of spatio–temporal models given in Section 2 were used to fit the data. To speed up the simulation times, and to stabilise the minimisation algorithm, the noise variance  $\sigma_\varepsilon^2$  was assumed known throughout the simulation experiment. Another advantage of assuming  $\sigma_\varepsilon^2$  to be known is that this is a key parameter, and may not be well estimated by MLE.

The prediction was obtained, using the simple-kriging predictors (see Eq. (1)), by plugging in the fitted spatio–temporal covariance structure, and the performances among various methods were compared using the average mean-squared-prediction-error (MSPE) criterion at the last time, i.e. time 10, as follows:

$$MSPE = \frac{1}{10^2} \sum_{i=1}^{10} \sum_{j=1}^{10} E\left(\hat{Y}(s(i, j), 10) - Y(s(i, j), 10)\right)^2. \quad (4)$$

Note that we can take advantage of the fact that the MSPE under the true model can be calculated exactly, and does not need to be simulated. Then, using standard properties of Gaussian distributions, we have that each term  $E\left(\hat{Y}(s(i, j), t) - Y(s(i, j), t)\right)^2$  appearing in (4) can be decomposed as

$$E\left(\hat{Y}(s(i, j), t) - Y(s(i, j), t)\right)^2 = E\left(\hat{Y}^*(s(i, j), t) - Y(s(i, j), t)\right)^2 + E\left(\hat{Y}(s(i, j), t) - \hat{Y}^*(s(i, j), t)\right)^2, \quad (5)$$

where  $\hat{Y}^*(s(i, j), t)$  is the optimal (best linear) predictor of  $Y(s(i, j), t)$  under the true model.

Since the first term on the right-hand side of (5) can be computed exactly using the usual formula for the MSPE of the best linear predictor (see Eqs. (2) and (4)), we only need to approximate the second term by using Monte Carlo simulation. Note that to obtain this second term via simulation, one only needs to simulate  $Y(s(i, j), t)$  at data locations, which clearly speeds up the simulation study. Incidentally, (5) proves that the true model must give smaller MSPE than any fitted model, and this will be confirmed by the simulation results. However, although the MSPE for the true model can be computed theoretically, it depends on the sampling locations. Thus, the MSPE for the true model is reported by averaging over 100 different sampling locations used in the simulations.

It is necessary to stress that even if starting with the optimal initial parameters, we may easily face the problem of convergence to local (but not global) minima, resulting in unusual large MSPE values. To check for convergence, we compared the four likelihood (one for each model) values or the corresponding AIC values. It gives a pretty precise indication of convergence problems if one is significantly larger than the others. In particular, an automatic but effective

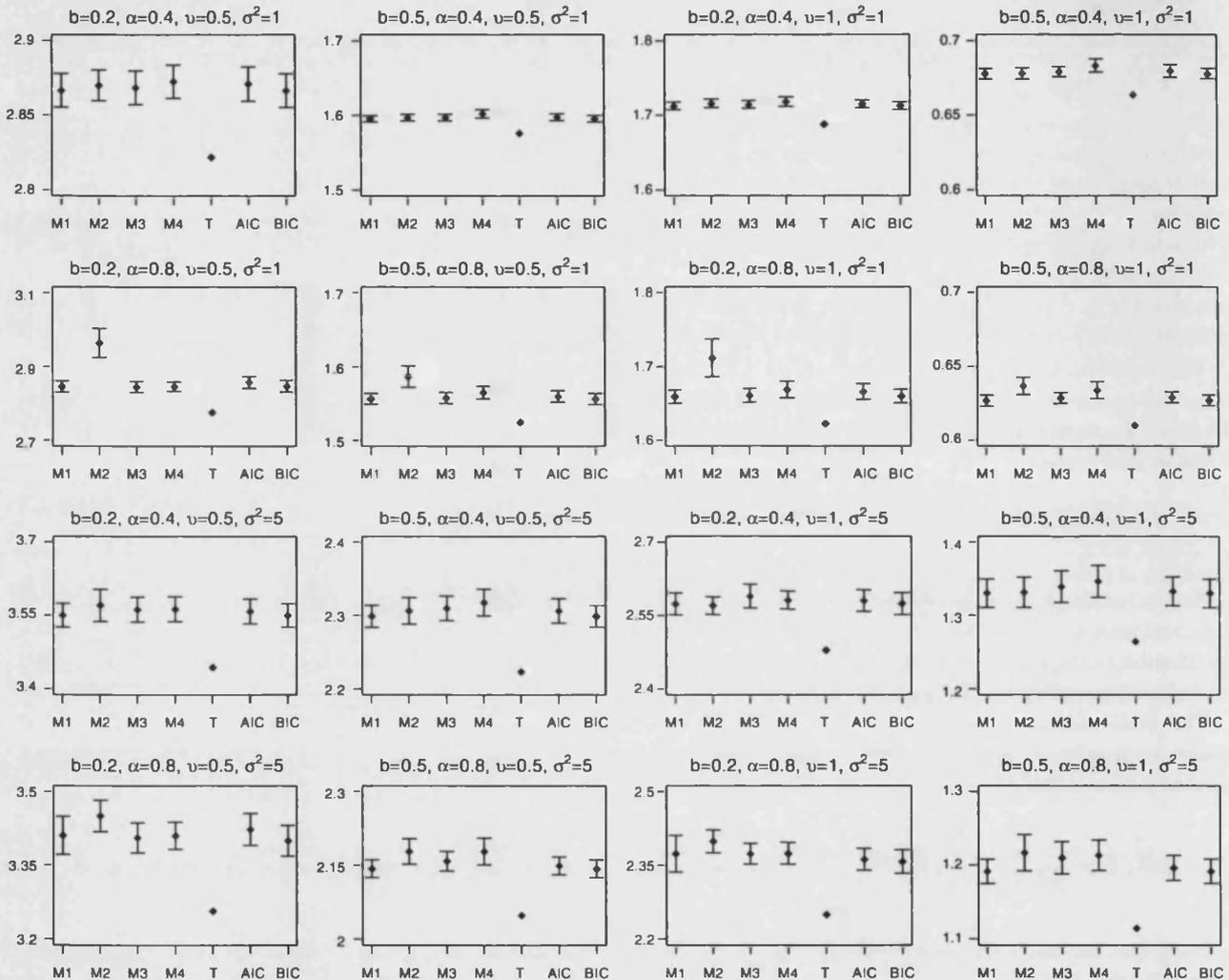


Fig. 1. Segment-shape plots of  $mean \pm 1.96 * standard\ error$  of MSPE values based on 100 replications of data generated from Model I with  $\sigma_Y^2 = 5$ . Within each plot, the order of the plots from left to right is: Model I, II, III, IV, True, AIC and BIC.

method is used here to remove those replicates with outlying MSPE values: a particular replicate is removed when the maximum MSPE values among the four models is 10 times larger than the corresponding minimum for each case.

An important message involved in estimating parameters of these kinds of models, is that the likelihood surface for each of these space–time models may contain many local minima, which is different from the likelihood surface of a purely spatial model using, for example, the Matérn class. Hence, to obtain the ML estimator, it is suggested to start with good several sets of initial parameters and check for goodness-of-fit.

The results are summarised in form of figures (the corresponding tables are available upon request for the interested reader). In particular, Figs. 1–4 show segment-shape plots of  $mean \pm 1.96 * standard\ error$  of MSPE values based on 100 replications corresponding to any of the above-mentioned parameter combinations. In particular, the MSPE performance of the various methods when data are generated from each one of the four considered models is analysed.

We considered AIC and BIC criteria to select among the four classes of models. For each replicate, the predicted surface corresponding to AIC (BIC) is equal to the one of the four predicted surfaces having the smallest AIC (BIC) value. The simulation was done on a collection of 16 Linux PCs (CPU: Athlon 1.6 GHz) using the R language.

The degree of smoothing and the noise level ( $\sigma_\epsilon^2$ ) defining a space–time process play a crucial role in terms of prediction. In particular, the larger (smaller) the degree of smoothing (noise level) is, the more reliable prediction

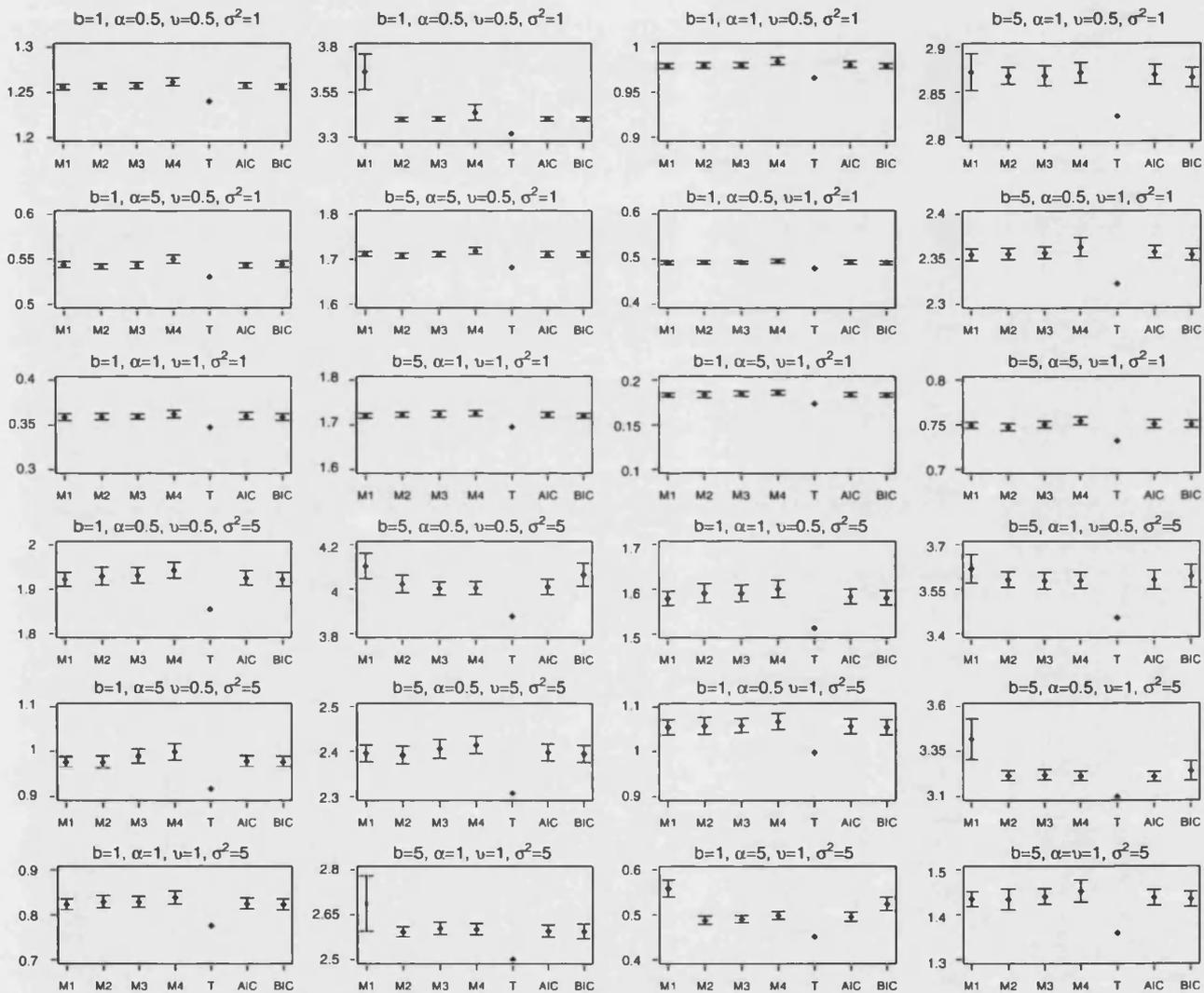


Fig. 2. Segment-shape plots of  $\text{mean} \pm 1.96 * \text{standard error}$  of MSPE values based on 100 replications of data generated from Model II with  $a = 1$  and  $\sigma_y^2 = 5$ . Within each plot, the order of the plots from left to right is: Model I, II, III, IV, True, AIC and BIC.

is. A separable structure, such as that of Model I, can provide good prediction results even for data generated from nonseparable structures, for example such structures of Models II and IV. In fact (see Figs. 2 and 4), and under many different parameter combinations, Model I provides the best prediction results when data is generated by the nonseparable structures defined by Models II and IV. This behaviour is due to the intrinsic flexibility of Model I, in which the spatial covariance follows a Matérn structure, and the temporal covariance is defined through an ARMA process. Indeed, Model I covers a large number of space–time covariance structures. However, Model I tended to perform worse under stronger spatial dependence.

As expected, see previous comments concerning (5), prediction based on the true model performed better than that based on a fitted model. However, the relative complexity of a class of models prevents prediction based on the correct class of models from providing, in all situations, better results than prediction based on a wrong class of models. For example, the parameters of Models II and IV are more difficult to estimate (compared to the other two models), and a direct consequence is that data initially generated by these two models is not better fitted by themselves.

Both AIC and BIC did as well as the true class of models, in fact AIC and BIC are hardly distinguishable. Maybe BIC performed slightly better than AIC. Note that the asymptotic properties of AIC and BIC have not been quite well

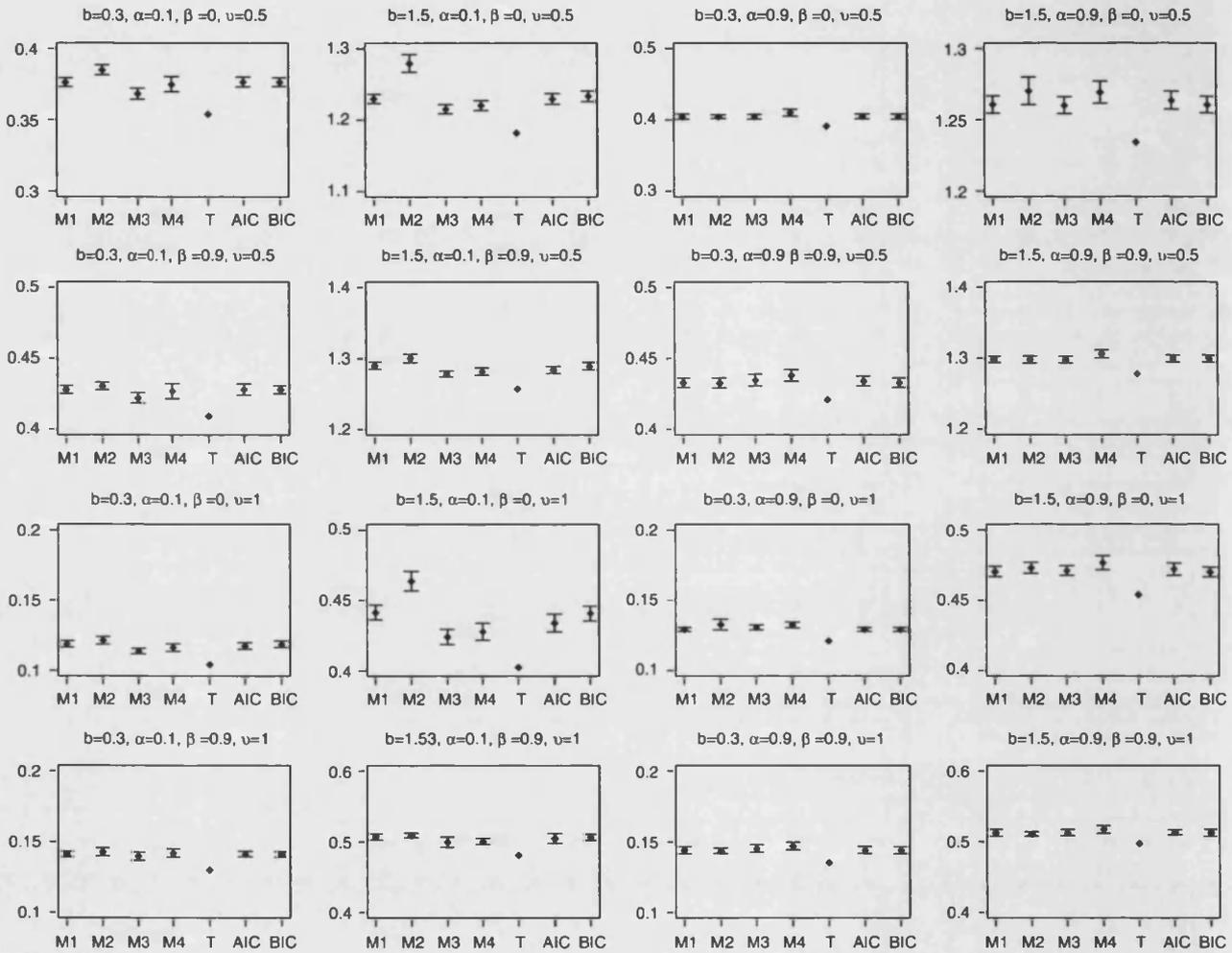


Fig. 3. Segment-shape plots of  $mean \pm 1.96 * standard\ error$  of MSPE values based on 100 replications of data generated from Model III with  $a = 0.5, \delta = 0.5, \sigma_y^2 = 5,$  and  $\sigma_\epsilon^2 = 1$ . Within each plot, the order of the plots from left to right is: Model I, II, III, IV, True, AIC and BIC.

studied for spatial or space–time models, especially when data locations are irregularly located. Some recent results can be found in Hoeting et al. (2006), where, in their conclusions section, it is pointed out the intrinsic difficulty in proving some asymptotic theory of AIC and BIC. If traditional results for regression and time-series hold in this context, then BIC is consistent, but AIC is not (c.f. McQuarrie and Tsai, 1998). This could be one reason why BIC outperforms AIC. However, these asymptotic results have yet to be developed for space–time model selection, in particular, for data that are sampled irregularly in space. Other reasons in favour of BIC are: (a) the sample size considered here is not large enough to satisfy asymptotic conditions; (b) in our setting, one of the four fitted candidate models has the same structure as the true model. Thus, a consistent criterion (BIC) should outperform a criterion which is not consistent (AIC).

It is generally difficult to see clearly the asymptotic properties of AIC and BIC just through a simulation experiment. Basically, AIC (BIC) tends to select a model with larger (smaller) number of parameters by having a smaller penalty. Hence, we see that BIC does better than AIC for Model I, whereas AIC does better than BIC for Model III. In fact, even if BIC is consistent, but AIC is not, we would still see this kind of behaviour unless the sample size is extremely large.

Finally, the underlying spatial process is rarely known in practice. Therefore, it is recommended to fit several different classes of models and use either AIC or BIC for model selection.

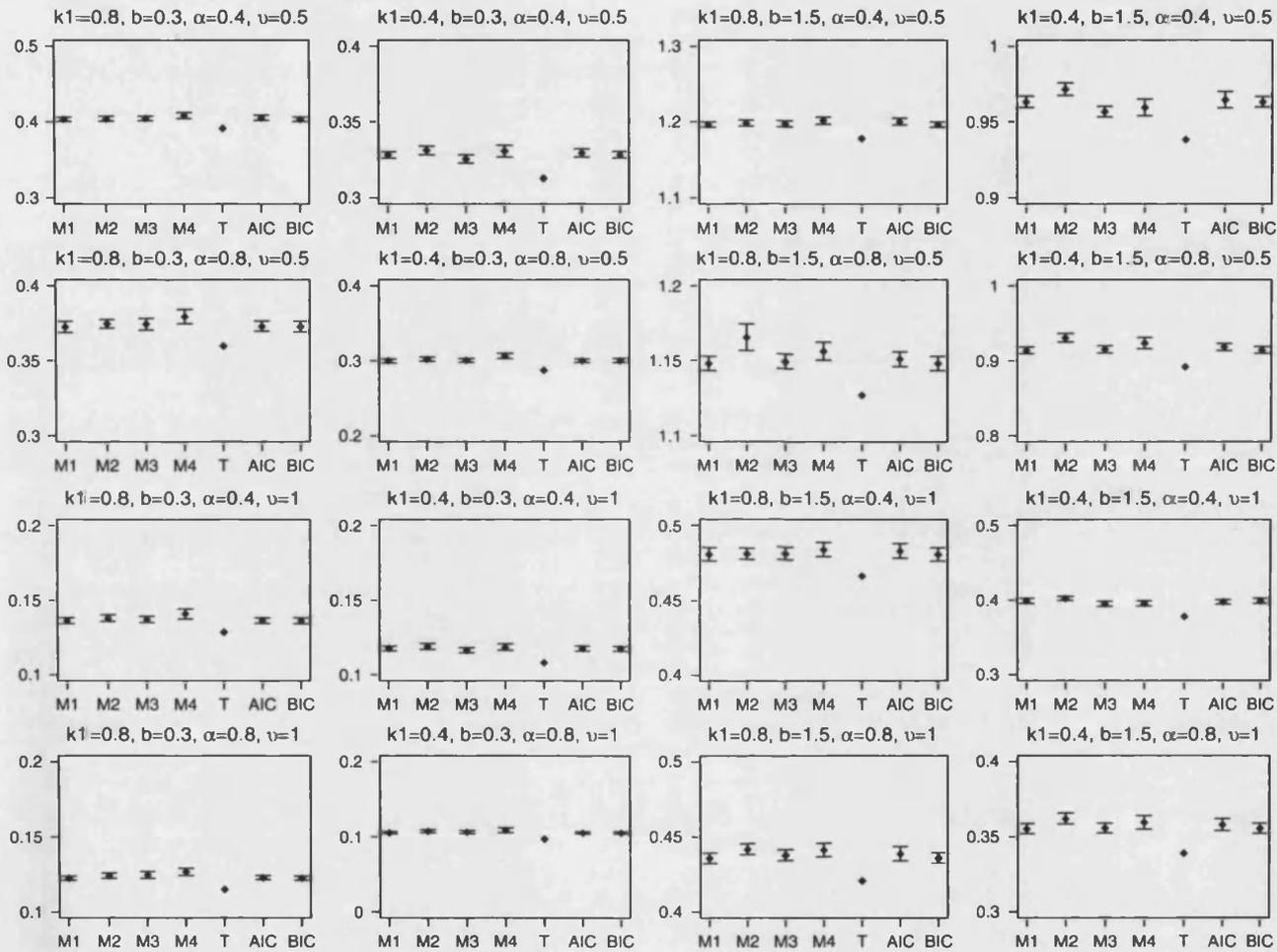


Fig. 4. Segment-shape plots of  $\text{mean} \pm 1.96 * \text{standard error}$  of MSPE values based on 100 replications of data generated from Model IV with  $k_2 = k_3 = (1 - k_1)/2$ ,  $\sigma_Y^2 = 5$ , and  $\sigma_\epsilon^2 = 1$ . Within each plot, the order of the plots from left to right is: Model I, II, III, IV, True, AIC and BIC.

#### 4. Analysis of the surface shortwave radiation budget (SSRB)

##### 4.1. The data

The SSRB is the amount of energy in the solar region of the electromagnetic spectrum ( $0.2\text{--}4.0\ \mu\text{m}$ ) absorbed at the surface. The best way of monitoring these data over extended regions is by means of satellites orbiting the Earth. Due to the fact that satellites provide top of atmosphere (TOA) measurements, the SSRB, measured in  $\text{W}/\text{m}^2$ , has to be modelled from those measurements, jointly with information about the state of the atmosphere and the surface. A global perspective of the SSRB, both in shortwave and in longwave parts of the spectrum has been achieved during the last years thanks to the SSRB Projects (Whitlock et al., 1995; Stackhouse et al., 1999). However, regional, high spatial resolution description of the SSRB is also of interest as indicator of changes and because of its relationship to aridification from well developed vegetation (Bolle, 1999). Besides, as a measurement of the available energy, SSRB is a key parameter for the derivation of the surface energy fluxes.

The SSRB database has been derived from satellite data at the top of the atmosphere and a physical model that relates the shortwave radiation budget at the top of the atmosphere with that at the surface, making use of ancillary information of precipitable water content in the atmospheric column (Li et al., 1993). For the derivation of the SSRB over the selected study area in the Iberian Peninsula (Fig. 5), we used the Mediterranean Extended Daily One km AVHRR Data



Fig. 5. Selected region  $D$ , with latitudes between  $37.60^\circ\text{N}$  and  $40.09^\circ\text{N}$  and longitudes between  $3.92^\circ\text{W}$  and  $0.93^\circ\text{W}$ , where the SSRB data set is analysed.

Set (MEDOKADS) (Koslowsky, 1996). The MEDOKADS database is a high spatial resolution database of 10-day maximum NDVI (normalized difference vegetation index) composites, obtained from the AVHRR (advanced very high resolution radiometer) sensors, onboard NOAA (National Oceanic and Atmosphere Administration) satellites. The area consists of land locations of Castilla-La Mancha and the Comunidad Valenciana running southwest to northeast and northwest to southeast, with latitudes between  $37.60^\circ\text{N}$  and  $40.09^\circ\text{N}$ , and longitudes between  $3.92^\circ\text{W}$  and  $0.93^\circ\text{W}$ . Due to the high spatial resolution, a square of size  $0.01^\circ$ , the process of SSRB can be considered to be based on point support (and so neglecting any subgrid variability), rather than block support and, consequently, the selected area  $D$  configures a rectangular grid of  $300 \times 250$  points. A more detailed description of the derivation of the SSRB dataset can be consulted in Bodas-Salcedo et al. (2003).

With respect to the temporal component, the SSRB data were measured during year 1996. Five measures were taken, each corresponding to a period of 10 days, from April 1 to May 20. Fig. 6 shows the evolution of the stochastic process formed by the SSRB data measured over the 75,000 points ( $300 \times 250$ ) over the five analysed temporal instants.

#### 4.2. Spatio-temporal statistical analysis

Each observation of the SSRB data set is supposed to be a realisation from a stochastic process  $Z(\mathbf{s}, t)$ , which is observed at the spatial locations  $\mathbf{s}$  and at temporal instants,  $t$ , corresponding to the period of 10 days, with  $\mathbf{s} \in D$  and  $t \in \mathbb{Z} = \{10, \dots, 14\}$ . With the aim of using the statistical models described in this paper, we decomposed the process  $Z(\mathbf{s}, t)$ ,  $\mathbf{s} \in D$ ,  $t \in \mathbb{Z}$  into a sum of two components: a trend or large-scale variation  $\mu(\mathbf{s}, t) = \mu$ , representing the mean of the process over the observed region and during the temporal interval considered, and a residual process or small-scale variation  $\delta(\mathbf{s}, t)$ . Thus,

$$Z(\mathbf{s}, t) = \mu + \delta(\mathbf{s}, t).$$

The trend  $\mu$  was estimated as the mean of the observed values, allowing to estimate the residuals by removing them from the observations,  $\hat{\delta}(\mathbf{s}_1, t) = Z(\mathbf{s}_1, t) - \hat{\mu}$ . The resulting stochastic process,  $\hat{\delta}(\cdot, \cdot)$ , defined the target process over which the four spatio-temporal models used in this paper were applied.

The original region  $D$  in Fig. 5 was divided into  $30 \times 25$  subregions of  $10 \times 10$  points each, allowing us to be consistent with the regions used throughout the simulation study. Note that we do not require the stationarity assumption for the SSRB process over the whole region of  $300 \times 250$  points shown in Fig. 6. Since data are eventually analysed locally for each small subregion, it is enough to assume that the SSRB process is stationary over each subregion. In fact, we can even allow different means for different subregions. The SSRB process seems to be spatially nonstationary after looking at Fig. 6. Thus, the data can either be explained by a process with very strong temporal dependence, or perhaps more

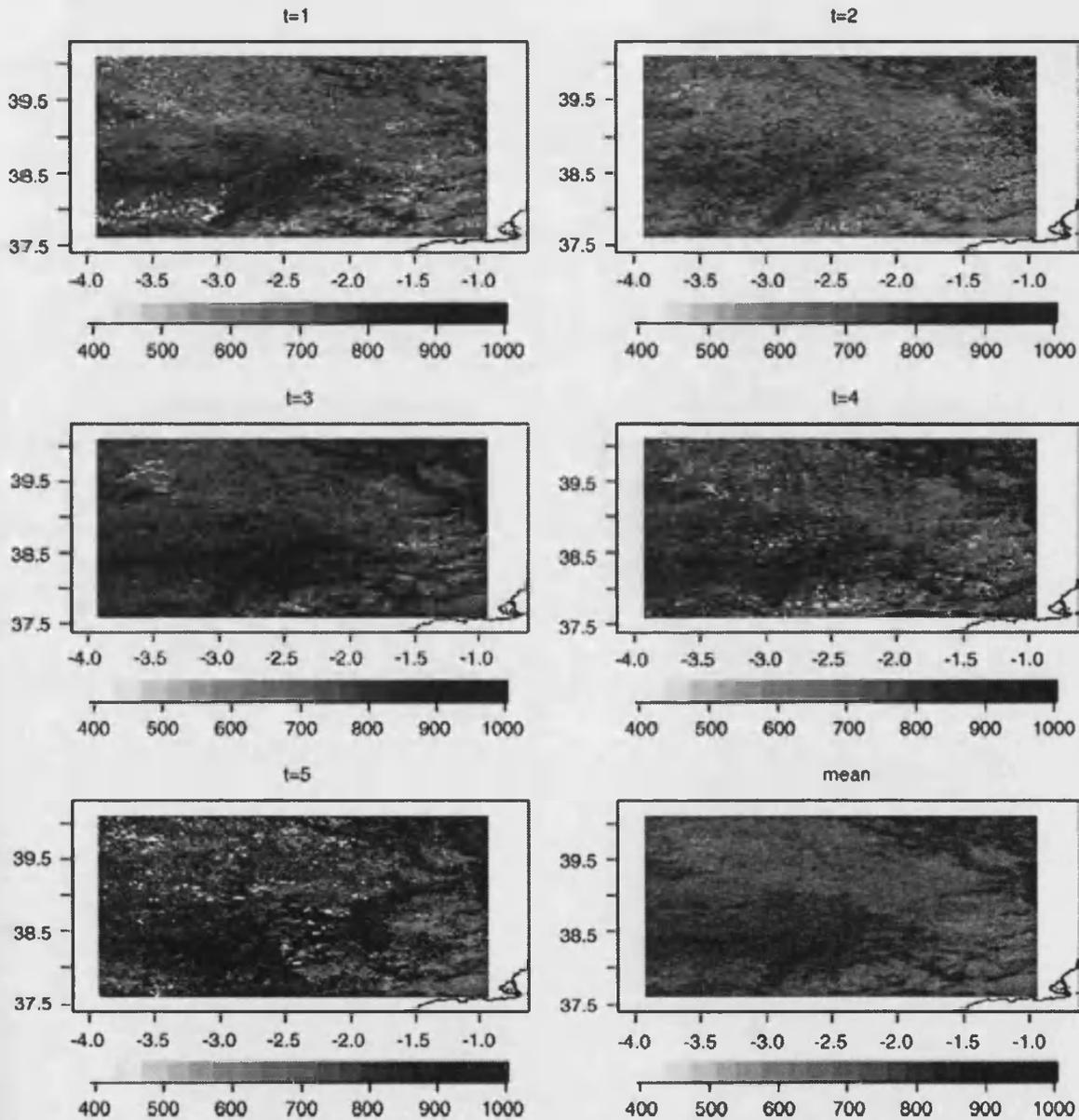


Fig. 6. Evolution of the stochastic process formed by the SSRB data measured over the  $300 \times 250$  points over the five analysed temporal instants and overall mean image.

appropriately, by a process having heterogeneous means over space. After subtracting the mean for each subregion, the process is expected to look less temporally correlated and more spatially stationary.

An exploratory analysis, not reported here, confirmed that the stochastic process  $\hat{\delta}(\cdot, \cdot)$  behaved in a similar way in each of these small subregions. As an example, Fig. 7 shows the evolution of the stochastic process  $Z(s, t)$  formed by the SSRB data measured over one of these  $30 \times 25$  subregions, randomly chosen from the whole region in  $D$ . This selected grid was located between the longitudes  $-3.92$  and  $-3.83$ , and the latitudes  $38.10$  and  $38.19$ . Moreover, Fig. 8 shows the standard deviation against the mean of the process over the five time instants analysed, each corresponding to a period of 10 days, from April 1 to May 20 of 1996.

The four models presented in this paper were used to estimate the SSRB data over each one of the  $30 \times 25$  subregions, above mentioned, at the last time  $T = 5$ . To compute the MSPE values in the data analysis, we note the following.

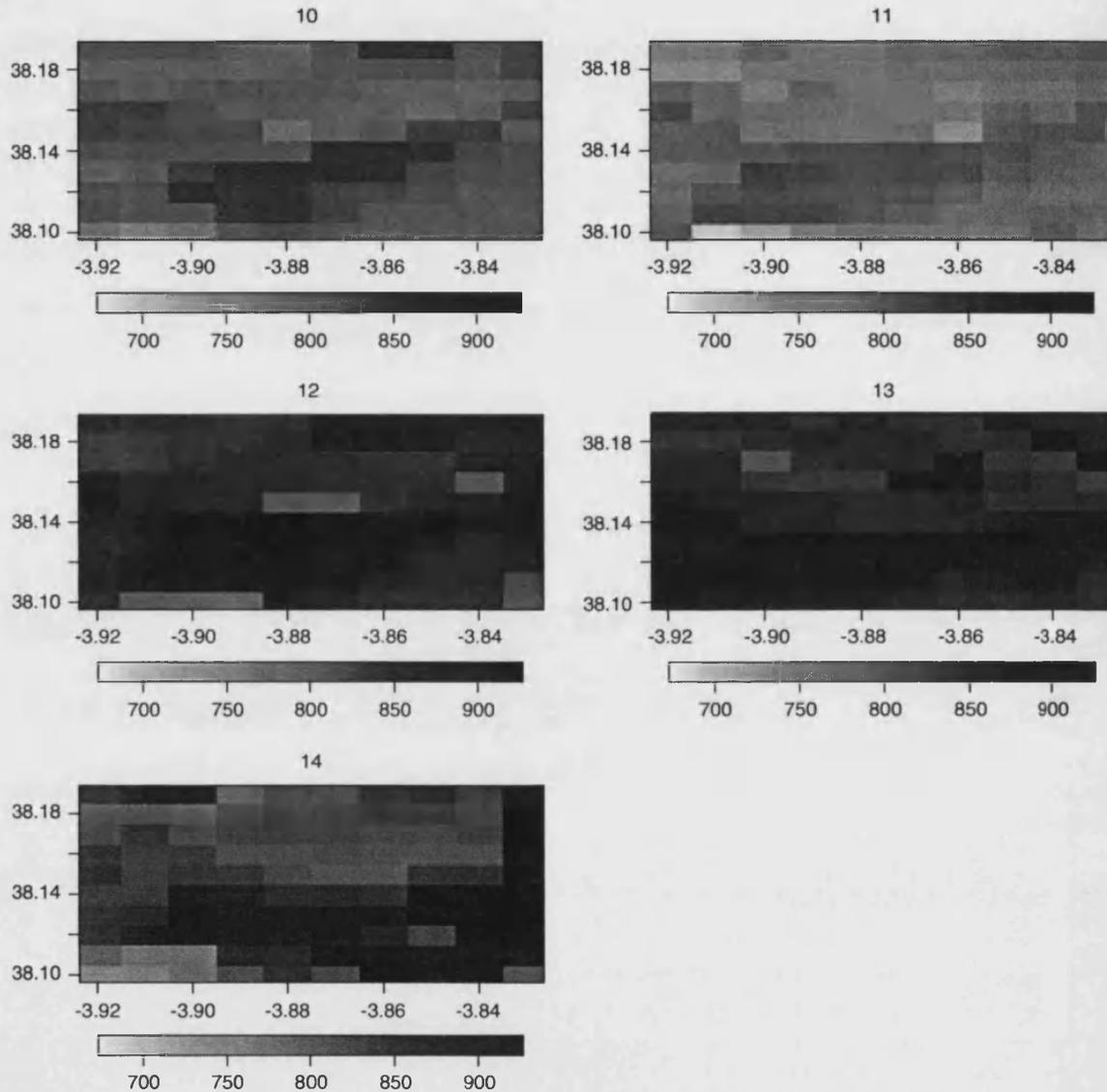


Fig. 7. Evolution of the stochastic process  $Z(s, t)$  formed by the SSRB data measured over one randomly subregion (located between the longitudes  $-3.92$  and  $-3.83$ , and the latitudes  $38.10$  and  $38.19$ ) of  $10 \times 10$  points over the five time instants analysed.

Under the model  $Z(s, t) = \mu + Y(s, t) + \varepsilon(s, t)$ , where  $\mu + Y(s, t)$  is the true SSRB value and  $\varepsilon(s, t)$  is the measurement error, we only have the noisy measurements  $Z(s, t)$  and the true values are unknown. Hence, we cannot compute  $(\hat{Y}(s, t) - Y(s, t))^2$  at the last time  $t = T$ . Thus, we used a widely known validation method by dividing the data into two parts: one part having sample size  $n = 20$  for model fitting, and the other part, with size  $80 = N^2 - n$ , is used for validation (testing). That is, we computed  $\hat{Z}(s, t)$  using the data observed at the  $n = 20$  sites, and computed the average of  $(\hat{Z}(s, t) - Z(s, t))^2$  over the remaining  $80 = N^2 - n$  sites for testing.

A summary, in terms of means and standard errors, of the MSPE results provided by these empirical estimations over 750 repetitions is given in Table 1. This table includes two additional columns for the AIC and BIC models, those showing the lowest values for AIC and BIC in each repetition. Fig. 9 shows segment-shape plots of  $mean \pm 1.96 * standard\ error$  of MSPE values for the six models. The MSPE values for the six models were quite similar, but a certain discrimination amongst the four models in favour of Model II is observed. In Table 1 it is also shown the percentage of cases in which

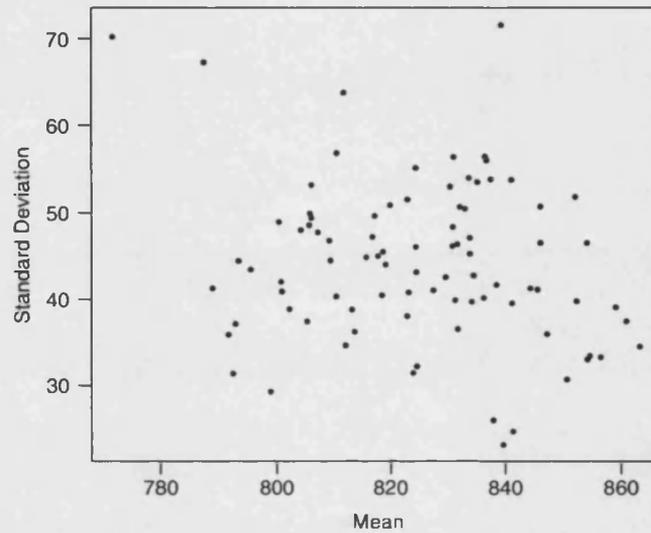


Fig. 8. Variation of standard deviation against the mean of the process over the five time instants analysed.

Table I  
Mean and standard error of MSPE values for the six models when fitted to SSRB data over the 750 repetitions

	Model I	Model II	Model III	Model IV	AIC	BIC
Mean	42.5379	41.5036	42.1175	42.6873	42.4536	42.4483
Std. error	0.5772	0.5532	0.5719	0.5842	0.5782	0.5782
Best	34.40%	42.93%	13.60%	9.07%		

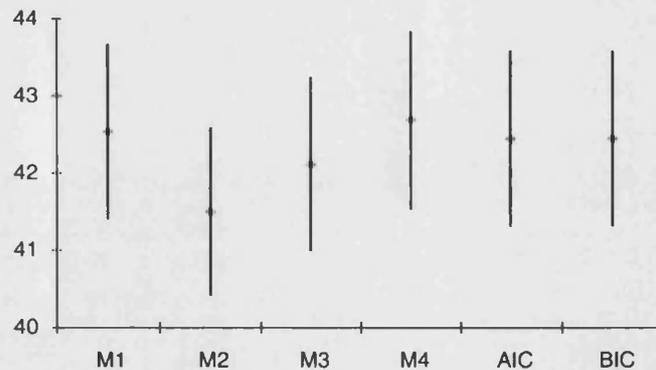


Fig. 9. Segment-shape plot of  $mean \pm 1.96 * standard\ error$  of MSPE values for the estimation of SSRB data at time  $T = 5$ .

each fitted model was the best, i.e. showed the lowest MSPE, amongst the 750 repetitions (subregions). Model II was the best in almost 43% of the occasions. Note that the data analysis, similar to the simulation study, did another model comparison via a validation method and this analysis supported the simulation results. In addition, Fig. 10 shows the best model (that model with a smallest MSPE) in each of the 750 grids of  $10 \times 10$  points, that partitioned the original region. We used a grey-scale code in which the white one represents Model I (M1) and the black one represents Model IV (M4).

Finally, Fig. 11 shows the estimation of SSRB data obtained for the four models over the  $10 \times 10$  points in the randomly selected subregion shown in Fig. 7. For comparison purposes, we also show the real values of the SSRB data for  $T = 5$  at the 20 sampling positions.

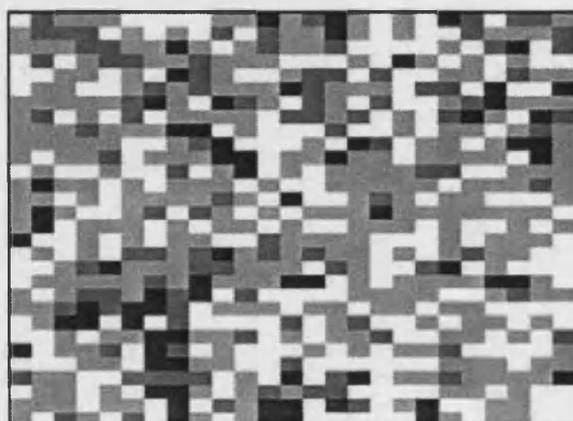


Fig. 10. Best model in each of the 750 grids of  $10 \times 10$  points that partitioned the original region (grey-scale: white = M1 to black = M4).

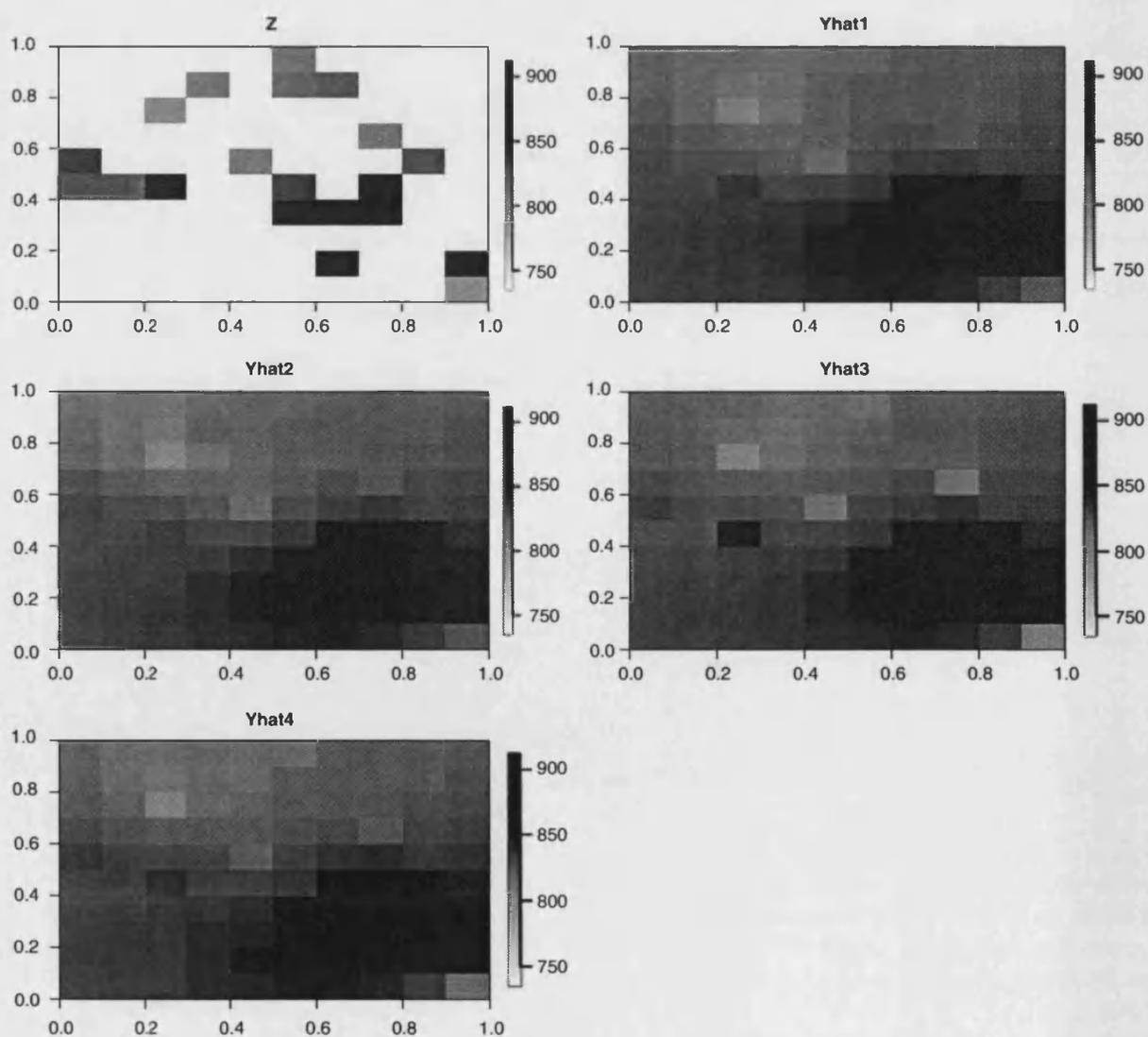


Fig. 11. Real values,  $Z$ , and prediction of SSRB data,  $Yhat1$  (Model I) to  $Yhat4$  (Model IV) in the randomly selected grid of  $10 \times 10$  (see details in Fig. 7).

## 5. Conclusions and discussion

There has been quite a bit of research in recent years on modelling of space–time processes, but very little direct work examining the consequences of using various types of models. We have described a through simulation study comparing four different classes of space–time models in terms of their predictive ability. Model selection is analysed through the use of AIC and BIC values. Thus, we fill a blank in the space–time statistical modelling context.

In practical modelling, we first suggest doing some exploratory data analysis for the data. If the data appear to be spatially and temporally stationary after removing a large-scale trend, the next step would be to check whether the spatio–temporal covariance function is separable. This can be done by simply looking at the empirical spatio–temporal covariances or, more formally, by using a test (e.g. Fuentes, 2003). Choose Model I if separability is appropriate, and consider fitting various spatial and temporal covariance models. Yet, if separability may not be appropriate, consider fitting also nonseparable spatio–temporal models. Finally, use either AIC or BIC to select the final model.

Based on our findings, we highlight the following points:

- (1) The degree of smoothing and the noise level ( $\sigma_\varepsilon^2$ ) defining a space–time process play a crucial role in terms of prediction. In particular, matching a tentative theoretical model with a model for fitting can be less tricky for processes with a larger degree of smoothing and/or with a small noise level.
- (2) A separable structure can provide good prediction results even for data generated from nonseparable structures. This can be simply due to the intrinsic flexibility of the separable model proposed. For example, a Matérn structure for the spatial covariance, and an ARMA structure for the temporal interactions, can well do this job.
- (3) Though overstated, it is important to stress that prediction based on the true model performs better than that based on a fitted model.
- (4) However, the relative complexity of a class of models prevents prediction based on the correct class of models from providing, in all situations, better results than prediction based on a wrong class of models.
- (5) Both AIC and BIC did as well as the true class of models, in fact AIC and BIC are hardly distinguishable. Maybe, BIC performed slightly better than AIC. This is probably due to the fact that the traditional consistent result of BIC in regression and time series model selection may still hold in space–time model selection considered here. However, as mentioned in Hoeting et al. (2006), even asymptotic theory of the MLE for geostatistical models has not been fully developed when data are irregularly sampled in space. Further investigation is needed to understand the asymptotic properties of AIC and BIC, but this is beyond the scope of this paper.
- (6) The underlying spatial process is rarely known in practice. Thus, it is usually advised trying various models after an appropriate exploratory data analysis, and doing some diagnostic checks afterwards or using either AIC or BIC for model selection.
- (7) In terms of computation and the ability to handle large space–time data sets, a separable and flexible model, such as our Model I, would be the choice. However, if nonseparability is guessed, then a flexible, easy-to-compute and interpret nonseparable structure, such as our Model II, should be clearly proposed.
- (8) Model II showed the best behaviour amongst the four models when the solar radiation data was analysed. This model presented the lowest values of both MSPE and corresponding standard errors. However, the separable structure showed the second best results. So, although the SSRB data evolves jointly in space and time, in a nonseparable way, we could also use other sufficiently flexible separable structures.

## Acknowledgements

The Associate Editor and two anonymous referees are acknowledged with thanks. Their precise comments and suggestions have clearly improved an earlier version of the manuscript.

## References

- Akaike, H., 1973. Information theory and the maximum likelihood principle. In: Petrov, V., Csáki, F. (Eds.), International Symposium on Information Theory. Akadémiai Kiadó, Budapest.
- Angulo, J.M., Ruiz-Medina, M.D., Anh, V.V., Grecksch, W., 2000. Fractional diffusion and fractional heat equation. *Adv. Appl. Probab.* 32, 1077–1099.
- Anh, V.V., Leonenko, N.N., 2001. Spectral analysis of fractional kinetic equations with random data. *J. Statist. Phys.* 104, 1349–1387.

- Anh, V.V., Leonenko, N.N., 2002. Renormalization and homogenization of fractional diffusion equations with random data. *Probab. Theory Relat. Fields* 124, 381–408.
- Bodas-Salcedo, A., López-Baeza, E., Martínez, F., Mateu, J., Montes, F., 2003. Spatio-temporal modeling and prediction of solar radiation. *J. Geophys. Res.* 108 (D24).
- Bolle, H.J. (Ed.), 1999. Synthesis of change detection parameters into a land-surface change indicator for long-term desertification studies in the Mediterranean area (RESYSMED). Final Report Contract No. ENV4-CT97-0683, Firenze and Munchen.
- Brown, P.E., Karesen, K.F., Roberts, G.O., Tonellato, S., 2000. Blur-generated nonseparable space-time models. *J. Roy. Statist. Soc., Ser. B* 62, 847–860.
- Christakos, G., 2000. *Modern Spatiotemporal Geostatistics*. Oxford University Press, New York.
- Christakos, G., Hristopulos, D.T., 1998. *Spatiotemporal Environmental Health Modelling: A Tractatus Stochasticus*. Kluwer Academic Press, Dordrecht, The Netherlands.
- Cressie, N., Huang, H.-C., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* 94, 1330–1340.
- Cressie, N., Wikle, C.K., 2002. Space-time Kalman filter. In: El-Shaarawi, A.H., Piegorisch, W.W. (Eds.), *Encyclopedia of Environmetrics*, vol. 4. Wiley, New York, pp. 2045–2049.
- de Cesare, S., Myers, D.E., Posa, D., 2001a. Estimating and modeling space-time correlation structures. *Statist. Probab. Lett.* 51, 9–14.
- de Cesare, L., Myers, D.E., Posa, D., 2001b. Product-sum covariance for space-time modeling: an environmental application. *Environmetrics* 12, 11–23.
- de Iaco, S., Myers, D.E., Posa, D., 2001. Space-time analysis using a general product-sum model. *Statist. Probab. Lett.* 52, 21–28.
- Dimitrakopoulos, R., Lou, X., 1994. Spatiotemporal modeling: covariances and ordinary kriging system. In: Dimitrakopoulos, R. (Ed.), *Geostatistics for the Next Century*. Kluwer Academic Publisher, Dordrecht, pp. 88–93.
- Fernández-Casal, R., 2003. *Geostatística Espacio-Temporal. Modelos Flexibles de Variogramas Anisotropicos no Separables*. Doctoral Thesis, Santiago de Compostela.
- Fuentes, M., 2003. Testing for separability of spatial-temporal covariance functions. Technical Report 2533, Department of Statistics, North Carolina State University.
- Gneiting, T., 2002. Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* 97, 590–600.
- Guttorp, P., Meiring, W., Sampson, P., 1994. A space-time analysis of ground-level ozone data. *Environmetrics* 5, 241–254.
- Heine, V., 1955. Models for two-dimensional stationary stochastic processes. *Biometrika* 42, 170–178.
- Hoeting, J.A., Davis, R.A., Merton, A., Thompson, S.E., 2006. Model selection for geostatistical models. *Ecol. Appl.* 16, 87–98.
- Huang, H.-C., Cressie, N., 1996. Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Comput. Statist. Data Anal.* 22, 159–175.
- Huang, H.-C., Hsu, N.-J., 2004. Modeling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics* 15, 251–268.
- Jones, R.H., Zhang, Y., 1997. Models for continuous stationary space-time processes. In: Gregoire, T.G., Brillinger, D.R., Diggle, P.J., Russek-Cohen, E., Warren, W.G., Wolfinger, R.D. (Eds.), *Modelling Longitudinal and Spatially Correlated Data*. Springer, New York, pp. 289–298.
- Kolovos, A., Christakos, G., Hristopulos, D.T., Serre, M.L., 2004. Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Adv. Water Resour.* 27, 815–830.
- Koslowsky, D., 1996. Mehrjährige validierte und homogenisierte Reihen des Reflexionsgrades und des Vegetationsindex von Landoberflächen aus täglichen AVHRR-Daten hoher Auflösung. FUB, Meteorologische Abhandlungen, Neue Folge, Serie A Monographien, Band 9/Heft 1, Berlin, Germany, 236pp.
- Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space-time models: a review. *Math. Geol.* 31, 651–684.
- Li, Z., Leighton, H.G., Masuda, K., Takashima, T., 1993. Estimation of SW flux absorbed at the surface from TOA reflected flux. *J. Climate* 6, 317–330.
- Ma, C., 2002. Spatio-temporal covariance functions generated by mixtures. *Math. Geol.* 34, 965–975.
- Ma, C., 2003a. Families of spatio-temporal stationary covariance models. *J. Statist. Plann. Inference* 116, 489–501.
- Ma, C., 2003b. Spatio-temporal stationary covariance models. *J. Multivariate Anal.* 86, 97–107.
- Ma, C., 2005a. Linear combinations of space-time covariance functions and variograms. *IEEE Trans. Signal Process.* 53, 857–864.
- Ma, C., 2005b. Semiparametric spatio-temporal covariance models with the ARMA temporal margin. *Ann. Inst. Statist. Math.* 57, 221–233.
- Mardia, K., Goodall, C., Redfern, E., Alonso, F., 1998. The kriged Kalman filter (with discussion). *Test* 7, 217–285.
- Matérn, B., 1960. *Spatial Variation, Meddelanden från Statens Skogsforskningsinstitut*, vol. 49(5). Stockholm, Almaenna Foerlaget. [Second edition (1986), Lecture Notes in Statistics, vol. 36. Springer, Berlin.]
- McQuarrie, A.D., Tsai, C.-L., 1998. *Regression and Time Series Model Selection*. World Scientific Publishing, Singapore.
- Renshaw, E., 1994. The linear spatio-temporal interaction process and its relation to  $1/\omega$ -noise. *J. Royal Statist. Soc. Ser. B* 56, 75–91.
- Rohuani, S., Hall, T.J., 1989. Space-time kriging of groundwater data. In: Armstrong, M. (Ed.), *Geostatistics*, vol. 2. Kluwer Academic Publishers, Dordrecht, pp. 639–651.
- Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of non-stationary spatial covariance structures. *J. Amer. Statist. Assoc.* 87, 108–119.
- Schwarz, G., 1978. Estimating the dimension of a model. *Ann. Statist.* 6, 461–464.
- Stackhouse, P.W., Jr., Cox, S.J., Gupta, S.K., DiPasquale, R.C., Brown, D.E., 1999. The WCRP/GEWEX Surface Radiation Budget Project Release 2: first results at 1 degree resolution. 10th Conference on Atmospheric Radiation: A Symposium with tributes to the works of Verner E. Suomi. American Meteorological Society, Madison, Wisconsin.
- Stein, M., 1999. *Interpolation of Spatial Data. Some Theory of Kriging*. Springer, New York.
- Stein, M., 2005. Space time covariance functions. *J. Amer. Statist. Assoc.* 100, 310–321.
- Stroud, J.R., Muller, P., Sansó, B., 2001. Dynamic models for spatiotemporal data. *J. Roy. Statist. Soc. Ser. B* 63, 673–689.

- Whittle, P., 1954. On stationary processes on the plane. *Biometrika* 41, 434–449.
- Whittle, P., 1956. On the variation of yield variance with plot size. *Biometrika* 43, 337–343.
- Whittle, P., 1962. Topographic correlation, power-law covariance functions, and diffusion. *Biometrika* 49, 305–314.
- Whitlock, C.H., Charlock, T.P., Staylor, W.F., Pinker, R.T., Laszlo, I., Ohmura, A., Gilgen, H., Konzelman, T., DiPasquale, R.C., Moats, C.D., LeCroy, S.R., Ritchey, N.A., 1995. First global WCRP shortwave surface radiation budget data set. *Bull. Amer. Meteorol. Soc.* 76, 905–922.
- Wikle, C.K., Cressie, N., 1999. A dimension reduction approach to space–time Kalman filtering. *Biometrika* 86, 815–829.
- Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* 99, 250–261.

#### 6.4.1. Modelización de la SSRB con el modelo suma de productos generalizado

Con el fin de evaluar la capacidad predictiva del modelo suma de productos generalizado introducido en la Sección 4.12, de entre las 750 rejillas en que se dividió la zona de observación en Huang et al. (2007), seleccionamos al azar 50 de ellas. Cada una de estas rejillas está formada por  $10 \times 10$  localizaciones espaciales en las que se ha observado la radiación solar en superficie (SSRB) durante 5 periodos de 10 días, a lo largo del año 1996. Los detalles puede consultarse en el texto anteriormente referido, que se reproduce al inicio de esta Sección.

	Modelo I	Modelo II	Modelo III	Modelo IV	Modelo V
Media	46,7270	44,7236	44,7088	44,7039	44,5630
Error estándar	2,3265	2,1237	2,1003	2,2463	2,1604
Mejor	8 %	20 %	26 %	32 %	14 %

Tabla 6.1: Media y error estándar del EPCM para los cinco modelos ajustados a los datos de SSRB sobre las 50 rejillas

La Tabla 6.1 es equivalente a la Tabla 1 de Huang et al. (2007), reproducida en las páginas precedentes. En ella se muestran los valores medios de los errores cuadráticos de las predicciones obtenidos sobre las  $n = 50$  rejillas consideradas. Con el fin de evaluar su variabilidad, se incluye también el error estándar asociado a cada una de estas medias, así como el porcentaje de casos en los que cada uno de los cinco modelos aquí considerados es mejor que el resto. En esta tabla podemos observar cómo el nuevo modelo, suma de productos generalizado, mejora los resultados obtenidos con los cuatro utilizados en el artículo citado. En cualquier caso, hay que señalar que los resultados son muy similares para todos los modelos, excepción hecha del modelo I, basado en la predicción mediante el filtro de Kalman. La explicación podría estar en que la brevedad del periodo temporal considerado, sólo 5 instantes de tiempo.

## Spatiotemporal modeling and prediction of solar radiation

A. Bodas-Salcedo,<sup>1</sup> E. López-Baeza,<sup>1</sup> F. Martínez,<sup>2</sup> J. Mateu,<sup>3</sup> and F. Montes<sup>2</sup>

Received 24 June 2003; revised 19 August 2003; accepted 15 September 2003; published 25 November 2003.

[1] The radiation budget in the Earth-atmosphere system is what drives Earth's climate, and thus measurements of this balance are needed to improve our knowledge of Earth's climate and climate change. In the present paper we focus on the analysis of the surface shortwave radiation budget (SSRB), which is the amount of energy in the solar region of the electromagnetic spectrum (0.2–4.0  $\mu\text{m}$ ) absorbed at the surface. The SSRB has to be modeled from the surface to the top of the atmosphere, jointly with information about the state of the atmosphere and the surface. These data come from satellites orbiting the Earth and are often missing or disturbed. Its interest is not only at global scales; rather, regional, high-spatial-resolution description is also of interest as an indicator of changes and because of its relationship to aridification from well-developed vegetation. The goal of this paper is to estimate and predict the spatiotemporal evolution of SSRB data at a regional scale in eastern Spain. Two different spatiotemporal models with covariates are considered: one is based on modeling the spatiotemporal semivariogram and the other uses the Kalman filter technique for spatiotemporal prediction. We present comparisons between these two models with respect to the simpler, purely spatial model. The results show that there is not a great benefit to use the more complicated models, although there is a marginal improvement with complexity. **INDEX TERMS:** 3210 Mathematical Geophysics: Modeling; 3220 Mathematical Geophysics: Nonlinear dynamics; 3230 Mathematical Geophysics: Numerical solutions; 3299 Mathematical Geophysics: General or miscellaneous; **KEYWORDS:** cross validation, Kalman filter, kriging, spatiotemporal model, SSRB data

**Citation:** Bodas-Salcedo, A., E. López-Baeza, F. Martínez, J. Mateu, and F. Montes, Spatiotemporal modeling and prediction of solar radiation, *J. Geophys. Res.*, 108(D24), 8777, doi:10.1029/2003JD003903, 2003.

### 1. Introduction

[2] The need to understand and predict the processes that influence the Earth's atmosphere is one of the grand scientific challenges of this century. Owing to the complexity of observational data and the variability of geophysical phenomena at many spatial and temporal scales, statistical science has an important role to play in this effort.

[3] There exists substantial understanding of the fundamental physics controlling our atmosphere. The Earth's atmosphere and oceans are fluids. In principle, the laws of physics should provide a prescription for the evolution of these fluids, but numerical and computational problems are so involved that important physical processes cannot be resolved by a numerical model. We are facing complicated sources of variation; combining these with the increasing size of observational data sets, we believe that the problems of weather and climate are best stated and answered in the language of statistics. Indeed, one might claim that the

Earth's climate system is a stochastic process; hence we should abandon the strong use of physics in favor of statistical prediction models. Our view is that ultimate success will involve a combination of viewpoints: a synthesis between statistics and physics.

[4] The simplest way to handle temporal replications of spatial data is to assume that they are temporally independent and that the spatial structure does not change through time. While this makes it possible to better estimate the spatial correlation structure, it is not generally efficient since it ignores the temporally dynamic aspect of these processes and neglects temporal correlation that can greatly aid in prediction. Also, virtually all atmospheric and oceanographic processes involve variability over space and time. Furthermore, observations of geophysical processes typically include measurement errors and are often temporally and spatially incomplete. Usual geostatistical approaches either treat time as another spatial dimension or they treat the temporal observations as separate variables. Neither approach is entirely satisfactory. Both approaches ignore the temporally dynamic nature of space-time processes. More generally, specification of realistic space-time covariance models is difficult. The standard models are separable in the sense that the space-time covariance function is either a product or sum of the spatial and temporal covariance functions. A recent innovation is due to *Cressie and Huang* [1999], who propose a flexible class of space-time covariance models.

<sup>1</sup>Departament de Termodinàmica, Universitat de València, Burjassot, Spain.

<sup>2</sup>Departament d'Estadística i Investigació Operativa, Universitat de València, Burjassot, Spain.

<sup>3</sup>Departament de Matemàtiques, Universitat Jaume I, Castelló, Spain.

[5] There is a rich and growing literature on space-time modeling. Fundamentally, it is clear that in the absence of a temporal component, second-order geostatistical models can be used to represent spatial variability. These are descriptive in the sense that although they model spatial correlation, there is no causative interpretation associated with them. Thus for space-time modeling the geostatistical paradigm assumes a descriptive structure for both space and time. For example, one can extend the geostatistical kriging methodology for spatial processes by assuming that time is just another spatial dimension. Alternatively, one can treat time slices of a spatial field as variables and apply a multivariate or cokriging approach [Kyriakidis and Journel, 1999]. Typically, simplifications such as separability between space and time are necessary to implement these approaches, although this does not cover some realistic covariance structures to accurately capture the complicated dynamical processes, as found in geophysical applications.

[6] If both temporal and spatial components are present, it is natural to combine the temporally dynamic state-space approach and the spatially descriptive approach. Such a model is referred to as a space-time dynamic model. Then, the state process is assumed to contain two space-time components, a spatial process evolving with time in a Markovian fashion and a nondynamic space-time component to account for spatial structure that does not evolve with time [Guttorp *et al.*, 1994; Goodall and Mardia, 1994; Wikle *et al.*, 1998].

[7] The intrinsic motivation of this paper is to analyze the evolution of the surface shortwave radiation budget in both space and time and to relate this evolution to possible factors of aridification. Moreover, the developed models will be useful to predict any input missing data, a particular common fact of solar radiation, with readings coming from satellites which only provide top of atmosphere (TOA) measurements.

[8] The plan of the paper is the following. Section 2 presents the data set. Section 3 is focused on the spatiotemporal methodology. We used two different strategies based on modeling the spatiotemporal dependencies by separable covariance structures [Cressie and Majure, 1997] and prediction by means of a Kalman filter procedure [Huang and Cressie, 1996]. Both methodologies can use information from explanatory variables when available. Section 4 describes the analysis of the surface shortwave radiation budget (SSRB), showing comparative results from the output of the spatiotemporal models used and, in turn, with the purely spatial model. Section 5 is devoted to conclusions and further developments.

## 2. The Data Set: Surface Shortwave Radiation Budget (SSRB)

[9] The SSRB is the amount of energy in the solar region of the electromagnetic spectrum (0.2–4.0  $\mu\text{m}$ ) absorbed at the surface. The best way of monitoring this data over extended regions is by means of satellites orbiting the Earth. Owing to the fact that satellites provide TOA measurements, the SSRB, measured in  $\text{W m}^{-2}$ , has to be modeled from those measurements jointly with information about the state of the atmosphere and the surface. A global perspective of the surface radiation budget (SRB), both in short-

wave and in longwave parts of the spectrum, has been achieved during thanks to the SRB Projects [Whitlock *et al.*, 1995; Stackhouse *et al.*, 1999]. However, a regional, high-spatial-resolution description of the SRB is also of interest as an indicator of changes and because of its relationship to aridification from well-developed vegetation [Bolte, 1999]. Besides, as a measurement of the available energy, SRB is a key parameter for the derivation of the surface energy fluxes.

[10] For the derivation of the SSRB over the selected study area in the Iberian Peninsula, the Mediterranean Extended Daily 1 km Advanced Very High Resolution Radiometer (AVHRR) Data Set (MEDOKADS) [Koslowsky, 1996] and the algorithm of Li *et al.* [1993a] have been used. The model proposed by Li *et al.* [1993a] is a linear relationship between TOA-reflected flux and SSRB, with the offset and the slope parameterized as functions of the solar zenith angle, precipitable water content, and cloud type. This allows one to obtain the SSRB under clear or overcast conditions. Model results have been validated extensively by means of comparisons with surface measurements [Li *et al.*, 1993b; Hollmann and Bodas, 2001].

[11] The MEDOKADS database [Koslowsky, 1996] is a high-spatial-resolution database of 10-day maximum Normalized Difference Vegetation Index composites, obtained from the AVHRR sensors, onboard National Oceanic and Atmosphere Administration (NOAA) satellites. Its spatial resolution is  $1 \times 1 \text{ km}^2$ , which means that each pixel represents an average over a squared grid of area  $1 \text{ km}^2$ , centered on the considered pixel. Owing to the high spatial resolution, the process of SSRB considered is based on point support rather than on block support. Note that solar radiation provides very smooth (sometimes even no) change in a square region of  $1 \times 1 \text{ km}^2$ . The data are interactively navigated, being estimated with a position error of 2 km.

[12] Several problems have to be overcome in order to work with multiannual databases. In principle, NOAA satellites are Sun-synchronous satellites. However, their orbits are not exactly Sun synchronous due to the perturbing force of the Sun and the orbit precesses [Price, 1991]. This causes a progressive delay in the equator (and the rest of the latitudes) crossing time, which is negligible for short-term periods but of great impact for multiannual studies. Therefore, in order to make the multiannual comparisons reliable, the database has to be normalized. Thus the SSRB has been computed by means of the Li *et al.* [1993a] model. Only clear-sky pixels have been processed. This causes the variability observed in SSRB to be mainly driven by changes in surface properties and latitudinal effects, the latter being small due to the small range of latitudes of the study area. The Li *et al.* [1993a] model needs several inputs, all of them provided by the derived database.

## 3. Statistical Models for the Spatiotemporal Evolution

### 3.1. Introduction

[13] If both temporal and spatial components are present, we combine both approaches within a statistical space-time dynamic model, i.e., a model which is temporally dynamic and spatially descriptive. Furthermore, with the addition of a measurement error component, the space-time dynamic

processes can be described through a state-space formulation. It is then natural to consider prediction via a Kalman filter [Kalman, 1960]. Although Kalman filter approaches to space-time prediction have been considered by the scientific community for nearly 2 decades [Ghil et al., 1981], the statistics community only recently considered space-time processes from this perspective.

[14] In this section we propose and present two competing methodologies for the analysis of a spatiotemporal process. We first model the large-scale variation of the stochastic process, following Cressie and Majure [1997], and then, assuming the data and the processes we are dealing with have been suitably detrended, we model the small-scale variation. The first model refers to prediction of the small-scale spatiotemporal variation using ordinary kriging, and the second one uses a Kalman filter for prediction. A particular case of the first model, when analyzing only the spatial evolution, is also considered.

[15] Notice that for ordinary kriging, only data obtained in the current week is used to predict the unknown SSRB data at an unsampled location. When using a spatiotemporal model and a Kalman filter, we incorporate past data as well as current data into the prediction of SSRB. Huang and Cressie [1996] were perhaps the first to successfully demonstrate a full implementation of a space-time dynamic model with a Kalman filter, using a separable space-time structure.

3.2. Setup

[16] Let  $Z_t(\mathbf{s})$  denote a stochastic process (for example, the surface solar radiation) at location  $\mathbf{s}$  and time  $t$ . We suppose that we have the data  $\{Z_t(\mathbf{s}), \mathbf{s} \in D, t \in T\}$ , and we wish to predict the original process value  $Z_{t_0}(\mathbf{s}_0)$  at the spatial location  $\mathbf{s}_0 \in D$  (sampled or not sampled) for the time point  $t_0$ . We assume that the data  $Z$  are observed without measurement error.

[17] A very general class of spatiotemporal models assumes that, possibly after some initial transformation, the observed process  $Z_t(\mathbf{s})$  is a Gaussian process with mean  $E(Z_t(\mathbf{s})) = \mu_t(\mathbf{s})$  and positive definite covariance function  $\text{Cov}\{Z_{t_1}(\mathbf{s}_1), Z_{t_2}(\mathbf{s}_2)\} = C(\mathbf{s}_1, t_1, \mathbf{s}_2, t_2)$ . Evidently, it will simplify the analysis if we can work with a more restricted class of spatiotemporal covariances, as will be commented on in section 4.3.3.

[18] Commonly, in environmental processes, stationarity is not an appropriate assumption to impose on the random field  $Z_t(\mathbf{s})$  that generated the data. Following Cressie and Majure [1997], we further assume that we can decompose  $Z_t(\mathbf{s})$  into a deterministic large-scale variation (trend) plus a stochastic small-scale variation (error). The model can thus be written as

$$Z_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \delta_t(\mathbf{s}), \tag{1}$$

where  $E(Z_t(\mathbf{s})) = \mu_t(\mathbf{s})$  and  $\delta_t(\cdot)$  is a zero-mean intrinsically stationary stochastic Gaussian process with variogram  $2\gamma_u(\mathbf{h}) = \text{Var}(Z_{t+u}(\mathbf{s} + \mathbf{h}) - Z_t(\mathbf{s})) = \text{Var}(\delta_{t+u}(\mathbf{s} + \mathbf{h}) - \delta_t(\mathbf{s}))$  that characterizes the spatial and temporal dependence. The large-scale variation  $\mu_t(\mathbf{s})$  and the small-scale variation  $\delta_t(\cdot)$  are modeled as deterministic and stochastic processes, respectively, but there is no way of making the decomposition identifiable

[19] The mean of this process, represented by  $\mu_t(\mathbf{s})$ , is expressed as a linear function of the  $k$  regressors  $\mathbf{x}_t(\mathbf{s}) = (x_{t,1}(\mathbf{s}) \dots x_{t,k}(\mathbf{s}))$  as  $\mu_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})'\beta$ , where  $(x_{t,1}(\mathbf{s}) \dots x_{t,k}(\mathbf{s}))$  are the values of explanatory variables at location  $\mathbf{s}$  and time  $t$ . The use of a linear model here is a means to an end, the end being handy spatiotemporal prediction of SSRB data. More mechanistic, deterministic models are possible, but they do not allow quantification of the predictor's variability.

[20] The coefficients  $\beta$  of this equation are fitted using an appropriate method (ordinary or generalized least squares), and the large-scale process could be estimated by

$$\hat{\mu}_t(\mathbf{s}) = \mathbf{x}_t(\mathbf{s})'\hat{\beta}. \tag{2}$$

[21] The small-scale variation term is estimated by the residuals using the general expression residual = data - fit. In particular, from equations (1) and (2) we can estimate the error process  $\delta_t(\mathbf{s})$  with

$$\hat{\delta}_t(\mathbf{s}) = Z_t(\mathbf{s}) - \mathbf{x}_t(\mathbf{s})'\hat{\beta}. \tag{3}$$

[22] These residuals are treated as "data" and are analyzed to characterize spatial and temporal dependence. This dependence, once characterized, is used to make predictions in space and time with known confidence.

3.3. Spatiotemporal Model and Objective Analysis

[23] The process  $\hat{\delta}_t(\mathbf{s})$  will be considered a proxy for  $\delta_t(\cdot)$ . The spatiotemporal dependence can be characterized through the variogram, which is assumed to be a function only of the space-time lags. This intrinsic stationarity assumption has to be properly checked. Moreover, the above assumption allows the variogram  $2\gamma_u(\mathbf{h})$  to be estimated using, for example, the classical estimator of Cressie [1993]. Notice that for an isotropic variogram,  $2\gamma_u(\mathbf{h}) = 2\gamma_u(\|\mathbf{h}\|)$ .

[24] Once the semivariogram of the residuals is fitted, we can estimate the original process value  $Z_{t_0}(\mathbf{s}_0)$  with the large-scale process estimated value  $\hat{\mu}_{t_0}(\mathbf{s}_0)$  (equation (2)) plus the error process estimated value  $\hat{\delta}_{t_0}(\mathbf{s}_0)$  (equation (3)), obtaining an original process value prediction  $\hat{Z}_{t_0}(\mathbf{s}_0)$  and its mean-square prediction error.

[25] For the spatiotemporal prediction of  $\hat{\delta}_{t_0}(\mathbf{s}_0)$  from the  $n$  data  $\{\hat{\delta}_t(\mathbf{s}_i), \mathbf{s}_i \in D, i = 1, \dots, n, t \in T\}$ , derived from equation (3), we can use the objective analysis or the ordinary kriging approach [Cressie, 1993]. Then, we add back the residual predicted value to the estimated mean process (large-scale variation) to give a final predicted value and its associated root mean squared prediction error.

[26] This method uses a linear predictor  $(p_{\delta})_{t_0}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i \hat{\delta}_t(\mathbf{s}_i) = \lambda'\hat{\delta}$ , with  $\lambda'1 = \sum_{i=1}^n \lambda_i = 1$  to guarantee uniform unbiasedness. The  $\lambda'$  values are chosen as those that minimize the mean squared prediction error subject to this constraint.

[27] Finally, the prediction for the original process value is

$$(p_z)_{t_0}(\mathbf{s}_0) = \hat{Z}_{t_0}(\mathbf{s}_0) = \mathbf{x}_t(\mathbf{s})'\hat{\beta} + (p_{\delta})_{t_0}(\mathbf{s}_0), \tag{4}$$

and its standard error (ignoring the effect of the  $O(1/n)$  error from the  $\beta$ ) equals the kriging variance, i.e.,  $(\sigma_z)_{t_0}(\mathbf{s}_0) = (\sigma_{ok})_{t_0}(\mathbf{s}_0)$ . This procedure can be easily adapted to the

objective analysis for the purely spatial case by removing the temporal index.

[28] In computing the kriging variance we assume that the variogram is known, although, as is typically the case in geostatistical studies, it is actually estimated from the data. The effect of this is that sometimes the estimated prediction error is an underestimate of the kriging standard error [see *Cressie and Zimmerman*, 1992]. The above predictor is optimal assuming we know the covariance or variogram parameters. In our case, these parameters have to be estimated, and by substituting these estimates in place of the unknown parameters, we no longer obtain exact estimates. However, despite these drawbacks and as commented by *Cressie and Majure* [1997] and *Wikle and Cressie* [1999], the above estimator can be considered optimal for the original process.

[29] An approximate 95% prediction interval for  $\hat{Z}_{t_0}(\mathbf{s}_0)$  is  $(p_z)_{t_0}(\mathbf{s}_0) \pm 2(\sigma_z)_{t_0}(\mathbf{s}_0)$ . The prediction interval is approximate because  $(\sigma_z)_{t_0}(\mathbf{s}_0)$  can be an underestimate of the true mean squared prediction error and because the true error distributions are slightly heavier-tailed than Gaussian, although the choice of the constant 2, rather than 1.96, helps ameliorate this problem.

[30] An intrinsic problem associated with the above procedure is that as we are estimating the mean and covariance separately, we are somehow introducing some nonnegligible bias. A possible alternative approach would be estimating jointly the mean and covariance using universal kriging or a restrictive maximum likelihood approach. Here there is a trade-off between computational efficiency and statistical precision. However, our approach, based on the decomposition in the two processes, is motivated by the need to model large spatiotemporal data sets. Hence we focus on a computationally efficient, easy-to-use, and easy-to-understand procedure.

**3.4. Kalman Filter Procedure**

[31] In this section we adapt to our particular case the Kalman filter methodology proposed by *Huang and Cressie* [1996]. Let us consider again the spatiotemporal model equation (1). We suppose that the observable error process defining the small-scale variation has a component of measurement error expressed through the measurement equation

$$\delta_t(\mathbf{s}) = S_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \tag{5}$$

where  $\varepsilon_t(\mathbf{s}) \sim N(0, \sigma_\varepsilon^2 I)$  is a white noise process independent of  $S_t(\mathbf{s})$  and the unobserved state processes  $\{S_t(\mathbf{s}) : \mathbf{s} \in D, t \in T\}$  are assumed to be mean zero, spatially and temporally second-order stationary Gaussian processes. Suppose they evolve in a Markov manner. This has proved to be a very effective assumption for modeling atmospheric processes [*Guttorp et al.*, 1994] and, in particular, solar radiation [*Stackhouse et al.*, 1999]. Linearizing this temporal dependence leads to a  $p$ th-order autoregressive process, AR( $p$ ),

$$S_t(\mathbf{s}) = \alpha_1 S_{t-1}(\mathbf{s}) + \dots + \alpha_p S_{t-p}(\mathbf{s}) + \eta_t(\mathbf{s}), \tag{6}$$

where  $(\alpha_1, \dots, \alpha_p)'$  are chosen to achieve temporal stationarity and  $\{\eta_t(\mathbf{s}) : \mathbf{s} \in D, t \in T\}$  is a zero mean,

spatially and temporally stationary Gaussian process such that  $\eta_{t_1}(\cdot)$  and  $\eta_{t_2}(\cdot)$  are independent for all  $t_1 \neq t_2$ ,  $\eta_t(\cdot)$  and  $S_{t-1}(\cdot)$  are independent for all  $t$ ,  $\varepsilon_{t_1}$  and  $\eta_{t_2}(\cdot)$  are independent for all  $t_1 \neq t_2$ , and  $\alpha_1, \dots, \alpha_p$  are constants.

[32] The Kalman filter is a recursive algorithm for inference about the state variables of a model written in a state space form such as equations (1), (5), and (6). Since it can be shown that  $\hat{S}_{t|t}(\mathbf{s}) = E[S_t(\mathbf{s}) | \delta_1, \dots, \delta_t]$  is a linear function of  $\delta_t$ , with  $\delta_t = (\delta_t(\mathbf{s}_1), \dots, \delta_t(\mathbf{s}_n))$  and  $\hat{S}_{t-1|t-1}(\mathbf{s})$  for all  $t$ , the Kalman filter enables the state variable to be updated once new observations are available. In the derivation below we compute the optimal spatiotemporal predictor of  $\hat{S}_t(\mathbf{s}_0)$ , not only for sites where observations are taken but also for locations where no observations are available.

[33] The following notation will be established for the following procedure development:

$$\hat{\delta}_{t|t}(\mathbf{s}) = E[\delta_t(\mathbf{s}) | \delta_1, \dots, \delta_t];$$

$$\hat{\delta}_{t|t} = E[\delta_t | \delta_1, \dots, \delta_t];$$

$$\Sigma_{t_1, t_2|t}(\mathbf{s}, \mathbf{r}) = \text{cov}[S_{t_1}(\mathbf{s}), S_{t_2}(\mathbf{r}) | \delta_1, \dots, \delta_t];$$

$$\Sigma_{r|t}(\mathbf{s}, \mathbf{r}) = \text{cov}[S_r(\mathbf{s}), S_r(\mathbf{r}) | \delta_1, \dots, \delta_t];$$

$$\Sigma_{r|t}(\mathbf{s}, \mathbf{r}) = \text{cov}[(S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}))'$$

$$\cdot (S_{t-p+1}(\mathbf{r}), \dots, S_t(\mathbf{r}))' | \delta_1, \dots, \delta_t];$$

$$\mathbf{F}_t = \text{var}[\delta_t | \delta_1, \dots, \delta_{t-1}];$$

$$\mathbf{G}_t(\mathbf{s}) = \text{cov}[\delta_t, (S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}))' | \delta_1, \dots, \delta_{t-1}];$$

$$\alpha = (\alpha_p, \alpha_{p-1}, \dots, \alpha_1)';$$

$$\mathbf{A} = (e_2', \dots, e_p', \alpha')',$$

where, for example,  $e_2' = (0, 1, 0, \dots, 0)$ .

[34] All joint and conditional distributions are multivariate Gaussian. Hence we have

$$\begin{bmatrix} \hat{S}_{t-p+1|t}(\mathbf{s}) \\ \vdots \\ \hat{S}_{t|t}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \hat{S}_{t-p+1|t-1}(\mathbf{s}) \\ \vdots \\ \hat{S}_{t|t-1}(\mathbf{s}) \end{bmatrix} + (\mathbf{G}_t(\mathbf{s}))' \mathbf{F}_t^{-1} (\delta_t - \hat{\delta}_{t|t-1}) \tag{7}$$

and

$$\Sigma_{t|t}(\mathbf{s}, \mathbf{r}) = \Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) - (\mathbf{G}_t(\mathbf{s}))' \mathbf{F}_t^{-1} \mathbf{G}_t(\mathbf{r}). \tag{8}$$

Note that

$$\begin{aligned} \hat{\delta}_{t|t-1}(\mathbf{s}) &= E[\delta_t(\mathbf{s}) | \delta_1, \dots, \delta_{t-1}] \\ &= E[S_t(\mathbf{s}) | \delta_1, \dots, \delta_{t-1}] \\ &= \alpha' (\hat{S}_{t-p|t-1}(\mathbf{s}), \dots, \hat{S}_{t-1|t-1}(\mathbf{s}))' \end{aligned} \tag{9}$$

and

$$\Sigma_{t|t-1}(\mathbf{s}, \mathbf{r}) = \mathbf{A} \Sigma_{t-1|t-1}(\mathbf{s}, \mathbf{r}) \mathbf{A}' + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^n(\mathbf{s}, \mathbf{r}) \end{pmatrix}. \tag{10}$$

On the other hand, we have the following expressions for matrices  $F_t$  and  $G_t(\cdot)$ :

$$F_t = \text{var}[\delta_t | \delta_1, \dots, \delta_{t-1}] = \text{var}[S_t | \delta_1, \dots, \delta_{t-1}] + \sigma_e^2 I \quad (11)$$

and

$$G_t(\mathbf{s}) = \text{cov}[\delta_t, (S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}))' | \delta_1, \dots, \delta_{t-1}] \\ = \text{cov}[S_t, (S_{t-p+1}(\mathbf{s}), \dots, S_t(\mathbf{s}))' | \delta_1, \dots, \delta_{t-1}]. \quad (12)$$

[35] Now we consider parameter estimation for the spatiotemporal model given by equations (5) and (6). Let the spatially stationary error process be described by the flexible three-parameter exponential covariance function

$$C^n(\mathbf{s}, \mathbf{r}) = \begin{cases} \phi_1 \exp(-\phi_2 \| \mathbf{s} - \mathbf{r} \|), & \text{if } \mathbf{s} \neq \mathbf{r} \\ \phi_1 + \phi_3, & \text{if } \mathbf{s} = \mathbf{r}. \end{cases} \quad (13)$$

From equations (5), (6), and (13) the parameters to be estimated are  $\alpha_1, \dots, \alpha_p, \phi_1, \phi_2, \phi_3$ . The covariance between state variables  $S$  over the same spatial locations  $\mathbf{s}$  separated a temporal distance  $k$  can be estimated by a consistent estimator that averages over the whole set of  $n$  locations and the whole set of temporal index ( $J$  years and  $t_0$  weeks, for example):

$$\widehat{C}_k^{(S)} = \frac{1}{nJt_0} \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^{t_0} \{ \delta_{jt}(\mathbf{s}_i) - \bar{\delta} \} \{ \delta_{jt}(\mathbf{s}_i) - \bar{\delta} \},$$

where

$$\bar{\delta} = \frac{1}{nJt_0} \sum_{i=1}^n \sum_{j=1}^J \sum_{t=1}^{t_0} \delta_{jt}(\mathbf{s}_i).$$

[36] Note that  $(\alpha_1, \dots, \alpha_p)'$  can be estimated by the Yule-Walker equations as

$$\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \\ \vdots \\ \widehat{\alpha}_p \end{bmatrix} = \begin{bmatrix} \widehat{C}_0^{(S)} & \widehat{C}_1^{(S)} & \dots & \widehat{C}_{p-1}^{(S)} \\ \widehat{C}_{-1}^{(S)} & \widehat{C}_0^{(S)} & \ddots & \vdots \\ & \ddots & \ddots & \widehat{C}_1^{(S)} \\ \widehat{C}_{1-p}^{(S)} & \widehat{C}_1^{(S)} & \widehat{C}_0^{(S)} & \vdots \end{bmatrix}^{-1} \begin{bmatrix} \widehat{C}_1^{(S)} \\ \widehat{C}_2^{(S)} \\ \vdots \\ \widehat{C}_p^{(S)} \end{bmatrix}$$

and  $C_0^{(S)}(\mathbf{s}, \mathbf{r}) = f_p(\alpha_1, \dots, \alpha_p) C^{(n)}(\mathbf{s}, \mathbf{r})$  for a particular function  $f_p(\alpha_1, \dots, \alpha_p)$ , which has to be determined by satisfying this last equality.

[37] The variogram of  $\delta_t(\cdot)$  between  $\mathbf{s}_i$  and  $\mathbf{s}_j$ , given by

$$2\gamma(\mathbf{s}_i, \mathbf{s}_j) = -\text{var}[\delta_t(\mathbf{s}_i) - \delta_t(\mathbf{s}_j)] \\ = \text{var}[S_t(\mathbf{s}_i) - S_t(\mathbf{s}_j)] \\ = C_0^{(S)}(\mathbf{s}_i, \mathbf{s}_i) + C_0^{(S)}(\mathbf{s}_j, \mathbf{s}_j) - 2C_0^{(S)}(\mathbf{s}_i, \mathbf{s}_j), \quad (14)$$

can be estimated by fitting an empirical semivariogram to the two-parameter semivariogram model given by the combination of equations (13) and (14). Following *Cressie*

[1993], the fitting can be done using weighted nonlinear least squares or restricted maximum likelihood. Note that because

$$C_0^{(S)}(\mathbf{s}, \mathbf{s}) = (\phi_1 + \phi_3) f_p(\alpha_1, \dots, \alpha_p),$$

it follows that  $\phi_3$  can be estimated by

$$\widehat{\phi}_3 = \frac{\widehat{C}_0^{(S)}}{f_p(\widehat{\alpha}_1, \dots, \widehat{\alpha}_p)} - \widehat{\phi}_1.$$

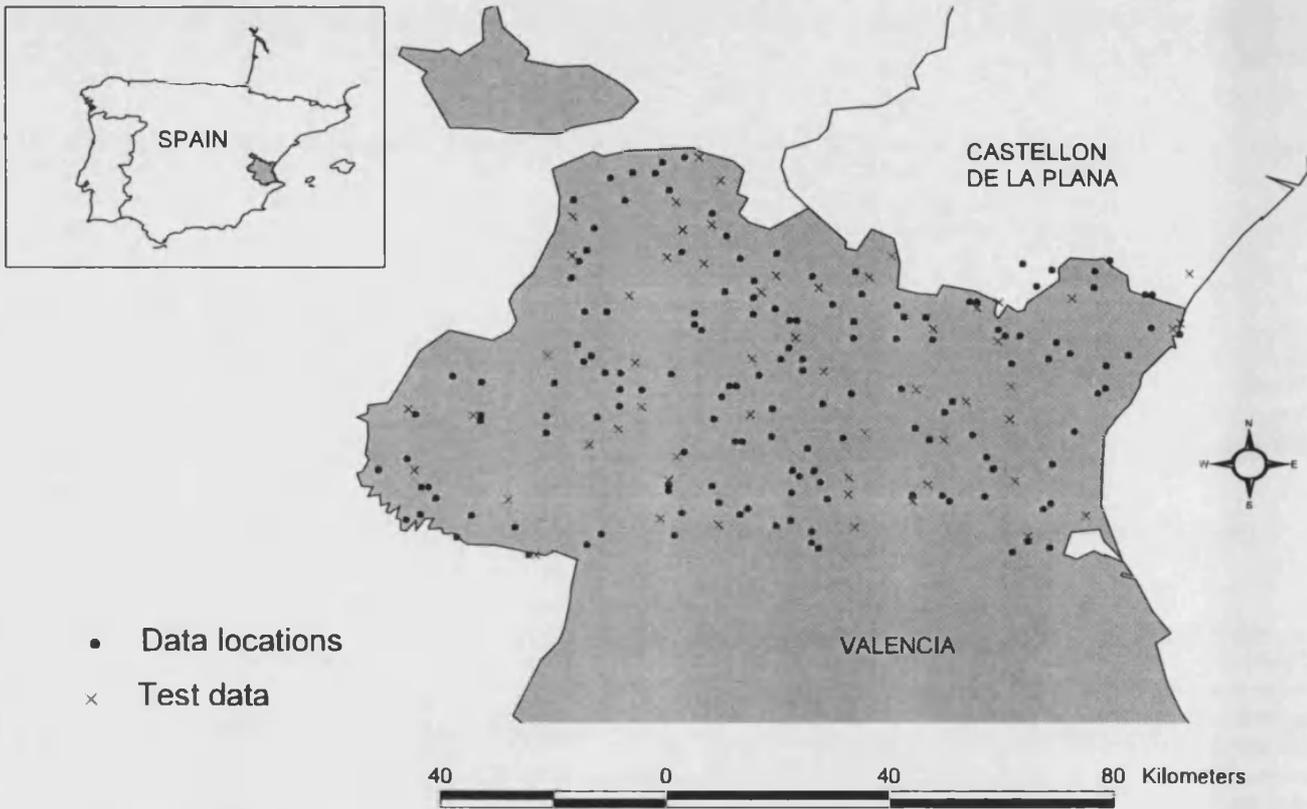
#### 4. Spatiotemporal Modeling of SSRB Data

[38] In this section we analyze the evolution of the surface shortwave radiation budget in both space and time and relate this evolution to certain soil variables having to do with aridification. We build and present statistical models that are able to predict the SSRB data in space and time and are thus able to input missing data, a common fact in these kind of measurements.

##### 4.1. Explanatory Variables

[39] During the years of 1994 and 1995 the public research center "Centro de Investigación Sobre Desertificación" carried out fieldwork to analyze soil characteristics at 205 spatial locations in a region located at the east of Iberian Peninsula, in particular, north of València province (Figure 1). These sample points were superimposed over a regular grid of resolution 1 km × 1 km, where the SSRB data was taken. A weighted mean (with weights being the inverse of the Euclidean distance) of SSRB data from the four closest grid points was associated to each one of the 205 sampled points. The result was a data set of soil characteristics together with SSRB measurements over 205 irregular locations. Note that the original support for the SSRB data was a regular fine grid. In this case, spectral methods [*Fuentes, 2002*] could be used for the spatiotemporal analysis. However, we restricted ourselves to the irregularly located 205 points where the soil characteristics were measured. At this point, it is worth noting that even in this case, when the source data were sampled over an irregular grid, taking into account the result in *Renshaw [2002]*, we could define an appropriate regular grid to approximate the irregular data locations and proceed normally with spectral analysis. However, in this paper we restrict ourselves to the spatial domain, where the three presented models provide a satisfactory prediction scheme for the SSRB data.

[40] The following variables were measured at each of the 205 locations: (1) altitude; (2) slope; (3) wet and dry soil color; (4) soil granulometric fractions (percent composition of superficial horizon according to the size of their particles; three fractions were considered, following the textural classification of USDA: (1) size particles as sand (2–0.05 mm), (2) size particles as silt (0.05–0.002 mm), (3) size particles as clay (<0.002 mm)); (5) soil texture (a qualitative classification summarizing the soil granulometric fractions); and (6) organic matter (percentage of soil organic matter). All these variables were used as explanatory variables in our analysis, taking into account that as granulometric fractions



**Figure 1.** The selected 205 locations used in our analysis. Dots define locations used for model development and crosses define locations for validation purposes.

are related by the equation sand + silt + clay = 100, only sand and silt were used.

**4.2. Exploratory Data Analysis**

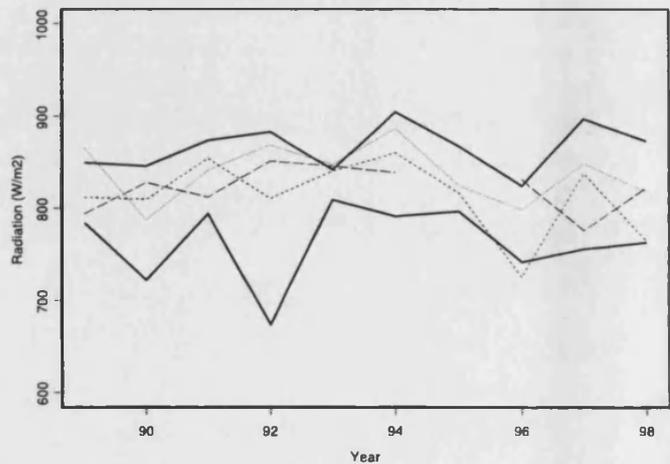
[41] We explored and described the SSRB data to analyze their behavior through time (weeks and years) and their relationship to the explanatory (covariates) variables having to do with soil. We selected 150 locations at random out of the initial 205 spatial locations to build the several statistical models. Then, the remaining 55 locations were used for model validation purposes. Revising the literature in the field of spatiotemporal modeling, as, for example, the work of *Huang and Cressie* [1996], *Cressie and Majure* [1997], or *Cressie and Huang* [1999], there is a trade-off between computational efficiency and statistical precision [*Wikle and Cressie*, 1999] when selecting the number of spatial locations to build a statistical model. In particular, the number of 150 spatial locations proved to be enough to get significant statistical conclusions, apart from focusing on computationally efficient statistical procedures.

[42] Recall that the SSRB data were measured for a period of 10 years, from 1989 to 1998. Within each year, five measures were taken, each corresponding to a period of 10 days, from 10 (1 April) to 14 (20 May), except for 1995, when the data for period 12 were missing. Thus the total number of data used in each model estimation was  $150 \times 5 - 150 = 7350$ .

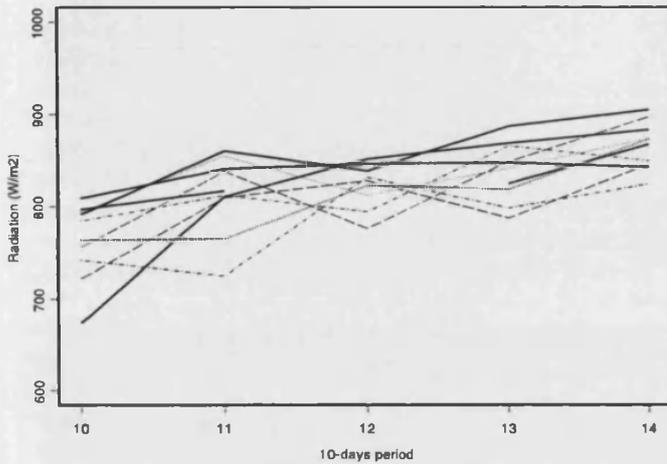
[43] Figure 2 shows the plot of the mean SSRB radiation, in  $W m^{-2}$ , calculated over the 150 locations as a function of

years for the five different 10-day periods. From this plot we see that the year effect was null and that the SSRB data behaved similarly through the years.

[44] Figure 3 shows the plot of the mean SSRB radiation, calculated over the 150 locations as a function of the 10-day period for the whole set of 10 years. We now see an increasing trend with the 10-day period, which was obvious as we came closer to the summer season.



**Figure 2.** Mean surface shortwave radiation budget (SSRB) radiation (over the 150 locations) along the years for the five different 10-day periods.



**Figure 3.** Mean SSRB radiation (over the 150 locations) along the 10-day periods for 10 years.

[45] We further analyzed the relationship between the SSRB data and the set of soil explanatory variables. Our aim was to reduce this set and to keep only those related to the radiation. We thus performed correlations and stepwise linear regression through the selection criterion given by *Ericksen et al.* [1989] to conclude that the only soil variables to be considered in the trend part of the models were altitude (alt), slope (slope), and organic matter (om). At this point, notice that, strictly speaking, it cannot be claimed that the coefficients were significantly different from zero due to spatial dependence among the regression errors. This would require knowledge of the errors' variance-covariance matrix, something that is unavailable at this phase of the analysis. In sections 4.3.3 and 4.3.4 we build such a matrix for the purposes of spatiotemporal prediction. Nevertheless, *Ericksen et al.*'s [1989] selection criterion can still be used as a means of large-scale variation model selection [*Cressie and Majure*, 1997].

### 4.3. Model-Fitting Results

[46] We implemented three alternative models, as shown in section 3: (1) the spatiotemporal model, with prediction based on ordinary kriging (STOK); (2) the state-space model, with an autoregressive signal of order 1 and with prediction based on the Kalman filter technique (KF-AR1); and (3) the purely spatial model, with prediction based on ordinary kriging (SOK). The original radiation data were transformed to stabilize the standard deviation, which tended to be larger for larger mean SSRB values, and at the same time to achieve Gaussianity. We used two data transformations. Each model was fitted with two different sets of covariates: considering only geographical coordinates and adding soil and temporal dependence variables to geographical coordinates. This gave 6 different combinations, which turned into 12 when using the two data transformations.

#### 4.3.1. Data Transformation

[47] In the following expressions the subindex  $t$  was replaced by a double subindex  $i, j$  to better describe the double temporal dependence of each observation. Thus  $j$  represents the year ( $j = 89, 90, \dots, 98$ ) and  $i$  represents the 10-day period ( $i = 10, 11, \dots, 14$ ).

[48] In normalization 1 the original radiation data were transformed by means of

$$Z_{j,i}^{(1)}(\mathbf{s}) = \frac{Z_{j,i}(\mathbf{s}) - \hat{\mu}}{\hat{\sigma}},$$

where  $\hat{\mu}$  and  $\hat{\sigma}^2$  are the mean and variance of the total set of data, respectively. In normalization 2 the original radiation data were transformed by means of

$$Z_{j,i}^{(2)}(\mathbf{s}) = \frac{Z_{j,i}(\mathbf{s}) - \hat{\mu}_i}{\hat{\sigma}_i},$$

where  $\hat{\mu}_i$  and  $\hat{\sigma}_i^2$  are the mean and variance of the data coming only from the 10-day period  $i$ , respectively:

$$\hat{\mu}_i = \frac{1}{150 \times 10} \sum_{s=1}^{150} \sum_{j=89}^{98} Z_{j,i}(\mathbf{s})$$

$$\hat{\sigma}_i^2(\mathbf{s}) = \frac{1}{150 \times 10} \sum_{s=1}^{150} \sum_{j=89}^{98} (Z_{j,i}(\mathbf{s}) - \hat{\mu}_i)^2.$$

Recall that in the above expressions the data corresponding to period 12 of 1995 were missing.

#### 4.3.2. Trend Analysis

[49] Recall that the theoretical model describing the solar radiation data, given by equation (1), is of the form  $Z_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \delta_t(\mathbf{s})$ , where  $\mu_t(\mathbf{s})$  defines the deterministic large-scale variation, which, in our case, is expressed as a linear function of  $k$  covariates. In a first class of models, referred to as "geocovariates" in the text, these covariates were defined by a quadratic form of the spatial geographical coordinates, longitude (long) and latitude (lat) (see equation (15)). A second class of models, referred to as "all covariates" in the text, was built by also considering soil explanatory and temporal dependence variables (equation (16)). The temporal dependence comes only through the 10-day periods, and so we introduced the (0-1) dummy variables  $x_{11}, \dots, x_{14}$ . Thus the expression defining the trend  $\mu_t(\mathbf{s})$  is given, for both classes of models, by

$$\begin{aligned} \mu_t(\mathbf{s}) = & \beta_0 + \beta_1 \text{long}(\mathbf{s}) + \beta_2 \text{lat}(\mathbf{s}) + \beta_3 \text{lat}^2(\mathbf{s}) \\ & + \beta_4 \text{long}^2(\mathbf{s}) + \beta_5 \text{long}(\mathbf{s})\text{lat}(\mathbf{s}) \end{aligned} \quad (15)$$

and

$$\begin{aligned} \mu_t(\mathbf{s}) = & \beta_0 + \beta_1 \text{long}(\mathbf{s}) + \beta_2 \text{lat}(\mathbf{s}) + \beta_3 \text{lat}^2(\mathbf{s}) + \beta_4 \text{long}^2(\mathbf{s}) \\ & + \beta_5 \text{long}(\mathbf{s})\text{lat}(\mathbf{s}) + \beta_6 \text{lat}(\mathbf{s}) + \beta_7 \text{slope}(\mathbf{s}) + \beta_8 \text{om}(\mathbf{s}) \\ & + \beta_9 x_{11}(\mathbf{s}) + \beta_{10} x_{12}(\mathbf{s}) + \beta_{11} x_{13}(\mathbf{s}) + \beta_{12} x_{14}(\mathbf{s}). \end{aligned} \quad (16)$$

#### 4.3.3. Parameter Estimations

[50] The trend in the linear model of equations (15) and (16) was fitted using ordinary least squares. Weighted least squares (WLS) or generalized least squares are other good possibilities. In particular, WLS was not used here as the matrix of the linear model variance was shown to be diagonal with equal elements. Table 1 shows

**Table 1.** Trend Analysis: Estimations and Standard Errors of the Coefficients for the Regression Models (15) and (16) Under the Two Data Normalizations\*

	Model Equation (15)		Model Equation (16)	
	Normalization 1	Normalization 2	Normalization 1	Normalization 2
$\beta_0$	-1.4897 <i>0.2286</i>	-1.7879 <i>0.2277</i>	-1.4488 <i>0.1883</i>	-0.7054 <i>0.2202</i>
$\beta_1$	-2.8894 <i>0.3173</i>	-3.4217 <i>0.3161</i>	0.0212 <i>0.2788</i>	0.0227 <i>0.3260</i>
$\beta_2$	0.7064 <i>0.5328</i>	0.9197 <i>0.5308</i>	1.0574 <i>0.4325</i>	1.3327 <i>0.5057</i>
$\beta_3$	1.4850 <i>0.4295</i>	1.6986 <i>0.4278</i>	-0.0698 <i>0.3551</i>	-0.1402 <i>0.4152</i>
$\beta_4$	-0.9104 <i>0.1202</i>	-1.0730 <i>0.1197</i>	-0.0076 <i>0.1026</i>	-0.0050 <i>0.1200</i>
$\beta_5$	2.4130 <i>0.2851</i>	2.8715 <i>0.2840</i>	1.5016 <i>0.2322</i>	1.7916 <i>0.2716</i>
$\beta_6$			0.0008 <i>0.0001</i>	0.0009 <i>0.0001</i>
$\beta_7$			0.0138 <i>0.0007</i>	0.0164 <i>0.0008</i>
$\beta_8$			0.0043 <i>0.0026</i>	0.0050 <i>0.0030</i>
$\beta_9$			0.7597 <i>0.0292</i>	
$\beta_{10}$			0.8976 <i>0.0300</i>	
$\beta_{11}$			1.1475 <i>0.0292</i>	
$\beta_{12}$			1.5643 <i>0.0292</i>	

\*Standard errors are in italics.

the estimations of the coefficients, together with the corresponding standard errors, in the two regression models for the trend under the two data normalizations. Note that the coefficients of the temporal variables are zero due to the intrinsic characteristics of the second normalization.

[51] It is tempting to use the regression models defined in Table 1 for purposes beyond those for which they are intended. However, the fitted regressions are merely a description of the large-scale variation of the data and should not be "overinterpreted." There is likely to be colinearity among the regressors. The regression models adopted are a means to an end, the end being prediction with known precision of SSRB data. In other environmental studies, where one wishes to interpret the regression coefficients, it would be efficient to use the spatiotemporal dependence model to obtain a weighted or generalized least squares estimator [Cressie and Majure, 1997].

[52] Once the trend was fitted, the residuals were examined using a normal probability plot. They were judged to be symmetric, with a tendency to be heavier tailed than the normal distribution. However, when we further modeled the spatial and spatiotemporal dependences in these residuals, we used a weighted least squares criterion to fit the variogram estimator and consequently, the influence of heavy tails was downweighted.

[53] Table 2 shows the results of the spatial semivariogram analysis carried out for the isotropic residuals after removing the deterministic trend. The four semivariograms models were exponential and fitted using the weighted nonlinear least squares criterion given by

Cressie [1993]. None of them presented a positive nugget effect, indicating that no measurement error and micro-scale variation were encountered (Figure 4). The spatial semivariogram model was better fitted (smaller standard errors) when using normalization 1 and when considering all covariates. This is a logical finding in the sense that the covariates are not evolving in time and so more information comes into the spatial model with covariates. Moreover, note that with normalization 1, no time dependence is considered.

[54] We next analyzed the spatiotemporal semivariogram. The dip in the variogram estimates near the origin (Figure 5) indicates the presence of spatial and temporal dependence. The variogram must be conditionally negative-definite [Cressie, 1993]; however, its estimators typically are not. Therefore, in order to carry out valid spatial prediction (with positive mean squared prediction errors), a valid variogram model must be fitted to the estimates shown in Figure 5.

[55] Many of the problems of space and time modeling, in particular in atmospheric sciences, can be overcome by using separable processes. This subclass of spatial-temporal processes has several advantages, including rapid fitting and simple extensions of many techniques developed and successfully used in time series and classical geostatistics. In particular, a major advantage of these processes is that the covariance matrix for a realization can be expressed as the Kronecker product of two smaller matrices that arise separately from the temporal and purely spatial processes, and hence its determinant and inverse are easily determined. Thus separability is a desirable property for spatial-temporal processes.

[56] Looking at the empirical spatiotemporal semivariograms, we saw that the fitted model should conform to the fact that the time series associated with two different spatial locations should have similar shapes. The product model is one of the simplest ways to model a separate covariance in space-time and satisfies the just-mentioned condition [De Cesare et al., 2001]. We thus fitted the spatiotemporal semivariogram with a separable exponential structure given by

$$\gamma_u(\|h\|) = n + s \left[ 1 - \exp\left(-\frac{u}{r_t} - \frac{\|h\|}{r_s}\right) \right], \quad (17)$$

if  $h \neq 0$  or  $u \neq 0$  or  $\gamma_u(\|h\|) = 0$  otherwise. In equation (17),  $n$  is the nugget,  $s$  is the sill, and  $r_t$  and  $r_s$  are the temporal

**Table 2.** Estimation Results and Standard Errors for the Spatial Semivariogram Analysis for the Residuals After Having Removed the Trend\*

	With Geocovariates		With All Covariates	
	Normalization 1	Normalization 2	Normalization 1	Normalization 2
	Nugget	0	0	0
Sill	0.2914 <i>0.0096</i>	0.5469 <i>0.0177</i>	0.1895 <i>0.0058</i>	0.3648 <i>0.0112</i>
Range	0.0720 <i>0.0107</i>	0.0732 <i>0.0105</i>	0.0452 <i>0.0089</i>	0.0523 <i>0.0092</i>

\*Standard errors are in italics.

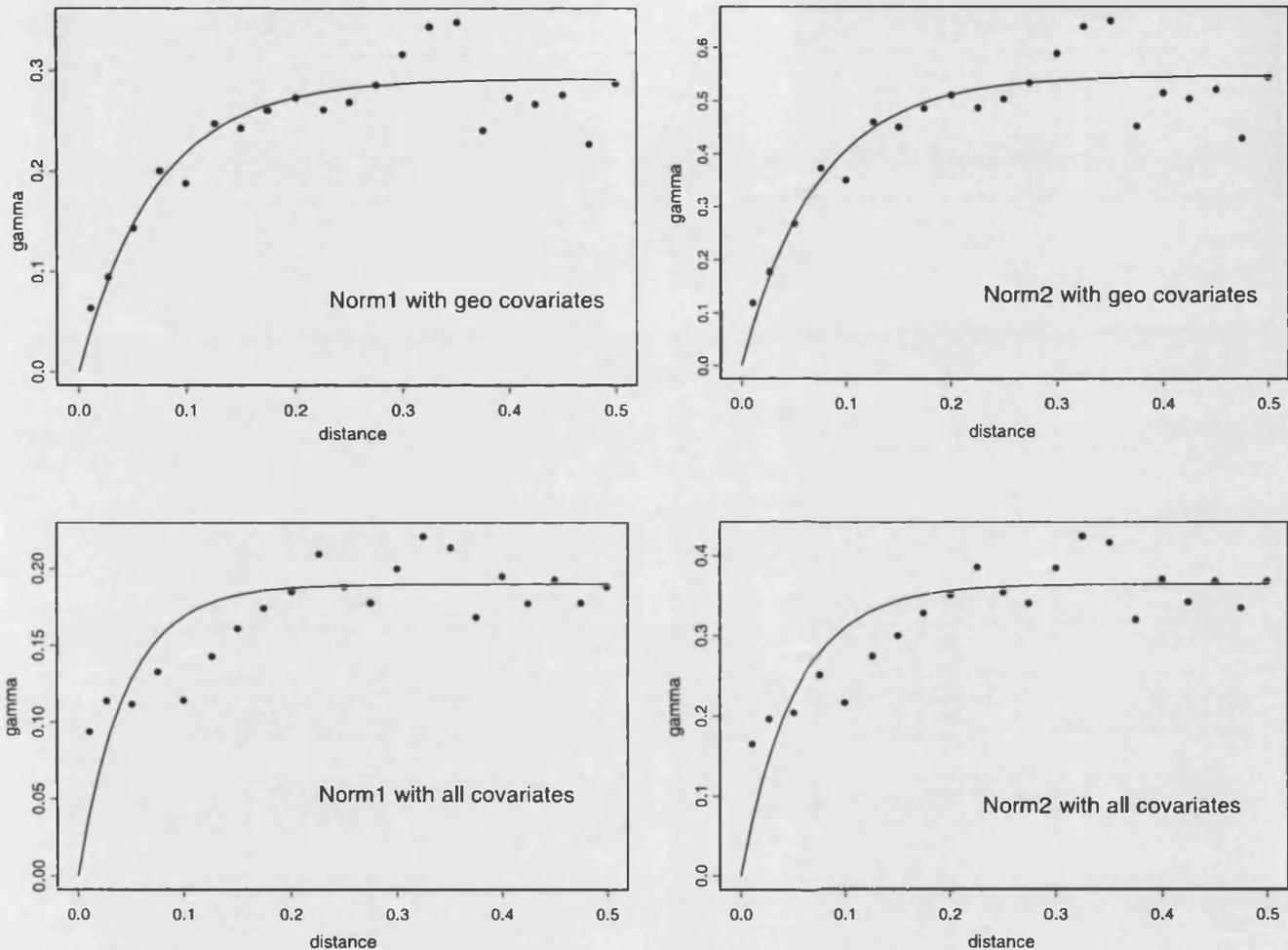


Figure 4. Empirical and fitted spatial semivariograms for the residuals after having removed the trend.

and spatial ranges, respectively. Again, we used weighted nonlinear least squares to fit the spatiotemporal semivariogram. Table 3 and Figure 5 show the corresponding results for the spatiotemporal semivariogram analysis. The spatiotemporal semivariogram model was better fitted (smaller standard errors) when using normalization 2 and when considering all covariates. Thus time dependence (as it is defined with normalization 2) has to be taken into account when normalizing SSRB data and using a spatiotemporal model. Again, the whole set of geographical and soil variables has to be considered for a better semivariogram fit. Note that when considering the whole set of covariates, the isotropic residual process seemed to be closer to a zero-mean process, and this was emphasized by the fact that the spatial and temporal ranges were markedly reduced with respect to the same ranges when considering geographical covariates only. This reinforces the use of soil covariates.

[57] Finally, Table 4 shows the results when using the state-space model and the Kalman filter procedure through an autoregressive structure of order 1, KF-AR1. This model was chosen in terms of the temporal dependence structure observed in the temporal range shown in the spatiotemporal semivariograms (Table 3). Note that these ranges, in Table 3, are close to 1, except for the case of normalization 1 with

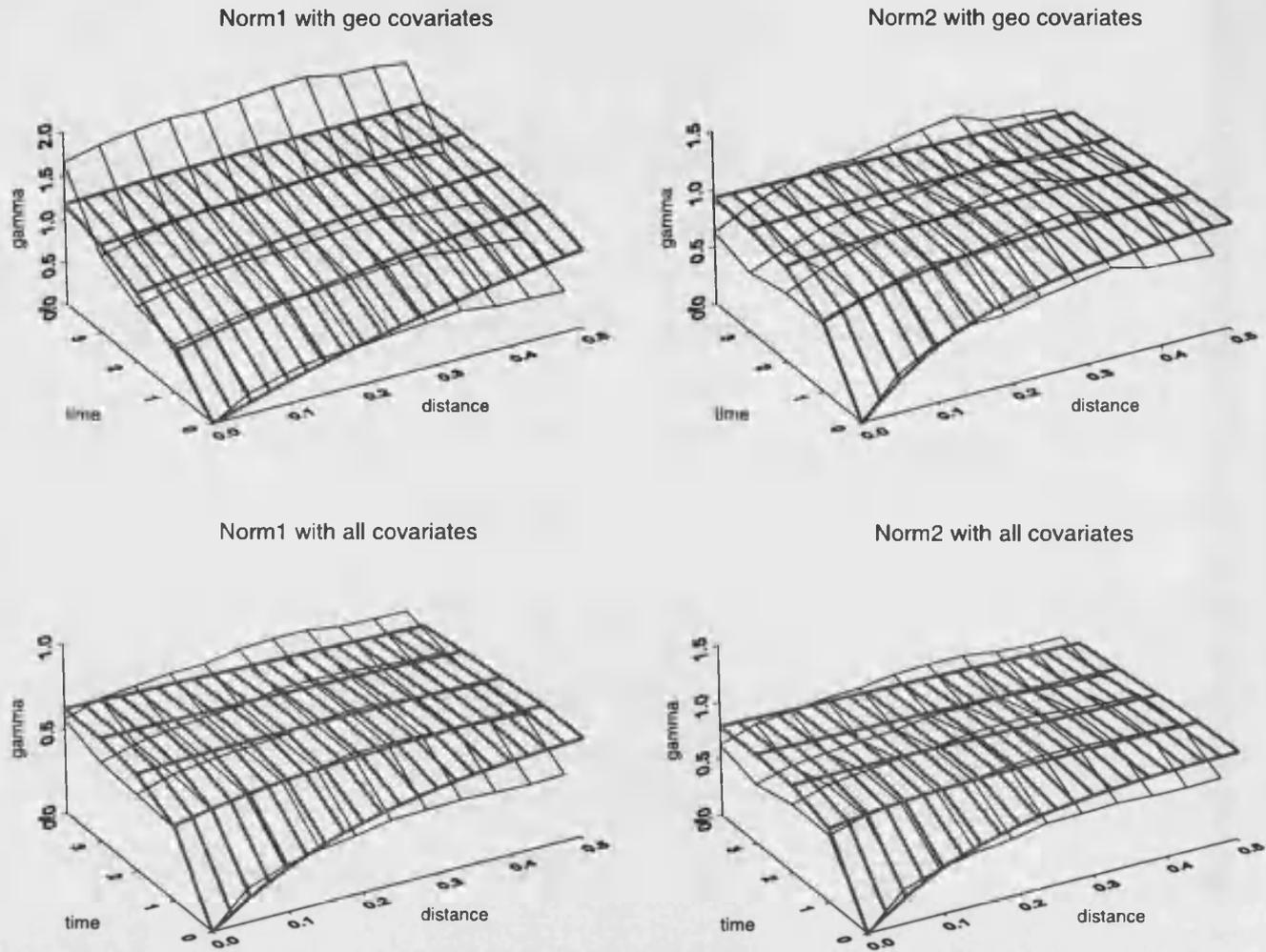
geocovariates. As the fitted semivariograms showed a zero-nugget value, we fixed  $\phi_3 = 0$  in Table 4.

#### 4.4. Prediction and Cross-Validation Results

[58] As goodness-of-fit techniques, we used both mean square error (MSE) obtained with the prediction at the 55 selected checkpoints (see Figure 1) and cross validation over the 150 locations used for model building (see Figure 1). The prediction refers to the last period of 10 days (named as 14) for the last year (1998). The temporal intervals used for the spatiotemporal models (STOK and KF-AR1) were the five 10-day periods corresponding to 1998.

[59] Table 5 can be used for comparison purposes among the 12 fitted models in terms of the MSE obtained with the prediction at the 55 checkpoints. For these locations, Figure 6 shows prediction and standard error (SE) of prediction maps for solar radiation for the fitted models SOK, STOK, and KF-AR1 when normalization 2 and geocovariates were considered.

[60] Table 5 shows that for both trend models, geocovariates and all covariates, the Kalman filter (KF-AR1) and the spatiotemporal (STOK) models provided smaller MSE than when we considered only the spatial dependence (SOK). Thus the temporal dependence has a role to play



**Figure 5.** Empirical (shaded line) and fitted (bold line) spatiotemporal semivariograms for the residuals after having removed the trend.

within the spatial interaction when modeling SSRB data. In particular, the MSE of the prediction under the trend models based on all covariates was interestingly smaller than those obtained with the trend based on geographical variables if normalization 2 is considered. So, concluding from Table 5, there was a clear improvement when predicting SSRB by modeling both the spatial and temporal dependence. More-

over, there was a marginal improvement when considering the more complicated model, KF-AR1, with respect to the STOK if the second type of normalization is taken into account and if we model the trend term through geographical covariates.

[61] A further goodness-of-fit technique is based on cross-validation tools, as used by *Huang and Cressie*

**Table 3.** Estimation Results and Standard Errors for the Spatiotemporal Semivariogram Analysis for the Residuals After Having Removed the Trend<sup>a</sup>

	With Geocovariates		With All Covariates	
	Normalization 1	Normalization 2	Normalization 1	Normalization 2
Nugget	0	0	0	0
Sill	1.4978	0.9443	0.6247	0.7927
	<i>0.1935</i>	<i>0.0206</i>	<i>0.0178</i>	<i>0.0171</i>
Spatial range	0.5317	0.1360	0.1897	0.1402
	<i>0.0982</i>	<i>0.0136</i>	<i>0.0184</i>	<i>0.0139</i>
Temporal range	2.6268	0.9667	0.8505	0.6045
	<i>0.8539</i>	<i>0.2267</i>	<i>0.2217</i>	<i>0.2104</i>

<sup>a</sup>Standard errors are in italics.

[1996]. The quantity CRV1 checks the unbiasedness of the predictor and should be approximately equal to zero; CRV2 checks the accuracy of the standard deviation of the prediction error and should be approximately equal to 1; and CRV3 is a measure of goodness of prediction. The results of these cross-validation statistics, calculated over the 150 locations selected for model building, are shown in Table 6.

[62] An inspection of Table 6 indicates that the values of CRV1 and CRV2 are all reasonable, though those values coming from models based on the second type of normalization were slightly better. Moreover, with this normalization we also got smaller values of CRV3, which indicates that independently of the trend model, a proper analysis of SSRB data should be carried out by considering a normalization which takes into account the temporal variation, as does normalization 2.

5. Conclusions

[63] When considering only the spatial dependencies (model SOK), the spatial covariance was better fitted when the SSRB data was transformed through type 1 normalization (normalization 1) and when the trend term of the model was modeled by means of the whole set of covariates (all covariates model). The time dependence (as it is defined with normalization 2) has to be taken into account when normalizing SSRB data and using a spatiotemporal model. Again, the whole set of geographical and soil variables has to be considered for a better covariance fit.

[64] In general, and for both trend models, geocovariates and all covariates, the Kalman filter (KF-AR1) and the spatiotemporal (STOK) models showed smaller MSE in the prediction of SSRB data than those showed by the spatial dependence model (SOK). Thus we claim that the temporal dependence has a role to play together with the spatial interaction when modeling SSRB data. In particular, the MSE of the prediction under the trend models based on all covariates was smaller than the one obtained with the trend based on geographical variables. So, there was a clear improvement when predicting SSRB by modeling both the spatial and temporal dependence.

[65] Moreover, there was a marginal improvement when considering the more complicated model, KF-AR1, with respect to the STOK if the second type of normalization was taken into account and if we modeled the trend term through geographical covariates.

[66] Considering the second type of normalization, we got smaller values of CRV3, which indicated that indepen-

Table 5. MSE for the 12 Fitted Models and Mean Standard Errors for Prediction Over the 55 Checkpoints<sup>a</sup>

	With Geocovariates		With All Covariates	
	Normalization 1	Normalization 2	Normalization 1	Normalization 2
	SOK	24.7191 <i>23.5896</i>	24.5968 <i>23.5384</i>	30.1845 <i>22.4457</i>
STOK	24.0856 <i>21.0068</i>	24.0917 <i>23.6430</i>	29.0381 <i>22.4396</i>	27.8575 <i>21.3614</i>
KF-AR1	24.0997 <i>21.6831</i>	24.0652 <i>23.8721</i>	29.0673 <i>22.4518</i>	27.8636 <i>21.5260</i>

<sup>a</sup>Mean standard errors are in italics.

dently of the trend model, a proper analysis of SSRB data should be carried out by considering a normalization which takes into account the temporal variation.

[67] So, taking into account the above considerations, we would recommend modeling both the spatial and temporal dependence of the SSRB data, previously transformed with normalization 2, in which each datum was standardized by the mean and variance of its corresponding 10-day period. As with respect to the statistical models, we recommend using the more complex one based on the Kalman filter as it provided slightly better results than the spatiotemporal model based on ordinary kriging, and also, and most important, it is more general and can be used in more and different environmental problems. However, we might expect a bigger difference in favor of the spatiotemporal models due to the nature of the solar radiation process. So, these results could be indicating that more data within each year should be taken into account.

[68] The spatiotemporal models proposed in this paper are statistical and, therefore, provide estimates of the precision of predictions. Large prediction standard errors, and thus prediction intervals, simply indicate that the quantity is difficult to predict accurately in time and space. Another appealing feature is the models' ability to take physical conditions into account through easy quantification of physical landscape characteristics in a geographic information system. Although the models presented here have been developed for the analysis of SSRB data, they should also prove useful for many other spatiotemporal prediction problems.

[69] It is important to note that in this paper we have only used separable covariance structures. Also, though its implementation has been successful in our (and many other's) application, there are fundamental differences between space and time, and it is not likely that realistic covariance structures can be specified that accurately capture the complicated dynamical processes as found in geophysical applications. In this sense, nonseparable structures could be further considered. A test such as the one developed by Fuentes [2003] would be of great help in this context. Also, space-time dynamic models could improve the presented methodology.

[70] Uncertainty in geostatistics plays a crucial role. We have not considered it, and it would be nice if Bayesian approaches could be used to capture different sources of uncertainty.

[71] Finally, as stated in the paper, spectral methods could have been used to analyze data on a regular or irregular grid.

Table 4. Estimation Results and Standard Errors for the Parameters of the Autoregressive Kalman Filter Model (KF-AR1)<sup>a</sup>

	With Geocovariates		With All Covariates	
	Normalization 1	Normalization 2	Normalization 1	Normalization 2
$\alpha_1$	0.3640	0.3942	0.3255	0.3257
$\phi_1$	0.1065	0.2353	0.1139	0.1946
$\phi_2$	3.4169	6.1285	5.1708	5.4160
	<i>0.3293</i>	<i>0.5250</i>	<i>0.5534</i>	<i>0.5821</i>
$\phi_3$	0	0	0	0

<sup>a</sup>Standard errors are in italics.

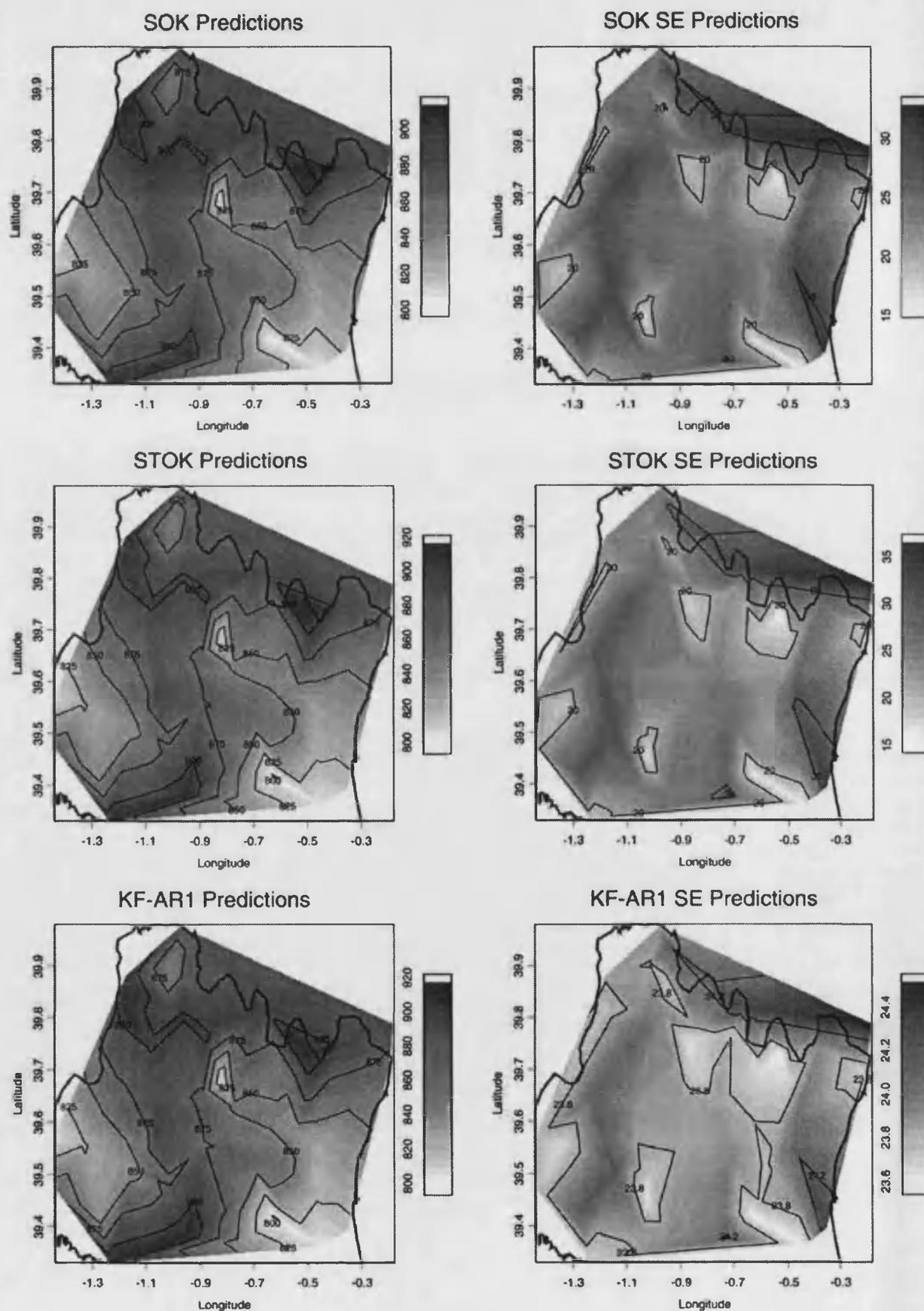


Figure 6. Predictions and standard errors over the 55 checkpoints.

**Table 6.** Summary of Cross-Validation Results for the 12 Fitted Models\*

	SOK	STOK	KF-AR1
<i>Normalization 1</i>			
CRV1	-0.0054	-0.0010	0.0018
	<i>-0.0014</i>	<i>0.0087</i>	<i>0.0037</i>
CRV2	1.0643	1.2107	1.1857
	<i>1.2012</i>	<i>1.2255</i>	<i>1.2012</i>
CRV3	25.5078	25.7212	25.7426
	<i>26.5723</i>	<i>27.0512</i>	<i>26.9925</i>
<i>Normalization 2</i>			
CRV1	-0.0159	0.0096	-0.0063
	<i>-0.0028</i>	<i>0.0149</i>	<i>0.0024</i>
CRV2	1.1352	1.0660	1.0685
	<i>1.2095</i>	<i>1.2444</i>	<i>1.2208</i>
CRV3	26.5699	25.6004	25.5739
	<i>26.0338</i>	<i>26.3249</i>	<i>26.3063</i>

\*Models with all covariates are in italics; other values are for models with geocovariates.

It would be interesting to compare spectral modeling to the modeling presented here.

[72] **Acknowledgments.** We acknowledge the Institute for Meteorology of the Free University of Berlin for the use of the MEDOKADS data used in this work. One of the authors, A.B.-S., holds a predoctoral grant funded by the Conselleria de Educació i Cultura de la Generalitat Valenciana. This work has been partially supported by DGICYT grants ESP-2001-4525-PE and BFM-2001-3286. We are grateful to the two referees and to the editor for their suggestions, which have improved the first draft of the paper.

## References

- Bolle, H. J., (Ed.), *Synthesis of Change Detection Parameters Into a Land-Surface Change Indicator for Long-Term Desertification Studies in the Mediterranean Area (RESYSMED)*, Final Rep. ENV4-CT97-0683, Inst. Meteorol. Climatol. and Remote Sens., Florence, Italy, Dec. 1999.
- Cressie, N., *Statistics for Spatial Data*, rev. ed., John Wiley, Hoboken, N. J., 1993.
- Cressie, N., and H. C. Huang, Classes of nonseparable, spatio-temporal stationary covariance function, *J. Am. Stat. Assoc.*, **94**, 239–252, 1999.
- Cressie, N., and J. Majure, Spatio-temporal statistical modelling of livestock waste in streams, *J. Agric. Biol. Environ. Stat.*, **2**, 24–47, 1997.
- Cressie, N., and D. L. Zimmerman, On the stability of the geostatistical method, *Math. Geol.*, **24**, 45–59, 1992.
- De Cesare, L., D. E. Myers, and D. Posa, Estimating and modeling space-time correlation structures, *Stat. Probab. Lett.*, **51**, 9–14, 2001.
- Ericksen, E. P., J. B. Kadane, and J. W. Tukey, Adjusting the 1980 Census of Population and Housing, *J. Am. Stat. Assoc.*, **84**, 927–944, 1989.
- Fuentes, M., Spectral methods for nonstationary spatial processes, *Biometrika*, **89**, 197–210, 2002.
- Fuentes, M., Testing for separability of spatial-temporal covariance functions, in *Proceedings of the ISI International Conference on Environmen-*

*tal Statistics and Health*, edited by J. Mateu, D. Holland, and W. Gonzalez-Manteiga, pp. 67–76, Int. Stat. Inst., Santiago de Compostela, Spain, 2003.

- Ghil, M., S. E. Cohn, J. Tavantzis, K. Bube, and E. Isaakson, Applications of estimation theory to numerical weather prediction, in *Dynamic Meteorology: Data Assimilation Methods*, edited by L. Bengtsson, M. Ghil, and E. Kallén, pp. 139–224, Springer-Verlag, New York, 1981.
- Goodall, C., and K. V. Mardia, Challenges in multivariate spatio-temporal modeling, paper presented at the 17th International Biometric Conference, Int. Biometric Soc., Hamilton, Ontario, Canada, 1994.
- Guttorp, P., W. Meiring, and P. D. Sampson, A space-time analysis of ground-level ozone data, *Environmetrics*, **5**, 241–254, 1994.
- Hollmann, R., and A. Bodas, Comparison of the surface solar netflux from ScaRaB with observations, in *IRS 2000: Current Problems in Atmospheric Radiation*, edited by W. L. Smith and Y. M. Timofeyev, pp. 536–539, A. Deepak, Hampton, Va., 2001.
- Huang, H. C., and N. Cressie, Spatio-temporal prediction of snow water equivalent using the Kalman filter, *Comput. Stat. Data Anal.*, **22**, 159–175, 1996.
- Kalman, R. E., A new approach to linear filtering and prediction problems, *J. Basic Eng.*, **82**, 35–45, 1960.
- Koslowsky, D., Mehrjaehrige validierte und homogenisierte Reihen des Reflexionsgrades und des Vegetationsindexes von Landoberflächen aus taeglichen AVHRR-Daten hoher Aufloesung, *Meteorol. Abh. Neue Folge, Ser. A Monogr.*, **9**(1), 236 pp., 1996.
- Kyriakidis, P. C., and A. G. Journel, Geostatistical space-time models: A review, *Math. Geol.*, **31**, 651–684, 1999.
- Li, Z., H. G. Leighton, K. Masuda, and T. Takashima, Estimation of SW flux absorbed at the surface from TOA reflected flux, *J. Clim.*, **6**, 317–330, 1993a.
- Li, Z., H. G. Leighton, and R. D. Cess, Surface net solar radiation estimated from satellite measurements: Comparisons with tower observations, *J. Clim.*, **6**, 1764–1772, 1993b.
- Price, J. C., Timing of NOAA afternoon passes, *Int. J. Remote Sens.*, **12**, 193–198, 1991.
- Renshaw, E., Two-dimensional spectral analysis for marked point processes, *Biometrical J.*, **6**, 718–745, 2002.
- Stackhouse, P. W., Jr., S. J. Cox, S. K. Gupta, R. C. DiPasquale, and D. E. Brown, The WCRP/GEWEX Surface Radiation Budget Project Release 2: First results at 1 degree resolution, paper presented at the 10th Conference on Atmospheric Radiation: A Symposium With Tributes to the Works of Verner E. Suomi, Am. Meteorol. Soc., Madison, Wis., 28 June to 2 July 1999.
- Whitlock, C. H., et al., First Global WCRP Shortwave Surface Radiation Budget data set, *Bull. Am. Meteorol. Soc.*, **76**, 905–922, 1995.
- Wikle, C. K., and N. Cressie, A dimension-reduced approach to space-time Kalman filtering, *Biometrika*, **86**, 815–829, 1999.
- Wikle, C. K., L. M. Berliner, and N. Cressie, Hierarchical Bayesian space-time models, *Environ. Ecol. Stat.*, **5**, 117–154, 1998.

A. Bodas-Salcedo and E. López-Baeza, Departament de Termodinàmica, Universitat de València, E-46100 Burjassot, Spain. (bodas@uv.es; lopez@uv.es)

F. Martínez and F. Montes, Departament d'Estadística i Investigació Operativa, Universitat de València, E-46100 Burjassot, Spain. (pmartinez@uv.es; montes@uv.es)

J. Mateu, Departament de Matemàtiques, Universitat Jaume I, E-08071 Castelló, Spain. (mateu@mat.uji.es)

# Dynamic Life Tables: a Geostatistical Approach

**Ana Debón**

*Dpto. de Estadística e I. O. Aplicadas y Calidad  
Universidad Politécnica de Valencia, Spain*

**F. Martínez-Ruiz**

*Oficina de Estadística  
Ayuntamiento de Valencia, Spain*

**F. Montes\***

*Dpt. d'Estadística i I. O.  
Universitat de València, Spain*

## Abstract

The aim of mortality data graduation is to estimate death probabilities at age  $x$ ,  $q_x$ , using different models. Experience shows that standard life tables overestimate death probabilities. This fact has a negative consequence for insurance companies because premiums are lower than they would be otherwise. The reason for this overestimation is that standard life tables, as they are computed for a specific period of time, cannot take into account the decreasing mortality trend over time. Dynamic life tables overcome this problem by incorporating the influence of the calendar when graduating mortality. Recent papers on the topic look for the development of new methods to deal with this dynamism.

The majority of methods used in dynamic tables are parametric ones, because parametric graduation allows the comparison of death probabilities for different years. Most of these methods apply traditional mortality laws and then analyze the evolution of estimated parameters with time series techniques.

Our contribution consists in applying spatial statistical methods, because a dynamic table could be seen as a two-way table with dependent values or as data observed in a regular grid. In particular, we have applied geostatistical techniques for estimating the dependence structure of the data and for prediction purposes.

*Keywords:* Standard Life Tables, Dynamic Life Tables, Spatio-temporal Variogram, Kriging.

---

\*Correspondence to: F. Montes, [montes@uv.es](mailto:montes@uv.es)

Research partially supported by a grant from MEyC (Ministerio de Educación y Ciencia, Spain, project MTM-2004-06231). The research of Francisco Montes has also been partially supported by a grant from DGITT (Direcció General d'Investigació i Transferència Tecnològica de la Generalitat Valenciana, project GRUPOS03/189).

## 1 Introduction

The graduation of mortality data by means of parametric methodology has been widely addressed in papers such as Forfar et al. (1988) and Renshaw (1991) and by means of non-parametric methodology in Gavin et al. (1993, 1994, 1995). All them study the influence of age on data graduation. However, the latest actuarial literature recognises the fact that mortality progresses over the years, but earlier methods do not take this fact into account, as they were designed to analyse data corresponding to one year in particular or, in the case of several years, they worked with aggregated data.

The concept of a dynamic table seeks to solve this problem by jointly analysing mortality data corresponding to a series of consecutive years. This approach allows the calendar effect's influence on mortality to be studied. A compilation of a sample of dynamic models and their classification can be found in Benjamin and Soliman (1995). The majority of these methods adapt traditional laws to the new situation.

We aim to use tools from other statistical fields to study dynamic tables, concretely, spatial statistic techniques and, more precisely, geostatistical techniques. To this end, we estimate the covariance structure of the residuals obtained on eliminating any of the traditional mortality measurements, by means of an appropriate variogram model. By determining this structure, we will be able to make predictions by applying kriging techniques.

In order to justify the use of statistical methods designed for a context apparently far apart of that one considered here, we point out that the methods can be applied with no problem. As a matter of fact, the purpose of the geostatistical techniques is to model the dependence structure that governs a set of data influenced by its neighbour. In this context we need to define a neighbourhood relationship as well as a distance. They are straightforward in the case of spatial data but also possible in other kind of data. The analysis of sudden infant's deaths syndrome (SIDS) in North Carolina and the spatial experimental designs application in Cressie (1993) as well as the analysis using spatial techniques of the 1970 US Draft Lottery (Mateu et al., 2004) support this assessment. Moreover, as in previous studies, we will show that these methods provide better solutions than the classical methods since they simultaneously take into account the effect of age and time, while the other treat both effects separately.

This article is structured as follows: Section 2 briefly presents the geostatistical methodology to be used and quickly summarises the two traditional methods adapted to the analysis of dynamic tables: Lee-Carter's and the Gompertz-Makeham Laws. Section 3 presents the results of the application of the three methods to the analysis of mortality data in the Comunidad Valenciana, a region located on the east coast of Spain, corresponding to the period 1980-1999, and also the results of the prediction of  $q_{x,2000}$ , death probability at age  $x$  in year 2000, obtained by means of the adjusted models. Section 4 establishes the conclusions to be drawn from the results in the previous section.

## 2 Adjustment and Prediction of $q_{x,t}$

### 2.1 Lee-Carter Model

The Lee-Carter Model, developed in Lee and Carter (1992), consists in adjusting the following function to the mortality measurement,

$$q_{x,t} = \exp(a_x + b_x k_t) + \epsilon_{xt}$$

or, equally, the function

$$\ln(q_{x,t}) = a_x + b_x k_t + \epsilon'_{xt}, \quad (2.1)$$

applied to its logarithm transform. In (2.1), the double subscript refers to the age,  $x$ , and to the year or unit of time,  $t$ ,  $q_{x,t}$  are the probabilities of death at age  $x$  in the year  $t$ ,  $a_x$  and  $b_x$  are age-dependent parameters and  $k_t$  is a specific mortality index for each year or unit of time. The errors  $\epsilon_{xt}$ , with a zero average and variance  $\sigma_\epsilon^2$ , reflect the historical influences of each specific age that are not captured by the model. Notice that the model cannot be adjusted by the usual regression methods because the index values  $k_t$  are not observable. The solution is found by means of an iterative method, which is detailed in Lee and Carter (1992).

Once the series of values  $k_t$  have been estimated, the last step of the Lee-Carter Model consists in finding a model for the series using Box-Jenkins methodology. In many of these applications, a good model for the  $k_t$  is

$$k_t = c + k_{t-1} + a_t.$$

where  $c$  is constant and  $a_t$  is white noise. With this model, the prediction of  $k_t$  varies in a linear way and each death rate predicted varies at a constant exponential rate.

Some authors have made suggestions about and modified this method, including Carter and Lee (1992), Wilmoth (1993) and Lee (2000) himself, in a paper where he compares his method to other alternatives, such as that of McNown and Rogers (1989, 1992).

## 2.2 Gompertz-Makeham Functions Concerning Age and Time

Renshaw et al. (1996) adapts Gompertz-Makeham functions to the new situation by including time as a variable. The model built for  $q_{x,t}$  from the Renshaw et al. (1996) proposal is as follows,

$$\log \left( \frac{q_{x,t}}{1 - q_{x,t}} \right) = \beta_0 + \sum_{j=1}^s \beta_j L_j(x') + \sum_{i=1}^r \left[ \left( \alpha_i + \sum_{j=1}^s \gamma_{ij} L_j(x') \right) t'^i \right], \quad (2.2)$$

subject to the convention that some of the  $\gamma_{ij}$  terms may be zero. In (2.2),  $x'$  and  $t'$  are transformations of age and time, respectively, in such a way that their values are within the interval  $[-1, 1]$  and  $L_j(x')$  are Legendre's Polynomials generated by,

$$L_{n+1}(x) = xL_n(x) - nL_{n-1}(x),$$

where  $n \geq 1$ ,  $L_0(x) = 1$  and  $L_1(x) = x$ .

If the equation(2.2) is rewritten as

$$\frac{q_{x,t}}{1 - q_{x,t}} = \left\{ \exp \left[ \beta_0 + \sum_{j=1}^s \beta_j L_j(x') \right] \right\} \left\{ \exp \left[ \sum_{i=1}^r \left( \alpha_i + \sum_{j=1}^s \gamma_{ij} L_j(x') \right) t'^i \right] \right\},$$

the first term can be interpreted as a Gompertz-Makeham function  $LGM(0, s+1)$  that corresponds to the graduation. The second term can be interpreted as time effect adjustment, in such a way that when at least one of the  $\gamma_{ij}$  is not null, it also depends on age.

### 2.3 Geostatistical Methods

The mortality data we want to analyze can be considered as a set of spatio-temporal data, with the *age* as the spatial unidimensional component and the *year* as the temporal one. In the following we use the methodology described by Cressie and Majure (1997).

Let  $Z(x, t)$  denote the mortality measure at age  $x$  and time  $t$ . We suppose that we have the data  $\{Z(x, t), x \in D, t \in T\}$ , and we wish to predict the process value  $Z(x_0, t_0)$  at the age  $x_0 \in D$  for the year  $t_0$ .

In order to achieve stationarity, we assume that the data can be decomposed into deterministic large-scale variation (trend) plus stochastic small-scale variation (error). The model can be written as,

$$Z(x, t) = \mu(x, t) + \delta(x, t) \quad (2.3)$$

where  $E(Z(x, t)) = \mu(x, t)$  and  $\delta(\cdot, \cdot)$  is a zero-mean intrinsically stationary Gaussian process with variogram  $2\gamma(h, u) = \text{Var}(Z(x+h, t+u) - Z(x, t)) = \text{Var}(\delta(x+h, t+u) - \delta(x, t))$  that characterizes the spatial and temporal dependence. The large-scale variation  $\mu(x, t)$  and the small-scale variation  $\delta(\cdot, \cdot)$  are modelled respectively as deterministic and stochastic processes, but there is no way of making the decomposition identifiable.

The mortality data in a dynamic life table can be viewed as a two-way table. In fact, if the observations were independent a two factors (age and year) ANOVA could be applied in order to analyze them. An alternative is to consider this two-way table as a grid equally spaced in either the vertical (age) or horizontal (year) direction. In this context, the deterministic component in (2.3),  $\mu(x, t)$ , is expressed as the sum

$$\mu(x, t) = \mu + r_x + c_t, \quad (2.4)$$

where  $\mu$  is an overall effect,  $r_x$  is a row effect due to age and  $c_t$  is a column effect due to year. Their ordinary-least-square estimators are the corresponding sample means, but some problems of bias can arise when estimating the variogram. These problems can be overcome using sample medians. A median-polish algorithm producing overall effect,  $\tilde{\mu}$ , row effects,  $\tilde{r}_x$ ,  $x \in D$ , and column effects,  $\tilde{c}_t$ ,  $t \in T$  can be found in Cressie (1993).

Then, from (2.3) and (2.4) the error process  $\delta(x, t)$  can be estimated by

means of

$$\hat{\delta}(x, t) = Z(x, t) - \bar{\mu}(x, t), \quad (2.5)$$

and then the variogram  $2\bar{\gamma}(h, u)$  can be estimated from the robust estimator (Cressie (1993), sec.2.4),

$$2\bar{\gamma}(h, u) = \frac{\left\{ \frac{1}{|N(h, u)|} \sum_{|N(h, u)|} |\hat{\delta}(x_i, t_i) - \hat{\delta}(x_j, t_j)|^{1/2} \right\}^4}{0,457 + \frac{0,494}{|N(h, u)|}} \quad (2.6)$$

where,

$$N(h, u) = \{(x_i, t_i), (x_j, t_j), x_i - x_j = h, t_i - t_j = u\} \quad (2.7)$$

and  $|N(h, u)|$  is the number of distinct pairs in  $N(h, u)$ .

A valid model of variogram is fitted to the empirical variogram using the weighted least squares criterion. Once this is done, we can estimate the original process value  $Z(x_0, t_0)$  with the large-scale process estimated value  $\bar{\mu}(x_0, t_0)$  (2.4) plus the error process estimated value  $\hat{\delta}(x_0, t_0)$  (2.5), obtaining an original process value prediction  $\hat{Z}(x_0, t_0)$  and its mean-squared prediction error.

For the prediction of  $\hat{\delta}(x_0, t_0)$  from the  $n$  data  $\{\hat{\delta}(x, t), x \in D, t \in T\}$ , we used the ordinary kriging approach. This method use a linear predictor,

$$\hat{p}_\delta(x_0, t_0) = \sum_{i=1}^n \lambda_i \hat{\delta}(x_i, t_i) = \boldsymbol{\lambda}' \hat{\boldsymbol{\delta}} \quad (2.8)$$

with  $\boldsymbol{\lambda}' \mathbf{1} = \sum_{i=1}^n \lambda_i = 1$  to guarantee uniform unbiasedness.

The  $\boldsymbol{\lambda}$  values that minimize the mean-squared prediction error subject to this constraint are,

$$\boldsymbol{\lambda}'_{ok} = \left( \boldsymbol{\gamma} + \mathbf{1} \frac{\mathbf{1}' \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}}{\mathbf{1}' \boldsymbol{\Gamma}^{-1} \mathbf{1}} \right)' \boldsymbol{\Gamma}^{-1} \quad (2.9)$$

where  $\boldsymbol{\gamma} = (\gamma(|x_0 - x_1|, |t_0 - t_1|) \dots \gamma(|x_0 - x_n|, |t_0 - t_n|))'$ , and  $\boldsymbol{\Gamma}$  is the  $n \times n$  matrix whose  $(i, j)$ -th element is  $\gamma(|x_i - x_j|, |t_i - t_j|)$ .

The minimized mean-squared prediction error (or kriging variance) is given by,

$$\sigma_{ok}^2(x_0, t_0) = \gamma' \Gamma^{-1} \gamma - \frac{(\mathbf{1}' \Gamma^{-1} \gamma - 1)^2}{\mathbf{1}' \Gamma^{-1} \mathbf{1}}. \quad (2.10)$$

Finally, the optimal prediction for the original process value is,

$$\hat{p}_z(x_0, t_0) = \hat{Z}(x_0, t_0) = \bar{\mu}(x_0, t_0) + \hat{p}_\delta(x_0, t_0) \quad (2.11)$$

and its standard error,

$$\sigma_z(x_0, t_0) = \sigma_{ok}(x_0, t_0).$$

When we want to extrapolate for a year  $t_0$  beyond the grid of observations, the expression for  $\bar{\mu}(x, t_0)$  is

$$\bar{\mu}(x, t_0) = \bar{\mu} + \bar{r}_x + \bar{c}_{t_{max}} + (t_{max} - t_0)(\bar{c}_{t_{max}} - \bar{c}_{(t_{max}-1)}), \quad \forall x \in D. \quad (2.12)$$

### 2.3.1 Data transformation

The mortality data, proportions or probabilities, do not satisfy the normality and stationarity conditions and a transformation of the original process is needed for achieving them. When this occurs, and the transformation is

$$Y(x, t) = \phi(Z(x, t)),$$

with  $\phi^{-1}$  twice-differentiable, a bias corrected optimal prediction  $\hat{Z}(x_0, t_0)$  is given by

$$\hat{p}_z(x_0, t_0) = \phi^{-1}(\hat{p}_y(x_0, t_0)) + (\phi^{-1})''(\bar{\mu}_y(x_0, t_0)) \{ \sigma_y^2(x_0, t_0) / 2 - m_y \}, \quad (2.13)$$

where  $m_y$  are the Lagrange multiplier of the ordinary kriging equation,

$$m_y = - \frac{1 - \mathbf{1}' \Gamma^{-1} \gamma}{\mathbf{1}' \Gamma^{-1} \mathbf{1}}.$$

The mean-square prediction error is then approximated by

$$\sigma_z^2(x_0, t_0) = \{ (\phi^{-1})'(\bar{\mu}_y(x_0, t_0)) \}^2 \sigma_y^2(x_0, t_0).$$

### 3 Analysis of Mortality Data from the Comunidad Valenciana

#### 3.1 Model Adjustment

The three methods described in Section 3 have been used to adjust mortality data for men and women separately, from the Comunidad Valenciana corresponding to the period 1980-1999. The Comunidad Valenciana is a region located on the east coast of Spain and is made up of three provinces, which from north to south are Castellón, Valencia and Alicante. The population of the region was 4,162,776 in the last census carried out in 2001. As insurance policies are not taken on at ages below 18, we have only considered the population of 18 to 96 years old. Moreover, the necessity of introducing dynamism in the analysis of mortality data is justified by the research carried out by the authors of this paper in a previous study Debón (2004).

The crude estimates of  $q_{x,t}$ , necessary for the three methods under study, have been obtained with the process used by the Instituto Nacional de Estadística (INE, Spanish National Institute of Statistics),

$$\hat{q}_{x,t} = \frac{1/2(D_{x,t} + D_{x,t+1})}{P_{x,t} + 1/2D_{x,t}}, \quad (3.1)$$

where  $D_{x,t}$  are deaths in the year  $t$  at age  $x$ ,  $D_{x,t+1}$  are deaths in the year  $t + 1$  at age  $x$ , and  $P_{x,t}$  population that on December 31 of year  $t$  are  $x$  years old.

##### 3.1.1 Lee-Carter Model

The high number of parameters estimated in this model,  $79 \times 2 + 20 = 178$  for men and a similar number for women, dissuade us from presenting them. Instead, they are presented in the form of a graph in Figure 1.

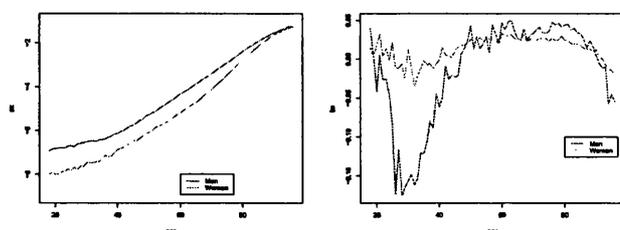


Figure 1.- Estimated values for  $a_x$  and  $b_x$ .

Estimate values for  $k_t$  are displayed in Figure 2. As far as its modelling by means of temporal series is concerned, the series of first differences is seen to be white noise in both sexes. The adjusted models are

$$k_t = -0.5257 + k_{t-1} + a_t \quad \text{y} \quad k_t = -1.0022 + k_{t-1} + a_t$$

for men and women respectively.

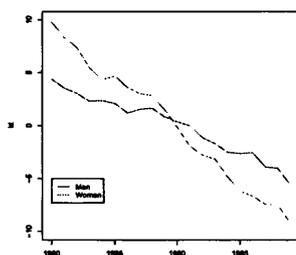


Figure 2.- Estimated values for  $k_t$ .

### 3.1.2 Gompertz-Makeham Functions

The number of parameters in (2.2) is the trade-off point between the goodness-of-fit and the least complexity possible. One criterion to this end consists in studying the improvement in Deviance for successive increases

in  $r$  and  $s$  and comparing them to the value of a  $\chi_1^2$ . We will choose those values above which the increase is no longer significant. By applying this criterion, the model chosen is the same for both sexes

$$\text{logit}(q_{x,t}) = \beta_0 + \sum_{j=1}^9 \beta_j L_j(x') + \alpha_1 t' + \sum_{j=1}^3 \gamma_{1j} L_j(x') t'. \quad (3.2)$$

The coefficient estimates for both sexes are displayed in the Table 1, all of them being significant with the exception of  $\hat{\beta}_8$  for men.

	Men			Women		
	coef	std error	t-value	coef	std error	t-value
$\beta_0$	-4.2994	0.0030	-1411.698	-4.9895	0.0045	-1092.509
$\beta_1$	3.2354	0.0059	544.941	3.7458	0.0088	422.760
$\beta_2$	0.5153	0.0066	78.048	0.6700	0.0095	70.152
$\beta_3$	-0.2449	0.0073	-33.190	-0.1064	0.0105	-10.097
$\beta_4$	-0.0377	0.0078	-4.833	-0.0789	0.0112	-6.991
$\beta_5$	0.0361	0.0081	4.431	-0.1053	0.0119	-8.821
$\beta_6$	-0.0737	0.0085	-8.594	-0.0317	0.0125	-2.519
$\beta_7$	0.0006	0.0087	0.073	0.0350	0.0131	2.659
$\beta_8$	0.0419	0.0082	5.102	0.0355	0.0122	2.892
$\beta_9$	-0.0146	0.0080	-1.817	-0.0187	0.0096	-1.935
$\alpha_1$	-0.0232	0.0048	-4.781	-0.1176	0.0074	-15.740
$\gamma_{11}$	-0.0715	0.0093	-7.633	-0.0480	0.0145	-3.310
$\gamma_{12}$	0.0878	0.0087	10.014	0.1369	0.0133	10.254
$\gamma_{13}$	0.1584	0.0095	16.575	0.1455	0.0113	12.790

Table 1: Gompertz-Makeham Function Coefficients

### 3.1.3 Geostatistical Model

Instead of working with crude estimates,  $\hat{q}_{x,t}$ , we have used  $Z(x,t) = \text{logit}(\hat{q}_{x,t})$  to reduce the influence of exceptional observations and to achieve the normality and stationarity.

Figure 3 shows the estimate in the scale  $q_{x,t}$  of the trend,  $\hat{\mu}(x,t)$ , obtained by the median-polish algorithm. Mortality is seen to grow with age, logically, but decreases with respect to the year and is more pronounced in the case of women.

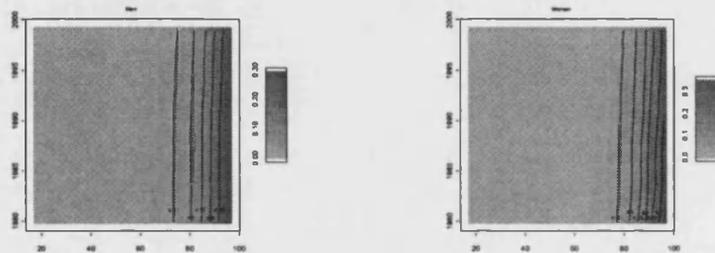
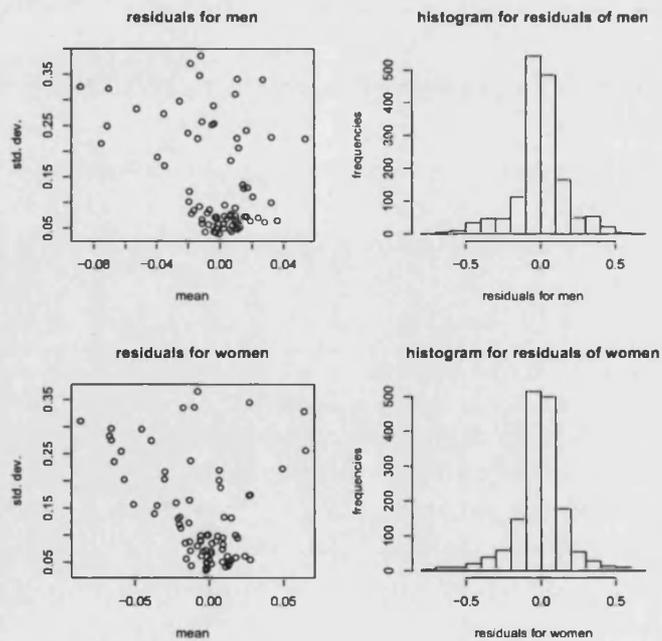


Figure 3.- Estimation of  $\hat{\mu}(x, t)$  for men (left) and women (right).

Once the trend has been estimated, (2.5) allows the residuals  $\hat{\delta}(x, t)$  to be obtained. Figure 4 summarized the exploratory analysis of these values for men and women, from the figure the stationarity and normality hypothesis can be accepted as there is not trend when standard deviation for each age is plot against its mean and the histograms of residuals shows a normal behaviour.



**Figure 4.-** Exploratory analysis of residuals.

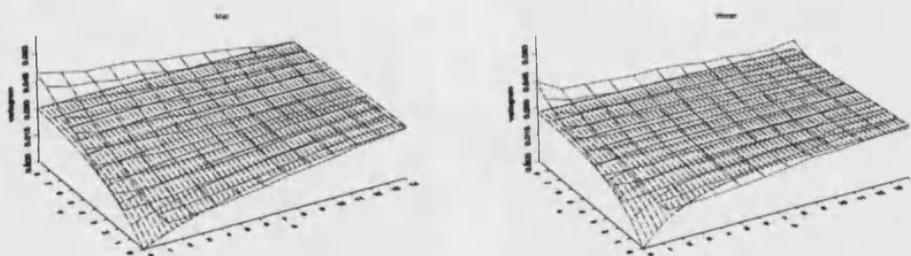
We use these residuals to estimate the empirical variogram. The valid model of variogram fitted to the empirical one is an exponential model, a separable variogram expressed by

$$\gamma(h, u; \theta) = \begin{cases} 0, & \text{if } h = u = 0, \\ c_0 + \sigma^2 (1 - \exp\{-\frac{h}{a} - \frac{u}{b}\}), & \text{if } h \neq 0, \text{ or } u \neq 0, \end{cases}$$

where  $\theta = (c_0, \sigma^2, a, b)'$  with  $c_0 \geq 0$  (nugget effect),  $\sigma^2 > 0$  (partial sill),  $a \geq 0$  (age scale parameter) and  $b \geq 0$  (year scale parameter). The adjusted values for these parameters were

	$c_0$	$\sigma^2$	a	b
Men	0	0.0327	7.0238	3.7919
Women	0	0.0277	2.2467	2.2717

Table 2: Variogram parameter estimations for men and women

**Figure 4.-** Empirical and adjusted variograms.

### 3.2 Predictions for year 2000

In order to measure the goodness-of-fit of each method and compare them, the predictions of  $q_{x,2000}$  corresponding to 2000 have been obtained. The

crude estimates,  $\hat{q}_{x,t}$ , are known and we also have mortality data for the year 2001. A summary of these predictions is presented in graph form in Figure 5, where both, predictions and crude estimates, have been presented together in order to better appreciate the accuracy of each method's predictions.

The MSE are a numerical summary of the accuracy of the predictions, which are displayed in Table 3

	Lee-Carter	Gompertz-Makeham	M-P kriging
Men	0.0270	0.0144	0.0089
Women	0.0148	0.0099	0.0053

Table 3: MSE for  $q_{x,2000}$  predictions

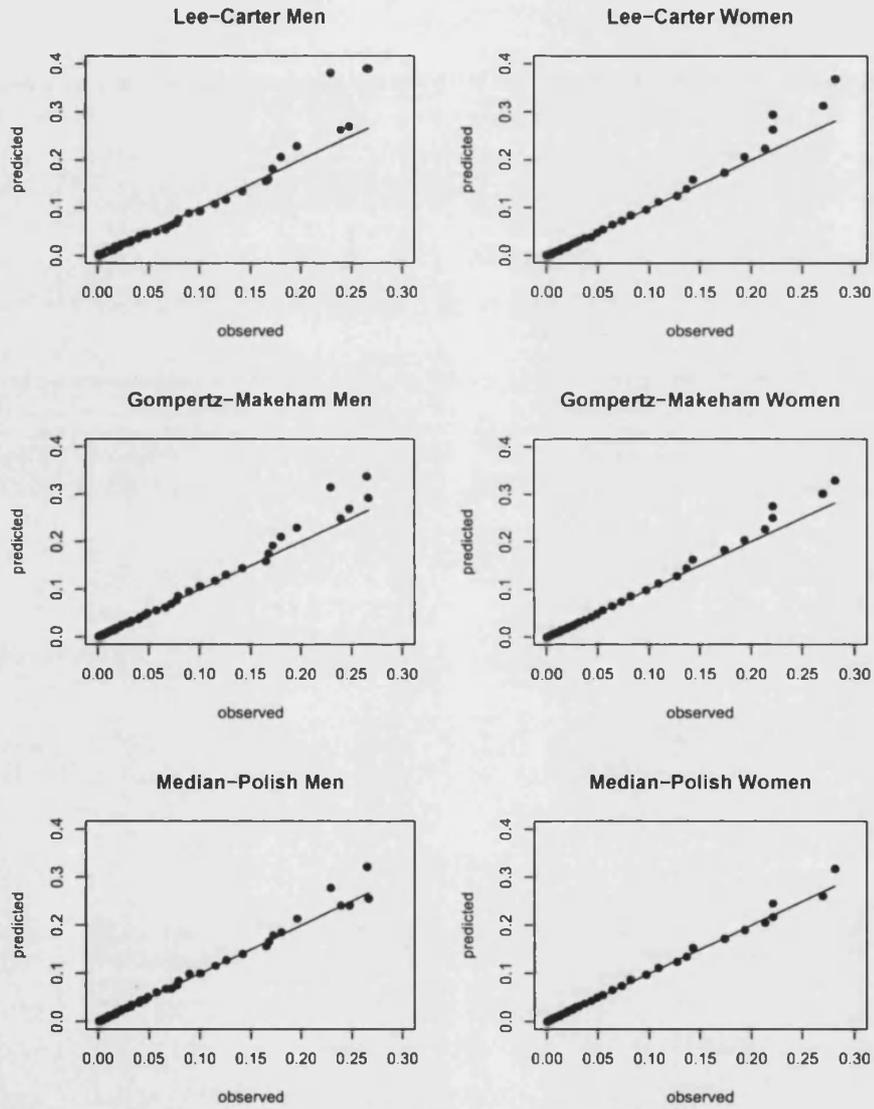


Figure 5.- Predicted vs observed values for each model.

#### 4 Conclusions

Figure 5 and the MSE table show that the predictions made by median-polish kriging are better than those provided by the other two methods, both in the case of men and women. As far as the Lee-Carter method is concerned, the decrease in MSE is 66,87% for men and 64,13% for women. These percentages are 38,06% and 46,22% respectively, when compared to the Gompertz-Makeham method.

While there is an important decrease in error achieved with the model (2.3), which combines the trend adjustment by means of a median-polish algorithm and the modelling of residuals by means of spatio-temporal variogram, more sophisticated spatio-temporal methods, such as those developed by Huang and Cressie (1996) and Mardia et al. (1998) would supposedly capture the momentary progress of the phenomenon better and should improve on the results obtained up to now.

#### References

- Benjamin, B. and Soliman, A. (1995). *Mortality on the Move*. Actuarial Education Service.
- Carter, L. and Lee, R. (1992). Modeling and forecasting u.s. sex differentials in mortality. *International Journal of Forecasting*, 8(3):393–411.
- Cressie, N. (1993). *Statistics for Spatial Data, Revised Edition*. John Wiley, New York.
- Cressie, N. and Majure, J. (1997). *Spatio-temporal Statistical Modelling of Livestock Waste in Streams*. *J. of Agricultural, Biological, and Environmental Statistics*, 2(1):24–47.
- Debón, A. (2004). Verificación de la evolución de la mortalidad a través de los años. *Recta*, 5:65–82.
- Forfar, D., McCutcheon, J., and Wilkie, A. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115 part I(459):1–149.

- Gavin, J., Haberman, S., and Verrall, R. (1993). Moving weighted average graduation using kernel estimation. *Mathematics and Economics*, 12(2):113–126.
- Gavin, J., Haberman, S., and Verrall, R. (1994). On the choice of bandwidth for kernel graduation. *Journal of the Institute of Actuaries*, 121:119–134.
- Gavin, J., Haberman, S., and Verrall, R. (1995). Graduation by kernel and adaptive kernel methods with a boundary correction. *Transactions. Society of Actuaries*, XLVII:173–209.
- Huang, H.C. and Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis*, 22:159–175.
- Lee, R. (2000). The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, 4(1):80–91.
- Lee, R. and Carter, L. (1992). Modelling and forecasting u.s. mortality. *Journal of the American Statistical Association*, 87(419):659–671.
- Mardia, K.V., Goodall, C., Redfern, E.J. and Alonso, F.J. (1998). The Kriged Kalman filter. *Test*, 7(2):217–286.
- Mateu, J., Montes, F. and Plaza, M. (2004). The 1970 US Draft Lottery Revisited: a Spatial Analysis. *JRSS Series C (Applied Statistics)*, 53:1–11.
- McNown, R. and Rogers, A. (1989). Forecasting mortality: A parametrized time series approach. *Demography*, 26(4):645–660.
- McNown, R. and Rogers, A. (1992). Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, 8:413–432.
- Renshaw, A., Haberman, S., and Hatzopoulos, P. (1996). The modelling of recent mortality in united kingdom male assured lives. *British Actuarial Journal*, 2(II):449–477.
- Renshaw, E. (1991). Actuarial graduation practice and generalised. *Journal of the Institute of Actuaries*, 118(II):295–312.

Wilmoth, J. (1993). Computational methods for fitting and extrapolating the lee-carter model of mortality change. *Technical Report, Department of Demography, University of California, Berkeley.*

## Capítulo 7

# Resumen y futuras líneas de investigación

### 7.1. Resumen

Dos son las aportaciones que queremos destacar de nuestro trabajo. La primera, recogida en la Sección 4.12, aporta un nuevo modelo de función de covarianza que hemos denominado modelo de suma de productos generalizado, que genera modelos espacio-temporales a partir de combinaciones lineales de modelos espaciales y temporales. Hay que señalar como rasgo distintivo respecto de los modelos existentes, el hecho de permitir que la covarianza del modelo resultante tome valores negativos o valores que oscilen entre negativos y positivos, lo que resulta muy útil en determinadas aplicaciones prácticas (Mateu et al., 2007b).

La segunda aportación es el estudio de simulación descrito en el Capítulo 5. Se trata de un exhaustivo estudio de simulación que compara la capacidad predictiva de un conjunto de modelos de función de covarianza, descritos en el Capítulo 4, y que cubren la mayoría de situaciones que podemos encontrar en la práctica. Este estudio es una ampliación, al incluir el nuevo modelo de suma de productos generalizado, del que hemos publicado recientemente (Huang et al., 2007) y que se reproduce en el Capítulo 6. Las conclusiones extraídas del estudio podemos resumirlas de la siguiente forma:

1. Desde el punto de vista de modelizar en la práctica un conjunto de observaciones provenientes de un fenómeno espacio-temporal, sugerimos como primera acción algún tipo de análisis exploratorio de los datos, con el fin de comprobar su estacionariedad espacial y temporal.
2. Conseguida ésta, si fuera necesario mediante la eliminación de la tendencia a gran escala, caso de estar presente, la siguiente etapa consiste en contrastar la separabilidad de la función de covarianza espacio-temporal. Para ello puede bastar una simple inspección visual de la covarianza empírica o, más formalmente, se puede utilizar alguno de los test citados en la Sección 4.4.

3. Si la separabilidad del proceso es aceptable, los modelos espacio-temporales separables y flexibles, como el modelo dinámico basado en el filtro de Kalman utilizado en nuestro estudio de simulación, son una buena elección probando diferentes modelos de covarianzas espaciales y temporales con el fin de capturar mejor el comportamiento del proceso. Si un modelo separable no fuera apropiado, conviene recurrir a alguno de los cuatro restantes, utilizando los criterios *AIC* o *BIC* para elegir el modelo final, si bien el modelo suma-producto y particularmente el modelo suma de productos generalizado son los que mejor comportamiento han mostrado.
4. En términos de tiempo de computación y de capacidad de manejo de grandes conjunto de datos, los modelos espacio-temporales basados en covarianzas separables son la opción más adecuada. Sin embargo, si se sospecha que la no separabilidad está presente, el modelo de Cressie y Huang utilizado en nuestro estudio de simulación ofrece una estructura no separable, flexible y de fácil cómputo. No obstante, si el tiempo de cálculo no es problema porque, como ha sido nuestro caso, se tiene acceso a un máquina de cálculo intensivo, modelos más complejos, como el modelo suma-producto y el modelo suma de productos generalizado, ofrecen mejores resultados y se adaptan a una mayor diversidad de situaciones.

No podemos caer en la petulancia de pensar que hemos culminado nuestro trabajo y agotado el objetivo que nos planteamos al comienzo de nuestra investigación, semejante actitud iría en contra de la esencia misma de lo que supone la tarea de investigar. Sí podemos decir que aquel objetivo se ha visto razonablemente cumplido. Hemos aportado un nuevo modelo de función de covarianza espacio-temporal y hemos llevado a cabo un estudio comparativo de los distintos modelos con la mayor exhaustividad posible. No obstante, y volvemos a la frase inicial de este párrafo, mucho queda por hacer y una parte de ello, que nos gustaría desarrollar, se expone a continuación.

## 7.2. Futuras líneas de trabajo

1. Desde el punto de vista teórico, parece interesante proseguir con el estudio de las propiedades del modelo suma de productos generalizado, comprobar su bondad a la hora de ajustar datos reales y sus ventajas cuando el fenómeno analizado requiera covarianzas negativas u oscilantes.
2. Las dificultades computacionales que aparecen en el ajuste de los modelos espacio-temporales, especialmente cuando el conjunto de datos manejado es grande, exigen un esfuerzo investigador en la dirección de encontrar métodos de ajuste que supongan un equilibrio entre el tiempo de computación y la pérdida de eficiencia. La aproximación basada en la verosimilitud compuesta ponderada (weighted composite likelihood) que los compañeros de la Universitat Jaume I están desarrollando parece una idea prometedora.
3. El trabajo de simulación del Capítulo 5, aun cuando ha supuesto una larga dedicación, debido al coste computacional entre otras razones, dista mucho de ser todo lo

exhaustivo que hubiéramos deseado. Las memorias de doctorado deben terminar en algún momento y no agotan, afortunadamente, el tema sobre el que versan. Un futuro, pero inmediato trabajo, será completar este estudio de simulación permitiendo mayor variabilidad en los parámetros y simulando estructuras con un mayor número de puntos.

4. Hemos de insistir, como se ha señalado en las conclusiones anteriores, lo poco que sabemos acerca del comportamiento y propiedades asintóticas de los criterios *AIC* y *BIC* en el contexto espacial. Esto también requiere la atención de una futura línea de investigación.
5. En el campo de las aplicaciones queremos destacar la emprendida con la modelización de los residuos en las tablas de mortalidad dinámicas. Un trabajo preliminar se recoge en Debón et al. (2007), pero constituye tan sólo una primera incursión en un problema que preocupa al mundo actuarial, porque los métodos clásicos de graduación de la mortalidad suponen independencia de los ratios de mortalidad entre edades y años de observación. Se impone por tanto una continuación de este trabajo que incorpore diferentes modelos de covarianza espacio-temporales (edad-año) y compare los resultados con los obtenidos con otros métodos. Particularmente importante es la predicción de la mortalidad futura y de la esperanza de vida, por ser factores decisivos a la hora de valorar las primas de los seguros de vida.



# Bibliografía

- Abramowitz, M. & Stegun, I.A. *Handbook of Mathematical Functions*. Dover, New York, 1965.
- Akaike, H. Information theory and the maximum likelihood principle. En *International Symposium on Information Theory*, Akademiai Kiado, Budapest, 1973.
- Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723, 1974.
- Anderson, T.W. *An introduction to multivariate statistical analysis. Second edition*. John Wiley and Sons Inc., New York, 1984.
- Andrieu, H., Creuting, J.D., Delrieu, J., Leousof, J. & Pointin, Y. Radar data processing for hydrology in the Cevennes region. Remote Sensing and Large-Scale Global Processes. En *Proceedings of the IAHS Third Int. Assembly*, Baltimore, MD, 1989.
- Angulo, J.M., Ruiz-Medina, M.D., Anh, V.V. & Grecksch, W. Fractional diffusion and fractional heat equation. *Advances in Applied Probability*, **32**, 1077–1099, 2000.
- Anh, V.V. & Leonenko, N.N. Spectral analysis of fractional kinetic equations with random data. *Journal of Statistical Physics*, **104**, 1349–1387, 2001.
- Anh, V.V. & Leonenko, N.N. Renormalization and homogeneization of fractional diffusion equations with random data. *Probability Theory and Related Fields*, **124**, 381–408, 2002.
- Armador, J.L., Murillo, J.M. & Rodríguez, L. Recarga artificial en el acuífero de Jávea mediante el aprovechamiento de los excedentes hídricos procedentes del río Jalón-Gorgos (Alicante). En *V Congreso Geológico de España*, Alicante, 2000.
- Armstrong, M., Chetboun, G. & Hubert, P. Kriging the rainfall in Lesotho. En *Geostatistics Tróia '92*, **2**, 661–672, Kluwer Academic Publ., Dordrecht, 1993.
- Atkinson, P.M., Foody, G.M., Curran, P.J. & Boyd, D.S. Assessing the ground data requirements for regional scale remote sensing of tropical forest biophysical properties. *International Journal of Remote Sensing*, **21**, 2571–2887, 2000.
- Atteia, O., Dubois, J.P. & Webster, R. Geostatistical analysis of soil contamination in the Swiss Jura. *Environmental Pollution*, **86**, 315–327, 1994.

- Azimi-Zonooz, A., Krajewski, W.F., Bowles, D.S. & Seo, D.J. Spatial rainfall estimation by linear and non-linear co-kriging of radar-rainfall and raingauge data. *Stochastic Hydrology and Hydraulics*, **3**, 51–67, 1989.
- Banerjee, S., Carlin, B.P. & Gelfand, A.E. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton-Florida, 2004.
- Bardossy, A., Haberlandt, U. & Grimm-Strele, J. Interpolation of groundwater quality parameters using additional information. En *GeoENV I – Geostatistics for Environmental Applications*, 189–200, Kluwer Academic Publ., Dordrecht, 1997.
- Bartlett, M.S. The approximate recovery of information from field experiments with large blocks. *Journal of Agricultural Science*, **28**, 418–427, 1938.
- Bartlett, M.S. *The statistical analysis of spatial pattern*. Chapman & Hall, London, 1976.
- Bartlett, M.S. Nearest neighbour models in the analysis of field experiments (with discussion). *Journal of the Royal Statistical Society, B*, **40**, 147–175, 1978.
- Benjamin, B. & Soliman, A.S. *Mortality on the Move*. Institute of Actuaries, Oxford, 1993.
- Berke, O. On spatiotemporal prediction for on-line monitoring data. *Communications in Statistics - Theory and Methods*, **27**, 2343–2369, 1998.
- Bernstein, S.N. Sur les fonctions absolument monotones. *Acta Mathematica*, 1–66, 1928.
- Besag, J.E. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, 192–236, 1974.
- Bevilacqua, M., Gaetan, C., Mateu, J. & Porcu, E. Estimating space and space-time covariance functions: a weighted composite likelihood approach. *Technical Report 117*, Universitat Jaume I, 2007.
- Blasco, M.P. *Estudio hidroquímico del acuífero del cuaternario de Jávea (Alicante)*. Tesis Doctoral. Universidad de Alicante. 1988.
- Bochner, S. Monotone Funktionen, Stieltjssche Integrale und harmonische Analyse. *Math. Ann.*, **108**, 378–410, 1933.
- Bodas, A., López, E., Martínez-Ruiz, F., Mateu, J. & Montes, F. Spatio-temporal modeling and prediction of solar radiation. *Journal of Geophysical Research-Atmospheres*, **108**(D24), 8777, 2003.
- Bolle, H.J. Synthesis of Change Detection Parameters into a Land-Surface Change Indicator for Long-Term Desertification Studies in the Mediterranean Area (RESYSMED). *Final Report Contract No. ENV4-CT97-0683*, 1999.
- Bolle, H.J., Eckardt, M., Koslowsky, D., Maselli, F. & Melia J. *Mediterranean Land-surface Processes Assessed from Space*. Springer-Verlag, Berlin, 2006.

- Brown, P.E., Karesen, K.F., Roberts, G.O. & Tonellato, S. Blur-generated non-separable space-time models. *Journal of the Royal Statistic Society, B*, **62**, 847–860, 2000.
- Burgess, T.M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. *Journal of Soil Science*, **31**, 315–331, 1980a.
- Burgess, T.M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *Journal of Soil Science*, **31**, 333–341, 1980b.
- Burgess, T.M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. *Journal of Soil Science*, **31**, 505–524, 1980c.
- Burgess, T.M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science*, **31**, 643–659, 1981.
- Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, **16**, 1190–1208, 1995.
- Carter, L. & Lee, R. Modeling and forecasting U.S. sex differentials in mortality. *International Journal of Forecasting*, **8**, 393–411, 1992.
- Charlock, T.P. & Alberta, T.L. The CERES/ARM/GEWEX Experiment (CAGEX) for the Retrieval of Radiative Fluxes with Satellite Data. *Bulletin of the American Meteorological Society*, **77**, 2673–2683, 1996.
- Chen, L., Fuentes, M. & Davis, J.M. Spatial temporal statistical modelling and prediction of environmental processes. En *Hierarchical Modelling for the Environmental Sciences*, 121–144, Oxford Univ. Press, 2006.
- Chien, Y.J., Lee, D.Y., Guo H.Y. & Houg H.H. Geostatistical analysis of soil properties of mid-west Taiwan soils. *Soil Science*, **162**, 291–298, 1997.
- Chilès, J.P. How to adapt kriging to non-classical problems: the case studies. En *Advanced Geostatistics in the Mining Industry*, 69–90, Springer, 1976.
- Chilès, J.P. & Liao, H.T. Estimating the recoverable reserves of gold deposits: Comparison between disjunctive kriging and indicator kriging. En *Geostatistics Tróia '92*, **2**, 1053–1064, Kluwer, Dordrecht, 1993.
- Chilès, J.P. & Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 1999.
- Christakos, G. & Hristopulos, D.T. *Spatiotemporal environmental health modelling: a tractatus stochasticus*. Kluwer Academic Publ., Boston, 1998.
- Christakos, G. *Modern Spatiotemporal Geostatistics*. Oxford University Press, Oxford, 2000.

- Christakos, G. On a deductive logic-based spatiotemporal random field theory. *Probability Theory and Mathematical Statistics*, **66**, 54–65, 2002.
- Christensen, O.F., Diggle, P.J. & Ribeiro, P.J. Analysing positive valued spatial data: the transformed Gaussian model. En *GeoENV III – Geostatistics for Environmental Applications*, 287–298, Kluwer Academic Publ., Dordrecht, 2001.
- Cliff, A.D. & Ord, J.K. *Spatial Processes: Models & Applications*. Pion, London, 1981.
- Cox, D.R. & Isham, V. A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London, A*, **415**, 317–328, 1988.
- Cressie, N. & Hawkins, D.M. Robust estimation of the variogram. *Mathematical Geology*, **12**, 115–125, 1980.
- Cressie, N. & Horton, R. A robust-resistant spatial analysis of soil water infiltration. *Water Resources Research*, **23**, 911–917, 1987.
- Cressie, N. & Zimmerman, D.L. On the stability of the geostatistical method. *Mathematical Geology*, **24**, 45–59, 1992.
- Cressie, N. *Statistics for Spatial Data, Revised Edition*. John Wiley and Sons Inc., New York, 1993.
- Cressie, N. & Majure, J. Spatio-temporal Statistical Modelling of Livestock Waste in Streams. *Journal of Agricultural, Biological & Environmental Statistics*, **2**, 24–47, 1997.
- Cressie, N. & Huang, H.C. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330–1340, 1999 y **96**, 784, 2001.
- Cressie, N. & Wikle, C.K. Space-time Kalman filter. En *Encyclopedia of Environmetrics*, **4**, 2045–2049, John Wiley and Sons Ltd., Chichester, 2002.
- Cressie, N. & Kornak, J. Spatial statistics in the presence of location error with an application to remote sensing of the environment. *Statistical Science*, **18**, 436–456, 2003.
- Curreiro, F.C. & Lele, S. A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological and Environmental Statistics*, **4**, 9–28, 1999.
- De Cesare, L., Myers, D.E. & Posa, D. Spatial temporal modeling of SO<sub>2</sub> in the Milan district. En *Geostatistics Wollongong '96*, **2**, 1031–1042, Kluwer Academic Publ., Dordrecht, 1997.
- De Cesare, L., Myers, D.E. & Posa, D. Estimating and modeling space-time correlation structures. *Statistics and Probability Letters*, **51**, 9–14, 2001a.
- De Cesare, L., Myers, D.E. & Posa, D. Product-sum covariance for space-time modeling: an environmental application. *Environmetrics*, **12**, 11–23, 2001b.

- De Cesare, L., Myers, D.E. & Posa, D. FORTRAN programs for space-time modeling. *Computers & Geosciences*, **28**, 205–212, 2002.
- De Iaco, S., Myers, D.E. & Posa, D. Space-time analysis using a general product-sum model. *Statistics and Probability Letters*, **52**, 21–28, 2001.
- De Iaco, S., Myers, D.E. & Posa, D. Space-time variograms and a functional form for total air pollution measurements. *Computational Statistics & Data Analysis*, **41**, 311–328, 2002a.
- De Iaco, S., Myers, D.E. & Posa, D. Nonseparable space-time covariance models: some parametric families. *Mathematical Geology*, **34**, 23–42, 2002b.
- De Iaco, S., Myers, D.E. & Posa, D. The linear coregionalization model and the product-sum space-time variogram. *Mathematical Geology*, **35**, 25–38, 2003.
- De Kwaadsteniet, J.W. On some fundamental weak spots of kriging technique and their consequences. *Journal of Hydrology*, **114**, 277–284, 1990.
- Debón, A. Verificación de la evolución de la mortalidad a través de los años. *Recta*, **5**, 65–82, 2004.
- Debón, A., Martínez-Ruiz, F. & Montes, F. Dynamic life tables: a geostatistical approach. Enviado a *Test*, 2007.
- DeGroot, M. *Probability and statistics. Second edición*. Addison-Wesley Publ. Company, 1989.
- Dibiasi, A. & Bowman, A.W. On the use of the variogram in checking for independence in spatial data. *Biometrics*, **57**, 211–218, 2001.
- Diggle, P.J., Tawn, J.A. & Moyeed, R.A. Model-based geostatistics. *Applied Statistics*, **47**, 299–326, 1998.
- Dimitrakopoulos, R. & Luo, X. Spatiotemporal modeling: covariances and ordinary kriging systems. En *Geostatistics for the next century*, Kluwer Academic Publ., 88–93, 1994.
- Dimitrakopoulos, R. & Luo, X. Joint space-time modeling in the presence of trends. En *Geostatistics Wollongong '96*, **1**, 138–149, Kluwer Academic Publ., Dordrecht, 1997.
- Diputación Provincial de Alicante. *Recarga artificial de acuíferos*. Ed. DPA, 1999.
- Dirks, K.N., Hay, J.E., Stow, C.D. & Harris, D. High-resolution studies of rainfall on Norfolk Island. Part II: Interpolation of rainfall data. *Journal of Hydrology*, **208**, 187–193, 1998.
- Ericksen, E.P., Kadane, J.B. & Tukey, J.W. Adjusting the 1980 Census of Population and Housing. *Journal of the American Statistical Association*, **84**, 927–944, 1989.

- Fairfield Smith, H. An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science*, **28**, 1-23, 1938.
- Faraway, J. *Linear Models with R*. Chapman and Hall, London, 2005.
- Férrnandez-Casal, R. *Geostatística Espacio-Temporal. Modelos Flexibles de Variogramas Anisotropicos no Separables*. Tesis doctoral, Santiago de Compostela, 2003.
- Férrnandez-Casal, R., González-Manteiga, W. & Febrero-Bande, M. Flexible spatio-temporal stationary variograms models. *Statistics and Computing*, **13**, 127-136, 2003.
- Flowerdew, R. & Green, M. Developments in Areal Interpolating Methods and GIS. *Annals of Regional Sciences*, **26**, 67-78, 1992.
- Flowerdew, R. & Green, M. Areal Interpolation and Types of Data. *Spatial analysis and GIS*, 121-146, 1994.
- Forfar, D., McCutcheon, J. & Wilkie, A. On graduation by mathematical formula. *Journal of the Institute of Actuaries*, **115**, 1-149, 1988.
- Fuentes, M. A high frequency kriging approach for nonstationary environmental processes. *Environmetrics*, **12**, 469-483, 2001.
- Fuentes, M. & Smith, R. A new class of nonstationary models. *Technical report no. 2534*, North Carolina State University, 2001.
- Fuentes, M. Modeling and prediction of nonstationary spatial processes. *Statistical Modelling: An International Journal*, **2**, 281-298, 2002a.
- Fuentes, M. Spectral methods for nonstationary spatial processes. *Biometrika*, **89**, 197-210, 2002b.
- Fuentes, M. Testing for separability of spatial-temporal covariance functions. *Technical Report no. 2533*, North Carolina State University, 2003.
- Fuentes, M., Chen, L., Davis, J. & Lackmann, G. Modeling and predicting complex space-time structures and patterns of coastal wind fields. *Environmetrics*, **16**, 449-464, 2005.
- Fuentes, M. Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference*, **136**, 447-466, 2005a.
- Gavin, J., Haberman, S. & Verrall, R. Moving weighted average graduation using kernel estimation. *Mathematics and Economics*, **12**, 113-126, 1993.
- Gavin, J., Haberman, S. & Verrall, R. On the choice of bandwidth for kernel graduation. *Journal of the Institute of Actuaries*, **121**, 119-134, 1994.
- Gavin, J., Haberman, S. & Verrall, R. Graduation by kernel and adaptive kernel methods with a boundary correction. *Transactions. Society of Actuaries*, **XLVII**, 173-209, 1995.

- Ghil, M., Cohn, S.E., Tavantzis, J., Bube, K. & Isaakson, E. Applications of estimation theory to numerical weather prediction. En *Dynamic Meteorology: Data Assimilation Methods*, 139–224, Springer-Verlag, New York, 1981.
- Gleckler, P.J. Surface energy balance errors in AGCMs: Implications for ocean-atmosphere model coupling. *Geophysical Research Letters*, **32**, L15708, 2005.
- Gneiting, T. & Sasvári, Z. The characterization problem for isotropic covariance functions. *Mathematical Geology*, **31**, 105–111, 1999.
- Gneiting, T., Sasvári, Z. & Schlather, M. Analogies and correspondences between variograms and covariance functions. *Advances in Applied Probability*, **33**, 617–630, 2001.
- Gneiting, T. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97**, 590–600, 2002.
- Gneiting, T. & Schlather, M. Space-time covariance models. En *Encyclopedia of Environmental Metrics*, 4, 2041–2045, Wiley, Chichester, 2002.
- Gneiting, T. Should weather and climate models be deterministic? En *Joint Statistical Meeting 2003*, San Francisco, California, 2003.
- Gneiting, T., Genton, M.G. & Guttorp, P. Geostatistical space-time models, stationarity, separability, and full symmetry. *Technical Report no. 475*, University of Washington, 2005.
- Godambe, V. An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, **31**, 1208–1212, 1960.
- Goodall, C. & Mardia, K.V. Challenges in multivariate spatio-temporal modeling. En *Proceedings of the XVIIth International Biometric Conference*, Hamilton, Ontario, Canada, 1994.
- Goovaerts, P. Factorial kriging analysis: A useful tool for exploring the structure of multivariate spatial soil information. *Journal of Soil Science*, **43**, 597–619, 1992.
- Goovaerts, P. & Webster, R. Scale-dependent correlation between topsoil copper and cobalt concentration in Scotland. *European Journal of Soil Science*, **45**, 79–95, 1994.
- Goovaerts, P. Study of spatial relationships between two sets of variables using multivariate geostatistics. *Geoderma*, **62**, 93–107, 1994.
- Goovaerts, P. & Journel, A.C. Integrating soil map information in modelling the spatial variation of continuous soil properties. *European Journal of Soil Science*, **46**, 397–414, 1995.
- Goovaerts, P. *Geostatistics for natural resources evaluation*. Oxford University Press, New York, 1997a.

- Goovaerts, P. Kriging vs stochastic simulation for risk analysis in soil contamination. En *GeoENV I - Geostatistics for Environmental Applications*, 247–258, Kluwer Academic Publ., Dordrecht, 1997b.
- Grimes, D.I.F., Coppola, E., Verdecchia, M. & Visconti, G. A neural network approach to real-time rainfall estimation for Africa using satellite data. *Journal of Hydrometeorology*, **4**, 1119–1133, 2003.
- Gupta, S.K, Ritchey, N.A., Wilber, A.C., Whitlock, C.H., Gibson, G.G. & Stackhouse P.W. A climatology of surface radiation budget derived from satellite data. *Journal of Climate*, **12**, 2691–2710, 1999.
- Guttorp, P., Meiring, W. & Sampson, P.D. A space-time analysis of ground-level ozone data. *Environmetrics*, **5**, 241–254, 1994.
- Guttorp, P. & Sampson, P.D. Methods for estimating heterogeneous spatial covariance functions with environmental applications. En *Handbook of Statistics*, **12**, 661–689, Elsevier Science B.V., 1994.
- Haas, T.C. Lognormal and moving window methods of estimating acid deposition. *Journal of the American Statistical Association*, **85**, 950–963, 1990.
- Haas, T.C. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association*, **90**, 1189–1199, 1995.
- Haas, T.C., Van Tooren, & Brettschneider, V. Optimisation of a cleanup using moving window kriging. En *Contaminated Soil '95*, 205–206, Kluwer Academic Press, 1995.
- Harries, J.E., Russell, J.E., Hanafin, J.A., Brindley, H., Futyan, J., Rufus, J., Kellock, S., Matthews, G., Wrigley, R., Last, A., Mueller, J., Mossavati, R., Ashmall, J., Sawyer, E., Parker, D., Caldwell, M., Allan, P.M., Smith, A., Bates, M. J., Coan, B., Stewart, B.C., Lepine, D.R., Cornwall, L.A., Corney, D. R., Ricketts, M.J., Drummond, D., Smart, D., Cutler, R., Dewitte, S., Clerbaux, N., Gonzalez, L., Ipe, A., Bertrand, C., Joukoff, A., Crommelynck, D., Nelms, N., Llewellyn-Jones, D.T., Butcher, G., Smith, G.L., Szewczyk, Z.P., Mlynczak, P.E., Slingo, A., Allan, R.P. & Ringer, M.A. The Geostationary Earth Radiation Budget Project. *Bulletin of the American Meteorological Society*, **86**, 945–960, 2005.
- Higdon, D., Swall, J. & Kern, J. Non-stationary spatial modeling. En *Bayesian Statistics*, 761–768, Oxford University Press, Oxford, 1999.
- Heine, V. Models for two-dimensional stationary stochastic processes. *Biometrika*, **42**, 170–178, 1955.
- Hoeting, J.A., Davis, R.A., Merton, A. & Thompson, S.E. Model selection for geostatistical models. *Ecological Applications*, **16**, 87–98, 2006.

- Holland, D., Saltzman, N., Cox, L.H. & Nychka, D. Spatial prediction of sulfur dioxide in the eastern United States. En *GeoENV II - Geostatistics for Environmental Applications*, 65–76, Kluwer Academic Publ., Dordrecht, 1999.
- Hollmann, R. & Bodas, A. Comparison of the surface solar netflux from ScaRaB with observations. En *IRS 2000: Current Problems in Atmospheric Radiation*, 536–539, A. Deepak Publ., Hampton, Virginia, 2001.
- Huang, H.C. & Cressie, N. Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics & Data Analysis*, **22**, 159–175, 1996.
- Huang, H.C. & Hsu, N. J. Modeling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics*, **15**, 251–268, 2004.
- Huang, H.C., Martínez-Ruiz, F., Mateu, J. & Montes, F. Model comparison and selection for stationary space-time models. *Computational Statistics & Data Analysis*, **51**, 4577–4596, 2007.
- Isaaks, E. & Srisvastava, R. *An introduction to Applied Geostatistics*. Oxford Univ. Press, New York, 1989.
- Jacques, D., Vanderborcht, J., Mallants, D., Mohanty, B.P. & Feyen J. Analysis of solute redistribution in heterogeneous soil: I Geostatistical approach. En *First European Conference on Geostatistics for Environmental Applications*, Lisboa, 1996.
- Jones, R.H. & Zhang, Y. Models for continuous stationary space-time processes. En *Modelling Longitudinal and Spatially Correlated Data*, 289–298, Springer Verlag, New York, 1997.
- Journel, A.G. & Huijbregts, C.J. *Mining Geoestistics*. Academic Press, New York, 1978.
- Juang, K.W. & Lee, D.Y. A comparison of three kriging methods using auxiliary variables in heavy-metal contaminated soils. *Journal of Environmental Quality*, **27**, 355–363, 1998.
- Kalman, R.E. A new approach to linear filtering & prediction problems. *Journal of Basic Engineering (ASME)*, **82**, 35–45, 1960.
- Kanesvskey, M., Arutyuyan, R., Bolshov, L., Chernov, S., Demyanov, V., Linge, I., Koptelova, N. Savelieva, E., Haas, T. & Maignan, M. Chernobyl Fallout: Review of advanced spatial data analysis. En *GeoENV I – Geostatistics for Environmental Applications*, 389–400, Kluwer Academic Publ., Dordrecht, 1997.
- Kitanidis, P. Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*, **19**, 909–921, 1983.
- Kitanidis, P. *Introduction to geostatistics: applications in hydrogeology*. Cambridge University Press, 1997.

- Kolovos, A., Christakos, G., Hristopulos, D.T. & Serre, M.L. Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources*, **27**, 815–830, 2004.
- Koslowsky, D. Mehrjaehrige validierte und homogenisierte Reihen des Reflexionsgrades und des Vegetationsindex von Landoberflaechen aus taeglichen AVHRR-Daten hoher Aufloesung. *Tech. Rep. Neue Folge, Serie A Monographien*, Freie Universität Berlin, Meteorologische Abhandlungen. 1996.
- Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, **52**, 119–139, 1951.
- Kyriakidis, P.C. & Journel, A.G. Geostatistical space-time models: a review. *Mathematical Geology*, **31**, 651–684, 1999 y **32**, 893, 2000.
- Laslett, G.M. & McBratney, A.B. Further comparison of spatial methods for predicting soil pH. *Journal of Soil Science American Society*, **54**, 1553–1558, 1990.
- Lebel, T. & Amani, A. Rainfall Estimation in the Sahel: What Is the Ground Truth. *Journal of Applied Meteorology*, **35**, 555–568, 1999.
- Lee, R. & Carter, L. Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, **87**, 659–671, 1992.
- Lee, R. The Lee-Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, **4**, 80–91, 2000.
- Li, Z., Leighton, H.G., Masuda, K. & Takashima, T. Estimation of SW flux absorbed at the surface from TOA reflected flux. *Journal of Climate*, **6**, 317–330, 1993a.
- Li, Z., Leighton, H.G. & Cess, R.D. Surface net solar radiation estimated from satellite measurements: Comparisons with tower observations. *Journal of Climate*, **6**, 1764–1772, 1993b.
- Li, Z., Whitlock, C.H. & Charlock, T.P. Assessment of the Global Monthly Mean Surface Insolation Estimated from Satellite Measurements Using Global Energy Balance Archive Data. *Journal of Climate*, **8**, 315–328, 1995.
- Lindsay, B. Composite likelihood methods. *Contemporary Mathematics*, **80**, 221–239, 1988.
- López-Baeza, E., Velázquez, A., Antolín, C., Bodas-Salcedo, A., Gimeno, J.F., Saleh, K., Ferrer, F., Domenech, C., Castell, N. & Sánchez, M.A. The Valencia Anchor Station, a Cal/val reference area for large-scale low spatial resolution remote sensing missions. *En 3a Conferencia Internacional sobre Experiencias con Estaciones Meteorológicas Automáticas*, Torremolinos, Málaga, Spain, 2001.

- Ma, C. Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology*, **34**, 965–975, 2002.
- Ma, C. Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference*, **116**, 489–501, 2003a.
- Ma, C. Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, **86**, 97–107, 2003b.
- Ma, C. Nonstationary covariance functions that model space-time interactions. *Statistics and Probability Letters*, **61**, 411–419, 2003c.
- Ma, C. Linear combinations of space-time covariance functions and variograms. *IEEE Trans. Signal Processing*, **53**, 857–864, 2005a.
- Ma, C. Semiparametric spatio-temporal covariance models with the ARMA temporal margin. *Annals of the Institute of Mathematical Statistics*, **57**, 221–233, 2005b.
- Ma, C. Spatio-temporal variograms and covariance models. *Advances in Applied Probability*, **37**, 706–725, 2005c.
- Maglione, D.S. & Diblasi, A. M. Exploring a valid model for the variogram of an isotropic spatial process. En *Proceedings of the Annual Conference of the International Association for Mathematical Geology*, Cancún, Méjico, 2001.
- Maglione, D.S. & Diblasi, A. M. Exploring a valid model for the variogram of an isotropic spatial process. *Stochastic Environmental Research and Risk Assessment*, **18**, 366–376, 2004.
- Mardia, K.V., Goodall, C., Redfern, E.J. & Alonso, F.J. The kriged Kalman filter (with discussion). *Test*, **7**, 217–286, 1998.
- Martínez-Ruiz, F., Mateu, J., Montes, F. & Morell, I. Análisis espacio-temporal del acuífero del cuaternario de Xàbia. *Boletín Geológico y Minero*, **114**, 323–332, 2003.
- Martínez-Ruiz, F., Mateu, J. & Montes, F. A comparative analysis of different spatial sampling esquemes: Modeling of SSRB data. Aceptado en *International Journal of Remote Sensing*, 2007.
- Matérn, B. Spatial Variation. *Meddelanden från Statens Skogsforskningsinstitut*, **49**, 5, Stockholm, Almaenna Foerlaget, 1960.
- Mateu, J., Montes, F. & Fuentes, M. Recent advances in space-time statistics with applications to atmospheric data: An overview. *Journal of Geophysical Research*, **108** (D24), 2003.
- Mateu, J., Montes, F. & Plaza, M. The 1970 US Draft Lottery Revisited: a Spatial Analysis. *Journal of The Royal Statistical Society, C*, **53**, 1–11, 2004.

- Mateu, J., Porcu, E. & Gregori, P. Recent advances to model anisotropic space-time data. Aceptado en *Statistical Methods and Applications*, 2007a.
- Mateu, J., Porcu, Christakos, G. & Bevilacqua, M. Fitting negative spatial covariances to geothermal field temperatures in Nea Kessani (Greece). Aceptado en *Environmetrics*, 2007b.
- Matheron, G. *Traité de géostatistique appliquée*. Editions Technip, Paris, 1962.
- Matheron, G. Principles of geostatistics. *Economic Geology*, **58**, 1246–1266, 1963.
- Matheron, G. The theory of regionalized variables and its applications. *Cahiers du Centre de Morphologie Mathématique*, **5**, Fontainebleau, France, 1971.
- McQuarrie, A.D. & Tsai, C.L. Regression and Time Series Model Selection. *World Scientific Publishing*, Singapore, 1998.
- McNown, R. & Rogers, A. Forecasting mortality: A parametrized time series approach. *Demography*, **26**, 645–660, 1989.
- McNown, R. & Rogers, A. Forecasting cause-specific mortality using time series methods. *International Journal of Forecasting*, **8**, 413–432, 1992.
- Meiring, W., Guttorp, P. & Sampson, P.D. Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. *Environmental and Ecological Statistics*, **5**, 197–222, 1998.
- Miller, K.S. & Samko, S.G. Completely monotonic functions. *Integral Trans. Spec. Funct.*, **12**, 389–402, 2001.
- Mitchell, M.W., Genton, M.G. & Gumpertz, M.L. Testing for separability of space-time covariances. *Environmetrics*, **16**, 819–831, 2005.
- Mitchell, M.W., Genton, M.G. & Gumpertz, M.L. A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis*, **97**, 1025–1043, 2006.
- Mohammadian Mosamam, A. & Kent, J.T. Asymmetric space-time models. En *Interdisciplinary Statistics and Bioinformatics*, 138–141, Leeds University Press, Leeds, 2006.
- Mugglin, A.S., Carlin, B.P. & Gelfand, A.E. Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association*, **95**, 877–887, 2000.
- Myers, D.E. & Journel, A.G. Variograms with zonal anisotropies and noninvertible kriging systems. *Mathematical Geology*, **22**, 779–785, 1990.
- Myers, D.E., De Iaco, S., Posa, D. & De Cesare, L. Space-Time Radial Basis Functions. *Computers and Mathematics with Applications*, **43**, 539–549, 2002.
- Nychka, D. & Saltzman, N. Design of air quality networks. En *Case Studies in Environmental Statistics*, 51–76, Springer Verlag, New York, 1998.

- Ohmura, A., Gilgen, H., Hegner, H., Muller, G., Wild, M., Dutton, E., Forgan, B., Frohlich, C., Philocon, R., Heimo, A., Konig-Langlo, G., McArthur, B., Pinker, B., Whitlock, C. & Dehne, K. Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, **79**, 2115–2136, 1998.
- Oliver, M.A. Geostatistics and its application to soil science. *Soil Use and Management*, **3**, 8–19, 1987.
- Papadakis, J.S. Méthode statistique pour les expériences du champ. *Bull. Inst. Amél. Plantes à Salonique*, **23**, 1937.
- Papritz, A. & Moyeed, R.A. Linear and non-linear kriging methods: Tools for monitoring soil pollution. En *Statistics for the Environment*, **4**, 303–336, Wiley, 1998.
- Pardo-Iguzguiza, E. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*, **210**, 206–220, 1998.
- Persicani, D. Evaluation of soil classification and kriging for mapping herbicide leaching simulated by two models. *Soil Technology*, **8**, 17–30, 1995.
- Price, J.C. Timing of NOAA afternoon passes. *International Journal of Remote Sensing*, **12**, 193–198, 1991.
- Porcu, E., Gregori, P. & Mateu, J. Nonseparable stationary anisotropic space-time covariance functions. *Stochastic Environmental Research and Risk Assessment*, **21**, 113–122, 2006.
- Porcu, E., Mateu, J., Zini, A. & Pini, R. Modelling spatio-temporal data: A new variogram and covariance structure proposal. *Statistics and Probability Letters*, **77**, 83–89, 2007.
- Porcu, E. & Mateu, J. Mixture-based modeling for space-time data. *Environmetrics*, **18**, 285–302, 2007.
- Pulido-Bosch, A. Salinización y recarga artificial en el acuífero de Jávea (Alicante). *Hydrogeología y Recursos Hidráulicos*, **2**, 772–787, 1976.
- Renshaw, A.E. Actuarial graduation practice and generalised. *Journal of the Institute of Actuaries*, **118**, 295–312, 1991.
- Renshaw, A.E., Haberman, S. & Hatzopoulos, P. The modelling of recent mortality in United Kingdom male assured lives. *British Actuarial Journal*, **2**, 449–477, 1996.
- Renshaw, E. The linear spatio-temporal interaction process and its relation to  $1/\omega$ -noise. *Journal of the Royal Statistical Society, B*, **56**, 75–91, 1994.
- Renshaw, E. Two-dimensional spectral analysis for marked point processes. *Biometrical Journal*, **44**, 718–745, 2002.

- Riesz, F. Über Sätze von Stone und Bochner. *Acta Sci. Math.*, **6**, 184–198, 1934.
- Ripley, B.D. *Spatial Statistics*. Wiley, New York, 1981.
- Rodriguez-Iturbe, I. & Mejia, J.M. The design of rainfall networks in time and space. *Water Resources Research*, **10**, 713–728, 1974.
- Rouhani, S. & Hall, T.J. Space-time kriging of groundwater data. En *Geostatistics*, **2**, 639–650, Kluwer Academic Publ., Dordrecht, 1989.
- Rouhani, S. & Myers, D.E. Problems in space-time kriging of hydrogeological data. *Mathematical Geology*, **22**, 611–623, 1990.
- Russo, D., Zaidel, J. & Laufer, A. Stochastic analysis of solute transport in partially saturated heterogeneous soil. I. Numerical experiments. *Water Resources Research*, **30**, 769–779, 1994.
- Sahu, S.K. & Mardia, K.V. Recent Trends in Modeling Spatio-Temporal Data. En *Proceedings of the special meeting on Statistics and Environment*, Messina, Italy, 2005.
- Sampson, P.D. & Guttorp, P. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of American Statistical Association*, **87**, 108–119, 1992.
- Savaliev, A.A., Mucharamova, S.S. & Piliugin, G.A. Modeling of the daily rainfall values using surface under tension and kriging. *Journal of Geographic Information and Decision Analysis*, **2**, 1998.
- Schlather, M. Introduction to positive definite functions and to unconditional simulation of random fields. *Technical report ST 99-10*, Lancaster University, 1999.
- Schoenberg, I.J. Metric spaces and completely monotone functions. *Annals of Mathematics*, **39**, 811–841, 1938.
- Schwarz, G. Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464, 1978.
- Seo, F.J., Krajewski, W.F. & Bowles, D.S. Stochastic interpolation of rainfall data from rain gauges and radar using cokriging. 1. Design of experiments. *Water Resources Research*, **26**, 469–477, 1990a.
- Seo, F.J., Krajewski, W.F. & Bowles, D.S. Stochastic interpolation of rainfall data from rain gauges and radar using cokriging. 2. Results. *Water Resources Research*, **26**, 915–924, 1990b.
- Smith, M.L. & Williams, R.E. Examination of methods for evaluationg remining a mine waste site, part II: indicator kriging for selective remediation. *Engineering Geology*, **43**, 23–30, 1996.

- Stackhouse, Jr. P.W., Cox, S., Gupta, S., DiPasquale, R. & Brown, D. The WCRP/GEWEX Surface Radiation Budget Project Release 2: first results at 1 degree resolution. En *10th Conference on Atmospheric Radiation: A Symposium with tributes to the works of Verner E. Suomi*, American Meteorological Society, Madison, Wisconsin, USA, 1999.
- Stein, M.L. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer-Verlag, New York, 1999.
- Stein, A., Van der Meer, F. & Gorte, B.G.H. *Spatial Statistics for Remote Sensing*. Kluwer Academic Publ., Dordrecht, 1999.
- Stein, M.L. Space-time covariance functions. *Journal of the American Statistical Association*, **100**, 310–321, 2005a.
- Stein, M.L. Nonstationary Spatial Covariance Functions. *Technical report No. 21*, The University of Chicago, 2005b.
- Stroud, J.R., Muller, P. & Sansó, B. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, B*, **63**, 673–689, 2001.
- Tanner, M. *Tools for Statistics Inference*. Springer, New York, 1996.
- Taylor, G.I. The spectrum of turbulence. *Proceedings of the Royal Society of London, A*, **164**, 476–490, 1938.
- Tillotson, P.M. & Nielsen, D.R. Scale factors in soil science. *Soil Science Society of American Journal*, **48**, 953–959, 1984.
- Trangmar, B.B. Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy*, **38**, 45–94, 1985.
- Tukey, J.W. *Exploratory Data Analysis*. Addison-Wesley, Reading, 1977.
- Van Groeningen, J.W., Siderius, W. & Stein, A. Constrained optimization of soil sampling for minimization of the kriging variance. *Geoderma*, **87**, 239–259, 1999.
- Vanderborght, J., Jacques, D., Mallants, D., Tseng, P. & Feyer, J. Analysis of solute redistribution in heterogeneous soil: II. Numerical simulation of solute transport. En *First European Conference on Geostatistics for Environmental Applications*, Lisboa, 1996.
- Ver Hoef, J.M. Applied Spatial Statistics. En *XVII Seminario Internacional de Estadística*, EUSTAT, Bilbao, 2002.
- Verly, G. & Sullivan, J. Multigaussian and probability kriging-application to the Jerritt Canyon deposit. *Mining Engineering*, **37**, 568–574, 1985.
- Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications. 2nd Revised Edition*. Springer-Verlag, New York, 1998.

- Webster, R. & Oliver, M.A. *Statistical Methods for Soil and Land Resources Survey*. Oxford University Press, 1990.
- Webster, R. Atteia, O., & Dubois, J.P. Coregionalization of trace metals in the soil in the Swiss Jura. *European Journal of Soil Science*, **45**, 205–218, 1994.
- Whittle, P. On stationary processes on the plane. *Biometrika*, **41**, 434–449, 1954.
- Whittle, P. On the variation of yield variance with plot size. *Biometrika*, **43**, 337–343, 1956.
- Whittle, P. Topographic correlation, power-law covariance functions, and diffusion. *Biometrika*, **49**, 305–314, 1962.
- Whitlock, C.H., Charlock, T.P., Staylor, W.F., Pinker, R.T., Laszlo, I., Ohmura, A., Gilgen, H., Konzelman, T., DiPasquale, R.C., Moats, C.D., LeCroy, S.R. & Rittley, N.A. First Global WCRP Shortwave Surface Radiation Budget data set. *Bulletin of the American Meteorological Society*, **76**, 905–922, 1995.
- Wielicki, B.A., Barkstrom, B.R., Harrison, E.F., Lee III, R.B., Smith, G.L. & Cooper, J.E. Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bulletin of the American Meteorological Society*, **77**, 853–868, 1996.
- Wigneron, J.P., Waldteufel, P., Chanzy, A., Calvet, J.C. & Kerr, Y. Two-dimensional microwave interferometer retrieval capabilities over land surfaces (SMOS mission). *Remote Sensing of Environment*, **73**, 270–282, 2000.
- Wikle, C.K., Berliner, L.M. & Cressie, N. Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, **5**, 117–154, 1998.
- Wikle, C.K., & Cressie, N. A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815–829, 1999.
- Wilmoth, J.R. Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. *Technical report*, University of California, Berkeley, 1993.
- Zhang, H. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250–261, 2004.
- Zhang, Y., Rossow, W.B., Lacis, A.A., Oinas, V. & Mishchenko, M.I. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *Journal of Geophysical Research*, **109**, D19105, 2004.
- Zhu, Z. & Stein, A. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 24–49, 2006.



UNIVERSITAT DE VALÈNCIA  
FACULTAT DE MATEMÀTIQUES

reunido el tribunal que suscribe, en el día de la fecha,  
acordó otorgar a esta Tesis Doctoral de D.....  
*Francisco Martínez Ruiz*.....  
la calificación de *Sobresaliente cum laude*  
Valencia, a *19*.....de *Febrero*.....*2008*..



El Presidente/a,

El Secretario/a,

