

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
Departamento de Estatística e Investigación Operativa

GEOESTADÍSTICA ESPACIO-TEMPORAL

**Modelos flexibles
de variogramas anisotrópicos no
separables**

Rubén Fernández Casal
Septiembre de 2003

Realizado el acto público de defensa y mantenimiento de esta tesis doctoral el día 19 de Diciembre de 2003 en la Facultad de Matemáticas de la Universidad de Santiago de Compostela, ante el tribunal formado por:

Presidente: Dr. D. José Manuel Prada Sánchez

Vocales: Dr. D. Ricardo Cao Abad
Dr. D. Jorge Mateu Mahiques
Dr. D. José Miguel Angulo Ibáñez

Secretario: Dr. D. Manuel Febrero Bande

siendo director de la misma el Dr. D. Wenceslao González Manteiga, obtuvo la máxima calificación de SOBRESALIENTE CUM LAUDE.

AGRADECIMIENTOS

Son numerosas las personas a las que tengo que agradecer la ayuda y el apoyo prestado durante el desarrollo de este trabajo de investigación.

En primer lugar tengo que darle las gracias al director de esta memoria Wenceslao González Manteiga por muchos motivos, empezando por haberme acogido como alumno, iniciarme en el campo de la estadística espacial y por su constante asistencia y estímulo.

También tengo que agradecer especialmente a Manuel Febrero Bande su colaboración en numerosos temas relacionados con este trabajo, incluyendo desde comentarios y discusiones sobre diversas cuestiones, hasta por ejemplo, su ayuda en la resolución de problemas concretos de programación.

En segundo lugar tengo que darles las gracias a los demás compañeros con los que tuve la suerte de trabajar en diversos proyectos en los últimos años y de los cuales pude aprender muchas cosas: José Manuel Prada Sánchez, Tomás R. Cotos Yáñez, Javier Roca Pardiñas y Belén M. Fernández de Castro.

Igualmente a mis compañeros del Departamento de Estadística e Investigación Operativa de la Universidad de Vigo, especialmente a Ana Pérez González, Angeles Saavedra González y Amalia Jácome Pumar.

Finalmente quiero darles las gracias a mis antiguos compañeros del Campus de Lugo, que me ayudaron a iniciarme en la docencia de la estadística, y a las personas del Servicio Galego de Saúde Pública y de Medio Ambiente de E.N.D.E.S.A.-As Pontes con las que tuve el placer de comenzar a trabajar como becario de investigación.

Este trabajo ha sido financiado en parte por los siguientes proyectos: PGIDT99MA20701 y PGIDT00PXI20704PN de la Xunta de Galicia, PB98-0182-C02-02 y BFM2002-03213 del DGEIC, Ministerio de Ciencia y Tecnología y por el Vicerrectorado de Investigación de la Universidad de Vigo.

*A Ana
y a mis padres*

ÍNDICE

CAPÍTULO 1	
INTRODUCCIÓN	1
1.1 Aplicaciones de la metodología espacial y espacio-temporal	2
1.2 Procesos espaciales y espacio-temporales	3
1.3 Geoestadística	4
1.4 Modelos clásicos y modelos espaciales	5
Ventajas de la aproximación espacial (y espacio-temporal).....	8
1.5 Esquema de la monografía	8

PARTE I: PRINCIPIOS DE LA GEOESTADÍSTICA

CAPÍTULO 2	
PROCESOS ESPACIALES ESTACIONARIOS	13
2.1 Introducción	13
Proceso estrictamente estacionario.....	14
Proceso estacionario de segundo orden.....	15
Proceso intrínsecamente estacionario	15
Relación entre los distintos tipos de procesos estacionarios.....	15
Procesos espaciales ergódicos.....	17
Procesos agregados.....	19
2.2 Propiedades del covariograma y del variograma	20
2.2.1 Propiedades elementales	20
Propiedades elementales del covariograma.....	20
Propiedades elementales del variograma	22
Características del variograma (covariograma)	24

2.2.2 Representación espectral.....	27
Representación espectral del covariograma	28
Representación espectral del semivariograma.....	31
2.2.3 Anisotropía	32
Anisotropía geométrica	33
Anisotropía zonal	34
CAPÍTULO 3	
PREDICCIÓN ESPACIAL LINEAL ÓPTIMA. KRIGING.....	35
3.1 Introducción	35
Kriging con error de medida	37
Kriging con covarianzas y variogramas no estacionarios.....	37
3.2 Kriging con media conocida. kriging simple	37
Detalles prácticos y computacionales	39
3.3 Kriging con media desconocida. kriging universal.....	40
Ecuaciones en función del covariograma	42
Tendencia polinómica.....	43
Detalles prácticos y computacionales	44
3.4 Algunas consideraciones acerca de los métodos kriging	46
3.4.1 Kriging como interpolador	46
3.4.2 Efecto del variograma (covariograma) en el kriging.....	48
Efecto de la estimación del variograma.....	50
3.4.3 Elección del vecindario.....	51
3.4.4 Kriging log-normal y trans-normal	54
CAPÍTULO 4	
MODELADO DE LA DEPENDENCIA ESPACIAL.....	57
4.1 Estimación del semivariograma	58
4.1.1 Estimación del (co)variograma bajo media constante.....	58
Estimadores clásicos del semivariograma y del covariograma	58
Ventajas de la estimación del variograma.....	60
Estimadores robustos	61
Estimadores no paramétricos	62

4.1.2 Estimación del variograma en el KU	64
Estimación del variograma basada en los residuos	65
4.2 Modelos paramétricos de variogramas	67
4.2.1 Obtención de modelos paramétricos.....	68
4.2.2 Modelos paramétricos isotrópicos.....	69
4.2.3 El modelo lineal de regionalización	72
4.3 Principales métodos para el ajuste de un modelo válido	73
4.3.1 Mínimos cuadrados	73
Detalles prácticos y computacionales	76
4.3.2 Estimación basada en máxima verosimilitud	78
Máxima verosimilitud	78
Máxima verosimilitud restringida.....	80
Detalles prácticos y computacionales	81
4.3.3 Algunos comentarios sobre los distintos métodos.....	82
4.4 Validación cruzada del modelo ajustado	84
Detalles prácticos y computacionales	86
Estimación de los parámetros del variograma mediante validación cruzada.....	88
4.5 Aproximaciones flexibles para el modelado de la dependencia espacial .	91
4.5.1 Discretización de la distribución espectral (Shapiro-Botha).....	91
4.5.2 Otros métodos basados en la representación espectral.....	96
4.5.3 Representación en medias móviles	98
4.5.4 Extensión de la aproximación de Shapiro-Botha para el caso de anisotropía	99
 CAPÍTULO 5	
GEOESTADÍSTICA MULTIVARIANTE	103
5.1 Introducción	104
5.1.1 Procesos espaciales multivariantes estacionarios.....	104
Estacionariedad de segundo orden	104
Estacionariedad intrínseca.....	105
Relación entre los distintos tipos de estacionariedad.....	105

5.1.2 Algunas propiedades de los covariogramas y variogramas cruzados ..	106
Representación espectral	109
5.2 Predicción espacial multivariante. cokriging	110
5.2.1 Métodos cokriging tradicionales	112
Cokriging simple	112
Cokriging ordinario	113
Influencia de la información secundaria	114
5.2.2 Variantes de los métodos cokriging tradicionales	115
Cokriging ordinario centrado y estandarizado	115
Cokriging co-posicionado (colocalizado)	116
5.3 Modelado de la dependencia espacial multivariante.....	118
5.3.1 Estimación de los covariogramas y variogramas cruzados	118
5.3.2 Modelos paramétricos	120
Modelo intrínseco	120
Modelo lineal de co-regionalización	121
Modelos de Markov	123
Otros modelos paramétricos	123
5.3.3 Modelado no paramétrico	125
Generalización de la aproximación de Shapiro-Botha	125
Otros métodos	126
5.3.4 Modelado de asimetría	127
5.3.5 Observaciones	128

PARTE II: GEOESTADÍSTICA ESPACIO-TEMPORAL

CAPÍTULO 6

MODELOS ESPACIO-TEMPORALES..... 133

6.1 Introducción 133

6.1.1 Aproximación espacial multivariante..... 135

6.1.2 Aproximación espacio-temporal univariante

6.1.3 Observaciones

6.2 Modelos paramétricos de semivariogramas no separables 142

6.3 Modelos flexibles de semivariogramas espacio-temporales	145
6.3.1 Modelos anisotrópicos en dos componentes.....	146
6.3.2 Ajuste	149
6.3.3 Selección de los puntos de discretización	152
CAPÍTULO 7	
ESTIMACIÓN DE UN SEMIVARIOGRAMA ESPACIO-TEMPORAL.....	155
7.1 Estimador empírico.....	156
7.1.1 Selección de los saltos y las regiones de tolerancia.....	157
7.1.2 Un ejemplo de la eficiencia del estimador empírico.....	158
7.2 Estimadores no paramétricos	163
Estimador Nadaraya-Watson	164
Estimador lineal local.....	165
Estimador lineal local robusto.....	166
Detalles computacionales	167
Selección de la ventana óptima	167
CAPÍTULO 8	
APLICACIÓN A DATOS REALES.....	171
8.1 Estimación y ajuste de los modelos	173
8.2 Predicciones	175
CAPÍTULO 9	
ESTUDIO DE SIMULACIÓN	179
9.1 Estimaciones piloto del semivariograma.....	181
9.2 Modelos ajustados	185
9.3 Predicciones	191
9.4 Conclusiones	195
BIBLIOGRAFÍA.....	199

CAPÍTULO 1

INTRODUCCIÓN

Es bien sabido que al utilizar en la práctica métodos estadísticos no siempre es adecuado suponer que las observaciones del fenómeno de interés han sido tomadas bajo condiciones idénticas e independientes unas de otras (i.e. que los datos son independientes e idénticamente distribuidos). Esta falta de homogeneidad en los datos suele ser modelada a través de la suposición de media no constante (por ejemplo suponiendo que ésta es una combinación lineal de ciertas variables explicativas) pero con la consideración de que los errores son independientes e idénticamente distribuidos. Sin embargo, como se muestra más adelante en la sección 1.4, esta suposición puede influir crucialmente en la inferencia; pudiendo ser en ocasiones preferible la suposición más realista de errores correlados.

Frecuentemente los datos tienen una componente espacial y/o temporal asociada a ellos (ver sección 1.1) y es de esperar que datos cercanos en el espacio o en el tiempo sean más semejantes que aquellos que están más alejados; en cuyo caso no deben ser modelados como estadísticamente independientes, siendo más conveniente emplear modelos que exploten adecuadamente dicha componente espacial o espacio-temporal. De forma natural surge la hipótesis de que los datos cercanos en el espacio o en el tiempo están correlados y que la correlación disminuye al aumentar la separación entre ellos, por lo que se puede pensar en la presencia de una dependencia espacial o espacio-temporal.

En general, cuando se considera que la componente espacial o espacio-temporal puede ser importante en el modelado y el análisis de la información es necesaria una aproximación estadística distinta a la tradicionalmente usada. Esto lleva al concepto de proceso espacial o espacio-temporal (sección 1.2) y a la investigación de sus propiedades. La geoestadística (sección 1.3) es una de las ramas de la estadística que se centra en el estudio de este tipo de procesos.

1.1 APLICACIONES DE LA METODOLOGÍA ESPACIAL Y ESPACIO-TEMPORAL

La metodología espacial y espacio-temporal ha sido utilizada de forma creciente (especialmente durante los últimos 25 o 30 años) para solventar problemas en muchos campos. Son muchas las disciplinas en las que interesa estudiar algún tipo de propiedad que tiene cierta característica espacial o espacio-temporal. En estas disciplinas la metodología espacial puede ser de ayuda en alguna o en muchas etapas del estudio, desde el diseño inicial del muestreo hasta la representación final de los resultados obtenidos (p.e. para la generación de mapas o animaciones). Algunos de los campos en los que se trabaja con datos espaciales o espacio-temporales por naturaleza son los relacionados con la geología, hidrología, ecología, ciencias medioambientales, meteorología, epidemiología, recursos mineros, geografía, astronomía, proceso de imágenes, experimentos agrícolas, etc.

Por ejemplo, algunos trabajos recientes en los que se ha aplicado la metodología espacial a distintos campos (principalmente métodos geoestadísticos espacio-temporales) son los siguientes:

- Ciencias medioambientales: Høst et al. (1995), Angulo et al. (1998), Fuentes y Smith (2001) (concentración de SO_2); Niu (1996), Carrol et al. (1997) (ozono); Kolenikov et al. (2002) (partículas en suspensión); Haas (1990, 1995) (lluvia ácida); De Iaco et al. (2002b) (índices generales de contaminación atmosférica).
- Meteorología: Sampson y Guttorp (1992) (radiación solar); Handcock y Wallis (1994) (cambio climático); Chao (1998) (pluviosidad); Cressie y Huang (1999), Wikle y Cressie (1999) (velocidad de viento); Bengtsson y Nychka (2001) (predicción del clima).
- Ecología: Fernandes y Rivoirard (1998) (distribución espacial de especies marinas); Mateu y Ribeiro (1998) (análisis espacial de especies forestales); Zhu et al. (2002) (índices de foliación de árboles).
- Hidrogeología: Rouhani y Hall (1989), Rouhani y Myers (1990), Samper y Carrera (1990) (hidrología subterránea); Müller (2003) (diseño de redes).
- Epidemiología: Cressie et al. (2000), Benes y Bodlák (2003) (tasas epidemiológicas); Mugglin et al. (2002) (dinámica de epidemias).
- Otros campos: edafología: Egbert y Lettenmaier (1986); economía: Gelfand et al. (2001), Albert et al. (2002); análisis de imágenes: Stoica et al. (2002);

desarrollo de medicamentos: Wang et al. (2003); control de operaciones militares: Irwin et al. (2002).

1.2 PROCESOS ESPACIALES Y ESPACIO-TEMPORALES

Sea $\mathbf{s} \in \mathbb{R}^d$ una posición genérica en el espacio euclideo d -dimensional y supongamos que el valor potencial $Z(\mathbf{s})$ en la posición espacial \mathbf{s} es una cantidad aleatoria. Entonces, si \mathbf{s} varía dentro del conjunto índice $D \subset \mathbb{R}^d$ se obtiene el proceso aleatorio multivariante:

$$\{Z(\mathbf{s}) : \mathbf{s} \in D\} \quad (1.1)$$

que denominaremos *proceso espacial* (también se suele denominar en la literatura función aleatoria, campo espacial aleatorio o variable regionalizada). Una realización del proceso espacial (1.1) se denotará por $\{z(\mathbf{s}) : \mathbf{s} \in D\}$.

En algunos casos, utilizando una notación más formal y enfatizando la aleatoriedad del proceso, (1.1) se suele escribir como $\{Z(\mathbf{s}, \omega) : \mathbf{s} \in D, \omega \in \Omega\}$, siendo (Ω, A, P) un espacio de probabilidad, así la realización $\{z(\mathbf{s}) : \mathbf{s} \in D\}$ se correspondería con un valor particular $\omega_0 \in \Omega$.

Si $\{Z_j(\mathbf{s}) : j = 1, \dots, k; \mathbf{s} \in D\}$ son k procesos espaciales univariantes (y que se suponen en principio interdependientes), el vector $\mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_k(\mathbf{s}))'$ lo denominaremos *proceso espacial multivariante* (también se denomina proceso espacial vectorial, campo vectorial espacial o vector regionalizado).

Se suele distinguir entre distintos casos dependiendo de las suposiciones acerca del conjunto D . En este trabajo nos centraremos en el caso de *datos geoestadísticos*: D es un subconjunto fijo de \mathbb{R}^d que contiene un rectángulo d -dimensional de volumen positivo (i.e. índice espacial continuo); $\mathbf{Z}(\mathbf{s})$ es un vector aleatorio en la posición $\mathbf{s} \in D$ (también se suele denominar este tipo de procesos como procesos espaciales continuos). El caso de posiciones espaciales discretas se considerará como resultado de la discretización de un proceso continuo.

La definición (1.1) es válida también para el caso espacio-temporal, por ejemplo podríamos considerar posiciones de la forma $\mathbf{s} = (s_1, \dots, s_{d-1}, t) \in \mathbb{R}^{d-1} \times \mathbb{R}^{+,0}$, donde $\mathbb{R}^{+,0} = \{t \in \mathbb{R} : t \geq 0\}$. Es importante destacar por tanto que las definiciones y resultados mostrados en la primera parte de este trabajo son aplicables también al caso espacio-temporal. Sin embargo, en muchos casos para enfatizar el carácter temporal se utilizará una notación de la forma:

$$\{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in D \times T\}$$

donde $D \times T \subset \mathbb{R}^d \times \mathbb{R}^{+,0}$, para referirse a un *proceso espacio-temporal*. Esta notación será la principalmente empleada a partir del capítulo 6 (parte II).

En algunos casos los procesos espacio-temporales son modelados también como procesos espaciales multivariantes (e.g. Egbert y Lettenmaier, 1986; Kyriakidis y Journel, 1999). Por ejemplo, se puede considerar una representación de la forma:

$$Z(\mathbf{s}, t) = \mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_k(\mathbf{s}))', \quad (1.2)$$

donde $Z_i(\mathbf{s}) = Z(\mathbf{s}, t_i)$, $i = 1, \dots, k$. O también:

$$Z(\mathbf{s}, t) = \mathbf{Z}(t) = (Z_1(t), \dots, Z_n(t))',$$

siendo $Z_j(t) = Z(\mathbf{s}_j, t)$, $j = 1, \dots, n$. Uno de los principales problemas al utilizar estas aproximaciones es que, utilizando los modelos geoestadísticos tradicionales, no es posible la predicción en todas las posiciones espacio-temporales sin algún tipo de modelado adicional. Por ejemplo, utilizando la representación (1.2) y los métodos geoestadísticos de predicción espacial multivariante, se pueden obtener en principio superficies de predicción solamente en los k instantes temporales t_i , $i = 1, \dots, k$, y no es posible la interpolación temporal sin modelado adicional (ver sección 5.3.5).

1.3 GEOESTADÍSTICA

La geoestadística (Matheron 1962, 1963) surgió como una mezcla de varias disciplinas: ingeniería de minas, geología, matemáticas y estadística, para dar respuesta a problemas como, por ejemplo, el de la estimación de los recursos de una explotación minera (se desarrolló principalmente a partir de los años 80). La diferencia (ventaja) respecto a otras aproximaciones es que, además de tener en cuenta la tendencia espacial (variación de gran escala), también tiene en cuenta la correlación espacial (variación de pequeña escala). Otros métodos sin embargo, sólo incluyen la variación de larga escala y suponen que los errores son independientes. Hoy en día se puede decir que la geoestadística es la parte de la estadística espacial que estudia los procesos con índice espacial continuo (las bases teóricas de la geoestadística se tratan en el capítulo 2).

Uno de los problemas iniciales más importantes de la geoestadística fue la predicción de la riqueza de un bloque minero a partir de una muestra observada. A este proceso Matheron (1963) lo denominó *kriging*¹, y también predicción espacial lineal óptima (estos métodos de predicción se muestran en el capítulo 3 y en la sección 5.3).

Al mismo tiempo que la geoestadística se desarrollaba en la ingeniería de minas por Matheron en Francia, la misma idea se desarrollaba en la meteorología por L.S. Gandin en la antigua Unión Soviética. El nombre que Gandin le dio a esta aproximación fue *análisis objetivo* y utilizó la terminología de *interpolación óptima* en lugar de kriging. Para más detalles sobre el origen del kriging ver p.e. Cressie (1990).

1.4 MODELOS CLÁSICOS Y MODELOS ESPACIALES

Uno de los modelos más utilizados en estadística para el caso de datos no homogéneos es el conocido modelo clásico de regresión lineal. Si $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$ es un proceso espacial, podemos suponer que:

$$Z(\mathbf{s}) = \sum_{j=0}^p x_j(\mathbf{s})\beta_j + \delta(\mathbf{s}), \quad \mathbf{s} \in D,$$

donde $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)' \in \mathbb{R}^{p+1}$ es un vector desconocido, $\{x_j(\cdot) : j = 0, \dots, p\}$ un conjunto de variables explicativas y $\delta(\cdot)$ un proceso de media cero incorrelado (i.e. $Cov(\delta(\mathbf{u}), \delta(\mathbf{v})) = 0$ si $\mathbf{u} \neq \mathbf{v}$) con $Var(\delta(\mathbf{s})) = \sigma^2$.

Supongamos por el momento que el objetivo es la estimación eficiente de la tendencia, o lo que es lo mismo la estimación óptima de los parámetros de la “variación de gran escala” $\boldsymbol{\beta}$, a partir de los datos observados en un conjunto de posiciones espaciales $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$. Bajo las hipótesis anteriores:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}$$

¹ D. G. Krige fue un ingeniero de minas de Sur África que desarrolló en los años 50 métodos empíricos para determinar la distribución de la riqueza de un mineral a partir de valores observados. Sin embargo la formulación de la predicción espacial lineal óptima no procede del trabajo de Krige.

siendo $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$, \mathbf{X} una matriz $n \times (p+1)$ con $X_{ij} = x_{j-1}(\mathbf{s}_i)$ y $\boldsymbol{\delta} = (\delta(\mathbf{s}_1), \dots, \delta(\mathbf{s}_n))'$; y el estimador lineal insesgado de $\boldsymbol{\beta}$ más eficiente resulta ser el estimador de mínimos cuadrados ordinarios (m.c.o.):

$$\hat{\boldsymbol{\beta}}_{mco} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}, \quad (1.3)$$

con $Var(\hat{\boldsymbol{\beta}}_{mco}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Sin embargo la suposición de que los errores son independientes e idénticamente distribuidos influye crucialmente en la inferencia. En el modelo anterior, en lugar de errores incorrelados, si suponemos que:

$$Var(\boldsymbol{\delta}) = \boldsymbol{\Sigma},$$

obtenemos el modelo lineal de regresión generalizado y en este caso el estimador lineal óptimo de $\boldsymbol{\beta}$ es el estimador de mínimos cuadrados generalizados (m.c.g.):

$$\hat{\boldsymbol{\beta}}_{mcg} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}. \quad (1.4)$$

Si $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_n$, siendo \mathbf{I}_n la matriz identidad $n \times n$, los estimadores (1.3) y (1.4) coinciden; pero en caso contrario las estimaciones basadas en el modelo anterior pueden llegar a ser altamente ineficientes. Puede verse fácilmente que en el caso general:

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}_{mcg}) &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}, \\ Var(\hat{\boldsymbol{\beta}}_{mco}) &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}, \end{aligned}$$

resultando además que $Var(\hat{\boldsymbol{\beta}}_{mco}) - Var(\hat{\boldsymbol{\beta}}_{mcg})$ es una matriz semidefinida positiva (e.g. Searle, 1971, sección 3.3).

En muchos casos el objetivo final es la predicción del proceso en una posición espacial \mathbf{s}_0 :

$$Z(\mathbf{s}_0) = \mathbf{x}'\boldsymbol{\beta} + \delta(\mathbf{s}_0),$$

donde $\mathbf{x} = (x_0(\mathbf{s}_0), \dots, x_p(\mathbf{s}_0))'$. Siguiendo la aproximación tradicional se puede pensar en utilizar como predictor el estimador de la tendencia. Por ejemplo, empleando (1.4) obtendríamos:

$$\hat{Z}(\mathbf{s}_0) = \hat{\mu}(\mathbf{s}_0) = \mathbf{x}'\hat{\boldsymbol{\beta}}_{mcg}. \quad (1.5)$$

Sin embargo (1.5) no es el predictor lineal óptimo; puede verse (e.g. Goldberger, 1962; sección² 3.3) que en este caso el mejor predictor lineal insesgado es:

$$\tilde{Z}(\mathbf{s}_0) = \mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + (\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c})'\hat{\boldsymbol{\beta}}_{mcg}, \quad (1.6)$$

siendo $\mathbf{c} = (Cov(\delta(\mathbf{s}_1), \delta(\mathbf{s}_0)), \dots, Cov(\delta(\mathbf{s}_n), \delta(\mathbf{s}_0)))'$, siendo la diferencia:

$$Var(\hat{Z}(\mathbf{s}_0)) - Var(\tilde{Z}(\mathbf{s}_0)) = \mathbf{c}'(\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1})\mathbf{c}$$

siempre positiva y en ocasiones significativamente mayor que cero (naturalmente utilizando el estimador (1.3) disminuye aún más la eficiencia de las predicciones).

Teniendo en cuenta los resultados anteriores podemos afirmar que al explotar la dependencia presente en los datos el incremento en eficiencia puede ser importante³. Sin embargo el principal inconveniente es que en la práctica normalmente la matriz $\boldsymbol{\Sigma}$ y el vector \mathbf{c} son desconocidos. El procedimiento habitual, para evitar la estimación de $n + n(n + 1)/2$ parámetros adicionales a partir del conjunto de n observaciones, suele ser la elección de un modelo paramétrico adecuado:

$$C(\mathbf{u}, \mathbf{v} | \boldsymbol{\theta}) \equiv Cov(\delta(\mathbf{u}), \delta(\mathbf{v})),$$

i.e. suponer que $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}(\boldsymbol{\theta})$ y $\mathbf{c} \equiv \mathbf{c}(\boldsymbol{\theta})$. Una hipótesis natural es suponer que los datos cercanos en el espacio o en el tiempo están correlados y que la correlación disminuye al aumentar la separación entre ellos; por tanto es normal pensar en errores espacialmente correlados. Por ejemplo, podemos considerar:

$$C(\mathbf{u}, \mathbf{v} | \boldsymbol{\theta}) = \sigma^2 \rho^{\|\mathbf{u}-\mathbf{v}\|},$$

con $\sigma^2 \geq 0$ y $0 < \rho < 1$; o en el caso espacio-temporal:

$$C(\mathbf{s}_1, t_1, \mathbf{s}_2, t_2 | \boldsymbol{\theta}) = \sigma^2 \rho_s^{\|\mathbf{s}_1 - \mathbf{s}_2\|} \rho_t^{|t_1 - t_2|},$$

con $0 < \rho_s < 1$ y $0 < \rho_t < 1$, ya que suele interesar distinguir entre salto espacial y salto temporal. De esta forma, si $\hat{\boldsymbol{\theta}}$ es un estimador de $\boldsymbol{\theta}$, podemos obtener por

² En esta sección se muestra también el predictor óptimo para un caso más general.

³ Por ejemplo, para un caso particular, Goldberger (1962, pp. 374-375) observó que la mejora en la predicción puede llegar a ser del 50%. En Cressie (1993, sección 1.3) se muestran también otros ejemplos del efecto de la presencia de correlación en la estimación.

ejemplo una estimación del predictor óptimo de $Z(\mathbf{s}_0)$ sustituyendo en (1.4) $\Sigma(\boldsymbol{\theta})$ por $\Sigma(\hat{\boldsymbol{\theta}})$ y $\mathbf{c}(\boldsymbol{\theta})$ por $\mathbf{c}(\hat{\boldsymbol{\theta}})$.

VENTAJAS DE LA APROXIMACIÓN ESPACIAL (Y ESPACIO-TEMPORAL)

Algunos de los beneficios de utilizar modelos espaciales para caracterizar y explotar la dependencia espacial de un conjunto de datos son los siguientes:

- Modelos más generales: en la mayoría de los casos, los modelos clásicos no espaciales son un caso particular de un modelo espacial.
- Estimaciones más eficientes: de la tendencia, de los efectos de variables explicativas, de promedios regionales,...
- Mejora de las predicciones: más eficientes, con propiedades de extrapolación más estables,...
- La variación espacial no explicada en la estructura de la media debe ser absorbida por la estructura del error, por lo que un modelo que incorpore la dependencia espacial puede decirse que está "protegido" frente a una mala especificación de este tipo. Esto en muchos casos tiene como resultado una simplificación en la especificación de la tendencia; en general los modelos con dependencia espacial suelen tener una descripción más parsimoniosa (en ocasiones con muchos menos parámetros) que los clásicos modelos de superficie de tendencia.

1.5 ESQUEMA DE LA MONOGRAFÍA

Como ya se indicó en la sección 1.2 podemos pensar en los procesos espacio-temporales como casos especiales de procesos espaciales. De hecho en el caso espacio-temporal se utiliza metodología desarrollada inicialmente en la geoestadística espacial, empleando en muchos casos las mismas herramientas o introduciendo algunas modificaciones. Este es uno de los motivos por los que se decidió organizar este trabajo en dos partes. La primera parte - *Principios de la Geoestadística* - consiste en una introducción a la geoestadística, donde se pretende realizar una revisión (superficial) de las distintas ideas y herramientas desarrolladas para el caso general de procesos espaciales (confiando en que también sea de utilidad para aquellas personas no familiarizadas con este tema).

El tema central de la segunda parte - *Geoestadística espacio-temporal* - es el modelado (flexible) de la dependencia espacio-temporal.

En los capítulos del 2 al 5 se intentará dar una visión global de los métodos geoestadísticos más conocidos aunque se tratará con mayor detalle las distintas propuestas relacionadas con el modelado no paramétrico de la dependencia espacial (modelos flexibles de variogramas y covariogramas espaciales; capítulo 4 y sección 5.3). Hay que insistir en que las definiciones y métodos mostrados a lo largo de estos capítulos son aplicables directamente para el caso espacio-temporal (por ejemplo los métodos de predicción espacial mostrados en el capítulo 3). Aunque algunas nociones (como por ejemplo el concepto de isotropía definido en el capítulo 2) pueden ser poco adecuadas para este caso, debido a que los procesos espacio-temporales normalmente no presentan el mismo comportamiento en el espacio que en el tiempo. Sin embargo, estas ideas pueden ser adaptadas de forma natural para el caso en que una de las coordenadas espaciales es temporal.

Los capítulos del 6 al 9 se centran en el caso espacio-temporal. En el capítulo 6 se tiene una revisión de los modelos geoestadísticos de dependencia espacio-temporal disponibles en la actualidad; al final de este capítulo se proponen algunas familias de modelos flexibles de semivariogramas espacio-temporales (presentadas inicialmente en Fernández-Casal et al., 2001). En el capítulo 7 se trata el problema de la obtención de estimaciones piloto de un semivariograma espacio temporal, proponiéndose el empleo de un estimador no paramétrico (el estimador lineal local considerado en Fernández-Casal et al., 2003b). En el capítulo 8 se tiene un ejemplo de aplicación a un conjunto de datos meteorológicos (datos de velocidad de viento). En el capítulo 9 se muestran los resultados obtenidos en un estudio de simulación y se realizan algunas observaciones finales sobre la metodología propuesta.

Para poder aplicar en la práctica cualquier método geoestadístico resulta casi imprescindible disponer de software adecuado. Por este motivo se ha invertido mucho tiempo en el desarrollo del código necesario para la implementación de los diferentes métodos descritos en este trabajo. Por motivos de extensión, y de utilidad práctica, se decidió incluir algunas librerías únicamente en formato electrónico en el cd-rom adjunto. Es importante destacar que actualmente muchas de estas rutinas son versiones preliminares y serán modificadas (también se irán añadiendo otras nuevas); las últimas

actualizaciones estarán disponibles en la página web <http://eiosun.uvigo.es/staff/rubenfc/rubenfc.html> (o solicitándolas en la dirección de correo electrónico rubenfc@uvigo.es). El fichero “léame.txt” contiene información sobre las diferentes herramientas y utilidades incluidas en el cd-rom.

PARTE I

**PRINCIPIOS DE LA
GEOESTADÍSTICA**

CAPÍTULO 2

PROCESOS ESPACIALES ESTACIONARIOS

Por geoestadística entendemos la parte de la estadística que estudia los procesos espaciales con índice espacial continuo (ver secciones 1.2 y 1.3). En la sección 2.1 se definen los principales procesos de este tipo habitualmente considerados en geoestadística y se introducen dos funciones relacionadas con estos procesos, el covariograma y el variograma. Algunas propiedades de estas funciones, que podríamos decir que son las herramientas fundamentales de la geoestadística, se muestran en la sección 2.2.

2.1 INTRODUCCIÓN

Supongamos que $\{z(\mathbf{s}) : \mathbf{s} \in D\}$ es una realización del proceso espacial:

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

con $D \subset \mathbb{R}^d$ fijo y volumen d -dimensional positivo. El proceso aleatorio (2.1) usualmente se caracteriza a través de las *distribuciones finito-dimensionales*:

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m) = P\{Z(\mathbf{s}_1) \leq z_1, \dots, Z(\mathbf{s}_m) \leq z_m\} \quad (2.2)$$

que deben verificar las condiciones habituales de simetría y consistencia:

- Simetría:

$$F_{\mathbf{s}_{i_1}, \dots, \mathbf{s}_{i_m}}(z_{i_1}, \dots, z_{i_m}) = F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m),$$

donde i_1, \dots, i_m es una permutación de los índices $1, \dots, m$.

- Consistencia:

$$F_{\mathbf{s}_1, \dots, \mathbf{s}_m, \mathbf{s}_{m+1}, \dots, \mathbf{s}_{m+k}}(z_1, \dots, z_m, \infty, \dots, \infty) = F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m).$$

Además se suele suponer la existencia de las funciones de densidad correspondientes a (2.2):

$$f_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m) = \frac{\partial^m}{\partial z_1 \dots \partial z_m} F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m).$$

Por ejemplo, un proceso espacial independiente (también denominado de memoria cero o ruido blanco) está completamente caracterizado por las funciones de densidad univariantes $f_{\mathbf{s}}(z)$, ya que en este caso:

$$f_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m) = \prod_{i=1}^m f_{\mathbf{s}_i}(z_i).$$

En general no se puede disponer de una realización completa del proceso $Z(\cdot)$ y solamente se observa un conjunto de valores $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$ en unas posiciones espaciales conocidas $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ (que por lo general van a ser irregulares, i.e. los datos no van a estar situados en una rejilla). Como los datos observados son una realización parcial es necesario hacer algunas suposiciones acerca del proceso de forma que sea posible la inferencia sobre el mismo. Por ejemplo, supondremos que existe:

$$\mu(\mathbf{s}) = E(Z(\mathbf{s})), \quad \forall \mathbf{s} \in D,$$

(la función $\mu(\cdot)$ se suele denominar tendencia), y en ocasiones también:

$$\sigma^2(\mathbf{s}) = Var(Z(\mathbf{s})), \quad \forall \mathbf{s} \in D.$$

Además se suele asumir algún tipo de estacionariedad del proceso, por ejemplo, a continuación se enumeran las habitualmente consideradas (de mayor a menor restricción).

PROCESO ESTRICTAMENTE ESTACIONARIO

El proceso $Z(\cdot)$ se dice *estrictamente estacionario* si se verifica que:

$$F_{\mathbf{s}_1 + \mathbf{h}, \dots, \mathbf{s}_m + \mathbf{h}}(z_1, \dots, z_m) = F_{\mathbf{s}_1, \dots, \mathbf{s}_m}(z_1, \dots, z_m), \quad \forall \mathbf{h} \in D, \forall m \geq 1,$$

(i.e. al trasladar en cualquier dirección una configuración de posiciones la distribución conjunta no varia).

PROCESO ESTACIONARIO DE SEGUNDO ORDEN

Si el proceso $Z(\cdot)$ verifica que:

$$E(Z(\mathbf{s})) = \mu, \forall \mathbf{s} \in D,$$

$$Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1 - \mathbf{s}_2), \forall \mathbf{s}_1, \mathbf{s}_2 \in D,$$

se dice que es *estacionario de segundo orden* (algunos autores también lo denominan proceso estacionario homogéneo o débilmente estacionario). La función $C(\cdot)$ se denomina *covariograma*. Si además $C(\mathbf{h}) = C^0(\|\mathbf{h}\|)$ (sólo depende de la magnitud y no de la dirección del salto) se dice que el covariograma es isotrópico.

En algunos casos para modelar la dependencia espacial de procesos estacionarios de segundo orden, en lugar del covariograma se suele utilizar el *correlograma*, definido por:

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})} \in [-1, +1],$$

suponiendo que $C(\mathbf{0}) > 0$.

PROCESO INTRÍNSECAMENTE ESTACIONARIO

El proceso $Z(\cdot)$ se dice *intrínsecamente estacionario* (también denominado proceso espacial de incrementos estacionarios u homogéneos) si se verifica que:

$$E(Z(\mathbf{s})) = \mu, \forall \mathbf{s} \in D,$$

$$Var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2), \forall \mathbf{s}_1, \mathbf{s}_2 \in D.$$

La función $2\gamma(\cdot)$ se denomina *variograma* y $\gamma(\cdot)$ semivariograma (aunque algunos autores utilizan también la denominación de variograma para referirse a $\gamma(\cdot)$). Si además $\gamma(\mathbf{h}) = \gamma^0(\|\mathbf{h}\|)$ (sólo depende de la distancia) se dice que el variograma es isotrópico.

RELACIÓN ENTRE LOS DISTINTOS TIPOS DE PROCESOS ESTACIONARIOS

Si un proceso es estrictamente estacionario y su momento de segundo orden es finito entonces es estacionario de segundo orden. Además, como es bien

conocido, en el caso de procesos normales ambas propiedades son equivalentes (ya que están caracterizados por su media y covarianza).

La clase de procesos intrínsecamente estacionarios es más general que la clase de procesos estacionarios de segundo orden. Si un proceso estacionario de 2° orden tiene covariograma $C(\cdot)$, como:

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

entonces su variograma viene dado por:

$$2\gamma(\mathbf{h}) = 2(C(\mathbf{0}) - C(\mathbf{h})), \quad (2.3)$$

y por tanto es un proceso intrínsecamente estacionario. El recíproco en general no es cierto (por ejemplo el caso de un movimiento Browniano isotrópico d -dimensional, ver p.e. Cressie, 1993, p. 68), aunque sí se verifica en muchos casos. Si el variograma está acotado¹ y:

$$\lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\mathbf{h}) = \gamma(\infty),$$

entonces podemos obtener² el covariograma correspondiente como:

$$C(\mathbf{h}) = \gamma(\infty) - \gamma(\mathbf{h}),$$

(ver p.e. Matheron, 1973, pp. 454-457). La relación entre el variograma y el covariograma se ilustra en la figura 2.1.

También se puede pensar en un proceso intrínsecamente estacionario como un proceso cuyos incrementos:

$$Y_{\mathbf{h}}(\mathbf{s}) = Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}),$$

son estacionarios. A partir del semivariograma $\gamma(\cdot)$ del proceso $Z(\cdot)$ podemos obtener cualquier covarianza entre incrementos, ya que:

¹ Además, en casos en los que el semivariograma parece no estar acotado (y por tanto el proceso no es aparentemente estacionario de segundo orden), resulta que el proceso no tiene media constante y puede ser modelado como una función de tendencia más residuos estacionarios.

² Suponiendo que $\lim_{\|\mathbf{h}\| \rightarrow \infty} C(\mathbf{h}) = 0$.

$$\begin{aligned} \text{Cov}(Y_{\mathbf{h}_1}(\mathbf{s}_1), Y_{\mathbf{h}_2}(\mathbf{s}_2)) &= \gamma(\mathbf{s}_1 + \mathbf{h}_1 - \mathbf{s}_2) + \gamma(\mathbf{s}_1 - \mathbf{s}_2 - \mathbf{h}_2) \\ &\quad - \gamma(\mathbf{s}_1 + \mathbf{h}_1 - \mathbf{s}_2 - \mathbf{h}_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2). \end{aligned} \quad (2.4)$$

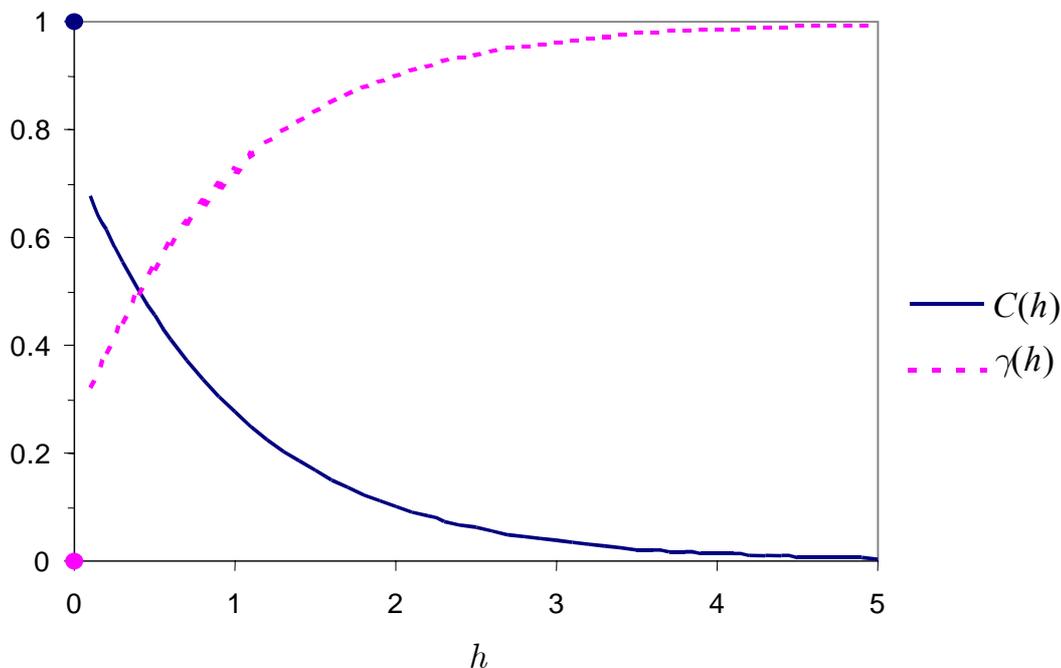


Figura 2.1: Relación entre el covariograma y el variograma (caso unidimensional).

PROCESOS ESPACIALES ERGÓDICOS

En general no se puede disponer de múltiples realizaciones del proceso de interés $Z(\cdot)$ (ni siquiera de una realización completa, ya que normalmente sólo se observa un conjunto de valores $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$) y por tanto deberíamos poder estimar los parámetros de interés del proceso a partir de una única realización. En otros campos, como por ejemplo en el caso de series de tiempo, los procesos aleatorios en los que esto es posible han sido ampliamente estudiados y son denominados *procesos ergódicos* (o se dice que poseen la propiedad de ergodicidad). Esta propiedad garantiza la convergencia (en media cuadrática) de los promedios muestrales de interés a sus correspondientes teóricos, y se habla por ejemplo, de ergodicidad en media (o covarianza) cuando la media (o covarianza) muestral convergen en media cuadrática a la media (o covarianza) teórica.

Se ha generalizado esta noción de ergodicidad para procesos espaciales en el caso en el que el dominio donde se realiza el promedio tiende a infinito (en todas direcciones). Por ejemplo, se dice que el proceso espacial $Z(\cdot)$ es *ergódico en media* si su media es constante $E(Z(\mathbf{s})) = \mu$ y además:

$$\lim_{L \rightarrow \infty} \frac{1}{(2L)^d} \int_{U_L} Z(\mathbf{s}) d\mathbf{s} = \mu,$$

donde el dominio de integración es el cubo d -dimensional $U_L = \{\mathbf{s} : |s_i| < L, i = 1, \dots, d\}$. Se han establecido también condiciones necesarias y suficientes para que esto ocurra (algunas basadas en representaciones espectrales, sección 2.2.3). Por ejemplo, una condición suficiente para que un proceso estacionario de segundo orden sea ergódico en media es que:

$$\lim_{\|\mathbf{h}\| \rightarrow \infty} C(\mathbf{h}) = 0, \quad (2.5)$$

aunque no es una condición necesaria.

De forma análoga al caso anterior se pueden definir otros tipos de ergodicidad, como la ergodicidad en la covarianza o en el semivariograma. Por ejemplo, en Yaglom (1986, vol. 1, cap. 3, secciones 16 y 17) y Christakos (1992, cap. 2, sección 12) se tiene una discusión más detallada sobre estas propiedades. Es importante destacar entre los resultados obtenidos que, en el caso de procesos normales, la condición (2.5) es suficiente para que el proceso sea ergódico.

En algunas ocasiones en geoestadística es imposible extender de forma infinita el dominio y, en lugar de considerar el comportamiento asintótico cuando el dominio tiende a infinito (“increase-domain asymptotics”), interesa estudiar el comportamiento asintótico del proceso cuando el dominio D donde está definido el proceso permanece fijo y la densidad de muestreo tiende a infinito³ (“infill asymptotics”). En estos casos solamente es de interés el comportamiento del variograma (o covariograma) cerca del origen y por lo tanto es suficiente con que se verifique algún tipo de ergodicidad cerca del origen. Este tipo de ergodicidad fue denominado por Matheron (1978) *micro-ergodicidad* (también se denomina quasi-ergodicidad). Para más detalles sobre esta clase de

³ Para más detalles sobre los dos tipos de conductas asintóticas tradicionalmente consideradas en geoestadística ver p.e. Cressie (1993, sección 5.8).

ergodicidad ver p.e. Christakos (1992, cap. 2, ejemplo 2) o Chilès y Delfiner (1999, sección 2.9.2).

Más recientemente se ha considerado una combinación de las dos conductas asintóticas tradicionales (e.g. Hall et al., 1994; García-Soidán et al., 2003a y 2003b; Lahiri et al., 2003), y se han estudiado comportamientos asintóticos cuando simultáneamente el dominio se extiende de forma infinita y el número de observaciones tiende a infinito, de forma que la densidad de muestreo sobre cualquier volumen fijo tiende también a infinito (podríamos decir que el relleno tiene prioridad sobre la expansión del dominio).

En la práctica, la hipótesis de ergodicidad no es realmente un problema y simplemente se asume que el proceso es ergódico en media y en covarianza o variograma, según convenga.

PROCESOS AGREGADOS

En algunos casos los datos pueden ser agregaciones espaciales en lugar de observaciones puntuales (e incluso observaciones sobre distintos soportes) o, por ejemplo, puede ser de interés la estimación de medias espaciales a partir de datos puntuales. Estas agregaciones pueden ser modeladas como el promedio de un proceso puntual, lo que permite deducir fácilmente las relaciones entre covariogramas y variogramas vinculados a diferentes soportes.

Supongamos que el proceso espacial $Z(\cdot)$ definido sobre $D \subset \mathbb{R}^d$ es integrable en media cuadrática. Entonces, si $B \subset D$ es un subconjunto acotado e integrable, se puede definir el *proceso espacial agregado* (también se denomina regularizado) como:

$$Z(B) \equiv \begin{cases} \frac{1}{|B|} \int_B Z(\mathbf{s}) d\mathbf{s} & \text{si } |B| > 0 \\ \text{media} \{ Z(\mathbf{s}) : \mathbf{s} \in B \} & \text{si } |B| = 0 \end{cases}$$

siendo $|B| = \int_B d\mathbf{s}$.

Si por ejemplo el proceso puntual es intrínsecamente estacionario con semivariograma $\gamma(\cdot)$, entonces a partir del variograma puntual podemos obtener el variograma del proceso agregado:

$$\begin{aligned} \text{Var}(Z(B_1) - Z(B_2)) &= -\frac{1}{|B_1|^2} \int_{B_1} \int_{B_1} \gamma(\mathbf{s} - \mathbf{u}) d\mathbf{s} d\mathbf{u} \\ &\quad - \frac{1}{|B_2|^2} \int_{B_2} \int_{B_2} \gamma(\mathbf{s} - \mathbf{u}) d\mathbf{s} d\mathbf{u} \\ &\quad + \frac{1}{|B_1||B_2|} \int_{B_1} \int_{B_2} 2\gamma(\mathbf{s} - \mathbf{u}) d\mathbf{s} d\mathbf{u}. \end{aligned}$$

Aunque nos centraremos en el caso de soporte puntual, los métodos descritos en este trabajo pueden ser extendidos para el caso de distintos soportes (para más detalles sobre el problema de cambio de soporte ver p.e. Cressie, 1993, sección 5.2 ó Chilès y Delfiner, 1999, sección 2.4).

2.2 PROPIEDADES DEL COVARIOGRAMA Y DEL VARIOGRAMA

El variograma y el covariograma son las funciones habitualmente consideradas en geoestadística para el modelado de la dependencia espacial o espacio-temporal, y son consideradas como un parámetro (de especial interés) del proceso. En la práctica normalmente se suele utilizar el variograma, no sólo porque es más general (puede existir en casos en que el covariograma no), sino por las ventajas en su estimación (sección 4.1.1). No obstante, en muchos casos se recurre a la relación (2.3) para construir modelos de variograma a partir de covariogramas, por lo que es importante también el estudio de estas funciones.

2.2.1 Propiedades elementales

El variograma y el covariograma deben verificar ciertas propiedades que sus estimadores no siempre verifican, entre ellas:

PROPIEDADES ELEMENTALES DEL COVARIOGRAMA

Si $Z(\cdot)$ es un proceso estacionario de segundo orden con covariograma $C(\cdot)$, entonces se verifica⁴ que:

⁴ Realmente sólo en el caso de procesos reales el covariograma es necesariamente una función par.

$$C(\mathbf{0}) = \text{Var}(Z(\mathbf{s})) \geq 0,$$

$$C(\mathbf{h}) = C(-\mathbf{h}),$$

y por la desigualdad de Cauchy-Schwarz:

$$|C(\mathbf{h})| \leq C(\mathbf{0}). \quad (2.6)$$

Además, el covariograma debe ser semidefinido positivo, es decir:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j C(\mathbf{s}_i - \mathbf{s}_j) \geq 0 \quad (2.7)$$

$$\forall m \geq 1, \forall \mathbf{s}_i \in D, \forall a_i \in \mathbb{R}; i = 1, \dots, m,$$

ya que:

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

La condición (2.7) es necesaria y suficiente para que exista un proceso estacionario de segundo orden con covariograma $C(\cdot)$ (se puede construir un proceso normal multivariante con covarianzas definidas por $C(\cdot)$). Por tanto la clase de covariogramas válidos en \mathbb{R}^d es equivalente a la clase de funciones semidefinidas positivas en \mathbb{R}^d .

Algunas propiedades adicionales que verifican los covariogramas son las siguientes⁵:

1. Si $C(\cdot)$ es un covariograma válido en \mathbb{R}^d , entonces $aC(\cdot)$, $\forall a \geq 0$, es también un covariograma válido en \mathbb{R}^d .
2. Si $C_1(\cdot)$ y $C_2(\cdot)$ son covariogramas válidos en \mathbb{R}^d , entonces $C_1(\cdot) + C_2(\cdot)$ es un covariograma válido en \mathbb{R}^d .
3. Si $C_1(\cdot)$ y $C_2(\cdot)$ son covariogramas válidos en \mathbb{R}^d , entonces $C(\cdot) = C_1(\cdot)C_2(\cdot)$ es un covariograma válido en \mathbb{R}^d .

⁵ Se han derivado también relaciones entre las propiedades de continuidad (y derivabilidad) del covariograma en el origen y las propiedades de continuidad (y diferenciabilidad) en media cuadrática del proceso $Z(\cdot)$. Tales como que un proceso estacionario de segundo orden es continuo en media cuadrática si y sólo si su covariograma es continuo en el origen. Para más detalles ver p.e. Stein (1999, sección 2.4).

4. Si $C_k(\cdot), \forall k \in \mathbb{N}$, son covariogramas válidos en \mathbb{R}^d y $\lim_{k \rightarrow \infty} C_k(\mathbf{h}) = C(\mathbf{h}), \forall \mathbf{h} \in \mathbb{R}^d$, entonces $C(\cdot)$ es un covariograma válido en \mathbb{R}^d .
5. Si $C(\mathbf{h}|\theta)$ es un covariograma válido en \mathbb{R}^d , $\forall \theta \in A \subset \mathbb{R}$ y $d\mu$ es una medida positiva definida sobre A , entonces:

$$\int_A C(\mathbf{h}|\theta) d\mu(\theta) \quad (2.8)$$

es un covariograma válido en \mathbb{R}^d , siempre que la integral exista $\forall \mathbf{h} \in \mathbb{R}^d$.

6. Un covariograma isotrópico válido en \mathbb{R}^d es también un covariograma isotrópico válido en \mathbb{R}^m , $\forall m \leq d$ (el recíproco no es en general cierto, ver p.e. Cressie, 1993, p. 84).

La propiedad 3 está relacionada con el concepto de separabilidad, equivale a suponer que el proceso $Z(\cdot)$ se obtiene como producto de dos procesos estacionarios de segundo orden independientes: $Z(\mathbf{s}) = Z_1(\mathbf{s})Z_2(\mathbf{s})$, con covariogramas $C_1(\cdot)$ y $C_2(\cdot)$ respectivamente. La propiedad 6 se deduce de forma inmediata teniendo en cuenta que si consideramos solamente m coordenadas de un proceso intrínsecamente estacionario de segundo orden en \mathbb{R}^d ($m < d$), se obtiene un proceso intrínsecamente estacionario de segundo orden en \mathbb{R}^m (esta propiedad se puede generalizar también para el caso no isotrópico). Las demás propiedades se deducen de la definición de función semidefinida positiva.

PROPIEDADES ELEMENTALES DEL VARIOGRAMA

Si $\gamma(\cdot)$ es el semivariograma de un proceso intrínsecamente estacionario $Z(\cdot)$, entonces se verifica que:

$$\gamma(\mathbf{0}) = 0,$$

$$\gamma(\mathbf{h}) = \gamma(-\mathbf{h}),$$

$$\gamma(\mathbf{h}) \geq 0.$$

El semivariograma debe ser además condicionalmente semidefinido negativo, es decir:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0 \tag{2.9}$$

$$\forall m \geq 1, \forall \mathbf{s}_i \in D, \forall a_i \in \mathbb{R}; i = 1, \dots, m, \text{ tales que } \sum_{i=1}^m a_i = 0.$$

Esto se deduce teniendo en cuenta que:

$$\left(\sum_{i=1}^m a_i Z(\mathbf{s}_i) \right)^2 = -\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m a_i a_j (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2,$$

y tomando esperanzas:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) = -Var \left\{ \sum_{i=1}^m a_i Z(\mathbf{s}_i) \right\} \leq 0,$$

utilizando en ambos casos que $\sum_{i=1}^m a_i = 0$.

La condición (2.9) es necesaria pero no suficiente (aunque pocas condiciones adicionales son necesarias para que el recíproco sea cierto), ver sección 2.2.3. Algunas propiedades adicionales que verifica un variograma son las siguientes:

1. Si $\gamma(\cdot)$ es un semivariograma válido en \mathbb{R}^d , entonces $a\gamma(\cdot)$, $\forall a \geq 0$, es también un semivariograma válido en \mathbb{R}^d .
2. Si $\gamma_1(\cdot)$ y $\gamma_2(\cdot)$ son semivariogramas válidos en \mathbb{R}^d , entonces $\gamma_1(\cdot) + \gamma_2(\cdot)$, es también un semivariograma válido en \mathbb{R}^d .
3. El semivariograma debe crecer más lentamente que $\|\mathbf{h}\|^2$ (Matheron, 1971), es decir:

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

4. Un variograma isotrópico válido en \mathbb{R}^d es también un variograma isotrópico válido en \mathbb{R}^m , $\forall m \leq d$.

La propiedad 2 equivale a suponer que el proceso $Z(\cdot)$ se obtiene como suma de dos procesos intrínsecamente estacionarios independientes: $Z(\mathbf{s}) = Z_1(\mathbf{s}) + Z_2(\mathbf{s})$, con semivariogramas $\gamma_1(\cdot)$ y $\gamma_2(\cdot)$ respectivamente⁶.

⁶ Se suele recurrir a esta idea para la obtención de modelos de variograma válidos, como p.e. en el caso de anisotropía zonal (sección 2.2.2) o en el caso del modelo lineal

CARACTERÍSTICAS DEL VARIOGRAMA (COVARIOGRAMA)

Hay ciertas características geométricas del variograma (o del covariograma) de especial importancia⁷, las principales se comentan a continuación.

Supongamos que $Z(\cdot)$ es un proceso intrínsecamente estacionario con semivariograma $\gamma(\cdot)$.

Umbral

Si el variograma está acotado y :

$$\lim_{\|\mathbf{h}\| \rightarrow \infty} \gamma(\mathbf{h}) = \sigma^2,$$

a σ^2 se le denomina *umbral del semivariograma*.

Si el proceso $Z(\cdot)$ es estacionario de segundo orden y $\lim_{\|\mathbf{h}\| \rightarrow \infty} C(\mathbf{h}) = 0$ (por ejemplo cuando se trata de un proceso normal ergódico) entonces $\sigma^2 = C(\mathbf{0})$.

Rango

Si σ^2 es el umbral del semivariograma (suponiendo que existe), se define el *rango del semivariograma* en la dirección $\mathbf{e}_0 = \mathbf{h}_0 / \|\mathbf{h}_0\| \in \mathbb{R}^d$ como:

$$r_0 = \min \{ r : \gamma(r(1 + \varepsilon)\mathbf{e}_0) = \sigma^2, \forall \varepsilon > 0 \}.$$

El rango en la dirección \mathbf{e}_0 puede interpretarse como el salto h a partir del cual no hay correlación entre $Z(\mathbf{s})$ y $Z(\mathbf{s} \pm h\mathbf{e}_0)$, por tanto está íntimamente ligado a la noción de “zona de influencia” (y va a tener un papel importante en la determinación de criterios de vecindad; sección 3.4.3).

de (co)regionalización (secciones 4.2.3 y 5.3.2).

⁷ Además de poder interpretar su influencia en la predicción espacial (sección 3.4.2), son utilizadas en la parametrización de la mayoría de los modelos de variogramas o covariogramas (sección 4.2.2).

Efecto Nugget (Pepita)

Como ya se comentó anteriormente, siempre se verifica que $\gamma(\mathbf{0}) = 0$, sin embargo puede ser que:

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \gamma(\mathbf{h}) = c_0 > 0,$$

entonces c_0 se denomina *efecto nugget*⁸ (Matheron, 1962). Además, si σ^2 es el umbral del semivariograma (suponiendo que existe), a $\sigma^2 - c_0$ se le denomina umbral parcial.

Una condición necesaria y suficiente para que el proceso $Z(\cdot)$ sea continuo en media cuadrática (i.e. $\lim_{\mathbf{h} \rightarrow \mathbf{0}} E(Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}))^2 = 0$) es que el variograma sea continuo en el origen⁹, ya que en el caso de procesos intrínsecamente estacionarios:

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

Entonces la presencia de efecto nugget indica que (en teoría) el proceso no es continuo y por tanto altamente irregular.

En general, podemos pensar que la presencia de un efecto nugget es debido a la existencia de un proceso de microescala con rango inferior a la mínima distancia entre las posiciones muestrales¹⁰ o a error de medida. Por ejemplo, puede ser de gran utilidad considerar el siguiente modelo (Cressie, 1993, pp. 112-113) para el proceso $Z(\cdot)$:

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + W(\mathbf{s}) + \eta(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad (2.10)$$

donde:

⁸ El origen de esta denominación está relacionado con la terminología minera. En algunos yacimientos de metal, como por ejemplo en el caso del oro, el mineral suele obtenerse como pepitas de material puro y estas pepitas normalmente son más pequeñas que el tamaño de la unidad de muestreo (lo que produce una variabilidad adicional en la muestra).

⁹ Análogamente al caso del covariograma, se han derivado también relaciones entre las propiedades de derivabilidad del variograma en el origen y las propiedades de diferenciabilidad en media cuadrática del proceso $Z(\cdot)$; por ejemplo en Chilès y Delfiner (1999, sección 2.3.1) se tienen algunas.

¹⁰ Cuando el rango del proceso de microescala es inferior al tamaño de la unidad de muestreo es el caso de un “verdadero” efecto nugget.

- $\mu(\mathbf{s}) = E(Z(\mathbf{s}))$ es la estructura de la media (determinística), también denominada variación de larga escala.
- $W(\mathbf{s})$ es un proceso intrínsecamente estacionario de media cero, continuo (en media cuadrática) y cuyo variograma tiene rango, si existe, mayor que $m = \min \{ \|\mathbf{s}_i - \mathbf{s}_j\| : 1 \leq i < j \leq n \}$. Este proceso se suele denominar variación de pequeña escala.
- $\eta(\mathbf{s})$ es un proceso intrínsecamente estacionario de media cero, independiente de $W(\mathbf{s})$ y cuyo variograma tiene rango (existe) menor que m . Este proceso se suele denominar variación de micro-escala.
- $\varepsilon(\mathbf{s})$ es un proceso ruido blanco de media cero, independiente de $W(\mathbf{s})$ y $\eta(\mathbf{s})$, y con varianza c_{EM} . Este proceso se denomina también error de medida o simplemente ruido.

A partir de la descomposición (2.10) obtenemos que:

$$\gamma_Z(\cdot) = \gamma_W(\cdot) + \gamma_\eta(\cdot) + c_{EM}.$$

Además, si el efecto nugget del proceso de micro-escala es c_{MS} , entonces:

$$c_0 = c_{MS} + c_{EM},$$

ya que el efecto nugget del proceso de pequeña escala es nulo. La mayoría de los autores (p.e. Matheron) generalmente suponen que la variación de micro-escala no es continua; sin embargo, si se espera la continuidad del proceso en la micro-escala (i.e. $c_{MS} = 0$) la única razón posible para $c_0 > 0$ es error de medida. En la práctica sólo se conocen los datos $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$ y no se sabe nada del variograma a distancias menores que $m = \min \{ \|\mathbf{s}_i - \mathbf{s}_j\| : 1 \leq i < j \leq n \}$. Por tanto, a no ser que algunas posiciones espaciales estén muy próximas, no se puede distinguir si la variación de micro-escala es continua o no.

Supongamos que σ_W^2 es el umbral del variograma del proceso $W(\cdot)$ y σ_η^2 el del proceso $\eta(\cdot)$ entonces:

$$\sigma_Z^2 = \sigma_W^2 + \sigma_\eta^2 + c_{EM}.$$

En la práctica no es fácil determinar el efecto nugget a partir de datos con posiciones espaciales muy separadas. Una estimación de c_0 se obtiene extrapolando un variograma experimental cerca del origen, aunque así estamos

estimando en realidad $\sigma_\eta^2 + c_{EM}$. Por este motivo, en muchos casos se suele denominar umbral parcial a $\sigma_W^2 = \sigma_Z^2 - \sigma_\eta^2 - c_{EM}$ en lugar de a $\sigma_Z^2 - c_0$.

La proporción del efecto nugget en el umbral total c_0/σ^2 proporciona mucha información acerca del grado de dependencia presente en los datos. Por ejemplo, en el caso en que toda la variabilidad es efecto nugget (i.e. $\gamma(\mathbf{h}) = c_0$, $\forall \mathbf{h} \neq \mathbf{0}$) entonces $Z(\mathbf{s}_1)$ y $Z(\mathbf{s}_2)$ son incorrelados $\forall \mathbf{s}_1, \mathbf{s}_2 \in D$ independientemente de lo cerca que estén (el proceso $Z(\cdot)$ es ruido blanco).

En la figura 2.2 se muestran las relaciones entre las distintas características del semivariograma.

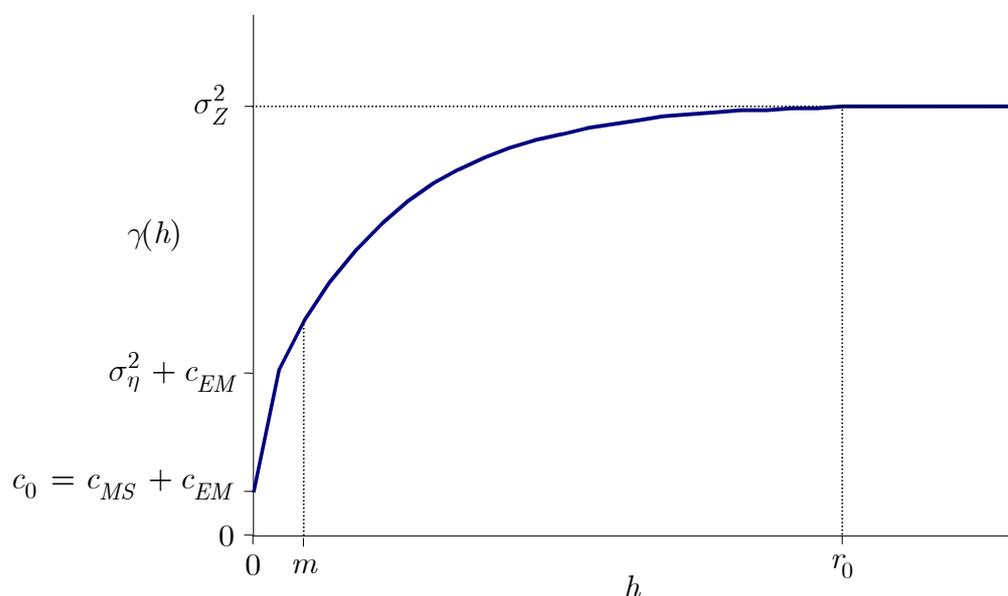


Figura 2.2: Semivariograma genérico en \mathbb{R}^1 .

2.2.2 Representación espectral

Los métodos espectrales son una herramienta de gran utilidad en el estudio de procesos espaciales (y espacio-temporales). En esta sección solamente se mostrarán algunos resultados relativos a la representación de variogramas y covariogramas continuos (centrándonos principalmente en este último caso). Un tratamiento más detallado se tiene por ejemplo en los libros de Yaglom (1986) y Christakos (1992), otras referencias donde se trata con bastante detalle este

tema son por ejemplo Stein (1999, secciones 2.5 a 2.10) y Chilès y Delfiner (1999, sección 2.3.3).

REPRESENTACIÓN ESPECTRAL DEL COVARIOGRAMA

Como ya se comentó en la sección 2.2.1, la clase de covariogramas válidos en \mathbb{R}^d es equivalente a la clase de funciones semidefinidas positivas en \mathbb{R}^d . Podemos caracterizar las funciones de este tipo a partir de su representación espectral (resultado que facilita notablemente la obtención de modelos válidos de covariogramas).

Supongamos por el momento que $Z(\cdot)$ es un proceso espacial (real) en \mathbb{R}^d estacionario de segundo orden con covariograma $C(\cdot)$ continuo en el origen.

Por el teorema de Bochner (Bochner, 1955), como $C(\cdot)$ es una función semidefinida positiva y continua en el origen, admite una representación de la forma:

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} e^{i(\boldsymbol{\omega} \cdot \mathbf{h})} dF(\boldsymbol{\omega}), \quad (2.11)$$

donde dF es una medida positiva finita, siendo el recíproco también cierto.

Si la función F es diferenciable, la ecuación (2.11) puede expresarse como:

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} e^{i(\boldsymbol{\omega} \cdot \mathbf{h})} f(\boldsymbol{\omega}) d\boldsymbol{\omega},$$

siendo $f(\boldsymbol{\omega})$ la función de densidad espectral¹¹:

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(\boldsymbol{\omega} \cdot \mathbf{h})} C(\mathbf{h}) d\mathbf{h},$$

verificando además que $f(\boldsymbol{\omega}) \geq 0$.

Como en el caso de procesos reales el covariograma es una función par (i.e. $C(\mathbf{h}) = C(-\mathbf{h})$), podemos sustituir en las expresiones anteriores los factores exponenciales por cosenos, obteniendo entonces que:

$$C(\mathbf{h}) = \int_{\mathbb{R}^d} \cos(\boldsymbol{\omega} \cdot \mathbf{h}) f(\boldsymbol{\omega}) d\boldsymbol{\omega}, \quad (2.12)$$

¹¹ A partir de esta expresión podemos pensar que $f/C(\mathbf{0})$ es la función de densidad de probabilidad de un vector aleatorio con función característica $C/C(\mathbf{0})$, lo que nos brinda otra alternativa para la obtención de covariogramas válidos.

y la función de densidad espectral:

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \cos(\boldsymbol{\omega} \cdot \mathbf{h}) C(\mathbf{h}) d\mathbf{h}, \quad (2.13)$$

es también una función par.

Si el covariograma es isotrópico:

$$C(\mathbf{h}) \equiv C(r); \quad r = \|\mathbf{h}\|,$$

entonces:

$$f(\boldsymbol{\omega}) \equiv f(\lambda); \quad \lambda = \|\boldsymbol{\omega}\|.$$

Las ecuaciones (2.12) y (2.13) pueden transformarse (pasando a coordenadas esféricas, ver p.e. Stein, 1999, pp. 43-44) en sus correspondientes expresiones isotrópicas:

$$C(r) = (2\pi)^{d/2} \int_0^\infty \frac{J_{(d-2)/2}(\lambda r)}{(\lambda r)^{(d-2)/2}} f(\lambda) \lambda^{d-1} d\lambda, \quad (2.14)$$

$$f(\lambda) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \frac{J_{(d-2)/2}(\lambda r)}{(\lambda r)^{(d-2)/2}} C(r) r^{d-1} dr,$$

donde $J_v(\cdot)$ es la función de Bessel de orden v , que puede expresarse (ver Abramovitz y Stegun, 1965, 9.1.20 y 9.1.10) como:

$$\begin{aligned} J_v(z) &= \frac{\left(\frac{1}{2}z\right)^v}{\pi^{1/2} \Gamma(v + \frac{1}{2})} \int_0^\pi \cos(z \cos \theta) \sin^{2v} \theta d\theta \\ &= \left(\frac{1}{2}z\right)^v \sum_{k=0}^\infty \frac{\left(-\frac{1}{4}z^2\right)^k}{k! \Gamma(v + k + 1)}. \end{aligned} \quad (2.15)$$

A partir de esto, podemos pensar en rescribir (2.14) para que sea más manejable definiendo:

$$\kappa_d(x) = \left(\frac{2}{x}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) J_{(d-2)/2}(x), \quad (2.16)$$

y obtendríamos:

$$\begin{aligned} C(r) &= \frac{2\pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \int_0^\infty \kappa_d(\lambda r) f(\lambda) \lambda^{d-1} d\lambda \\ &= \int_0^\infty \kappa_d(\lambda r) dG_d(\lambda), \end{aligned} \quad (2.17)$$

donde G_d es una función acotada no decreciente en $[0, \infty)$, de la forma:

$$G_d(\lambda) = \int_{\|\omega\| < \lambda} dF(\omega),$$

siendo dF una medida positiva simétrica.

Además, teniendo en cuenta (2.15) y (2.16), se puede ver que $\kappa_d(0) = 1$, y por tanto:

$$C(0) = \int_0^\infty dG_d(\lambda).$$

En general las expresiones de la ecuación (2.17) incluyen funciones trigonométricas cuando d es impar y funciones de Bessel de orden entero cuando d es par. Por ejemplo, simplificando la notación, en los casos particulares de $d = 1, 2$ y 3 obtendríamos:

- $d = 1$: $\kappa_1(x) \equiv \cos x$, $C(r) = \int_0^\infty \cos(\lambda r) dG(\lambda)$.
- $d = 2$: $\kappa_2(x) \equiv J_0(x)$, $C(r) = \int_0^\infty J_0(\lambda r) dG(\lambda)$.
- $d = 3$: $\kappa_3(x) \equiv \frac{\text{sen } x}{x}$, $C(r) = \int_0^\infty \frac{\text{sen } \lambda r}{\lambda r} dG(\lambda)$.

Como se vio en la sección 2.2.1, un covariograma isotrópico válido en \mathbb{R}^d es también un covariograma isotrópico válido en \mathbb{R}^m , $\forall m \leq d$. Además, se puede deducir de los resultados anteriores (teniendo en cuenta que $\kappa_d((2d)^{1/2}x) \rightarrow e^{-x^2}$ uniformemente en x cuando $d \rightarrow \infty$), que una función es un covariograma válido en cualquier dimensión si y sólo si admite una representación de la forma¹²:

$$C(r) = \int_0^\infty e^{-\lambda^2 r^2} dG(\lambda),$$

siendo G una función acotada no decreciente en $[0, \infty)$ (ver p.e. Stein, 1999, pp. 44-45). Por lo tanto denotaremos:

$$\kappa_\infty(x) \equiv e^{-x^2}. \quad (2.18)$$

¹² Otra condición necesaria y suficiente es que $g(r) = C(r^{1/2})$ sea completamente monótona para $r \geq 0$ (Schoenberg, 1938, p.821); en la sección 6.2 se tiene la definición de este tipo de funciones.

En general, el covariograma puede no ser continuo en el origen (presencia de efecto nugget); en ese caso, podemos expresarlo como:

$$C(\mathbf{h}) = c_0 \mathcal{I}_{\{\mathbf{0}\}}(\mathbf{h}) + C^0(\mathbf{h}), \quad (2.19)$$

donde $c_0 \geq 0$ es el efecto nugget, $\mathcal{I}_{\{\mathbf{0}\}}(\cdot)$ es la función indicadora del origen y $C^0(\cdot)$ es un covariograma continuo en el origen que admite la correspondiente representación espectral. A partir de (2.19) y (2.3) se deduce que la expresión del correspondiente semivariograma es

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C^0(\mathbf{h}) = c_0 + (C^0(\mathbf{0}) - C^0(\mathbf{h})), \quad (2.20)$$

para $\mathbf{h} \neq \mathbf{0}$ (hay que recordar que siempre se verifica que $\gamma(\mathbf{0}) = 0$).

REPRESENTACIÓN ESPECTRAL DEL SEMIVARIOGRAMA

Como se mostró en la sección 2.2.1 el semivariograma de un proceso intrínsecamente estacionario es necesariamente condicionalmente semidefinido negativo. Para determinar las condiciones en las que el recíproco es cierto, es de utilidad el siguiente resultado:

Si $g(\cdot)$ es una función continua en \mathbb{R}^d verificando $g(\mathbf{0}) = 0$, entonces las siguientes afirmaciones son equivalentes (Schoenberg, 1938; Neuman y Schoenberg, 1941):

1. $g(\mathbf{h})$ es condicionalmente semidefinida negativa.
2. $e^{-tg(\mathbf{h})}$ es semidefinida positiva (i.e. un covariograma) $\forall t > 0$.
3. $g(\cdot)$ es de la forma:

$$g(\mathbf{h}) = \int_{\mathbb{R}^d} \frac{1 - \cos(\boldsymbol{\omega} \cdot \mathbf{h})}{\|\boldsymbol{\omega}\|^2} dF(\boldsymbol{\omega}) + Q(\mathbf{h}), \quad (2.21)$$

donde $Q(\cdot) \geq 0$ es una forma cuadrática y dF es una medida positiva, simétrica y continua en el origen que verifica:

$$\int_{\mathbb{R}^d} \frac{1}{1 + \|\boldsymbol{\omega}\|^2} dF(\boldsymbol{\omega}) < \infty.$$

A partir de este resultado, puede deducirse (p.e. Cressie 1993, pp. 87-88) que si $\gamma(\cdot)$ es una función condicionalmente semidefinida negativa con¹³ $Q(\cdot) = 0$

¹³ Algunos autores definen el semivariograma como $\gamma(\mathbf{h}) = \frac{1}{2} E (Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}))^2$, en cuyo caso puede ser que $Q(\cdot) > 0$.

en la representación (2.21), entonces es el variograma de un proceso intrínsecamente estacionario en \mathbb{R}^d .

Análogamente al caso anterior, si el semivariograma es isotrópico:

$$\gamma(\mathbf{h}) \equiv \gamma(r); \quad r = \|\mathbf{h}\|,$$

la representación espectral isotrópica correspondiente a (2.21) es:

$$\gamma(r) = \int_0^\infty \frac{1 - \kappa_d(\lambda r)}{\lambda^2} dG(\lambda), \quad (2.22)$$

donde $\kappa_d(\cdot)$ está definida por (2.16) y G es una función acotada no decreciente en $[0, \infty)$ que verifica:

$$\int_0^\infty \frac{1}{1 + \lambda^2} dG(\lambda) < \infty. \quad (2.23)$$

Además, en las condiciones anteriores de la representación (2.22), se puede ver (Christakos, 1984) que es equivalente a la restricción (2.23) que se verifique que:

$$\lim_{r \rightarrow \infty} \frac{\gamma(r)}{r^2} = 0.$$

Para el caso general en el que el semivariograma no es necesariamente continuo en el origen, basta tener en cuenta que podemos expresarlo (utilizando la misma notación que en (2.19)) como:

$$\gamma(\mathbf{h}) = c_0(1 - \mathcal{I}_{\{\mathbf{0}\}}(\mathbf{h})) + \gamma^0(\mathbf{h}),$$

siendo $\gamma^0(\cdot)$ un semivariograma continuo en el origen.

2.2.3 Anisotropía

La hipótesis de isotropía simplifica notablemente el modelado de la dependencia espacial por lo que la mayoría de los modelos (básicos) de semivariogramas considerados en geoestadística son isotrópicos (ver sección 4.2). Sin embargo, en muchos casos no se puede asumir que la dependencia es igual en cualquier dirección (uno de los ejemplos más claros es el caso espacio-temporal, donde en principio no es lógico pensar que un salto espacial es equivalente a un salto temporal). En esos casos se suelen considerar ligeras variaciones de la hipótesis de isotropía para modelar la dependencia espacial. En

esta sección se comentan brevemente las distintas aproximaciones tradicionalmente consideradas en geoestadística (para más detalles ver p.e. Chilès y Delfiner, 1999, sección 2.5.2, o Goovaerts, 1997, sección 4.2.2), otras aproximaciones distintas se tratan en la sección 4.5 y en el capítulo 6 (caso espacio-temporal).

Cuando el variograma es función de la dirección además de la magnitud del salto, se dice que es el variograma es *anisotrópico* (no isotrópico). Los tipos de anisotropía habitualmente considerados son los siguientes:

ANISOTROPÍA GEOMÉTRICA

En algunos casos, cuando aparentemente el umbral permanece constante mientras que el rango varía con la dirección, se puede corregir esta clase de anisotropía mediante una transformación lineal del vector de salto \mathbf{h} , i.e.:

$$\gamma(\mathbf{h}) = \gamma^0(\|\mathbf{A}\mathbf{h}\|), \forall \mathbf{h} \in \mathbb{R}^d,$$

siendo \mathbf{A} una matriz $d \times d$ y $\gamma^0(\cdot)$ un semivariograma isotrópico. En este caso se dice que el variograma es geoméricamente anisotrópico. Por ejemplo, en el caso bidimensional, se suelen considerar una matriz de la forma:

$$\mathbf{A} = \begin{pmatrix} b_1 & 0 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} \cos \theta & \text{sen } \theta \\ -\text{sen } \theta & \cos \theta \end{pmatrix},$$

que se corresponde con las direcciones principales de anisotropía θ y $\theta + \pi/2$ (normalmente se toma θ igual a la dirección de máximo rango). Esto puede extenderse fácilmente para el caso tridimensional (ver p.e. Chilès y Delfiner, 1999, pp. 94-95).

Esta idea (que el espacio euclídeo no es apropiado para medir distancias entre posiciones espaciales pero una transformación lineal de él sí) ha sido también generalizada para el caso de deformaciones no lineales del espacio. Por ejemplo, Sampson y Guttorp (1992) consideraron transformaciones no lineales $f(\cdot)$ (obtenidas mediante técnicas de escalamiento óptimo multidimensional) de forma que:

$$\gamma(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{2} \text{Var}(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = \gamma^0(|f(\mathbf{s}_1) - f(\mathbf{s}_2)|), \forall \mathbf{s}_1, \mathbf{s}_2 \in D,$$

siendo $\gamma^0(\cdot)$ un semivariograma isotrópico.

ANISOTROPÍA ZONAL

Cuando el umbral del semivariograma varía con la dirección se dice que hay anisotropía zonal (también se denomina anisotropía estratificada). En estos casos se suele considerar una combinación de un semivariograma isotrópico más otros “zonales” que depende solamente de la distancia en ciertas direcciones (o componentes del vector de salto). Por ejemplo, en el caso bidimensional, si ϕ es la dirección de mayor varianza se suele considerar una combinación de la forma:

$$\gamma(\mathbf{h}) = \gamma_1(\|\mathbf{h}\|) + \gamma_2(h_\phi),$$

siendo $\gamma_1(\cdot)$ y $\gamma_2(\cdot)$ semivariogramas isotrópicos y $h_\phi = \cos(\phi)h_1 + \sin(\phi)h_2$ el salto en la dirección ϕ , para $\mathbf{h} = (h_1, h_2) \in \mathbb{R}^2$.

Es importante destacar que este tipo de anisotropías pueden causar la aparición de problemas al realizar predicción espacial (ver p.e. Myers y Journel, 1990; y Rouhani y Myers, 1990), como por ejemplo dar lugar a sistemas kriging no válidos con ciertas configuraciones de los datos. Hay que tener un especial cuidado cuando el covariograma es expresado como suma de covariogramas unidimensionales, en cuyo caso el resultado puede ser únicamente condicionalmente semidefinido positivo sobre un dominio multidimensional (en el capítulo 3 se tienen más detalles).

Una variante de la anisotropía zonal es el caso de covariogramas separables (también denominados factorizables) en componentes del vector de salto. Por ejemplo, un covariograma completamente separable en \mathbb{R}^d es de la forma:

$$C(\mathbf{h}) = \prod_{i=1}^d C_i(h_i),$$

para $\mathbf{h} = (h_1, \dots, h_d)' \in \mathbb{R}^d$ y siendo $C_i(\cdot), i = 1, \dots, d$, covariogramas en \mathbb{R}^1 . En este caso se puede pensar que el proceso espacial se obtiene como producto de d procesos unidimensionales independientes definidos sobre cada uno de los ejes de coordenadas

CAPÍTULO 3

PREDICCIÓN ESPACIAL LINEAL ÓPTIMA. KRIGING

En este capítulo se comentan brevemente los métodos más conocidos de predicción espacial denominados métodos kriging¹ (ver sección 1.3 para un resumen del origen de esta terminología), centrándonos únicamente en el caso de predicción lineal puntual univariante (el caso multivariante se trata en el capítulo 5). Una revisión más completa de estos métodos se tiene por ejemplo en Cressie (1993, capítulo 3 y secciones 5.1, 5.4 y 5.9.1) o Chilès y Delfiner (1999, capítulos 3, 4 y 6).

3.1 INTRODUCCIÓN

Sea $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$ un proceso espacial y $\{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$ n valores observados de este proceso, y supongamos² que a partir de estas observaciones se trata de predecir $Z(\mathbf{s}_0)$.

Si denotamos por $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$, los distintos métodos kriging proporcionan un predictor $p(\mathbf{Z}, \mathbf{s}_0)$ de $Z(\mathbf{s}_0)$ verificando que:

- es lineal (homogéneo o heterogéneo) en \mathbf{Z} :

¹ Podríamos definir los métodos kriging como algoritmos de predicción de mínimo error en media cuadrática que tienen en cuenta la estructura de 2º orden del proceso.

² El caso general sería la predicción de $g(Z(\cdot)) = \int_{\mathbb{R}^d} Z(\mathbf{s}) dp(\mathbf{s})$ para una medida integrable $dp(\cdot)$ (p.e. $g(Z(\cdot)) = Z(B_0)$) a partir de $\{z(B_1), \dots, z(B_n)\}$, que se resuelve de forma análoga (ver comentarios al final de la sección 2.1).

$$p(\mathbf{Z}, \mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) + \lambda_0, \quad (3.1)$$

- es uniformemente insesgado, para cualquier $\mu(\cdot)$:

$$E(p(\mathbf{Z}, \mathbf{s}_0)) = \mu(\mathbf{s}_0), \quad (3.2)$$

- y minimiza el error en media cuadrática (e.m.c.) de predicción³:

$$E\left((p(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0))^2\right), \quad (3.3)$$

(al hablar de predicción óptima nos referiremos a que se verifican estas dos últimas condiciones).

Dependiendo de las suposiciones acerca de la función de tendencia $\mu(\cdot)$, se distingue principalmente entre tres métodos kriging:

1. Kriging simple (KS): se supone que la media es conocida (algunos autores suponen también que es constante o incluso cero). Además se asume que el covariograma existe y es conocido.
2. Kriging ordinario (KO): se supone que la media es constante (i.e. $E(Z(\mathbf{s})) = \mu, \forall \mathbf{s} \in D$) y desconocida. Además se asume que por lo menos existe el variograma y es conocido.
3. Kriging universal (KU; también denominado kriging con modelo de tendencia): se supone que la media es desconocida y no constante, pero que es una combinación lineal (desconocida) de $p + 1$ funciones (o variables explicativas) conocidas $\{f_j(\cdot) : j = 0, \dots, p\}$:

$$\mu(\mathbf{s}) = \sum_{j=0}^p f_j(\mathbf{s})\beta_j$$

donde $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)' \in \mathbb{R}^{p+1}$ es un vector desconocido. Se asume también que por lo menos existe el variograma y es conocido (siempre que una de las funciones explicativas sea idénticamente 1, p.e.

³ Como es bien sabido, en el caso de normalidad el predictor óptimo (tomando como función de pérdidas el error cuadrático) es lineal y va a coincidir con los predictores kriging. Pero si el proceso no es normal no tiene porque serlo, lo que ha motivado el desarrollo de métodos kriging no lineales (ver p.e. Rivoirard, 1994) y del kriging trans-normal (ver sección 3.4.4).

$f_0(\cdot) \equiv 1$, en caso contrario las ecuaciones kriging sólo pueden expresarse en función del covariograma).

Por simplicidad el kriging ordinario se tratará en este capítulo como un caso particular del kriging universal (aunque en la práctica se suele pensar en el KO como un método distinto al KU, principalmente por ciertos inconvenientes que presenta este último; ver sección 4.1.2) .

KRIGING CON ERROR DE MEDIDA

Teniendo en cuenta la descomposición (2.10) del proceso $Z(\cdot)$ mostrada al final de la sección 2.2.1, se va a asumir también que no hay error de medida (i.e. $c_{EM} = 0$). En caso contrario sería de interés el proceso libre de ruido $S(\mathbf{s}) = Z(\mathbf{s}) - \varepsilon(\mathbf{s})$, y habría que modificar ligeramente las ecuaciones mostradas en las siguientes secciones para la predicción⁴ de $S(\mathbf{s}_0)$. Para más detalles, ver por ejemplo Cressie (1993, pp. 128-130) o Chilès y Delfiner (1999, sección 3.7.1).

KRIGING CON COVARIANZAS Y VARIOGRAMAS NO ESTACIONARIOS

La suposición de que el variograma (o el covariograma) sólo dependa del salto es conveniente para facilitar el modelado de la dependencia espacial, pero para la predicción espacial no es necesaria esta consideración. Por tanto en las expresiones de las ecuaciones de los distintos métodos kriging se utilizará la notación más general no estacionaria:

$$C(\mathbf{s}_1, \mathbf{s}_2) = Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2)),$$

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

en lugar de suponer que son funciones de $\mathbf{s}_1 - \mathbf{s}_2$.

3.2 KRIGING CON MEDIA CONOCIDA. KRIGING SIMPLE

Supongamos que el proceso $Z(\cdot)$ admite una descomposición de la forma:

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

⁴ En este caso una diferencia interesante es que los correspondientes predictores kriging no son interpoladores exactos; ver sección 3.4.1.

siendo $\mu(\cdot)$ la función de tendencia conocida y $\delta(\cdot)$ un proceso espacial⁵ de media cero con covariograma (no necesariamente estacionario) conocido $C(\mathbf{s}_1, \mathbf{s}_2) = Cov(\delta(\mathbf{s}_1), \delta(\mathbf{s}_2))$.

El predictor óptimo será de la forma (3.1) y tal que minimiza el e.m.c. (3.3), que puede expresarse como:

$$\begin{aligned} E\left((p(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0))^2\right) &= Var(p(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0)) + (E(p(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0)))^2 \\ &= Var\left(Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right) + \left(\mu(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i \mu(\mathbf{s}_i) - \lambda_0\right)^2, \end{aligned}$$

de donde se deduce que:

$$\lambda_0 = \mu(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i \mu(\mathbf{s}_i), \quad (3.4)$$

(por tanto el sesgo debe ser nulo, verificándose la condición (3.2)). Entonces el predictor es de la forma:

$$p(\mathbf{Z}, \mathbf{s}_0) = \mu(\mathbf{s}_0) + \sum_{i=1}^n \lambda_i (Z(\mathbf{s}_i) - \mu(\mathbf{s}_i)),$$

(por tanto se puede pensar que se trata de la estimación lineal homogénea de un proceso de media cero) y el e.c.m. de predicción es igual a:

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

Para minimizar esta función se igualan a cero las derivadas parciales respecto a los pesos, obteniéndose las ecuaciones del kriging simple:

$$\begin{aligned} \sum_{j=1}^n \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - C(\mathbf{s}_i, \mathbf{s}_0) &= 0, \\ i &= 1, \dots, n, \end{aligned}$$

que pueden expresarse en forma matricial como:

$$\Sigma \lambda = \mathbf{c}, \quad (3.5)$$

⁵ En la práctica se supone además (por lo menos en la mayoría de los casos) que es un proceso estacionario de segundo orden.

siendo $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$, $\mathbf{c} = (C(\mathbf{s}_1, \mathbf{s}_0), \dots, C(\mathbf{s}_n, \mathbf{s}_0))'$ y $\boldsymbol{\Sigma}$ la matriz $n \times n$ de varianzas-covarianzas de los datos (i.e. $\Sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$). A partir de (3.4) y (3.5) se obtiene el predictor del kriging simple:

$$p_{KS}(\mathbf{Z}, \mathbf{s}_0) = \mu(\mathbf{s}_0) + \mathbf{c}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\mu}), \quad (3.6)$$

donde $\boldsymbol{\mu} = (\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_n))'$, y el correspondiente mínimo e.m.c. de predicción, también denominado *varianza kriging*:

$$\sigma_{KS}^2(\mathbf{s}_0) = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c}.$$

Una de las principales utilidades de la varianza kriging es la construcción de intervalos de confianza (basados en la hipótesis de normalidad u obtenidos a partir de desigualdades tipo Tchebycheff; ver p.e. Chilès y Delfiner, 1999, sección 3.4.4).

DETALLES PRÁCTICOS Y COMPUTACIONALES

Para que exista una única solución del sistema (3.5) la matriz $\boldsymbol{\Sigma}$ debe ser no singular. Una condición suficiente para que esto ocurra es que el covariograma $C(\cdot, \cdot)$ sea una función semidefinida positiva (hay que tener cuidado con la anisotropía zonal, ver sección 2.2.2) y las posiciones de los datos sean distintas.

En la práctica suele interesar la predicción en múltiples posiciones. Teniendo en cuenta que la matriz del sistema no depende de la posición de predicción, el procedimiento recomendado a seguir sería el siguiente:

- Calcular la factorización Cholesky de la matriz $\boldsymbol{\Sigma}$ (p.e. utilizando la rutina LFTDS de la librería IMSL).
- Para cada posición de predicción \mathbf{s}_0 resolver el sistema (3.5) utilizando la factorización obtenida en el primer paso (p.e. con las rutinas LFSDS o LFIDS de la librería IMSL).

Hay que destacar también que tanto los pesos kriging como la varianza kriging no dependen de los datos observados, solamente de las posiciones y del covariograma (lo que por ejemplo, entre otras cosas, facilita el diseño de la configuración espacial de muestreo). Además, en el caso de que el covariograma sea estacionario, los pesos y varianza kriging son invariantes frente a traslaciones de la configuración espacial (lo que permite, por ejemplo, disminuir

significativamente el tiempo de computación cuando se realizan predicciones locales con datos regularmente espaciados).

3.3 KRIGING CON MEDIA DESCONOCIDA. KRIGING UNIVERSAL

El kriging universal se basa en el siguiente modelo para el proceso $Z(\cdot)$:

$$Z(\mathbf{s}) = \sum_{j=0}^p f_j(\mathbf{s})\beta_j + \delta(\mathbf{s}), \quad (3.7)$$

donde $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)' \in \mathbb{R}^{p+1}$ es un vector desconocido, $\{f_j(\cdot) : j = 0, \dots, p\}$ son funciones conocidas y $\delta(\cdot)$ un proceso espacial de media cero con variograma conocido $2\gamma(\mathbf{s}_1, \mathbf{s}_2) = \text{Var}(\delta(\mathbf{s}_1) - \delta(\mathbf{s}_2))$ (aunque en la práctica se suele suponer estacionario). Supondremos también que $f_0(\cdot) \equiv 1$, de esta forma además en el caso particular de $p = 0$, (3.7) se corresponderá con el modelo del kriging ordinario (ver sección 3.1) muy utilizado en la práctica.

Utilizando una notación matricial podemos escribir:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta},$$

siendo $\boldsymbol{\delta} = (\delta(\mathbf{s}_1), \dots, \delta(\mathbf{s}_n))'$ y \mathbf{X} una matriz $n \times (p+1)$ con $X_{ij} = f_{j-1}(\mathbf{s}_i)$, y:

$$Z(\mathbf{s}_0) = \mathbf{x}'\boldsymbol{\beta} + \delta(\mathbf{s}_0),$$

con $\mathbf{x} = (f_0(\mathbf{s}_0), \dots, f_p(\mathbf{s}_0))'$.

El predictor del kriging universal será de la forma (3.1) y en este caso como:

$$E\left(\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) + \lambda_0\right) = \boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta} + \lambda_0,$$

siendo $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$, una condición necesaria y suficiente para que el predictor sea uniformemente insesgado, i.e. $E(p(\mathbf{Z}, \mathbf{s}_0)) = E(Z(\mathbf{s}_0)) = \mathbf{x}'\boldsymbol{\beta}$, $\forall \boldsymbol{\beta} \in \mathbb{R}^{p+1}$, es que $\lambda_0 = 0$ y:

$$\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'. \quad (3.8)$$

Además como $f_0(\cdot) \equiv 1$, una de estas restricciones es:

$$\sum_{i=1}^n \lambda_i = 1, \quad (3.9)$$

que es la única condición que deben verificar los pesos en el caso del kriging ordinario.

Por tanto el predictor del kriging universal será de la forma:

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

verificando (3.8) y tal que minimiza el e.m.c. de predicción. Entonces se trata de minimizar:

$$E \left(Z(\mathbf{s}_0) - \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) \right)^2 - 2 \sum_{j=0}^p m_j \left(\sum_{i=1}^n \lambda_i f_j(\mathbf{s}_i) - f_j(\mathbf{s}_0) \right) \quad (3.10)$$

respecto a $\{\lambda_i : i = 1, \dots, n\}$ y $\{m_j : j = 0, \dots, p\}$, multiplicadores de Lagrange que garantizan (3.8). Teniendo en cuenta que el predictor es insesgado y que los pesos verifican (3.9), entonces:

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

y podemos escribir (3.10) como:

$$-\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j) + 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_i, \mathbf{s}_0) - 2 \sum_{j=0}^p m_j \left(\sum_{i=1}^n \lambda_i f_j(\mathbf{s}_i) - f_j(\mathbf{s}_0) \right)$$

Derivando respecto a $\{\lambda_i : i = 1, \dots, n\}$ y $\{m_j : j = 0, \dots, p\}$ e igualando a cero se obtienen las $n + p + 1$ ecuaciones del kriging universal que, expresadas en forma matricial, resultan ser:

$$\mathbf{\Gamma}_U \boldsymbol{\lambda}_U = \boldsymbol{\gamma}_U, \quad (3.11)$$

con:

$$\mathbf{\Gamma}_U = \begin{pmatrix} \mathbf{\Gamma} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\lambda}_U = \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{m} \end{pmatrix}, \quad \boldsymbol{\gamma}_U = \begin{pmatrix} \boldsymbol{\gamma} \\ \mathbf{x} \end{pmatrix},$$

donde $\boldsymbol{\gamma} = (\gamma(\mathbf{s}_1, \mathbf{s}_0), \dots, \gamma(\mathbf{s}_n, \mathbf{s}_0))'$, $\mathbf{m} = (m_0, \dots, m_p)'$ y $\mathbf{\Gamma}$ es una matriz $n \times n$ con $\Gamma_{ij} = \gamma(\mathbf{s}_i, \mathbf{s}_j)$. Además el e.m.c. mínimo de predicción:

$$\sigma_{KU}^2(\mathbf{s}_0) = 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0, \mathbf{s}_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j)$$

se puede obtener (teniendo en cuenta (3.11)) como:

$$\sigma_{KU}^2(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0, \mathbf{s}_i) + \sum_{j=0}^p m_j f_j(\mathbf{s}_0) = \boldsymbol{\lambda}'_U \boldsymbol{\gamma}_U.$$

En el caso particular del kriging ordinario ($p = 0$), la expresión de la varianza kriging resulta ser:

$$\sigma_{KO}^2(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_0, \mathbf{s}_i) + m_0.$$

ECUACIONES EN FUNCIÓN DEL COVARIOGRAMA

Cuando existe el covariograma $C(\mathbf{s}_1, \mathbf{s}_2) = Cov(\delta(\mathbf{s}_1), \delta(\mathbf{s}_2))$ del proceso $\delta(\cdot)$ y es conocido (una suposición más fuerte), podemos expresar las ecuaciones del kriging universal (o del KO) en función de $C(\cdot, \cdot)$. Además, si ninguna de las funciones explicativas es idénticamente 1, las ecuaciones del kriging universal sólo pueden expresarse en función del covariograma. El proceso sería análogo al caso anterior, el sistema del kriging universal equivalente a (3.11) resultante es:

$$\boldsymbol{\Sigma}_U \boldsymbol{\lambda}_U = \mathbf{c}_U, \quad (3.12)$$

donde (utilizando la notación de la sección 3.2):

$$\boldsymbol{\Sigma}_U = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\lambda}_U = \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{m} \end{pmatrix}, \quad \mathbf{c}_U = \begin{pmatrix} \mathbf{c} \\ \mathbf{x} \end{pmatrix},$$

y la varianza kriging es:

$$\sigma_{KU}^2(\mathbf{s}_0) = C(\mathbf{s}_0, \mathbf{s}_0) - \sum_{i=1}^n \lambda_i C(\mathbf{s}_0, \mathbf{s}_i) + \sum_{j=0}^p m_j f_j(\mathbf{s}_0) = C(\mathbf{s}_0, \mathbf{s}_0) - \boldsymbol{\lambda}'_U \mathbf{c}_U.$$

Otra forma de obtener estas ecuaciones (que puede ser más interesante) es a partir del predictor del kriging simple. Suponiendo que $\boldsymbol{\beta}$ es conocido en el modelo (3.7), el predictor lineal “óptimo” de $Z(\mathbf{s}_0)$ es el predictor del kriging simple (3.6), que en este caso resulta ser:

$$p_{KS}(\mathbf{Z}, \mathbf{s}_0) = \mathbf{c}' \boldsymbol{\Sigma}^{-1} \mathbf{Z} + (\mathbf{x} - \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{c})' \boldsymbol{\beta}.$$

Cuando $\boldsymbol{\beta}$ no es conocido es lógico pensar en utilizar en su lugar su estimador lineal óptimo $\hat{\boldsymbol{\beta}}_{mcg} = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{Z}$, obteniéndose el predictor:

$$p^*(\mathbf{Z}, \mathbf{s}_0) = \mathbf{c}' \boldsymbol{\Sigma}^{-1} \mathbf{Z} + (\mathbf{x} - \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{c})' \hat{\boldsymbol{\beta}}_{mcg}. \quad (3.13)$$

Puede verse (Goldberger, 1962) que el predictor (3.13), lineal e insesgado, es óptimo (en el sentido de que minimiza el e.m.c. sobre todos los predictores lineales e insesgados) y por tanto coincide con el predictor del kriging universal⁶.

Entonces:

$$p_{KU}(\mathbf{Z}, \mathbf{s}_0) = p_{KS}(\mathbf{Z}, \mathbf{s}_0) + (\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c})'(\hat{\boldsymbol{\beta}}_{mcg} - \boldsymbol{\beta}), \quad (3.14)$$

y además, teniendo en cuenta que el error $(p_{KS}(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0))$ tiene covarianza nula con cualquier combinación lineal de \mathbf{Z} (ver p.e. Chilès y Delfiner, 1999, p. 161), la relación (3.14) se extiende también a la varianza kriging:

$$\sigma_{KU}^2(\mathbf{s}_0) = \sigma_{KS}^2(\mathbf{s}_0) + (\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c})'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c}), \quad (3.15)$$

donde el segundo termino cuantifica la precisión en la estimación de la media. Las expresiones (3.14) y (3.15) son conocidas como la relación de aditividad entre el KS y el KU.

Los resultados anteriores permiten pensar en la predicción lineal con media desconocida como un proceso de dos etapas: en la primera estimar la media desconocida, y en la segunda realizar la predicción lineal óptima con media supuestamente conocida.

TENDENCIA POLINÓMICA

En muchos casos se supone que las funciones $\{f_j(\cdot) : j = 0, \dots, p\}$ son los monomios $x_1^{i_1} \cdots x_d^{i_d}$ donde $\mathbf{s} = (x_1, \dots, x_d)'$ e i_1, \dots, i_d son números naturales cuya suma es menor o igual que k , siendo k un numero natural dado. En ese caso (si se consideran todas las combinaciones posibles, i.e. $p = \binom{k+d}{d} - 1$) se dice que $\mu(\cdot)$ es una superficie de tendencia polinómica de grado⁷ k .

⁶ Tendiendo en cuenta (3.13), Gotway y Cressie (1993) consideraron un estimador lineal de $\boldsymbol{\beta}$ no insesgado que da lugar a un predictor (tipo James-Stein) no insesgado pero con un error de predicción en media cuadrática menor que la varianza del kriging universal (ver también Cressie, 1993, p. 175).

⁷ Esta suposición es también la base del kriging con funciones intrínsecas de orden k (ver p.e. Cressie, 1993, sección 5.4 o Chilès y Delfiner, 1999, cap. 4), donde se extiende la noción de incrementos (que filtran una media constante) estacionarios al caso de incrementos generalizados de orden k (que filtran una tendencia polinómica de orden k)

Por ejemplo, en el caso bidimensional la expresión de este tipo de tendencias sería:

$$\mu(\mathbf{s}) = \sum_{0 \leq l+m \leq k} \alpha_{lm} x^l y^m; \quad \mathbf{s} = (x, y)' \in \mathbb{R}^2,$$

que se corresponde con $p = \{(k+1)(k+2)/2\} - 1$ y:

$$f_0(\mathbf{s}) = 1, f_1(\mathbf{s}) = x, f_2(\mathbf{s}) = y, \dots, f_p(\mathbf{s}) = y^k; \quad \mathbf{s} = (x, y)' \in \mathbb{R}^2.$$

DETALLES PRÁCTICOS Y COMPUTACIONALES

Para que exista una única solución del sistema (3.11) la matriz $\mathbf{\Gamma}$ debe ser no singular (o $\mathbf{\Sigma}$ en el caso de (3.12)) y \mathbf{X} de rango máximo. Análogamente al caso del kriging simple, la primera condición se verifica normalmente si $\gamma(\cdot, \cdot)$ es un semivariograma válido (salvo posibles problemas con la anisotropía zonal) y las posiciones de los datos son distintas. La segunda condición (sólo de interés en el KU) implica que las $p+1$ funciones explicativas deben ser linealmente independientes por lo menos sobre el conjunto de datos, i.e.:

$$\sum_{j=0}^p a_j f_j(\mathbf{s}_i) = 0, \quad i = 1, \dots, n \Leftrightarrow a_j = 0, \quad j = 0, \dots, p.$$

Esta condición (ya conocida en la estimación por mínimos cuadrados) suele satisfacerse normalmente cuando los datos están irregularmente espaciados ya que p suele ser mucho menor que n . Sin embargo cuando los datos están regularmente espaciados pueden aparecer problemas, especialmente al utilizar vecindarios locales.

Al contrario que en el KS, las matrices de los sistemas (3.11) o (3.12) no son semidefinidas positivas (y no se puede utilizar en principio la “ideal” factorización de Cholesky). Aunque reformulando el sistema (3.12) puede obtenerse su solución a partir de dos subsistemas con matrices semidefinidas positivas⁸ (ver p.e. Chilès y Delfiner, 1999, p. 169). Esta aproximación sin

estacionarios. Una idea similar es la utilizada en la estimación por máxima verosimilitud restringida mostrada en la sección 4.3.2.

⁸ Muchos de los algoritmos utilizados para la solución de los sistema kriging están diseñados y optimizados para covariogramas (por ejemplo tienen en cuenta que no es

embargo sólo es de especial interés cuando la dimensión de la matriz es grande. Además, en la práctica el número de observaciones involucradas en el sistema kriging es normalmente pequeño (p.e. menor que 100 o incluso mucho menor), ya que cuando se trabaja con grandes muestras se suele optar por vecindarios locales. Por tanto, por lo menos en esos casos, puede ser preferible utilizar la matriz original del sistema y el procedimiento recomendado a seguir sería el siguiente:

- Obtener la factorización \mathbf{UDU}' de la matriz simétrica $\mathbf{\Gamma}_U$, donde \mathbf{U} es una matriz triangular superior y \mathbf{D} una matriz diagonal (p.e. utilizando la rutina LFTSF de la librería IMSL).
- Para cada posición de predicción \mathbf{s}_0 resolver el sistema (3.11) utilizando la factorización obtenida en el primer paso (p.e. con las rutinas LFSSF o LFISF de la librería IMSL); ya que la matriz del sistema, al igual que en el KS, tampoco depende de la posición de predicción.

Para evitar problemas numéricos en el procedimiento anterior, se puede tener en cuenta también que la solución al sistema (3.11) es la misma cambiando $\gamma(\cdot, \cdot)$ por $c + \gamma(\cdot, \cdot)$, para cualquier $c \in \mathbb{R}$.

Análogamente al caso del KS, tanto los pesos kriging como la varianza kriging no dependen de las observaciones (solamente de sus posiciones a través del variograma y de las funciones explicativas en el KU). Además, en el caso del KO y de que el variograma sea estacionario, los pesos y varianza kriging son también invariantes frente a traslaciones de la configuración espacial⁹.

necesario pivotar la matriz de covarianzas ya que los valores máximos están en la diagonal). En el caso de variogramas no acotados se trabaja en ocasiones con pseudo-covariogramas (ver sección 2.4.1).

⁹ Puede verse también (p.e. Chilès y Delfiner, 1999, sección 3.4.6) que el predictor y varianza del KU son invariantes frente a transformaciones lineales de las funciones explicativas, y por tanto invariantes frente a traslaciones cuando por ejemplo la tendencia es polinómica de orden k .

3.4 ALGUNAS CONSIDERACIONES ACERCA DE LOS MÉTODOS

KRIGING

3.4.1 Kriging como interpolador

Las ecuaciones kriging tienen en cuenta varios aspectos del problema de interpolación¹⁰ (ver p.e. Chilès y Delfiner, 1999, secciones 3.3.2 y 3.4.2):

- La configuración espacial de los datos, a través de las matrices Σ o Γ (donde el covariograma o variograma actúa como una “distancia estadística” entre las observaciones y de forma que se tiene en cuenta la información redundante presente en los datos).
- La situación de la posición de predicción respecto a los datos, a través de \mathbf{c} o γ .
- La presencia de una función determinística de tendencia.

Adicionalmente también tienen en cuenta propiedades estadísticas del proceso $Z(\cdot)$, a través del variograma o el covariograma (que como se comentó en la sección 2.2.1, entre otras cosas, determina las propiedades de continuidad del proceso). Esta es la principal diferencia con otros métodos de interpolación que no tienen en cuenta la estructura de segundo orden del proceso.

Una propiedad importante de los predictores kriging es que son interpoladores exactos (suponiendo que no hay error de medida, i.e. $c_{EM} = 0$), en el sentido de que $p(\mathbf{Z}, \mathbf{s}_0) = Z(\mathbf{s}_0)$ cuando $\mathbf{s}_0 = \mathbf{s}_i$ (la solución de los sistemas es $\lambda_i = 1$ y $\lambda_j = 0, \forall j \neq i$), y naturalmente en ese caso la estimación de la varianza es 0. Además, por lo general no son continuos en las posiciones de los datos¹¹; para que lo sean, teniendo en cuenta la descomposición (2.10) mostrada en la sección 2.2.1, es necesario que $c_{EM} = 0$ y $c_{MS} = 0$ (i.e. el efecto nugget debe ser nulo).

¹⁰ Puede verse también fácilmente, a partir de las denominadas ecuaciones duales del KU, que hay una equivalencia teórica (aunque no en la práctica) entre los métodos kriging y la predicción tipo spline (ver p.e. Cressie, 1993, pp. 180-183).

¹¹ Un ejemplo muy ilustrativo de la dependencia de las predicciones kriging del comportamiento del variograma cerca del origen se tiene p.e. en Chilès y Delfiner (1999, p. 161).

Otra característica de la interpolación kriging (y que no aparece en otros métodos como los que asignan pesos inversamente proporcionales a la distancia) es el denominado *efecto pantalla*. En general se observa (ver comentarios en la siguiente sección) que la influencia de un valor es menor si está oculto detrás de otro valor (e.g. Journel y Huijbregts, 1978, p. 346). Esto produce, por ejemplo en el caso del KO (incluso si el variograma es estacionario), que puntos situados a la misma distancia de la posición de predicción puedan tener distintos pesos y que los datos cercanos no apantallados reciban los mayores pesos, reduciéndose considerablemente (llegando a ser negativos) los pesos de los datos que quedan ocultos¹².

La aparición de pesos negativos (o mayores que 1) en el KO como consecuencia del efecto pantalla, puede provocar (incluso suponiendo media constante) que el predictor kriging no esté necesariamente comprendido entre el valor mínimo y máximo de los datos. Esto que en principio puede ser una propiedad muy interesante puede conducir a resultados extraños en ciertas ocasiones, como por ejemplo dar lugar a predicciones negativas en casos en los que la variable considerada es necesariamente positiva. Para solucionar estos problemas se han propuesto numerosas alternativas, entre ellas la inclusión de restricciones adicionales sobre los pesos de forma que sean positivos¹³ (lo cual puede dar lugar a un incremento considerable del e.m.c. de predicción) o sobre el predictor (ver p.e. Goovaerts, 1997, sección 7.4.2). Otra alternativa que puede ser preferible es la transformación del proceso $Z(\cdot)$ a otra escala (de forma que se aproxime a la normalidad), realizar la predicción kriging del proceso transformado y volver a la escala original (pero asegurándose de que al hacer la transformación inversa el resultado tenga las propiedades de optimalidad deseadas); más detalles sobre este tema se tienen en la última sección de este capítulo.

¹² En la literatura se muestran numerosos ejemplos sobre el comportamiento de los pesos kriging en distintos escenarios; una colección bastante completa se tiene en Wakernagel (1998, capítulo 13).

¹³ Las ecuaciones kriging con pesos positivos se tiene en Barnes y Johnson (1984) y también en Chilès y Delfiner (1999, sección 3.9.1).

3.4.2 Efecto del variograma (covariograma) en el kriging

El variograma (o el covariograma) tiene un efecto determinante en la predicción espacial. Como ejemplo, a continuación se tienen algunas observaciones acerca de la influencia en el kriging de las tres principales características de un variograma estacionario¹⁴ (normalmente tratadas como parámetros) definidas en la sección 2.2.1.

Rango

Supongamos que la posición de predicción \mathbf{s}_0 está a una distancia mayor que el rango de las posiciones de los datos $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ (i.e. la posición de predicción está fuera de la zona de influencia de los datos), entonces $\mathbf{c} = \mathbf{0}$ en (3.6) y (3.13), obteniéndose que:

$$p_{KS}(\mathbf{Z}, \mathbf{s}_0) = \mu(\mathbf{s}_0), \quad p_{KU}(\mathbf{Z}, \mathbf{s}_0) = \mathbf{x}'\hat{\boldsymbol{\beta}}_{mcg},$$

por lo tanto la predicción kriging se reduce a la media¹⁵ (estimada en el caso del KU).

Nugget y umbral

Está claro que las estimaciones obtenidas de la varianza de los predictores kriging dependen en gran medida de estos parámetros. Es importante destacar que la escala del variograma (covariograma) no influye en las predicciones obtenidas, solamente en la varianza kriging. Si se multiplica el variograma (o el covariograma, según el caso) por una constante, las ecuaciones kriging (3.5) o (3.11) quedan invariantes y consecuentemente los pesos kriging no cambian, aunque la varianza kriging resulta multiplicada por esa constante. Para estudiar su influencia en la predicción resulta de utilidad la proporción del efecto nugget en el

¹⁴ Una discusión bastante detallada sobre este tema (centrada en el caso del KO) se tiene en Isaaks y Srivastava (1989, pp. 299-313); ver también Cressie (1993, sección 3.2.1) o Fernández-Casal (1995, sección 3.1.1).

¹⁵ Puede verse fácilmente que las ecuaciones del kriging universal con $\mathbf{c} = \mathbf{0}$, son las obtenidas en el denominado método kriging de estimación de la tendencia (para más detalles sobre este método ver p.e. Wackernagel, 1998, pp. 212-213, o Chilès y Delfiner, 1999, sección 3.4.5).

umbral total c_0/σ^2 (que como ya se comentó al final de la sección 2.2.1, proporciona mucha información acerca del grado de dependencia presente en los datos). Por ejemplo, en el caso en que toda la variabilidad es efecto nugget (i.e. el proceso $Z(\cdot)$ es ruido blanco) entonces $\Sigma = \sigma^2 \mathbf{I}_n$, siendo \mathbf{I}_n la matriz identidad $n \times n$, y $\mathbf{c} = \mathbf{0}$ (suponiendo que $\mathbf{s}_0 \neq \mathbf{s}_i, \forall i$), y los predictores kriging se reducen a la estimación por m.c.o. de la media:

$$p_{KU}(\mathbf{Z}, \mathbf{s}_0) = \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} = \mathbf{x}'\hat{\boldsymbol{\beta}}_{mco}.$$

En el caso del KO se obtiene la media muestral:

$$p_{KO}(\mathbf{Z}, \mathbf{s}_0) = \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z(\mathbf{s}_i),$$

predictor bien conocido que asigna igual peso a todos los datos (por tanto el efecto pantalla es nulo); además, teniendo en cuenta (3.15) y que $\sigma_{KS}^2(\mathbf{s}_0) = \sigma^2$ (caso más desfavorable), entonces:

$$\sigma_{KO}^2(\mathbf{s}_0) = \sigma^2 + \frac{\sigma^2}{n},$$

i.e. la varianza del KO para el caso de procesos incorrelados es igual a la varianza del proceso más la varianza de la media muestral. En general, se puede ver que al aumentar el porcentaje de efecto nugget en el umbral total disminuye el efecto pantalla¹⁶ y aumenta la varianza kriging (ver p.e. Isaaks y Srivastava, 1989, pp. 305-306).

Teniendo en cuenta los comentarios anteriores, podemos afirmar que todos los parámetros (características) del variograma influyen en el kriging (aunque quizás el rango es el que tiene un menor efecto, ya que pequeñas variaciones en este parámetro producen resultados casi idénticos). Como observación adicional acerca de la influencia del variograma en la predicción espacial, hay que destacar la importancia del comportamiento del variograma cerca del origen. Por ejemplo, la forma del variograma cerca del origen determina las propiedades de continuidad y regularidad de las predicciones kriging (ver p.e. Chilès y Delfiner, 1999, pp. 160-161 y observación sobre

¹⁶ Cuando la media no es constante puede ocurrir incluso lo contrario del efecto pantalla y observaciones alejadas pueden tener una gran influencia en la estimación (como es bien conocido en la estimación de la tendencia).

continuidad realizada en el apartado anterior). Además, Stein (1988) probó que para una predicción asintóticamente (de relleno) eficiente lo que se necesita generalmente es capturar la conducta del variograma cerca del origen. Este resultado no contradice las observaciones anteriores sobre la influencia de las características del variograma. Es difícil determinar cuando los datos están suficientemente cerca como para tener sólo en cuenta el efecto nugget y la pendiente del variograma cerca del origen.

EFEECTO DE LA ESTIMACIÓN DEL VARIOGRAMA

En los métodos descritos en las secciones anteriores se suponía que el variograma (covariograma) era conocido, sin embargo en la práctica en realidad el variograma es estimado (kriging estimado). Supongamos que $\hat{p}_K(\mathbf{Z}, \mathbf{s}_0)$ y $\hat{\sigma}_K^2(\mathbf{s}_0)$ son el predictor y varianza kriging obtenidos utilizando alguno de los métodos descritos anteriormente con un variograma estimado.

En cuanto al predictor $\hat{p}_K(\mathbf{Z}, \mathbf{s}_0)$ se puede ver que, bajo condiciones muy generales, converge al valor correcto si la densidad de datos tiende a infinito incluso cuando la estimación del variograma no es muy buena (Yakowitz. y Szidarovszky, 1985). El factor más importante, en concordancia con las observaciones realizadas en el apartado anterior, es que la aproximación al variograma verdadero cerca del origen no sea muy mala (Stein, 1988).

Al contrario que en el caso del predictor kriging, la estimación del variograma afecta directamente a la varianza kriging y en general $\hat{\sigma}_K^2(\mathbf{s}_0)$ no es un estimador consistente del error en media cuadrática del predictor (hay que recordar también que la escala del variograma no influye en la predicción pero sí en la varianza kriging). En muchos casos, por ejemplo si la distribución del proceso espacial es normal, se espera que:

$$\hat{\sigma}_K^2(\mathbf{s}_0) \leq E(p_K(\mathbf{Z}, \mathbf{s}_0) - Z(\mathbf{s}_0))^2,$$

(ver p.e. Christensen, 1991, pp. 283-284), y por tanto la estimación de la varianza kriging debería ser incrementada para que tenga en cuenta el efecto de la estimación del variograma. En la sección 4.4 se propone una corrección para la varianza kriging obtenida a partir del método de validación cruzada.

Un tratamiento bastante detallado de este tema con una revisión de los “escasos” resultados teóricos obtenidos hasta la fecha se tiene en Cressie (1993, sección 5.3). Otras referencias que pueden también ser de interés son: Christensen (1991, sección 6.5), Stein (1999, sección 6.8) o Chilès y Delfiner (1999, sección 3.4.3).

3.4.3 Elección del vecindario

Una práctica habitual es geoestadística es la inclusión en las ecuaciones kriging solamente los $n(\mathbf{s}_0)$ datos más “próximos” a la posición de predicción \mathbf{s}_0 en lugar de considerar todas las observaciones disponibles. Esto puede justificarse por varias razones:

- Utilizar todos los datos puede dar lugar a un sistema de difícil solución debido a problemas numéricos. Por ejemplo, entre otras cosas, el tiempo de computación (aproximadamente proporcional a $n(\mathbf{s}_0)^3$) aumenta drásticamente al aumentar el número de datos.
- Las estimaciones del variograma son normalmente eficientes (o incluso el propio modelo geoestadístico válido) solamente sobre pequeñas distancias.
- El uso de vecindarios locales permite la relajación de las hipótesis del modelo (como la estacionariedad intrínseca en el caso del KO) o su simplificación (por ejemplo, en el caso del KU se puede suponer que localmente la estructura de la tendencia es más simple o incluso constante y utilizar en su lugar KO local).
- En muchos casos los datos cercanos apantallan a los más alejados reduciendo su influencia (aunque no siempre de forma significativa; ver observaciones siguientes).

La selección “óptima” de un vecindario local resulta no obstante un problema algo complejo. Por ejemplo, se ha pensado que el rango del variograma permite determinar por sí solo un criterio de vecindad, como incluir en las ecuaciones sólo los datos que estén dentro del rango de \mathbf{s}_0 , sin embargo esto puede no ser adecuado en muchos casos (aunque en la práctica el valor de este parámetro puede ser de gran utilidad como referencia inicial). Teniendo en cuentas las observaciones realizadas en secciones anteriores, cuando aumenta la proporción de

efecto nugget disminuye el efecto pantalla, el predictor kriging se reduce a la media y observaciones a más distancia que el rango de la posición de predicción contribuyen (a veces de forma significativa) a la estimación de la tendencia. Se ha observado también (p.e. Journel y Huijbregts, 1978, p. 346 o Cressie, 1993, pp. 132-133) que eliminar un valor fuera del rango puede ser inapropiado incluso cuando esos valores están ocultos por otros más cercanos y la media es constante. Aunque el variograma sea estacionario, es un error pensar que el peso λ_i del kriging ordinario es sólo función de $\|\mathbf{s}_0 - \mathbf{s}_i\|$ ya que, teniendo en cuenta la expresión de los pesos del KO, depende también de la conducta del proceso en todas las demás posiciones espaciales (ver p.e. Cressie, 1993, pp. 131 y 132, para una formulación explícita). Por tanto, observaciones fuera del rango pueden tener influencia en la predicción a través de su correlación con observaciones dentro del rango (esto es conocido en la literatura como efecto “relay”, que podríamos traducir por efecto transmisión).

Se han propuesto algunos criterios para la selección de un vecindario óptimo, por ejemplo basados en la idea de que la varianza del KO (Rivoirard, 1987) o del KU (Cressie, 1993, pp. 176-177) se aproxime a la del KS. Sin embargo esos criterios realmente solo son de utilidad cuando los datos están regularmente espaciados y el vecindario se puede fijar de antemano (en Cressie, 1993, pp. 158-162 se tiene un ejemplo). En esos casos puede ser preferible utilizar directamente la varianza kriging e incluir datos en el vecindario hasta que no se produzca una disminución “significativa” en la estimación de la varianza kriging.

En la práctica, la densidad de datos y su configuración espacial en torno a las posiciones de predicción pueden ser muy irregulares. Teniendo en cuenta esto se han desarrollado distintos algoritmos, algunos bastante sofisticados, para la selección de vecindarios. Para la selección de los datos típicamente se fija un radio máximo de búsqueda y solamente se consideran los datos dentro de una circunferencia (esfera) centrada en la posición de predicción. En el caso de anisotropía (sección 2.2.2), se considera una elipse (elipsoide) con el radio mayor orientado en la dirección de máxima variación. Además suele interesar disponer de observaciones en todas direcciones, por lo que se divide la zona de búsqueda en sectores angulares (p.e. cuadrantes en el caso bidimensional o octantes en el caso tridimensional) y se selecciona un número mínimo de datos en todos o en la mayoría de esos sectores (esto evita también que clusters de datos tengan una

excesiva influencia sobre predicciones en su entorno). Si se sospecha de la presencia de una tendencia en los datos (KU), puede ser deseable la inclusión de observaciones más alejadas de la posición de predicción para poder estimarla de forma más eficiente.

Utilizando un algoritmo de búsqueda que tenga en cuenta todas o alguna de las consideraciones anteriores, típicamente se selecciona un número pequeño de datos como vecindario (p.e. entre 8 y 20 observaciones) para cada posición de predicción. Sin embargo esto puede causar que las superficies de predicción presenten discontinuidades (especialmente cuando los datos están regularmente espaciados y se utiliza búsqueda por octantes). Una aproximación distinta sería la selección un único vecindario más grande (p.e. de 20 a 40 datos) para pequeños conjuntos de posiciones de predicción, de esta forma en condiciones normales los vecindarios correspondientes a conjuntos de predicción próximos se solapan y es de esperar que no aparezcan discontinuidades. Además el incremento en tiempo de computación debido a la inclusión de un número mayor de observaciones se compensa por el hecho de que sólo es necesario factorizar una vez la matriz sistema kriging para obtener las predicciones en el conjunto considerado¹⁷. Por ejemplo, si se pretenden obtener predicciones en una rejilla bidimensional, el tiempo de computación considerando un vecindario distinto con 16 observaciones para cada nodo resulta similar a considerar un vecindario con 25 (o incluso más) observaciones para grupos de 4 nodos (dos por fila y columna). Otra forma de proceder que puede ser de interés en la práctica, es usar el semivariograma como distancia en lugar de la tradicional distancia euclídea; de esta forma los datos son seleccionados preferentemente en la dirección de máxima continuidad (y se evita tener que considerar elipsoides en el caso de anisotropía). Adicionalmente, cuando el número de datos es grande y sus posiciones irregulares, es recomendable utilizar la técnica conocida como *búsqueda por súper-bloques*. Esta técnica consiste en la clasificación inicial de los datos en “súper-bloques” definidos por una rejilla regular

¹⁷ Esto puede ser también de utilidad en el caso de validación cruzada (donde es de especial importancia la selección del vecindario), ya que la solución del sistema kriging para la predicción en una de las posiciones s_i del vecindario considerando el resto de observaciones, puede obtenerse a partir de la matriz correspondiente a todo el vecindario (ver sección 4.4).

de poca resolución espacial (de mucho mayor espaciado que la rejilla de predicción) superpuesta a los datos, de esta forma para determinar los datos próximos a una posición espacial \mathbf{s}_0 se puede limitar la búsqueda a los súper-bloques más cercanos (en lugar de verificar todos los datos).

Independientemente del algoritmo de búsqueda que se vaya a utilizar, puede ser recomendable realizar antes algunas pruebas utilizando por ejemplo la técnica de validación cruzada (sección 4.4). Para un discusión adicional sobre la selección del vecindario en la práctica, ver por ejemplo Isaaks y Srivastava (1989, capítulo 14) o Deutsch y Journel (1992, secciones II.4 y IV.6).

3.4.4 Kriging log-normal y trans-normal

En el caso de normalidad el predictor óptimo $E(Z(\mathbf{s}_0)|\mathbf{Z})$ de $Z(\mathbf{s}_0)$ es lineal y coincide con los predictores kriging. Pero si el proceso no es normal este predictor puede ser altamente no lineal, en esos casos la transformación del proceso $Z(\cdot)$ a otra escala puede producir que se aproxime a la normalidad¹⁸. De esta forma se puede pensar en realizar la predicción lineal en la escala transformada y posteriormente hacer la transformación inversa (pero asegurándose de que el resultado tenga las propiedades de optimalidad deseadas). En esta sección simplemente se muestran algunos resultados sobre este tema, para un tratamiento más detallado ver p.e. Cressie (1993, sección 3.2.2) o Chilès y Delfiner (1999, sección 3.4.10).

Una de las transformaciones más utilizadas en geoestadística es la transformación logarítmica, i.e. el caso en el que se asume que el proceso $Z(\cdot)$ sigue una distribución lognormal. Un proceso aleatorio lognormal es un proceso $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ (que toma valores positivos) tal que:

$$Y(\mathbf{s}) = \log(Z(\mathbf{s})); \quad \mathbf{s} \in D,$$

es un proceso normal.

El *kriging simple lognormal* (KSL) se basa en la suposición adicional de que el proceso $Y(\cdot)$ verifica las hipótesis del kriging simple (media y

¹⁸ Lo cual también permite emplear otros métodos que asumen esta hipótesis, como p.e. los basados en máxima verosimilitud para la estimación del variograma; ver sección 4.3.2.

covariograma conocidos). En ese caso, utilizando el método del KS, a partir de $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'$ podemos obtener el predictor $p_{KS}(\mathbf{Y}, \mathbf{s}_0)$ de $Y(\mathbf{s}_0)$ y la correspondiente varianza kriging $\sigma_{KS}^2(\mathbf{s}_0)$. Si transformamos de nuevo este valor a la escala $Z(\cdot)$, obtenemos $\exp(p_{KS}(\mathbf{Y}, \mathbf{s}_0))$ que no es un predictor insesgado de $Z(\mathbf{s}_0)$ (es un predictor insesgado en mediana). El predictor óptimo de $Z(\mathbf{s}_0)$ resulta ser:

$$p_{KSL}(\mathbf{Z}, \mathbf{s}_0) = \exp\left(p_{KS}(\mathbf{Y}, \mathbf{s}_0) + \frac{1}{2}\sigma_{KS}^2(\mathbf{s}_0)\right), \quad (3.16)$$

y la correspondiente varianza kriging:

$$\sigma_{KSL}^2(\mathbf{s}_0) = p_{KS}(\mathbf{Y}, \mathbf{s}_0)^2 \left(\exp(\sigma_{KS}^2(\mathbf{s}_0)) - 1\right).$$

En el caso de media no conocida, i.e. suponiendo que el proceso $Y(\cdot)$ verifica las hipótesis del KU (kriging universal lognormal, KUL), se complica aún más el problema. No basta con sustituir la media teórica en (3.16) por su predictor óptimo ya que se obtendría un predictor sesgado. Si se hace una corrección para obtener un predictor insesgado de $Z(\mathbf{s}_0)$, el resultado sería:

$$p_{KUL}(\mathbf{Z}, \mathbf{s}_0) = \exp\left(p_{KU}(\mathbf{Y}, \mathbf{s}_0) + \frac{1}{2}\sigma_{KU}^2(\mathbf{s}_0) - \mathbf{m}'\mathbf{x}\right), \quad (3.17)$$

utilizando la notación de la sección 3.3 (y suponiendo también que una de las funciones explicativas es idénticamente 1). Hay que destacar que el predictor (3.17) no es un predictor óptimo en el sentido de que minimice el e.m.c. de predicción, este predictor minimiza $E(\log p(\mathbf{Z}, \mathbf{s}_0) - Y(\mathbf{s}_0))^2$ (sujeto a las correspondientes restricciones de insesgades y forma del predictor). La varianza kriging tiene una expresión considerablemente más complicada que en el caso de media conocida (ver Cressie, 1993, p. 136; para el caso de media constante). Sin embargo, si el objetivo es la construcción de intervalos de confianza, se pueden transformar directamente de la escala $Y(\cdot)$. Por ejemplo:

$$\left(\exp\left(p_K(\mathbf{Y}, \mathbf{s}_0) - 1.96\sigma_K^2(\mathbf{s}_0)\right), \exp\left(p_K(\mathbf{Y}, \mathbf{s}_0) + 1.96\sigma_K^2(\mathbf{s}_0)\right)\right),$$

es un intervalo de confianza al 95% para $Z(\mathbf{s}_0)$.

La aproximación anterior puede generalizarse para una transformación cualquiera:

$$Z(\mathbf{s}) = \phi(Y(\mathbf{s})); \quad \mathbf{s} \in D,$$

siendo $Y(\cdot)$ un proceso normal y $\phi(\cdot)$ una función medible dos veces diferenciable. En general no se dispone de expresiones exactas como en el caso del kriging lognormal¹⁹; a partir de un predictor kriging $p_K(\mathbf{Y}, \mathbf{s}_0)$ de $Y(\mathbf{s}_0)$ se pueden obtener un predictor aproximadamente insesgado de $Z(\mathbf{s}_0)$ teniendo en cuenta que:

$$\begin{aligned} \phi(Y(\mathbf{s}_0)) &\simeq \phi(p_K(\mathbf{Y}, \mathbf{s}_0)) + (Y(\mathbf{s}_0) - p_K(\mathbf{Y}, \mathbf{s}_0))\phi'(p_K(\mathbf{Y}, \mathbf{s}_0)) \\ &\quad + \frac{1}{2}(Y(\mathbf{s}_0) - p_K(\mathbf{Y}, \mathbf{s}_0))^2\phi''(p_K(\mathbf{Y}, \mathbf{s}_0)), \end{aligned}$$

si el error kriging $p_K(\mathbf{Y}, \mathbf{s}_0) - Y(\mathbf{s}_0)$ es pequeño. A partir de esto, si $\sigma_K^2(\mathbf{s}_0)$ es la correspondiente varianza kriging, se obtiene el predictor aproximadamente insesgado del *kriging trans-normal*²⁰ (KT):

$$p_{KT}(\mathbf{Z}, \mathbf{s}_0) = \phi(p_K(\mathbf{Y}, \mathbf{s}_0)) + \frac{1}{2}\sigma_K^2(\mathbf{s}_0)\phi''(p_K(\mathbf{Y}, \mathbf{s}_0)),$$

que en el caso del KS se aproxima al predictor óptimo $E(Z(\mathbf{s}_0)|\mathbf{Z})$ de $Z(\mathbf{s}_0)$. Como aproximación de la varianza kriging de este predictor se puede utilizar:

$$\sigma_{KT}^2(\mathbf{s}_0) = \sigma_K^2(\mathbf{s}_0)\phi'(p_K(\mathbf{Y}, \mathbf{s}_0))^2.$$

¹⁹ Shimizu y Iwase (1987) obtuvieron expresiones que permiten el cálculo de predictores (exactamente) insesgados de forma análoga al caso lognormal, cuando la transformación $\phi(\cdot)$ verifica $\phi''(x) = a + b\phi(x)$. En ese caso al proceso $Z(\cdot)$ lo denominaron proceso lognormal generalizado.

²⁰ Este predictor es ligeramente distinto al considerado por Cressie (1993, p.137) obtenido a partir de una aproximación similar (utilizando del método delta en torno a la media).

CAPÍTULO 4

MODELADO DE LA DEPENDENCIA ESPACIAL

Este capítulo se centra en el problema de la obtención de un modelo de variograma (o covariograma) que describa adecuadamente la dependencia espacial de los datos. En la resolución de este problema, frecuentemente denominado *análisis estructural*¹, se utilizan en la práctica desde procedimientos totalmente automáticos (tipo “black-box”) hasta otros totalmente manuales en los que el usuario (con la ayuda de herramientas estadísticas) especifica por completo la dependencia espacial (esto se suele denominar ajuste a sentimiento). Sin embargo, habitualmente se siguen los siguientes pasos (o por lo menos algunos):

1. Estimación inicial del semivariograma utilizando algún tipo de estimador experimental (sección 4.1).
2. Ajuste de un modelo válido de semivariograma a las estimaciones “piloto” obtenidas en el primer paso (secciones 4.2 y 4.3), ya que éstas normalmente no pueden ser utilizadas directamente en la predicción espacial.
3. Diagnóstico del modelo de semivariograma obtenido, utilizando por ejemplo la técnica de validación cruzada (sección 4.4).

Más recientemente se han propuesto diversas alternativas al modelado tradicional principalmente con la idea de disponer de mayor flexibilidad en el

¹ Como se muestra a continuación, la estimación del variograma está ligada al resto de parámetros del proceso espacial; por este motivo se suele denominar también de esta forma al procedimiento completo de determinación de todos los parámetros del modelo geoestadístico.

ajuste de un semivariograma válido, una revisión de estos métodos se tiene en la sección 4.5.

4.1 ESTIMACIÓN DEL SEMIVARIOGRAMA

Supongamos por el momento que el proceso $Z(\cdot)$ es intrínsecamente estacionario y que $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$ son los datos observados. En este caso, como la media es constante, se tiene que:

$$E(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2 = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2), \quad \forall \mathbf{s}_1, \mathbf{s}_2 \in D. \quad (4.1)$$

Las distintas propuestas para la obtención estimadores, mostradas a continuación, se basan en (4.1). Sin embargo, cuando la media no es constante estos estimadores no deben ser utilizados mientras que no se elimine la tendencia de los datos. Por ejemplo, si el proceso $Z(\cdot)$ verifica la descomposición:

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

siendo $\mu(\cdot)$ la función de tendencia y $\delta(\cdot)$ un proceso intrínsecamente estacionario de media cero, entonces:

$$E(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2 = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2) + (\mu(\mathbf{s}_1) - \mu(\mathbf{s}_2))^2, \quad (4.2)$$

y si $\mu(\cdot)$ no es constante (4.2) no es necesariamente función de $\mathbf{s}_1 - \mathbf{s}_2$, ni tiene porque verificar las propiedades de un variograma. En el caso de que la media sea conocida (KS) se puede eliminar la tendencia y estimar el variograma a partir del proceso $\delta(\cdot)$. Sin embargo, si $\mu(\cdot)$ es desconocida (p.e. en el caso del KU) aparecen ciertos inconvenientes. Este último caso se trata en la sección 4.1.2.

4.1.1 Estimación del (co)variograma bajo media constante

ESTIMADORES CLÁSICOS DEL SEMIVARIOGRAMA Y DEL COVARIOGRAMA

La expresión (4.1) sugiere la estimación del semivariograma mediante el método de los momentos, obteniendo de esta forma el denominado estimador empírico (o clásico) del semivariograma (Matheron, 1962):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, \quad \mathbf{h} \in \mathbb{R}^d, \quad (4.3)$$

donde:

$$N(\mathbf{h}) = \{(i, j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}; i, j = 1, \dots, n\}, \quad (4.4)$$

y $|N(\mathbf{h})|$ es el número de pares distintos en $N(\mathbf{h})$.

Realmente el estimador (4.3) sólo es de utilidad en el caso de datos regularmente espaciados. Cuando los datos están irregularmente espaciados generalmente se utiliza una versión suavizada de este estimador, por ejemplo considerando en lugar de (4.4):

$$N(Tol(\mathbf{h})) = \{(i, j) : \mathbf{s}_i - \mathbf{s}_j \in Tol(\mathbf{h})\}, \quad (4.5)$$

donde $Tol(\mathbf{h}) \subset \mathbb{R}^d$ es una región de tolerancia en torno a \mathbf{h} . Esta región debería ser lo suficientemente grande como para que no aparezcan inestabilidades, por lo que se recomienda (Journel y Huijbregts 1978, p. 194) que el número de aportaciones a la estimación en un salto \mathbf{h} sea por lo menos de treinta² (i.e. $|N(Tol(\mathbf{h}))| \geq 30$).

Análogamente al caso anterior, suponiendo estacionariedad de segundo orden, se obtiene el estimador clásico del covariograma:

$$\hat{C}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (Z(\mathbf{s}_i) - \bar{Z})(Z(\mathbf{s}_j) - \bar{Z}), \quad \mathbf{h} \in \mathbb{R}^d, \quad (4.6)$$

siendo:

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z(\mathbf{s}_i),$$

la media muestral y $N(\mathbf{h})$ dado por (4.4) (o reemplazándolo por (4.5) en el caso posiciones irregulares). El principal problema con este estimador es la necesidad de estimar la media μ del proceso (además de que la suposición de estacionariedad de segundo orden es menos general; sección 2.1).

Existen numerosos estudios en la literatura sobre las propiedades estadísticas de los estimadores (4.3) y (4.6). Por ejemplo, se ha probado (bajo

² Algunos detalles adicionales sobre la selección de las regiones de tolerancia se tienen en la sección 7.1.

condiciones “mixing” adecuadas) su normalidad asintótica³ (ver p.e. Davis y Borgman, 1979, para el caso del variograma) y se obtuvieron también expresiones que permiten obtener la matriz de varianzas-covarianzas de estimaciones a distintos saltos (p.e. las correspondientes a (4.3) se tienen en Cressie, 1985). A continuación se comentan algunas propiedades de estos estimadores que justifican el modelado de la dependencia espacial a través del semivariograma. Mas detalles sobre la distribución y propiedades de estos estimadores se tienen por ejemplo en Cressie (1993, pp. 71-74).

VENTAJAS DE LA ESTIMACIÓN DEL VARIOGRAMA

Algunas de las principales razones que justifican el empleo del estimador del variograma (4.3) en lugar del estimador (4.6) del covariograma son las siguientes:

- *Sesgo*: Cuando el proceso $Z(\cdot)$ es intrínsecamente estacionario $2\hat{\gamma}(\cdot)$ es un estimador insesgado de $2\gamma(\cdot)$, sin embargo, cuando el proceso es estacionario de segundo orden, $\hat{C}(\cdot)$ tiene sesgo $O(1/n)$ (ver p.e. Fuller, 1996, sección 6.2).
- *Estimación basada en residuos*: Si la media no es constante debe ser estimada y eliminada de los datos para poder utilizar estos estimadores (ver comentarios en el apartado anterior). En ese caso los estimadores obtenidos a partir de los residuos son sesgados, sin embargo Cressie y Grondona (1992) probaron, en un caso bastante general, que el sesgo del estimador del variograma es de menor orden que el del correspondiente estimador del covariograma (en la sección 4.1.2 se tienen más detalles sobre este tema).
- *Contaminación por tendencia*: Si la media no es constante y no es eliminada adecuadamente se pueden producir estimaciones desastrosas del covariograma al utilizar $\hat{C}(\cdot)$. Por ejemplo Cressie (1993, p. 72) probó, para un caso particular, que el efecto de introducir una

³ Aunque la distribución del estimador clásico del variograma para un salto fijo h se aproxima más a la lognormal (Baczkowski y Mardia, 1987).

contaminación lineal es mucho menor en el estimador del variograma y especialmente cerca del origen.

ESTIMADORES ROBUSTOS

Uno de los problemas con el estimador empírico del semivariograma (4.3) es su falta de robustez frente a observaciones atípicas. Por este motivo se han propuesto numerosas alternativas robustas en la literatura; por ejemplo, una recopilación y comparativa de este tipo de estimadores se tiene en Dutter (1996).

A continuación se exponen algunos de éstos estimadores (propuestos inicialmente por Hawkins y Cressie, 1984) basados en la siguiente idea⁴: si el proceso $Z(\cdot)$ es normal entonces $(Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h}))^2$ sigue una distribución $2\gamma(\mathbf{h})\chi_1^2$, sin embargo esta distribución es muy asimétrica y la transformación de potencia que hace que se aproxime más a la simetría (normalidad) es la raíz cuarta. Por tanto Hawkins y Cressie (1984) sugirieron trabajar con la raíz cuadrada de las diferencias⁵, i.e. $|Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})|^{\frac{1}{2}}$, y posteriormente transformar el resultado a la escala original pero asegurándose de obtener un estimador aproximadamente insesgado (utilizando del método delta). Basándose en esta idea se obtienen los estimadores:

$$2\bar{\gamma}(\mathbf{h}) = \left(\frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{\frac{1}{2}} \right)^4 \left/ \left(0.457 + \frac{0.494}{|N(\mathbf{h})|} + \frac{0.045}{|N(\mathbf{h})|^2} \right) \right., \quad (4.7)$$

o:

$$2\tilde{\gamma}(\mathbf{h}) = \frac{1}{0.457} \left(\text{mediana} \left\{ |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{\frac{1}{2}} : (i, j) \in N(\mathbf{h}) \right\} \right)^4,$$

donde $N(\mathbf{h})$ esta definido por (4.4) (o (4.5) en el caso posiciones irregulares). Aunque este último estimador es más robusto, diversos estudios de simulación (p.e. Cressie y Hawkins, 1980) mostraron que $\bar{\gamma}(\cdot)$ es un estimador más eficiente que $\tilde{\gamma}(\cdot)$.

⁴ Que será utilizada también en el capítulo 7 para la estimación no paramétrica robusta de un variograma espacio-temporal.

⁵ Otra ventaja de utilizar la raíz cuadrada de las diferencias es que, en general, están menos correladas que las diferencias al cuadrado (ver p.e. Cressie, 1993, p. 76)

Estos estimadores son utilizados muchas veces en la práctica junto a métodos robustos de predicción espacial (que por motivos de extensión no se tratan en este trabajo). Por ejemplo Cressie y Hawkins (1980) propusieron una robustificación del kriging ordinario conocida como *kriging robusto* y, para el caso de media no constante, Cressie (1986) propuso el método conocido como *median-polish kriging* (que se puede ser modificado para distintas situaciones, ver p.e. Fernández-Casal, 1995, sección 4.3). Más detalles sobre estos métodos se tienen también por ejemplo en Cressie (1993, secciones 3.3 y 3.5) y Chilès y Delfiner (1999, sección 3.7.2).

ESTIMADORES NO PARAMÉTRICOS

Los estimadores locales tipo núcleo son herramientas frecuentemente utilizadas en la estimación de curvas y superficies. Entre los más conocidos podemos señalar los estimadores tipo Nadaraya-Watson y los polinómicos locales (ver p.e. Wand y Jones, 1995, o Fan y Gijbels, 1996). Recientemente se han considerado también estas ideas para la estimación del covariograma (e.g. Hall et al., 1994; Hall y Patil, 1994) y del semivariograma (e.g. Febrero-Bande et al., 1998; Yu y Mateu, 2002; García-Soidán et al., 2003a y 2003b)

Suponiendo estacionariedad de segundo orden, Hall et al. (1994) consideraron el siguiente estimador tipo Nadaraya-Watson del covariograma en el caso unidimensional:

$$\hat{C}(s) = \frac{\sum_{i=1}^n \sum_{j=i}^n K\left(\frac{s-(s_i-s_j)}{h}\right) (Z(s_i) - \bar{Z})(Z(s_j) - \bar{Z})}{\sum_{i=1}^n \sum_{j=i}^n K\left(\frac{s-(s_i-s_j)}{h}\right)}, \quad (4.8)$$

siendo $K(\cdot)$ una función núcleo unidimensional y h el parámetro ventana. Además, estos autores estudiaron sus propiedades asintóticas (considerando una combinación de dominio creciente y relleno) suponiendo continuidad del covariograma teórico⁶. Un estudio similar ha sido realizado por Hall y Patil (1994) para el caso general multidimensional (no isotrópico).

⁶ Adicionalmente, estos autores propusieron un método para la obtención de covariogramas válidos a partir de estos estimadores, estudiando también las

En el caso de estacionariedad intrínseca, la expresión general de un estimador no paramétrico de un semivariograma isotrópico es:

$$\hat{\gamma}(r) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \omega_{ij}(r) (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2}{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \omega_{ij}(r)}, \quad (4.9)$$

donde $\omega_{ij}(r) \geq 0$, $\forall i, j$ y $\sum_{i,j} \omega_{ij}(r) > 0$. Dependiendo de la elección de estos pesos obtenemos distintos estimadores, por ejemplo tomando:

- $\omega_{ij}(r) = \mathcal{I}_{Tol(r)}(\|\mathbf{s}_i - \mathbf{s}_j\|)$, siendo $Tol(r) \subset \mathbb{R}$ una región de tolerancia en torno a r (y denotando por $\mathcal{I}_A(\cdot)$ función indicadora del conjunto A), obtenemos el estimador clásico (suavizado) del variograma.
- $\omega_{ij}(r) = K\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^{-r}}{h}\right)$, (4.9) es el estimador Nadaraya-Watson (Hall et al., 1994, p. 2123; Febrero-Bande et al., 1998; García-Soidán et al., 2003a).
- $\omega_{ij}(r) = K\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^{-r}}{h}\right) \times \sum_k \sum_l K\left(\frac{\|\mathbf{s}_k - \mathbf{s}_l\|^{-r}}{h}\right) (\|\mathbf{s}_k - \mathbf{s}_l\| - r) (\|\mathbf{s}_k - \mathbf{s}_l\| - \|\mathbf{s}_i - \mathbf{s}_j\|)$

se obtiene el estimador lineal local del semivariograma propuesto por García-Soidán et al. (2003b).

- $\omega_{ij}(r) = \frac{1}{\delta_0(\|\mathbf{s}_i - \mathbf{s}_j\|)} K\left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^{-r}}{h\delta_0(\|\mathbf{s}_i - \mathbf{s}_j\|)}\right)$, siendo $\delta_0(\cdot) > 0$ un parámetro adicional de suavizado, se obtiene el denominado estimador tipo “vecino más próximo” con parámetro de vecindario variable propuesto por Yu y Mateu (2002).

La expresión (4.9) mostrada aquí no incluye los índices $i = j$ ya que en general el semivariograma no es continuo en el origen; sin embargo, si se supone que hay continuidad (e.g. García-Soidán et al., 2003a), puede incrementarse significativamente la eficiencia de estos estimadores cerca del origen incluyendo los términos de la diagonal.

propiedades de la estimación final obtenida; en la sección 4.5.2 se describe brevemente esta metodología.

Las propiedades asintóticas (suponiendo continuidad) de los estimadores tipo Nadaraya-Watson y del estimador lineal local han sido estudiadas por García-Soidán et al. (2003a) y (2003b) respectivamente⁷. Además de realizar un estudio similar al de Hall et al. (1994) o Hall y Patil (1994); hay que destacar que, entre otros resultados⁸, estos autores obtuvieron la expresión de la ventana óptima h que minimiza asintóticamente el error en media cuadrática. Yu y Mateu (2002) realizaron un estudio similar sobre el estimador tipo vecino más próximo descrito anteriormente.

4.1.2 Estimación del variograma en el KU

En el caso del kriging universal no es apropiado utilizar directamente los estimadores del semivariograma mostrados en la sección anterior, ya que la media no es constante y por tanto:

$$E(Z(\mathbf{s}_1) - Z(\mathbf{s}_2))^2 = 2\gamma(\mathbf{s}_1 - \mathbf{s}_2) + \left(\sum_{j=0}^p \beta_j (f_j(\mathbf{s}_1) - f_j(\mathbf{s}_2)) \right)^2,$$

utilizando la notación de la sección 3.3 (que será la empleada a lo largo de este sección).

En ocasiones se puede “solventar” este problema analizando el comportamiento del proceso en distintas direcciones. Si se encuentra una dirección en los datos a través de la cual la media es aparentemente constante, se puede estimar el variograma en esa dirección (utilizando alguno de los estimadores anteriores) y posteriormente ajustar un modelo válido de variograma. Si el variograma se puede suponer isotrópico y el modelo ajustado es también un variograma isotrópico válido en \mathbb{R}^d , se obtiene de esta forma un modelo válido para cualquier dirección (ver p.e. Cressie, 1993, sección 3.4.1 o

⁷ En la sección 7.2 se realizan algunas observaciones adicionales sobre este tipo de estimadores (en el caso espacio-temporal).

⁸ Adicionalmente, propusieron el uso de estos estimadores piloto junto a los modelos flexibles de Shapiro-Botha (1991) y estudiaron las propiedades de la estimación final obtenida; en la sección 4.5.1 se tienen más detalles sobre esta aproximación.

Chilès y Delfiner, 1999, sección 2.7.4). Sin embargo esta solución no siempre es adecuada o se puede llevar a cabo en la práctica.

Si β es conocido (kriging ordinario) podemos estimar el variograma a partir del proceso:

$$\delta(\mathbf{s}) = Z(\mathbf{s}) - \sum_{j=0}^p f_j(\mathbf{s})\beta_j.$$

En el caso del KU β no es conocido, por lo que se puede pensar en estimarlo, calcular los residuos (estimados) y obtener una estimación del variograma a partir de ellos. Sin embargo para que β pueda ser estimado de forma eficiente sería necesario (en principio) conocer $Var(\mathbf{Z})$, i.e. se necesita conocer $\gamma(\cdot)$. Este problema circular es el principal inconveniente achacado al kriging universal (ver p.e. Armstrong, 1984, para más detalles).

ESTIMACIÓN DEL VARIOGRAMA BASADA EN LOS RESIDUOS

Supongamos por el momento que el proceso $\delta(\cdot)$ es estacionario de segundo orden con covariograma $C(\cdot)$ (más adelante se tratará el caso de que este proceso solamente sea intrínsecamente estacionario). Cuando $\Sigma = Var(\mathbf{Z})$ es conocida podemos calcular el estimador lineal óptimo de β :

$$\hat{\beta}_{mcg} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Z}.$$

A partir de este estimador, denotando por $\mathbf{P}_{mcg} = \mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}$ la matriz proyección, obtenemos el vector de residuos estimados:

$$\hat{\delta} = \mathbf{Z} - \mathbf{X}\hat{\beta}_{mcg} = (\mathbf{I}_n - \mathbf{P}_{mcg})\mathbf{Z},$$

cuya matriz de varianzas-covarianzas resulta ser:

$$\begin{aligned} Var(\hat{\delta}) &= (\mathbf{I}_n - \mathbf{P}_{mcg})\Sigma(\mathbf{I}_n - \mathbf{P}_{mcg})' \\ &= \Sigma - \mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'. \end{aligned} \tag{4.10}$$

Por tanto si utilizamos estimaciones de los residuos, incluso procediendo de la forma más eficiente, se introduce un sesgo en la estimación de la dependencia espacial. Explícitamente, teniendo en cuenta que $Var(\hat{\beta}_{mcg}) = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$, (4.10) es equivalente a:

$$\begin{aligned} C_{\hat{\delta}}(\mathbf{s}_i, \mathbf{s}_j) &= Cov(Z(\mathbf{s}_i) - \hat{\mu}(\mathbf{s}_i), Z(\mathbf{s}_j) - \hat{\mu}(\mathbf{s}_j)) \\ &= C(\mathbf{s}_i - \mathbf{s}_j) - Cov(\hat{\mu}(\mathbf{s}_i), \hat{\mu}(\mathbf{s}_j)), \end{aligned}$$

denotando por $\hat{\mu}(\mathbf{s})$ la estimación m.c.g. de la tendencia, y expresado en función del semivariograma obtenemos que:

$$\gamma_{\hat{\delta}}(\mathbf{s}_i, \mathbf{s}_j) = \gamma(\mathbf{s}_i - \mathbf{s}_j) - \frac{1}{2}Var(\hat{\mu}(\mathbf{s}_i) - \hat{\mu}(\mathbf{s}_j)). \quad (4.11)$$

Por tanto al utilizar alguno de los estimadores mostrados anteriormente con los residuos estimados obtenemos estimaciones sesgadas del semivariograma teórico. Matheron (1971, pp. 152-155) ya notó que, por lo general, el sesgo del estimador del semivariograma es pequeño en los saltos próximos al origen, pero más sustancial en saltos grandes. Para un caso particular, Cressie (1993, pp. 166-167) observó que los residuos basados en el estimador m.c.g. dan lugar a un estimador del variograma con sesgo negativo y cuadrático en \mathbf{h} . Parece ser que este problema provocó una desilusión con el kriging universal y la iniciativa hacia el kriging con funciones intrínsecamente estacionarias (ver p.e. Matheron, 1973; Cressie, 1993, sección 5.4; o Chilès y Delfiner, 1999, cap. 4).

En la práctica además Σ es desconocida; para solventar este problema Neuman y Jacobson (1984) propusieron una aproximación iterativa, empezar con el estimador m.c.o. de β , estimar al variograma a partir de los residuos, ajustar un modelo de variograma válido, calcular el estimador m.c.g. basado en el modelo ajustado y así sucesivamente hasta convergencia. En la práctica este procedimiento suele converger en pocas iteraciones (normalmente menos de 5). Hay que destacar que esta forma de proceder no soluciona el problema del sesgo, ya que incluso cuando Σ es conocida el sesgo no es nulo.

En cuanto a las consecuencias de que el estimador del variograma no sea insesgado en el kriging universal, hay que tener en cuenta que:

- Al ajustar un modelo de variograma por mínimos cuadrados ponderados o generalizados (sección 4.3.1), automáticamente los saltos pequeños reciben mayor peso en el ajuste.
- Además si la predicción espacial se lleva a cabo con un criterio de vecindad, el variograma sólo es evaluado en saltos pequeños, donde se tiene una buena estimación y un buen ajuste.

- También hay que tener en cuenta el resultado de Stein (1988), i.e. para una predicción eficiente generalmente sólo es necesario capturar la conducta del variograma cerca del origen

Por lo anterior, el desencanto con el kriging universal ha sido prematuro, el efecto del sesgo del estimador del variograma sobre el predictor del kriging universal es pequeño. Sin embargo la varianza del kriging universal se ve más afectada y es menor de lo que debería ser (para más detalles ver Cressie, 1993, pp. 296-299; ver también la sección 4.4. para una posible solución). Adicionalmente se han propuesto alternativas a los métodos de ajuste basados en mínimos cuadrados (mostrados en la sección 4.3.1) que tienen en cuenta el sesgo en la estimación del variograma (ver Beckers y Bogaert, 1998; y comentarios en sección 4.3.1).

Estimación en el caso de variogramas no acotados

Si el proceso $\delta(\cdot)$ sólo es intrínsecamente estacionario, no está disponible la matriz Σ y en principio es imposible estimar la tendencia. Sin embargo, normalmente se suele trabajar en un dominio acotado D y podemos encontrar una constante positiva A tal que $C^*(\mathbf{h}) = A - \gamma(\mathbf{h}) \geq 0, \forall \mathbf{h} \in D$ (y por tanto esta función es un covariograma válido en ese dominio); la función $C^*(\mathbf{h})$ se suele denominar “pseudo-covariograma” (o covarianza localmente equivalente; ver p.e. Chilès y Delfiner, 1999, sección 4.6.2). Si utilizamos $C^*(\mathbf{h})$ en lugar del covariograma en la estimación de la media (o en las ecuaciones del predictor del KU), la constante A se cancela y obtenemos los mismos resultados (sin embargo las varianzas si que dependen de esta constante).

4.2 MODELOS PARAMÉTRICOS DE VARIOGRAMAS

Los variogramas deben ser condicionalmente semidefinidos negativos, una propiedad que los estimadores normalmente no poseen. Tradicionalmente esto se remedia ajustando un modelo paramétrico válido al estimador piloto (sección 4.3). En la sección 4.2.1 se comentan algunas aproximaciones para la obtención de modelos válidos de semivariogramas y a continuación, sección 4.2.2, se tiene una revisión de los modelos isotrópicos tradicionalmente utilizados en

geoestadística. Estos modelos son empleados también en algunos casos como estructuras básicas a partir de las cuales se construyen modelos más complejos como los denominados *modelos lineales de regionalización*. Este tipo de modelos son tratados en la sección 4.2.3.

4.2.1 Obtención de modelos paramétricos

Entre las principales aproximaciones para la obtención de modelos de semivariograma (o covariograma) destacan las basadas en la representación espectral (sección 2.2). Por ejemplo, teniendo en cuenta (2.22), a partir de la función (isotrópica) de densidad espectral:

$$g(\lambda) = 4\pi^{2-d/2}\Gamma(\frac{d}{2})|\lambda|^{2-d},$$

se obtiene el modelo isotrópico de semivariograma:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + b\|\mathbf{h}\| & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases} \quad (4.12)$$

denominado *modelo lineal*; donde $\boldsymbol{\theta} = (c_0, b)'$, siendo $c_0 \geq 0$ el efecto nugget y $b \geq 0$ el parámetro de escala. Este modelo es válido en \mathbb{R}^d , $\forall d \geq 1$.

Otra alternativa que puede ser de interés es la basada en la representación del proceso $Z(\cdot)$ como la media móvil de un proceso espacial independiente:

$$Z(\mathbf{s}) = \int_{\mathbb{R}^d} f(\mathbf{x} - \mathbf{s})W(\mathbf{x})d\mathbf{x} + \varepsilon(\mathbf{s}),$$

siendo f una función real cuyo cuadrado es integrable en \mathbb{R}^d , $W(\cdot)$ un proceso espacial ruido blanco de media cero y varianza unidad y $\varepsilon(\mathbf{s})$ es un proceso ruido blanco de media cero, independiente de $W(\mathbf{s})$ y con varianza c_0 . A partir de esta expresión se deduce que el proceso $Z(\cdot)$ es estacionario de segundo orden y su variograma (acotado) viene dado por:

$$2\gamma(\mathbf{h}) = 2c_0 + \int_{\mathbb{R}^d} (f(\mathbf{x})f(\mathbf{x} - \mathbf{h}))^2 d\mathbf{x}, \quad (4.13)$$

para $\mathbf{h} \neq \mathbf{0}$ (ver p.e. Barry y Ver Hoef, 1996, apéndice A). Teniendo en cuenta este resultado se puede pensar en utilizar parametrizaciones de la función f de forma que la integral (4.13) admita una solución explícita. Por ejemplo, considerando la función definida en \mathbb{R}^1 :

$$f(x) = \sqrt{b}I_{\{-a \leq x \leq 0\}},$$

se obtiene el semivariograma lineal-acotado:

$$\gamma(h|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } h = 0 \\ c_0 + b|h| & \text{si } 0 < |h| \leq a \\ c_0 + ba & \text{si } |h| > a \end{cases} \quad (4.14)$$

siendo $\boldsymbol{\theta} = (c_0, a, b)'$ con $c_0 \geq 0$, $a \geq 0$ y $b \geq 0$.

Algunos de los resultados mostrados en la sección 2.2 (y otros similares) son utilizados también para la obtención de modelos válidos de variogramas o covariogramas (por ejemplo, en ocasiones se recurre a la propiedad (2.8) para obtener nuevos modelos a partir de los ya conocidos). Referencias con información adicional sobre este tema que pueden ser de interés son por ejemplo Yaglom (1986), Christakos (1992) o Chilès y Delfiner (1999, secciones 2.3 y 2.5).

4.2.2 Modelos paramétricos isotrópicos

A continuación se presentan algunos de los modelos paramétricos de semivariograma conocidos, una revisión más completa se tiene por ejemplo en Chilès y Delfiner (1999, sección 2.5.1). En la notación utilizada en las parametrizaciones $c_0 \geq 0$ representa el efecto nugget, $c_1 \geq 0$ el umbral parcial (en el caso de variogramas acotados) y $a > 0$ el rango (si existe) o el parámetro de escala. En el caso de semivariogramas acotados que alcanzan el umbral asintóticamente (rango infinito), el parámetro a representa el *rango práctico*, definido como la distancia en la que el valor del semivariograma es el 95% del umbral parcial.

Los modelos isotrópicos de semivariograma más utilizados en geoestadística son:

- Modelo esférico:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \|\mathbf{h}\| = 0 \\ c_0 + c_1 \left\{ \frac{3}{2} \frac{\|\mathbf{h}\|}{a} - \frac{1}{2} \left(\frac{\|\mathbf{h}\|}{a} \right)^3 \right\} & \text{si } 0 < \|\mathbf{h}\| \leq a \\ c_0 + c_1 & \text{si } \|\mathbf{h}\| > a \end{cases}$$

válido en \mathbb{R}^d , $d = 1, 2, 3$.

- Modelo exponencial:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + c_1 \left(1 - \exp\left(-\frac{3\|\mathbf{h}\|}{a}\right)\right) & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases}$$

válido en \mathbb{R}^d , $\forall d \geq 1$.

- Modelo racional cuadrático:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + c_1 \frac{\|\mathbf{h}\|^2}{\frac{1}{19}a^2 + \|\mathbf{h}\|^2} & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases}$$

válido en \mathbb{R}^d , $\forall d \geq 1$.

- Modelo potencial:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + a\|\mathbf{h}\|^\lambda & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases}$$

con $0 \leq \lambda < 2$ y válido en \mathbb{R}^d , $\forall d \geq 1$. En el caso de $\lambda = 1$ se obtiene el conocido modelo lineal (4.12).

- Modelo exponencial-potencial:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + c_1 \left(1 - \exp\left(-3\left(\frac{\|\mathbf{h}\|}{a}\right)^\lambda\right)\right) & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases} \quad (4.15)$$

con $0 \leq \lambda \leq 2$ y válido en \mathbb{R}^d , $\forall d \geq 1$. Cuando $\lambda = 2$ (4.15) es denominado modelo gaussiano; este modelo sin embargo no debería ser utilizado en la predicción espacial debido a las inestabilidades numéricas que produce en los algoritmos kriging (especialmente cuando el efecto nugget es grande; ver p.e. Wackernagel, 1998, pp. 120-123). El modelo exponencial se obtiene también como caso particular cuando $\lambda = 1$.

- Modelo oscilatorio:

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + c_1 \left(1 - \frac{a}{\|\mathbf{h}\|} \text{sen}\left(\frac{\|\mathbf{h}\|}{a}\right)\right) & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases}$$

válido en \mathbb{R}^d , $d = 1, 2, 3$. Este modelo con forma de onda (hay correlaciones negativas) alcanza su valor máximo ($c_0 + 1.218c_1$) cuando $\|\mathbf{h}\| \simeq 4.5a$, siendo a el parámetro de escala.

- Modelo de Matérn (o K-Bessel):

$$\gamma(\mathbf{h}|\boldsymbol{\theta}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ c_0 + c_1 \left(1 - \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\|\mathbf{h}\|}{a} \right)^\nu K_\nu \left(\frac{\|\mathbf{h}\|}{a} \right) \right) & \text{si } \mathbf{h} \neq \mathbf{0} \end{cases}$$

siendo $\nu \geq 0$ y K_ν la función de Bessel modificada de tercera clase de orden ν (ver p.e. Abramowitz y Stegun, 1965, pp. 374-379). Este modelo es válido en \mathbb{R}^d , $\forall d \geq 1$. El modelo exponencial se obtiene como caso particular cuando $\nu = \frac{1}{2}$ y en el límite $\nu \rightarrow \infty$ el modelo gaussiano.

En la figura 2.3 se tienen algunos ejemplos de las formas de algunos de estos semivariogramas.

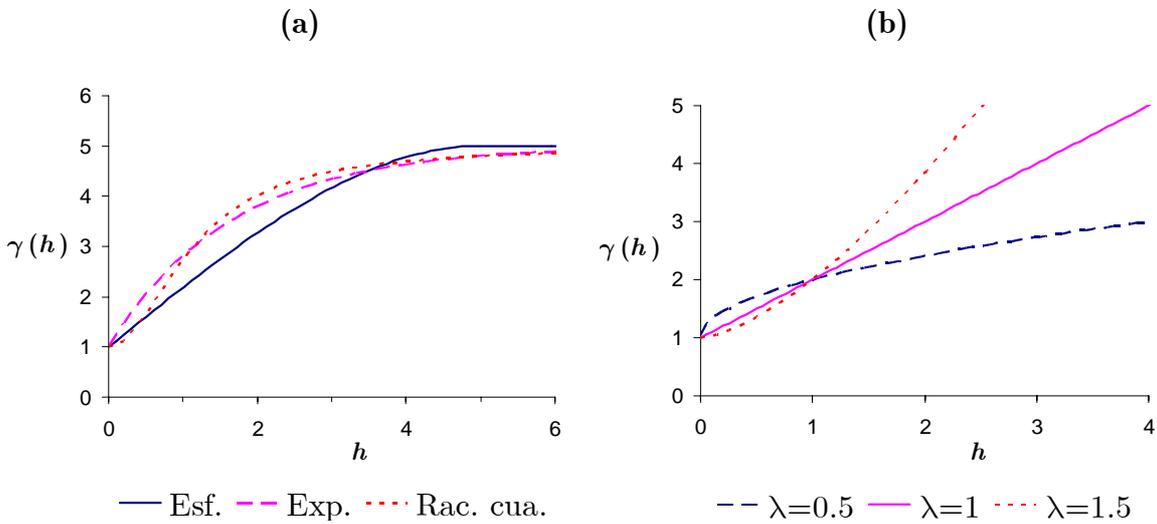


Figura 2.3: Representaciones de algunos los modelos paramétricos isotrópicos de semivariogramas: (a) esférico, exponencial y racional cuadrático con parámetros $c_0 = 1$, $c_1 = 4$ y $a = 5$; (b) modelo potencial con $c_0 = 1$, $a = 1$ y $\lambda = 0.5$, 1 (lineal) y 1.5.

4.2.3 El modelo lineal de regionalización

Estos modelos se basan en la suposición de que el proceso $Z(\cdot)$ puede expresarse como⁹:

$$Z(\mathbf{s}) = \sum_{k=0}^q a_k Y_k(\mathbf{s}) + \mu,$$

donde $Y_k(\mathbf{s})$ son procesos mutuamente independientes, de media cero e intrínsecamente estacionarios (o de segundo orden) con semivariograma $\gamma_k(\mathbf{h})$ (que se supone es un modelo paramétrico elemental). A partir de esto se deduce que el semivariograma del proceso $Z(\cdot)$ viene dado por:

$$\gamma(\mathbf{h}) = \sum_{k=0}^q a_k^2 \gamma_k(\mathbf{h}). \quad (4.16)$$

Los modelos de la forma (4.16) son denominados *modelos lineales de regionalización*, también conocidos en la literatura como “nested models” o *modelos anidados*. Uno de los modelos básicos $\gamma_k(\mathbf{h})$ suele ser un modelo efecto nugget y el resto algunos de los modelos mostrados en la sección anterior con efecto nugget nulo y umbral unidad. Además, cada modelo básico puede ser isotrópico o disponer de algún tipo de anisotropía (sección 2.2.2); por ejemplo en ocasiones se suponen con anisotropía geométrica y por tanto tendríamos que:

$$\gamma_k(\mathbf{h}) \equiv \gamma_k(\|\mathbf{A}_k \mathbf{h}\|)$$

siendo \mathbf{A}_k , $k = 0, \dots, q$ matrices $d \times d$. De esta forma estos modelos pueden ser lo suficientemente flexibles como para modelar la mayoría de situaciones que se pueden presentar en la práctica. Sin embargo, es difícil establecer un procedimiento automático (o semi-automático) para la selección y el ajuste de este tipo de modelos. Esto provoca que el proceso normalmente se realice en la práctica de forma interactiva por el usuario y utilizando principalmente herramientas gráficas; siendo por tanto poco recomendables para algunos casos. En primer lugar hay que especificar el número y tipo de estructuras básicas, y

⁹ Realmente no es necesario considerar esta representación (relacionada con el denominado kriging factorial, e.g. Goovaerts, 1997, sección 5.6); estos modelos también se pueden obtener simplemente considerando las propiedades 1 y 2 del semivariograma mostradas en la sección 2.2.1.

en segundo lugar (aunque se suele hacer en la práctica de forma simultánea) está el problema de la estimación de los parámetros, donde puede ser especialmente complicado la determinación de los rangos y los parámetros de anisotropía de los distintos componentes. Ver por ejemplo Goovaerts (1997, sección 4.2.4) para más detalles sobre el uso en la práctica de éste tipo de modelos.

4.3 PRINCIPALES MÉTODOS PARA EL AJUSTE DE UN MODELO

VÁLIDO

Como ya se comentó anteriormente, en general los estimadores del variograma no pueden ser usados directamente en la predicción espacial; no son condicionalmente semidefinidos negativos y eso puede causar por ejemplo estimaciones negativas de la varianza kriging. Este problema normalmente se remedia buscando un modelo paramétrico válido que describa adecuadamente la dependencia espacial presente en los datos. Supongamos que $P = \{2\gamma(\mathbf{h}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$, donde $2\gamma(\mathbf{h}; \boldsymbol{\theta})$ es un variograma válido en \mathbb{R}^d (normalmente isotrópico), es la familia parametrizada de variogramas escogida. Se trata de encontrar el mejor elemento de P , para lo que se han propuesto diversos criterios de bondad de ajuste (ver p.e. Cressie, 1993, sección 2.6). Entre ellos hay que destacar los basados en mínimos cuadrados y en máxima verosimilitud, descritos a continuación.

4.3.1 Mínimos cuadrados

Supongamos que $2\gamma(\mathbf{h}; \boldsymbol{\theta}_0)$ es el variograma teórico y que $\hat{\gamma}_i = \hat{\gamma}(\mathbf{h}_i)$, $i = 1, \dots, K$, son las estimaciones del semivariograma obtenidas utilizando algún tipo de estimador piloto (p.e. alguno de los mostrados en la sección 4.1.1). Normalmente, siguiendo las recomendaciones sugeridas por Journel y Huijbregts (1978, p. 194), solamente se consideran en el ajuste saltos menores o iguales que la mitad del máximo salto (i.e. $\|\mathbf{h}_i\| \leq \frac{1}{2} \max\{\|\mathbf{s}_k - \mathbf{s}_l\|\}$); y, si se utiliza el estimador empírico (o uno similar), de forma que el número de aportaciones a cada estimación sea por lo menos de treinta (i.e. $|N(\mathbf{h}_i)| \geq 30$).

Habitualmente (e.g. Cressie, 1993, p. 96-97) la estimación por mínimos cuadrados de $\boldsymbol{\theta}_0$ se obtiene al minimizar:

$$(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta}))' \mathbf{V}(\boldsymbol{\theta})(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta})), \quad (4.17)$$

siendo $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}(\mathbf{h}_1), \dots, \hat{\gamma}(\mathbf{h}_K))'$, $\boldsymbol{\gamma}(\boldsymbol{\theta}) = (\gamma(\mathbf{h}_1; \boldsymbol{\theta}), \dots, \gamma(\mathbf{h}_K; \boldsymbol{\theta}))'$ y $\mathbf{V}(\boldsymbol{\theta})$ una matriz $K \times K$ semidefinida positiva que puede depender de $\boldsymbol{\theta}$, considerando alguno de los siguientes casos:

- *Mínimos cuadrados ordinarios* (m.c.o.): cuando $\mathbf{V}(\boldsymbol{\theta}) = \mathbf{I}_K$, siendo \mathbf{I}_K la matriz identidad $K \times K$.
- *Mínimos cuadrados ponderados* (m.c.p.): si se considera que $\mathbf{V}(\boldsymbol{\theta}) = \text{diag}(w_1(\boldsymbol{\theta}), \dots, w_K(\boldsymbol{\theta}))$, con $w_i(\boldsymbol{\theta}) \geq 0$, $i = 1, \dots, K$. Normalmente se suele tomar estos pesos inversamente proporcionales a $\text{Var}(\hat{\gamma}(\mathbf{h}_i))$.
- *Mínimos cuadrados generalizados* (m.c.g.): si $\mathbf{V}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\theta})^{-1}$, siendo $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\theta})$ la matriz de covarianzas (asintótica) de $\hat{\boldsymbol{\gamma}}$ obtenida suponiendo que el variograma teórico es $2\gamma(\mathbf{h}; \boldsymbol{\theta})$.

Sin embargo esta aproximación ha recibido algunas críticas debido a que la matriz de pesos utilizada en el ajuste depende también del parámetro sobre el que se realiza la minimización. Otra alternativa sería intentar obtener $\hat{\boldsymbol{\theta}}_0$ de forma que minimizara:

$$(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta}))' \mathbf{V}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta})), \quad (4.18)$$

sin embargo esto no es en principio posible ya que (4.18) depende del verdadero valor $\boldsymbol{\theta}_0$ desconocido. Para solucionar este problema se puede pensar en proceder de forma iterativa; comenzando con una estimación inicial $\hat{\boldsymbol{\theta}}_0^{(0)}$, obtenida por ejemplo mediante m.c.o., y posteriormente en cada etapa k obtener una nueva aproximación $\hat{\boldsymbol{\theta}}_0^{(k)}$ al minimizar:

$$(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta}))' \mathbf{V}(\hat{\boldsymbol{\theta}}_0^{(k-1)})(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta})), \quad (4.19)$$

repetiendo este proceso hasta convergencia (realmente la mayoría de los algoritmos diseñados para el ajuste por mínimos cuadrados proceden de esta forma).

Es importante señalar también que al utilizar el criterio m.c.g. el cálculo de la matriz de covarianzas $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}}(\boldsymbol{\theta})$ generalmente no resulta fácil; por ejemplo en Cressie (1985) se tienen las expresiones (suponiendo normalidad) para el

estimador empírico (4.3) y el estimador robusto (4.7) (en Cressie 1993, p. 96 hay un pequeño sumario). Esto produce que la minimización de la función objetivo (4.17) (o (4.19)) sea computacionalmente prohibitiva en muchos casos. Por este motivo se han considerado simplificaciones al método de mínimos cuadrados generalizados; por ejemplo, Genton (1998) propuso utilizar en lugar de $\Sigma_{\hat{\gamma}}(\boldsymbol{\theta})$ una matriz de covarianzas obtenida suponiendo datos independientes y regularmente espaciados (dando la expresión explícita de esa matriz).

El método de mínimos cuadrados ponderados puede verse como un compromiso entre la eficiencia del método de m.c.g. y la simplicidad del método de m.c.o.. Además, suponiendo normalidad y que el variograma teórico es $2\gamma(\mathbf{h};\boldsymbol{\theta})$, Cressie (1985) probó que:

$$\text{Var}(\hat{\gamma}(\mathbf{h}_i)) \simeq 2 \frac{\gamma(\mathbf{h}_i;\boldsymbol{\theta})^2}{|N(\mathbf{h}_i)|}, \quad (4.20)$$

en el caso del estimador empírico (4.3); y para el estimador robusto (4.7):

$$\text{Var}(\bar{\gamma}(\mathbf{h}_i)) \simeq 2.885 \frac{\gamma(\mathbf{h}_i;\boldsymbol{\theta})^2}{|N(\mathbf{h}_i)|},$$

siendo esta aproximación incluso mejor que en el caso anterior¹⁰. Proponiendo en estos casos la minimización de:

$$\sum_{i=1}^K w_i(\boldsymbol{\theta})(\hat{\gamma}(\mathbf{h}_i) - \gamma(\mathbf{h}_i;\boldsymbol{\theta}))^2, \quad (4.21)$$

siendo $w_i(\boldsymbol{\theta}) = |N(\mathbf{h}_i)|/\gamma(\mathbf{h}_i;\boldsymbol{\theta})^2$, como aproximación al criterio m.c.p.. De esta forma sin embargo los pesos dependen del parámetro $\boldsymbol{\theta}$ (y al minimizar (4.21) en cierto sentido se están maximizando también las varianzas), por lo que puede ser preferible utilizar el algoritmo iterativo descrito anteriormente¹¹.

Estos criterios de ajuste tiene unas propiedades interesantes: cuanto mayor sea $|N(\mathbf{h}_i)|$ mayor peso recibe el residuo en el salto \mathbf{h}_i y además, cuanto más pequeño sea el valor del variograma teórico mayor peso recibe también el residuo correspondiente. Por este motivo, los saltos próximos al origen típicamente reciben mayor peso con lo que se consigue un buen ajuste del

¹⁰ Genton (1998) generalizó estas aproximaciones para el caso de M-estimadores.

¹¹ Otra posible alternativa sería tomar $w_i = |N(\mathbf{h}_i)|/\hat{\gamma}(\mathbf{h}_i)^2$, de esta forma se evitaría el tener que proceder de forma iterativa.

modelo de variograma cerca del origen (esto es especialmente importante; ver p.e. Stein, 1988, y comentarios en la sección 3.4.2). Adicionalmente estos métodos pueden ser implementados fácilmente en la práctica (de forma similar al m.c.o.).

Aunque para obtener las expresiones (o aproximaciones) de las varianzas y covarianzas de las estimaciones piloto se supone habitualmente que la distribución de los datos es normal, se puede probar fácilmente que los procedimientos de ajuste obtenidos son también válidos para el caso de datos normales transformados (ver p.e. Cressie, 1993, p. 98). Esta es una de las principales ventajas de los métodos m.c.p. o m.c.g. frente a otras alternativas; como utilizan solamente la estructura de segundo orden (asintótica) del estimador del variograma, no es necesario hacer suposiciones sobre la distribución completa de los datos. De hecho la distribución y eficiencia asintótica de los estimadores mínimo cuadráticos ha sido estudiada por Lahiri et al. (2003) (considerando una conducta asintótica que combina dominio creciente y de relleno; ver sección 2.1), demostrando su consistencia y normalidad asintótica bajo condiciones muy generales.

Como ya se comentó en la sección 4.1.2, en el caso del kriging universal el ajuste se realiza normalmente siguiendo el procedimiento iterativo propuesto por Neuman y Jacobson (1984). Sin embargo la estimación basada en los residuos introduce sesgos en la estimación de los parámetros del variograma. Otra alternativa es tener en cuenta el sesgo del estimador piloto del variograma en el ajuste. Beckers y Bogaert (1998) propusieron minimizar, utilizando el criterio de mínimos cuadrados, las diferencias entre las estimaciones piloto $\hat{\gamma}_R(\mathbf{h}_i)$ obtenidas a partir de residuos m.c.o. y sus valores esperados $E(\hat{\gamma}_R(\mathbf{h}_i))$ obtenidos de forma análoga a (4.11). Sin embargo, teniendo en cuenta las observaciones realizadas en la sección 4.1.2, en la mayoría de los casos no es realmente necesario utilizar esta aproximación.

DETALLES PRÁCTICOS Y COMPUTACIONALES

Para la minimización de la función objetivo (4.17) (o (4.19)) es necesario utilizar algoritmos de minimización multidimensional no lineal. Según la experiencia personal, después de haber considerado distintos métodos, es

recomendable utilizar un procedimiento iterativo que combine regresión lineal m.c.p. (o m.c.g., según el caso) con un algoritmo Levenberg-Marquardt no lineal con restricciones en los parámetros. Hay que tener en cuenta que los modelos paramétricos normalmente considerados (p.e. los mostrados en la sección 4.2) son funciones lineales del efecto nugget y del umbral parcial si se fijan el resto de parámetros.

Por ejemplo en el caso de m.c.p. el procedimiento sería el descrito a continuación; en la k -ésima iteración:

1. Se recalculan los pesos:

$$w_i^{(k)} = \frac{|N(\mathbf{h}_i)|}{\gamma(\mathbf{h}_i; \hat{\boldsymbol{\theta}}_0^{(k-1)})^2},$$

donde $\hat{\boldsymbol{\theta}}_0^{(k-1)}$ es la aproximación obtenida en la iteración anterior. En la primera iteración se igualan los pesos a 1 (m.c.o.).

2. Se estima el efecto nugget y el umbral parcial por regresión lineal m.c.p. (fijando los demás parámetros); si alguno de los valores resulta negativo, se iguala a cero y el otro parámetro se estima de nuevo.
3. Se estiman el resto de parámetros (fijando el efecto nugget y el umbral parcial) utilizando un algoritmo Levenberg-Marquardt modificado con restricciones en los parámetros (rutina BCLSF de la IMSL).

En el caso de la estimación por m.c.g. el algoritmo es similar; en la k -ésima iteración:

1. Se calcula $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\theta}}_0^{(k-1)})$, se obtiene la factorización de Cholesky $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\gamma}}}(\hat{\boldsymbol{\theta}}_0^{(k-1)}) = \mathbf{L}\mathbf{L}'$ y se calcula \mathbf{L}^{-1} . En la primera iteración se sustituye esta matriz por la matriz identidad \mathbf{I}_K (m.c.o.).
2. Se estima el efecto nugget y el umbral parcial por regresión lineal m.c.g. (para ello es recomendable utilizar la matriz \mathbf{L}^{-1} obtenida en el paso anterior y un algoritmo de m.c.o. lineal); procediendo de modo análogo al caso anterior si alguno de los valores resulta negativo.
3. Se estiman el resto de parámetros utilizando el algoritmo Levenberg-Marquardt para minimizar $\mathbf{F}(\boldsymbol{\theta})'\mathbf{F}(\boldsymbol{\theta})$, siendo $\mathbf{F}(\boldsymbol{\theta}) = \mathbf{L}^{-1}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\boldsymbol{\theta}))$.

Estos algoritmos requieren fijar unos valores iniciales de los parámetros, salvo para el efecto nugget y el umbral parcial; aunque varias pruebas

mostraron que los ajustes obtenidos no dependen aparentemente de los parámetros iniciales utilizados. En el caso de los parámetros de escala espacial una práctica habitual es fijar el rango (práctico) igual a la mitad del salto máximo considerado en el ajuste. Otra aproximación consiste en ajustar rectas por m.c.o. a las estimaciones piloto del variograma en ciertos saltos para obtener aproximaciones iniciales a los valores buscados (e.g. Fernández-Casal, 1995).

Hay que tener en cuenta los problemas y limitaciones de los algoritmos de minimización no lineal multidimensional. Uno de los problemas más frecuentes es la convergencia del método hacia un mínimo local; por ello, independientemente del algoritmo elegido para realizar el ajuste, es conveniente representar gráficamente el variograma ajustado junto a las estimaciones piloto.

4.3.2 Estimación basada en máxima verosimilitud

La estimación por máxima verosimilitud es un método muy conocido en la inferencia estadística paramétrica, aunque su empleo en geoestadística ha sido relativamente reciente (a partir de mediados de los 80). Si suponemos que la distribución de los datos es normal, se puede deducir fácilmente la expresión de la función de verosimilitud y obtener las estimaciones de los parámetros buscando los valores que la maximizan.

A lo largo de esta sección se supondrá por tanto que la distribución de los datos es normal:

$$\mathbf{Z} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad (4.22)$$

donde $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, utilizando la notación de secciones anteriores.

MÁXIMA VEROSIMILITUD (MV, ML)

Suponiendo (4.22), la función de densidad de los datos es:

$$f(\mathbf{z}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\right\}. \quad (4.23)$$

Además en la mayoría de los casos podemos reparametrizar el covariograma de forma que:

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{V}(\boldsymbol{\theta}), \quad (4.24)$$

siendo σ^2 la varianza (desconocida). A partir de (4.23) y (4.24) se deduce que la expresión del logaritmo negativo de la función de verosimilitud (LNV) es:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{Z}) = \frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln(\sigma^2) + \frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\theta})| + \frac{1}{2\sigma^2} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}), \quad (4.25)$$

donde $|\mathbf{V}(\boldsymbol{\theta})|$ denota el determinante de la matriz $\mathbf{V}(\boldsymbol{\theta})$.

Las estimaciones de los parámetros $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ se obtendrán minimizando (4.25). Un resultado bien conocido es que el mínimo de (4.25) se obtiene, independientemente de $\boldsymbol{\theta}$, para:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{Z}, \quad (4.26)$$

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}(\boldsymbol{\theta})^{-1} (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (4.27)$$

Por tanto la función a minimizar respecto a $\boldsymbol{\theta}$ es:

$$L(\boldsymbol{\theta} | \mathbf{Z}) = L(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}, \hat{\sigma}^2 | \mathbf{Z}) = \frac{n}{2} \ln(2\pi) + \frac{n}{2} \ln(\hat{\sigma}^2) + \frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\theta})| + \frac{n}{2}. \quad (4.28)$$

Para ello es necesario utilizar algoritmos de minimización multidimensional no lineal. Además hay un problema adicional en la minimización de (4.28) y es la posible multimodalidad de esta función (ver p.e. Mardia y Watkins, 1989). Por tanto habría que asegurarse de que el algoritmo elegido no converge a un mínimo local. Al final de esta sección se tienen algunos detalles computacionales para la minimización de esta función. Si $\hat{\boldsymbol{\theta}}$ es la estimación de $\boldsymbol{\theta}$ obtenida al resolver este problema, sustituyendo en (4.26) y (4.27) obtenemos las estimaciones de los demás parámetros.

El comportamiento asintótico (bajo dominio creciente) de estos estimadores ha sido estudiado por Mardia y Marshal (1984), dando condiciones (no muy fáciles de chequear en la práctica) para su consistencia y normalidad asintótica (ver también Cressie, 1993, sección 7.3.1). Sin embargo existen numerosos ejemplos en la literatura donde se observa que los estimadores de σ^2 y $\boldsymbol{\theta}$ pueden tener un sesgo considerable (especialmente cuando la tendencia no es constante), algo que es bastante conocido en la estimación de la varianza con datos independientes. Este es uno de los principales problemas de la estimación MV, que se puede resolver (por lo menos en parte) utilizando la variante de este método mostrada a continuación.

MÁXIMA VEROSIMILITUD RESTRINGIDA (MVR, REML)

El método de MVR se basa en la idea de filtrar los datos de forma que la distribución conjunta no dependa de β . Se trata de maximizar la verosimilitud de $m = n - p - 1$ contrastes de error linealmente independientes:

$$\mathbf{Y} = \mathbf{\Lambda Z}, \quad (4.29)$$

siendo $\mathbf{\Lambda}$ una matriz $m \times n$ de rango m y tal que $\mathbf{\Lambda X} = \mathbf{0}$ (i.e. $E(\mathbf{Y}) = \mathbf{0}$); estas combinaciones lineales se denominan también habitualmente incrementos generalizados. Por ejemplo, podemos construir una matriz $\mathbf{\Lambda}$ a partir de la matriz de proyección:

$$\mathbf{P} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad (4.30)$$

eliminando $p + 1$ filas (ya que esta matriz es de rango m).

Bajo la hipótesis de normalidad (4.22) se tiene que:

$$\mathbf{Y} \sim N(\mathbf{0}, \mathbf{\Lambda \Sigma \Lambda}'), \quad (4.31)$$

y por tanto la distribución de los incrementos generalizados no depende de β . A partir de (4.31) podemos obtener la expresión del LNV de forma análoga al caso anterior. Siguiendo el mismo proceso obtendríamos que la función objetivo a minimizar es igual a (4.28) reemplazando \mathbf{Z} por \mathbf{Y} , m por n , $\mathbf{V}(\theta)$ por $\mathbf{V}_Y(\theta) = \mathbf{\Lambda V}(\theta)\mathbf{\Lambda}'$ y eliminando $\mathbf{X}\beta$:

$$L(\theta | \mathbf{Y}) = L(\theta, \hat{\sigma}_Y^2 | \mathbf{Y}) = \frac{m}{2} \ln(2\pi) + \frac{m}{2} \ln(\hat{\sigma}_Y^2) + \frac{1}{2} \ln |\mathbf{V}_Y(\theta)| + \frac{m}{2}, \quad (4.32)$$

siendo:

$$\hat{\sigma}_Y^2 = \frac{1}{m} \mathbf{Y}' \mathbf{V}_Y(\theta)^{-1} \mathbf{Y}.$$

Adicionalmente Harville (1974) demostró que las funciones de verosimilitud de distintos incrementos generalizados son iguales salvo una constante y que se pueden obtener expresiones simplificadas seleccionando la matriz $\mathbf{\Lambda}$ de forma que $\mathbf{\Lambda}'\mathbf{\Lambda} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ y $\mathbf{\Lambda \Lambda}' = \mathbf{I}_m$; resultando en este caso:

$$\begin{aligned} L(\theta | \mathbf{Y}) = L(\theta, \hat{\sigma}_Y^2 | \mathbf{Y}) &= \frac{m}{2} \ln(2\pi) + \frac{m}{2} \ln(\hat{\sigma}_Y^2) - \frac{1}{2} |\mathbf{X}'\mathbf{X}| \\ &+ \frac{1}{2} |\mathbf{X}'\mathbf{V}(\theta)^{-1}\mathbf{X}| + \frac{1}{2} \ln |\mathbf{V}(\theta)| + \frac{m}{2}, \end{aligned} \quad (4.33)$$

y:

$$\hat{\sigma}_Y^2 = \frac{1}{m}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (4.34)$$

En general se da por hecho que la estimación MVR mejora, a veces significativamente, los resultados obtenidos con la estimación MV (sobre todo si p es grande comparado con n). Realmente numerosos estudios de simulación (e.g. Zimmerman y Zimmerman, 1991; Fernández-Casal et al., 2003b; capítulo 9) han demostrado que el sesgo en las estimaciones de los parámetros del variograma es en general menor; aunque no se han estudiado todavía las propiedades asintóticas teóricas de estos estimadores.

DETALLES PRÁCTICOS Y COMPUTACIONALES

Además del problema ya comentado de multimodalidad de la función de verosimilitud, hay que tener en cuenta que para evaluar las expresiones anteriores son necesarios cálculos matriciales. Como las dimensiones de estas matrices pueden ser grandes en la práctica, es especialmente aconsejable en este caso realizar los cálculos utilizando la máxima precisión posible; en caso contrario pueden aparecer problemas numéricos debido a la acumulación de errores de redondeo (esto es especialmente importante en la estimación MVR si se utiliza (4.30) y (4.32)).

En el caso de MV el procedimiento recomendado a seguir sería el siguiente:

1. Reparametrizar el covariograma de forma que se verifique (4.24). Por ejemplo en el caso de los semivariogramas mostrados en la sección 4.2.2, si c_0 es el efecto nugget y c_1 el umbral parcial, en lugar de éstos parámetros se considerarían la varianza (umbral) $\sigma^2 = c_0 + c_1$ y la proporción de nugget en el umbral total $c_0^* = c_0 / (c_0 + c_1)$ (lo que equivale a suponer en las expresiones del covariograma que $c_1 = 1$ y $0 \leq c_0 \leq 1$).
2. Utilizar un algoritmo micro-genético para la minimización de (4.28) (detalles sobre este tipo de algoritmos se tienen p.e. en Goldberg, 1989); de esta forma se evitan, entre otras cosas, los problemas relacionados con mínimos locales.

Los pasos a seguir en el cálculo de (4.28), para unos valores de $\boldsymbol{\theta}$, serían los siguientes (entre paréntesis las rutinas correspondientes de la librería IMSL):

3. Calcular $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta})$ y su factorización de Cholesky $\mathbf{V} = \mathbf{L}\mathbf{L}'$ (DLFTDS).
4. A partir de la factorización calcular $|\mathbf{V}|$ y \mathbf{L}^{-1} (DLFDDS).
5. Calcular $\mathbf{Z}^* = \mathbf{L}^{-1}\mathbf{Z}$ y $\mathbf{X}^* = \mathbf{L}^{-1}\mathbf{X}$. Obtener $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$, dados por (4.26) y (4.27), al minimizar $(\mathbf{Z}^* - \mathbf{X}^*\boldsymbol{\beta})'(\mathbf{Z}^* - \mathbf{X}^*\boldsymbol{\beta})$ utilizando un algoritmo m.c.o. lineal (DRLSE).
6. Finalmente evaluar (4.28).

Los pasos 3 a 6 habrá que repetirlos para cada iteración del algoritmo micro-genético por lo que el tiempo de computación puede ser grande (comparado con otros métodos más sencillos como el de m.c.p.). Además este algoritmo normalmente no depende de los valores iniciales elegidos; aunque es importante, para la velocidad de convergencia y la obtención de los valores óptimos, determinar de forma adecuada los valores mínimos y máximos de los parámetros.

En el caso de la estimación MVR disponemos de dos opciones: una consistiría en obtener en primer lugar la matriz $\boldsymbol{\Lambda}$ a partir de (4.30), calcular los correspondientes incrementos generalizados (4.29) y posteriormente minimizar (4.32); la otra opción consistiría en utilizar las expresiones (4.33) y (4.34). Aunque numerosos autores utilizan la primera de las opciones (e.g. Pardo-Igúzquiza, 1997), es recomendable utilizar la segunda opción por requerir normalmente un menor tiempo de computación y para evitar posibles problemas de redondeo (relacionados con el cálculo inicial de (4.30)), sobre todo si el número de observaciones es grande (p.e. más de 200).

Comparando las expresiones (4.28) y (4.33) se observa que la principal diferencia es el cálculo del determinante de la matriz $\mathbf{X}'\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X} = (\mathbf{X}^*)'\mathbf{X}^*$, utilizando la notación anterior. Por tanto el algoritmo será análogo al de MV, calculando en el paso 6 este determinante antes de evaluar (4.33).

4.3.3 Algunos comentarios sobre los distintos métodos

Podemos afirmar que los métodos de estimación basados en mínimos cuadrados son los utilizados con mayor frecuencia en geoestadística. Por el contrario la estimación por máxima verosimilitud ha sido objeto de debate en

los últimos años, con numerosos comentarios en la literatura a favor (e.g. Pardo-Igúzquiza, 1998) y en contra (e.g. Ripley, 1988) de este tipo de métodos.

Una ventaja de los métodos de máxima verosimilitud es que permiten estimar de forma conjunta β y θ directamente de los datos (y no es necesario calcular estimaciones piloto del variograma). Los problemas numéricos relacionados con este tipo de estimación se pueden resolver en la práctica utilizando por ejemplo el algoritmo propuesto anteriormente; aunque el tiempo de computación aumenta notablemente cuando el número de datos es grande. Además la estimación por máxima verosimilitud tiene una conexión directa con la estimación tipo Bayes (e.g. Handcock y Wallis, 1994). Sin embargo, uno de los principales inconvenientes normalmente achacados a estos métodos es que la hipótesis de normalidad es difícil (o más bien imposible) de chequear en la práctica a partir de una única realización parcial del proceso. Otro problema que también se debe tener en cuenta al utilizar estos métodos es su falta de robustez cuando hay valores atípicos (outliers) en los datos.

No obstante es de esperar que las estimaciones obtenidas con los métodos de máxima verosimilitud sean más eficientes cuando la distribución de los datos se aproxima a la normalidad y el modelo paramétrico está especificado correctamente (especialmente con la estimación MVR); aunque no está claro si esta mejora es realmente significativa comparada con otros métodos más sencillos como el de m.c.p.. De hecho, bajo la hipótesis de normalidad, Zimmerman y Zimmerman (1991) observaron, al comparar mediante simulación las estimaciones obtenidas utilizando distintos métodos (entre ellos MV, MVR, m.c.o., m.c.p. y m.c.g.), que el método de m.c.p. era a veces el mejor procedimiento y nunca resultaba malo (considerando el sesgo, el error en media cuadrática y la cobertura del intervalo de predicción al 95%). Además, los métodos de mínimos cuadrados sólo utilizan la estructura asintótica de 2º orden del estimador piloto (no es necesario hacer suposiciones sobre la distribución completa de los datos), por lo que resultan ser más robustos que los de máxima verosimilitud cuando no se conoce por completo la distribución de \mathbf{Z} (e.g. Carroll y Ruppert, 1982) y son adecuados incluso cuando la distribución de los datos no es normal (aunque en ese caso pueden no ser los óptimos). Los comentarios anteriores, además de su fácil implementación, justifican que el método m.c.p. sea el preferible para muchos autores.

Por otra parte, si se asume un modelo paramétrico habrá que asegurarse en la práctica de que esa suposición es adecuada. Diblasi y Bowman (2001) propusieron un método basado en la estimación no paramétrica para contrastar si el variograma teórico es constante (i.e. si no hay dependencia espacial) aunque lamentablemente todavía no están disponibles contrastes de hipótesis que permitan verificar si un determinado modelo paramétrico semivariograma es apropiado. Para evitar posibles problemas relacionados con la mala especificación del modelo de variograma se puede pensar en su estimación de forma no paramétrica (sección 4.5). En este caso los métodos de mínimos cuadrados serán claramente preferibles a los basados en máxima verosimilitud.

Para comparar el ajuste obtenido con distintos modelos se pueden considerar los correspondientes valores finales de la función objetivo utilizada; por ejemplo los valores m.c.p. (o m.c.g.) correspondientes a su ajuste al estimador piloto o los valores del LNV si se utiliza alguno de los métodos de máxima verosimilitud (en este caso también se emplean en ocasiones criterios para la selección de modelos que tienen en cuenta el número de parámetros, como p.e. el AIC – Aikaike Information Criterion). Sin embargo en muchas ocasiones el objetivo final es la predicción, por lo que se suele utilizar la técnica de validación cruzada descrita a continuación.

4.4 VALIDACIÓN CRUZADA DEL MODELO AJUSTADO

El método de validación cruzada es la técnica normalmente utilizada en geoestadística para verificar si un modelo de variograma describe adecuadamente la dependencia espacial de los datos (ver p.e. Cressie, 1993, sección 2.6.4). Aunque también es utilizada en ocasiones para otros fines (ver p.e. Isaaks y Srivastava, 1989, capítulo 15), entre ellos: comparar distintas hipótesis sobre el modelo geoestadístico (tipo de modelo, vecindarios, etc.), detectar observaciones atípicas o incluso para la estimación de los parámetros del variograma (último apartado de esta sección). La idea básica es eliminar una parte de los datos y utilizar el resto de los datos para predecir los datos eliminados, entonces el error de predicción puede deducirse del valor que se predice menos el observado; repitiendo esto sobre varios conjuntos de datos se tiene una idea sobre la variabilidad del error de predicción. No sólo suele interesar estudiar

las predicciones, en general son también de interés las estimaciones del error cuadrático de predicción (varianza kriging).

Supongamos que $\hat{Z}_{-j}(\mathbf{s}_j)$ es un predictor de $Z(\mathbf{s}_j)$ obtenido, utilizando alguno de los métodos de predicción espacial, a partir de $\{Z(\mathbf{s}_i) : i \neq j\}$ y el variograma ajustado $2\gamma(\cdot, \hat{\boldsymbol{\theta}})$ (calculado utilizando todos los datos), y que $\sigma_{-j}^2(\mathbf{s}_j)$ es el correspondiente error en media cuadrática de predicción. Hay varias formas de medir la aproximación de las predicciones a los verdaderos valores, por ejemplo:

- La media de los errores tipificados

$$\text{MET} = \frac{1}{n} \sum_{j=1}^n (\hat{Z}_{-j}(\mathbf{s}_j) - Z(\mathbf{s}_j)) / \sigma_{-j}(\mathbf{s}_j)$$

debería ser próxima a cero. Este no es un criterio muy adecuado (sobre todo en el caso del KO) ya que los predictores kriging son insesgados independientemente del modelo de variograma utilizado (ver p.e. Yakowitz y Szidarovski, 1985).

- El error cuadrático medio adimensional:

$$\text{ECMA} = \sqrt{\frac{1}{n} \sum_{j=1}^n \left((\hat{Z}_{-j}(\mathbf{s}_j) - Z(\mathbf{s}_j)) / \sigma_{-j}(\mathbf{s}_j) \right)^2}$$

debería ser próximo a uno. El valor de este estadístico puede interpretarse como una medida de la concordancia entre las varianzas kriging y las varianzas observadas. Teniendo en cuenta que si reescalamos el variograma multiplicándolo por una constante, las predicciones con el variograma reescalado son idénticas y las varianzas kriging serán las mismas multiplicadas por esa constante. Podemos pensar en “corregir” un modelo de variograma¹² de forma que ECMA sea igual a 1, multiplicándolo por ECMA^2 .

- El error cuadrático medio:

¹² En la predicción espacial se supone que el variograma teórico es conocido. En la práctica el variograma es desconocido y hay que estimarlo, por lo cual la varianza kriging debería ser aumentada en un término correspondiente a esa estimación (ver Cressie, 1993, sección 5.3, y comentarios anteriores en la sección 3.4.2); esto además puede agravarse en el caso del KU. Se podría pensar en utilizar este criterio para corregir las varianzas de predicción.

$$\text{ECM} = \frac{1}{n} \sum_{j=1}^n (\hat{Z}_{-j}(\mathbf{s}_j) - Z(\mathbf{s}_j))^2 \quad (4.35)$$

debería ser pequeño. El principal problema de este estadístico es que asigna igual peso a todos los datos y no tiene en cuenta las posiciones espaciales. Por lo general los errores son mayores en los puntos más alejados del resto de los datos (observaciones exteriores) y pueden tener un efecto dominante en la media global. Se podría pensar en calcular una media ponderada con pesos inversamente proporcionales a la varianza kriging o a alguna medida de la distancia de una posición al resto de los datos.

- Diversos criterios gráficos pueden ser también de interés como herramientas de diagnóstico, como p.e. el diagrama tallo-hoja de los residuos tipificados o gráficos de normalidad.

Después de la validación cruzada del variograma, si esta resultó ser satisfactoria, se puede confiar en que la predicción basada en el modelo ajustado $2\gamma(\cdot, \hat{\theta})$ es aproximadamente óptima y que las estimaciones del error en media cuadrática de predicción son bastante buenas (i.e. el modelo ajustado no es muy incorrecto).

Uno de los principales problemas de esta técnica es el elevado coste computacional (Cressie, 1993, p. 104). Con el método descrito en el siguiente apartado, se puede obtener la validación cruzada de un modelo de variograma de forma rápida y sencilla.

DETALLES PRÁCTICOS Y COMPUTACIONALES

En el caso de validación cruzada la matriz kriging no es la misma para la predicción en las distintas posiciones, por lo que si se realiza el cálculo de la forma habitual (resolviendo n sistemas de orden $n + p$) el proceso puede ser notablemente lento, sobre todo si el número de datos es grande. Teniendo en cuenta las expresiones de las ecuaciones kriging eliminando una observación y además que los métodos kriging son interpoladores exactos, se pueden deducir fácilmente las expresiones que permiten obtener las predicciones de validación cruzada a partir de la matriz kriging construida con el total de los datos.

Supongamos de momento por simplicidad que la media es constante (kriging ordinario). Si utilizamos el KO para predecir $Z(\mathbf{s}_0)$ a partir de $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ hay que resolver el sistema (3.11), que en este caso resulta ser:

$$\begin{pmatrix} \gamma(\mathbf{s}_1, \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_1, \mathbf{s}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_n, \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_n, \mathbf{s}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ m \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{s}_0, \mathbf{s}_1) \\ \vdots \\ \gamma(\mathbf{s}_0, \mathbf{s}_n) \\ 1 \end{pmatrix}, \quad (4.36)$$

$$\mathbf{\Gamma}_O \boldsymbol{\lambda}_O = \boldsymbol{\gamma}_O.$$

Supongamos que se elimina $Z(\mathbf{s}_1)$ y que:

$$\mathbf{\Gamma}_{-1} \boldsymbol{\lambda}_{-1} = \boldsymbol{\gamma}_{-1},$$

es el correspondiente sistema del KO para la predicción en \mathbf{s}_0 a partir de $\{Z(\mathbf{s}_i) : i \neq 1\}$. Entonces el sistema con todos los datos (4.36) puede escribirse de la forma:

$$\left(\begin{array}{c|c} \gamma(\mathbf{s}_1, \mathbf{s}_1) & \mathbf{g}' \\ \hline \mathbf{g} & \mathbf{\Gamma}_{-1} \end{array} \right) \begin{pmatrix} \lambda_1 \\ \lambda_{-1}^* \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{s}_0, \mathbf{s}_1) \\ \gamma_{-1} \end{pmatrix}.$$

A partir de esta expresión puede deducirse fácilmente que, si:

$$\mathbf{\Gamma}_O^{-1} = \begin{pmatrix} a & \mathbf{y}' \\ \mathbf{y} & \mathbf{A} \end{pmatrix},$$

entonces $(\mathbf{\Gamma}_{-1})^{-1} = \mathbf{A} - a^{-1} \mathbf{y} \mathbf{y}'$; por lo que los pesos y varianza kriging eliminando¹³ $Z(\mathbf{s}_1)$ resultan ser:

$$\begin{aligned} \boldsymbol{\lambda}_{-1} &= \boldsymbol{\lambda}_{-1}^* - \frac{\lambda_1}{a} \mathbf{y}, \\ \sigma_{-1}^2 &= \sigma_{KO}^2 - \frac{\lambda_1^2}{a}, \end{aligned}$$

¹³ Expresiones análogas pueden obtenerse para el caso de eliminar más de una observación y también las correspondientes a añadir nuevas observaciones. Estas últimas pueden ser de especial interés cuando se pretenden añadir nuevas posiciones a una red de observación.

donde σ_{KO}^2 es la varianza kriging correspondiente a la predicción con todos los datos.

Para el caso de que $\mathbf{s}_0 = \mathbf{s}_1$, tendríamos que:

$$\boldsymbol{\lambda}_O = (1, 0, \dots, 0)' \Rightarrow \begin{cases} \lambda_{-1} = -\frac{1}{a} \mathbf{y} \\ \sigma_{-1}^2 = -\frac{1}{a} \end{cases} \quad (4.37)$$

Además, para la obtención de a e \mathbf{y} no es necesario invertir la matriz $\boldsymbol{\Gamma}_O$, basta con resolver el sistema:

$$\boldsymbol{\Gamma}_O \begin{pmatrix} a \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}.$$

Estos resultados son válidos también para el caso del KU; además, reorganizando filas y columnas de la matriz del sistema obtendríamos los resultados para las demás posiciones. Resumiendo, el procedimiento a seguir sería similar al descrito en la sección 3.3:

- Obtener la factorización \mathbf{UDU}' de la matriz simétrica $\boldsymbol{\Gamma}_O$, donde \mathbf{U} es una matriz triangular superior y \mathbf{D} una matriz diagonal (rutina LFTSF de la librería IMSL).
- Para cada posición \mathbf{s}_i resolver el sistema:

$$\boldsymbol{\Gamma}_O \mathbf{y}^{(i)} = \mathbf{e}^{(i)},$$

siendo $\mathbf{e}_j^{(i)} = \delta_{ij}$ (delta de Kronecker), utilizando la factorización anterior (rutinas LFSSF o LFISF de la librería IMSL). Posteriormente calcular la predicción de validación cruzada $\hat{Z}_{-i}(\mathbf{s}_i)$ y la correspondiente varianza kriging $\sigma_{-i}^2(\mathbf{s}_i)$ según (4.37), tomando $a = y_i^{(i)}$ e \mathbf{y} el vector resultante de eliminar la fila i -ésima de $\mathbf{y}^{(i)}$.

Por tanto, cuando se calculan predicciones, se pueden obtener también las medidas de validación cruzada sin apenas coste computacional adicional.

ESTIMACIÓN DE LOS PARÁMETROS DEL VARIOGRAMA MEDIANTE VALIDACIÓN CRUZADA

Teniendo en cuenta las observaciones del apartado anterior se puede pensar también en utilizar la técnica de validación cruzada para la estimación

de los parámetros del modelo de variograma. Podríamos considerar por ejemplo el siguiente método:

1. Reparametrizar el modelo de semivariograma de forma que:

$$\gamma(\cdot, \boldsymbol{\theta}) = \alpha \gamma^*(\cdot, \boldsymbol{\theta}^*)$$

donde $\alpha > 0$ es el parámetro de escala y $\boldsymbol{\theta}^*$ es el vector de parámetros que determinan la forma de la dependencia espacial (hay que tener en cuenta que el parámetro de escala α no influye en el cálculo de las predicciones; ver sección 3.4.2)

2. Fijar el parámetro de escala a un valor inicial, por ejemplo $\alpha_0 = 1$.
3. Estimar los parámetros $\boldsymbol{\theta}^*$ de modo que minimicen el error cuadrático medio de validación cruzada (4.35) (Bastin y Gevers, 1985) o alguna versión ponderada (Fernández-Casal, 2000).
4. Reajustar el parámetro de escala de forma que las varianzas de predicción sean coherentes con lo observado en los datos (i.e. $ECMA = 1$), tomando:

$$\hat{\alpha} = \alpha_0 \frac{1}{n} \sum_{j=1}^n \left(\left(\hat{Z}_{-j}(\mathbf{s}_j) - Z(\mathbf{s}_j) \right) / \sigma_{-j}(\mathbf{s}_j) \right)^2.$$

En el campo de las series de tiempo se han considerado modificaciones del método tradicional de validación cruzada para la obtención de predicciones más eficientes bajo la presencia de dependencia temporal, eliminando en el cálculo de los errores de validación cruzada (además de la observación de predicción) varias observaciones cercanas (e.g. Härdle y Vieu, 1992). Estas modificaciones pueden ser implementadas también para el caso de dependencia espacial; si $\hat{Z}_{k,j}(\mathbf{s}_j)$ es el predictor de $Z(\mathbf{s}_j)$ y $\sigma_{k,j}^2(\mathbf{s}_j)$ la varianza kriging considerando el total de datos menos las k observaciones más “cercanas” a la posición espacial \mathbf{s}_j (seleccionadas utilizando por ejemplo el semivariograma como distancia), bastaría con utilizar estos valores en los pasos 3 y 4 del método anterior. Sin embargo, en este caso las observaciones exteriores incrementarán aún más su influencia en el error cuadrático medio, por lo se debería utilizar una versión ponderada; por ejemplo:

$$ECMP = \frac{1}{\sum_{j=1}^n w_{k,j}} \sum_{j=1}^n w_{k,j} \left(\hat{Z}_{k,j}(\mathbf{s}_j) - Z(\mathbf{s}_j) \right)^2, \quad (4.38)$$

siendo $w_{k,j} = 1/\sigma_{k,j}^2(\mathbf{s}_j)$. Sin embargo, como ya se comentó anteriormente, para una predicción (asintóticamente) eficiente es importante capturar la conducta del variograma cerca del origen, por lo que no se deberían eliminar observaciones próximas. Se puede pensar en una modificación adicional del método anterior que tenga también en cuenta esta información; por ejemplo, minimizando en el paso 3 una versión acumulada de (4.38):

$$\text{ECMA} = \frac{1}{\sum_{i=1}^k \sum_{j=1}^n w_{i,j}} \sum_{i=1}^k \sum_{j=1}^n w_{i,j} \left(\hat{Z}_{i,j}(\mathbf{s}_j) - Z(\mathbf{s}_j) \right)^2, \quad (4.39)$$

y en el paso 4 tomando:

$$\hat{\alpha} = \alpha_0 \frac{1}{nk} \sum_{i=1}^k \sum_{j=1}^n \left(\left(\hat{Z}_{i,j}(\mathbf{s}_j) - Z(\mathbf{s}_j) \right) / \sigma_{i,j}(\mathbf{s}_j) \right)^2.$$

Fernández-Casal (2000) comparó los métodos de validación cruzada correspondientes a (4.38) y (4.39) con el método de m.c.p. mediante simulación. En general, observó que los métodos de validación cruzada (al contrario de lo que podría pensarse en un principio) obtienen mejores resultados en el caso de mayor dependencia espacial y se ven muy poco afectados por la presencia de una tendencia (kriging universal). Además, la eliminación de más de una observación no mejora significativamente la eficiencia de las predicciones pero puede mejorar notablemente las estimaciones de las varianzas de predicción si se utiliza la versión acumulada. En cuanto a las estimaciones de los parámetros observó que, en general, producen estimaciones más eficientes del efecto nugget y del umbral parcial (sobre todo en el caso del kriging universal y al utilizar la versión acumulada), mientras que las estimaciones del rango son peores (debido quizás a que este parámetro influye menos en las predicciones obtenidas). Por tanto los métodos de validación cruzada pueden ser una alternativa al método de mínimos cuadrados, especialmente en el caso del kriging universal o de tamaños muestrales pequeños.

Otra aproximación distinta, propuesta por Samper y Neuman (1989), consiste en maximizar la verosimilitud de los errores de validación cruzada (suponiendo normalidad). Aunque este método ha recibido algunas críticas (e.g. Cressie, 1993, p. 103), debido a que los datos considerados en la estimación MV dependen también de los parámetros a estimar.

4.5 APROXIMACIONES FLEXIBLES PARA EL MODELADO DE LA DEPENDENCIA ESPACIAL

El principal problema al considerar modelos paramétricos es la posible mala especificación del modelo de variograma. En ocasiones puede ser difícil encontrar una familia parametrizada que se acomode a la forma del estimador piloto, por lo que el ajuste finalmente obtenido resulta poco satisfactorio. A continuación se hace una pequeña revisión de las distintas propuestas para solucionar este problema, basadas en la idea (no paramétrica) de disponer de mayor flexibilidad en el modelado espacial.

4.5.1 Discretización de la distribución espectral (Shapiro-Botha)

Shapiro y Botha (1991) propusieron una aproximación flexible y fácilmente implementable en la práctica para el ajuste de un modelo válido de variograma. Esta aproximación, que podríamos clasificar como no paramétrica, ha sido posiblemente la que mayor tratamiento ha recibido en la literatura de las comentadas en esta sección. A partir de la representación espectral (2.17) de un covariograma isotrópico obtuvieron una amplia familia de modelos de variogramas; por tanto solamente consideraron variogramas isotrópicos acotados¹⁴. Teniendo en cuenta la relación (2.20), la expresión del semivariograma (no nec. continuo) de un proceso estacionario de segundo orden isotrópico es de la forma:

$$\gamma(r) = \begin{cases} \nu_0 - \nu(r) & \text{si } r > 0 \\ 0 & \text{si } r = 0 \end{cases} \quad (4.40)$$

¹⁴ Es importante destacar que la hipótesis de estacionariedad de segundo orden realmente no es muy restrictiva en la práctica. Siempre se puede suponer que hay independencia a partir de una distancia a que puede ser mayor que el rango considerado. Por ejemplo, si consideramos el semivariograma lineal (4.12) y el semivariograma lineal con umbral (4.13), las predicciones kriging utilizando ambos modelos (con la misma pendiente y efecto nugget) coincidirán si no es necesario evaluar el semivariograma a distancias mayores que a .

siendo ν_0 una constante positiva y $\nu(r)$ una función semidefinida positiva continua en el origen (i.e. un covariograma), que admite la representación:

$$\nu(r) = \int_0^\infty \kappa_d(\lambda r) dG(\lambda) \quad (4.41)$$

donde $\kappa_d(\cdot)$ está definida por (2.16) y G es una función positiva, acotada y no decreciente en $[0, \infty)$. Recíprocamente, cualquier función verificando (4.40) y además $\nu_0 - \nu(r) \geq 0, \forall r \geq 0$, es un variograma isotrópico válido en \mathbb{R}^d . Adicionalmente, como una función semidefinida positiva está acotada en valor absoluto por su valor en el origen (ec. (2.6)), esta última condición es equivalente a:

$$\nu_0 - \nu(0) = \nu_0 - \int_0^\infty dG(\lambda) \geq 0. \quad (4.42)$$

El caso de que $c_0 = \nu_0 - \nu(0) > 0$ indicará la presencia de un efecto nugget.

En este caso, teniendo en cuenta los resultados anteriores, el problema del ajuste de un modelo de variograma se reduce a encontrar una constante positiva ν_0 y una función G acotada no decreciente de forma que el correspondiente variograma (4.40) describa adecuadamente la dependencia espacial de los datos. Sin embargo este problema puede ser demasiado complicado para resolverlo numéricamente, por lo que Shapiro y Botha (1991) consideraron una simplificación adicional que consiste en suponer que dG es una medida atómica, con un número finito de saltos positivos z_j en puntos $x_j, j = 1, \dots, J$; es decir, suponer que G es una función de la forma:

$$G(x) = \sum_{x_j \leq x} z_j.$$

Adicionalmente, Shapiro y Botha (1991) supusieron (por comodidad) posiciones equiespaciadas:

$$x_j = j\phi, j = 1, \dots, J, \quad (4.43)$$

siendo ϕ un número positivo fijado de antemano. Los modelos de variogramas obtenidos son de la forma (4.40), donde la expresión (4.41) se reduce a:

$$\nu(r) = \sum_{j=1}^J \kappa_d(x_j r) z_j,$$

y la restricción (4.42) se convierte en la restricción lineal:

$$\nu_0 - \sum_{j=1}^J z_j \geq 0.$$

Aunque Shapiro y Botha (1991) no consideraron los modelos correspondientes al caso límite $d = \infty$ (considerando $\kappa_\infty(\cdot)$ definida por (2.18)), éstos pueden ser suficientemente flexibles como para modelar muchos de los casos que se puedan presentar en la práctica.

Utilizando el criterio m.c.p. descrito en la sección 4.3.1, si $\hat{\gamma}_i = \hat{\gamma}(r_i)$, $i = 1, \dots, K$, son las estimaciones piloto del semivariograma, el problema del ajuste se reduce a encontrar el vector $\boldsymbol{\theta} = (z_1, \dots, z_J, \nu_0)'$ de dimensión $J + 1$ que minimiza la función:

$$Q(\boldsymbol{\theta}) = \sum_{i=1}^K w_i \left(\hat{\gamma}_i - \nu_0 + \sum_{j=1}^J \kappa_d(x_j r_i) z_j \right)^2, \quad (4.44)$$

sujeto a las restricciones lineales:

$$z_j \geq 0, \quad j = 1, \dots, J, \quad (4.45)$$

$$\nu_0 - \sum_{j=1}^J z_j \geq 0. \quad (4.46)$$

La función objetivo (4.44) es una función cuadrática que puede escribirse como:

$$Q(\boldsymbol{\theta}) = (\hat{\boldsymbol{\gamma}} - \mathbf{A}\boldsymbol{\theta})' \mathbf{W} (\hat{\boldsymbol{\gamma}} - \mathbf{A}\boldsymbol{\theta}),$$

siendo $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, \dots, \hat{\gamma}_K)'$, $\mathbf{W} = \text{diag}(w_1, \dots, w_K)$ la matriz de pesos $K \times K$ y \mathbf{A} la matriz de coeficientes $K \times (J + 1)$ definida por $a_{ij} = -\kappa_d(x_j r_i)$ para $i = 1, \dots, K$, $j = 1, \dots, J$ y $a_{i, J+1} = 1$ para $i = 1, \dots, K$. Como además las restricciones (4.45) y (4.46) son lineales, se trata de un problema de programación cuadrática que puede resolverse fácilmente aplicando alguna de las correspondientes técnicas de optimización lineal multidimensional (p.e. utilizando la rutina QPROG de la librería IMSL). Hay que tener en cuenta que para que este problema tenga solución única debe verificarse necesariamente que $K \geq J + 1$, por este motivo Shapiro y Botha (1991) propusieron tomar $J = K - 1$.

Adicionalmente, si en la elección de los pesos se sigue el criterio propuesto por Cressie (1985) de tomar $w_i = |N(r_i)|/\gamma(r_i)^2$, siendo $|N(r_i)|$ el número de aportaciones a la estimación del semivariograma en el salto i -ésimo,

habrá que proceder de forma iterativa, tomando $w_i = 1$ (m.c.o.) en el primer paso y recalculando los pesos en cada iteración hasta convergencia. Aunque Shapiro y Botha (1991) solo consideraron el criterio m.c.p., el algoritmo es también válido para el caso de m.c.g. (únicamente cambia el cálculo de la matriz de pesos \mathbf{W}).

Después de resolver el problema, si $\bar{\boldsymbol{\theta}} = (\bar{z}_1, \dots, \bar{z}_J, \bar{\nu}_0)'$ es la solución óptima obtenida, el modelo de semivariograma ajustado¹⁵ vendrá dado por:

$$\bar{\gamma}(r) = \bar{\nu}_0 - \sum_{j=1}^J \kappa_d(x_j r) \bar{z}_j, \quad (4.47)$$

para $r > 0$.

Si utilizamos el estimador clásico del semivariograma (u otro similar), las estimaciones $\hat{\gamma}_i$ pueden tener una gran variabilidad y como la familia de semivariogramas obtenida de esta forma es muy flexible, el semivariograma ajustado $\bar{\gamma}(r)$ puede tener una forma muy irregular (sobre todo si la dimensión espacial d es pequeña). Para evitar un “sobreajuste” de este tipo Shapiro y Botha (1991, pp. 91-94) consideraron la inclusión de restricciones (lineales) adicionales de suavidad, monotonía o convexidad¹⁶. Otra aproximación distinta, sugerida inicialmente por Febrero-Bande et al. (1998), es la estimación no paramétrica del semivariograma. Además de obtener estimaciones más eficientes, al ser estas más suaves, no es necesario añadir restricciones adicionales en el ajuste. García-Soidán et al. (2003a) y (2003b) estudiaron también las propiedades asintóticas del semivariograma ajustado (4.47) cuando se utilizan como estimador piloto alguno de los estimadores locales tipo núcleo mostrados en la sección 4.1.1.

¹⁵ Al contrario de lo que afirman algunos autores (e.g. Hall et al., 1994, p. 2115; Hall y Patil, 1994, p. 401) los modelos obtenidos con esta aproximación son semivariogramas válidos en el continuo.

¹⁶ La flexibilidad de estos modelos disminuye al aumentar la dimensión, por lo que otra alternativa a la inclusión de restricciones adicionales podría ser considerar de modelos correspondientes a dimensiones mayores (también válidos; ver propiedades elementales del variograma en la sección 2.2.1).

Aunque Shapiro y Botha (1991) no consideraron la estimación de las características del semivariograma teórico, a partir del modelo ajustado (4.47) se obtiene la estimación del efecto nugget:

$$\hat{c}_0 = \bar{v}_0 - \sum_{j=1}^J \bar{z}_j,$$

y también se puede pensar en utilizar \bar{v}_0 como estimación del umbral (varianza). Sin embargo, las estimaciones \bar{v}_0 son usualmente sesgadas (tienden a sobreestimar el umbral verdadero) y de gran variabilidad. Esto fue observado también por Cherry (1997) en un estudio de simulación, proponiendo la inclusión de un parámetro de penalización sobre los pesos en el ajuste para solventar este problema (aunque sólo consideró el criterio de m.c.o., i.e. $\mathbf{W} = \mathbf{I}_K$, con variogramas continuos). Por el contrario, en el caso del rango no se obtienen estimaciones de este parámetro de forma natural (aunque siempre se puede considerar algún tipo de criterio gráfico).

Un tema aún abierto de esta aproximación es la selección adecuada de los puntos de discretización $\{x_j : j = 1, \dots, J\}$. Si se sigue el criterio inicial propuesto por Shapiro y Botha (1991) de tomar estos puntos equiespaciados, es necesario fijar la distancia ϕ entre nodos en (4.43) (aunque solamente grandes cambios en el valor de este parámetro tienen realmente efecto en los ajustes obtenidos). Según la experiencia personal, se suelen obtener buenos resultados buscando en el intervalo $(0, 2/r_{\max}]$ el valor que produce un mejor ajuste, donde r_{\max} es el salto máximo considerado. La selección de nodos equidistantes ha sido sin embargo objeto de ciertas críticas (Cherry et al., 1996; Genton y Gorschich, 2003), debido a que pueden aparecer oscilaciones anómalas del variograma ajustado entre las estimaciones piloto (aunque estos autores no especificaron el criterio para la selección de los nodos); sin embargo en los ajustes realizados utilizando el criterio anterior no se han observado estas oscilaciones. Además, si se utiliza un estimador piloto no paramétrico del tipo (4.9), siempre se puede evitar este problema empleando en el ajuste un conjunto de estimaciones piloto suficientemente denso. Otra alternativa a tener en cuenta ha sido propuesta recientemente por Genton y Gorschich (2003), quienes sugirieron tomar¹⁷:

¹⁷ Esta idea esta basada en la siguiente propiedad bastante conocida de las funciones de

$$x_j = \frac{t_j}{r_{\max}}, j = 1, \dots, J,$$

siendo $\{t_j : j = 1, \dots, J\}$ los J primeros ceros de la función de Bessel $J_v(\cdot)$ de orden $v = (d - 2)/2$. Este criterio puede ser recomendable cuando el número de nodos es grande, ya que los valores t_j pueden calcularse rápidamente si se utiliza p.e. el algoritmo propuesto por Ball (2000). En cualquier caso hay que establecer también el número m de nodos. Si se sigue el criterio inicial de tomar $J = K - 1$, normalmente muchos de los valores \bar{z}_j obtenidos en el ajuste son cero o próximos a cero; lo que sugiere que en la mayoría de los casos un número bastante menor de parámetros puede ser suficiente (esto también ha sido observado por Genton y Gorsich, 2003), aunque todavía no están disponibles estudios que permitan determinar un valor adecuado de forma que se pueda optimizar el tiempo de computación.

4.5.2 Otros métodos basados en la representación espectral

Hall et al. (1994) propusieron un método no paramétrico para la estimación del covariograma en el caso unidimensional basado en la estimación tipo núcleo y en la representación espectral. El método sugerido consiste en primer lugar en estimar inicialmente el covariograma utilizando el estimador piloto tipo Nadaraya-Watson $\hat{C}(s)$ mostrado en la sección 4.1.1 (ec. (4.8)) y después obtener una estimación inicial $\hat{f}(\lambda)$ de la densidad espectral aplicando la transformación de Fourier sobre $\hat{C}(s)$ (asegurando antes la integrabilidad truncando (4.8) en caso de ser necesario). Posteriormente modificar $\hat{f}(\lambda)$ de forma que sea positiva (eliminando valores negativos e incluso realizando un

Bessel: las funciones $J_v(ax)$ y $J_v(bx)$ son ortogonales en el intervalo $x \in [0, 1]$ si a y b son distintos ceros de $J_v(\cdot)$. Esto produce que la matriz de coeficientes del algoritmo de programación cuadrática se aproxime a la ortogonalidad cuando se considera el covariograma en lugar del variograma en el ajuste. Sin embargo, teniendo en cuenta las observaciones realizadas en la sección 2.4.1 sobre la estimación piloto del covariograma, es preferible utilizar las ecuaciones basadas en el variograma. Otras ventajas sugeridas por Genton y Gorsich (2003) no son realmente de utilidad (ni recomendables) en la práctica.

suavizado adicional) y obteniendo la estimación final de la densidad espectral $\tilde{f}(\lambda)$. La estimación final del covariograma $\tilde{C}(s)$ se obtiene transformando de nuevo $\tilde{f}(\lambda)$ al dominio original. Una metodología similar ha sido propuesta por Hall y Patil (1994) para el caso general multidimensional y también se podría generalizar fácilmente esta aproximación para el caso multidimensional isotrópico. Esta forma de proceder sin embargo es poco recomendable en la práctica debido principalmente al coste computacional (cualquiera de las otras alternativas mostradas aquí son mucho más fáciles de implementar en la práctica).

Yao y Journel (1998) propusieron para el caso anisotrópico bidimensional (y considerando también el caso multivariante) una idea similar a la de Hall et al. (1994) pero operando de forma discreta, utilizando la transformación rápida de Fourier (FFT) en lugar de considerar transformaciones continuas. En el caso univariante (en la sección 5.3.3 se describe el caso multivariante), la aproximación propuesta consiste en primer lugar en la obtención de una tabla de estimaciones piloto del covariograma (suavizando y completando estimaciones obtenidas con el estimador clásico), posteriormente obtener una tabla de pseudo-densidad espectral (utilizando FFT y suavizando los valores obtenidos asegurándose de que el resultado es positivo) y en el último paso volver a transformar al dominio espacial. El resultado final es una tabla de covarianzas para un conjunto discreto de saltos (y no un covariograma válido para saltos continuos como sería deseable), por lo que en el caso general sugirieron tomar el valor más próximo de esta tabla (otra opción sería por ejemplo interpolar los valores cercanos).

Respecto a estas aproximaciones que utilizan directamente la representación espectral hay que señalar dos posibles fuentes de problemas. En primer lugar implican asumir la hipótesis de que el covariograma (variograma) es continuo, lo que puede producir una subestimación cerca del origen cuando el efecto nugget no es nulo (p.e. en el caso extremo de independencia). Si bien este problema (no considerado por los autores anteriores) se podría solucionar estimando y eliminando previamente este parámetro, aunque este primer paso tendría gran influencia sobre la estimación final cerca del origen. En segundo lugar, como al realizar las transformaciones se utiliza el rango completo de valores en el dominio original, es de esperar que estos métodos sean poco

robustos; por ejemplo, valores atípicos en la densidad espectral afectarán a la estimación del covariograma en todos los saltos.

4.5.3 Representación en medias móviles

Barry y Ver Hoef (1996) propusieron el uso de familias flexibles de semivariogramas anisotrópicos acotados basadas en la representación en medias móviles (4.13) comentada en la sección 4.2.1. En el caso unidimensional, fijado un número entero m y un rango $c > 0$, considerando la función:

$$f(x) = \sum_{i=1}^m a_i \mathcal{I}_{\left(\frac{(i-1)c}{m}, \frac{ic}{m}\right]}(x),$$

se obtiene la correspondiente familia a partir de (4.13) (denotando por $\mathcal{I}_A(\cdot)$ función indicadora del conjunto A). En el caso de que $|h| = jc/m$ se consigue fácilmente la expresión del semivariograma:

$$\gamma\left(\frac{jc}{m}\right) = c_0 + \frac{c}{m} \left(\sum_{i=1}^m a_i^2 - \sum_{i=j+1}^m a_i a_{i-j} \right).$$

El caso general de $0 < h < c$ se puede obtener mediante interpolación lineal:

$$\gamma(h) = (1 - V)\gamma\left(\frac{jc}{m}\right) + V\gamma\left(\frac{(j+1)c}{m}\right),$$

donde $j = [hm/c]$, denotando por $[\cdot]$ la parte entera (i.e. el entero menor o igual más próximo), y $V = (h - (jc/m))/(c/m)$. Para $|h| > c$ resulta:

$$\gamma(h) = c_0 + \frac{c}{m} \sum_{i=1}^m a_i^2.$$

El caso bidimensional es análogo; a partir de:

$$f(x, y) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \mathcal{I}_{\left(\frac{(i-1)c}{m}, \frac{ic}{m}\right)} \times \mathcal{I}_{\left(\frac{(j-1)d}{n}, \frac{jd}{n}\right)}(x, y),$$

se obtiene el correspondiente modelo de semivariograma anisotrópico. Si el salto coincide con un punto de discretización resulta:

$$\gamma\left(\frac{kc}{m}, \frac{ld}{n}\right) = c_0 + \frac{cd}{mn} \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 - \sum_{i=k+1}^m \sum_{j=l+1}^n a_{ij} a_{i-k, j-l} \right),$$

y el caso general se puede obtener por interpolación plana.

El principal inconveniente de este tipo de modelos es el elevado número de parámetros que contribuyen de forma no lineal al modelo. Este problema se agrava cuando aumenta el número de dimensiones, por lo que sólo es recomendable su uso cuando la dimensión espacial es menor o igual que dos (únicos casos considerados por Barry y Ver Hoef, 1996). Incluso para dimensiones pequeñas, el ajuste de estos modelos a estimaciones piloto del variograma resulta ser bastante complicado. Principalmente debido a que es necesaria la resolución de un problema de minimización no lineal multidimensional con un número de parámetros normalmente elevado. Además, Barry y Ver Hoef (1996) descubrieron que al estimar simultáneamente el efecto nugget y los demás parámetros del variograma se obtienen sobreestimaciones del efecto nugget, por lo que recomendaron realizar una estimación inicial del efecto nugget antes de estimar el resto de parámetros del modelo (ajustando un semivariograma lineal por m.c.p. a estimaciones piloto próximas del origen).

Los problemas computacionales de esta aproximación fueron también manifestados por Ver Hoef et al. (2003) (donde extendieron esta aproximación para el caso multivariante), proponiendo el uso de la transformación rápida de Fourier (FFT) para su evaluación. Utilizando la FFT se reduce significativamente el número de operaciones para el cálculo de los valores del variograma cuando el número de parámetros es grande, sin embargo persisten los inconvenientes de la minimización no lineal multidimensional.

4.5.4 Extensión de la aproximación de Shapiro-Botha para el caso de anisotropía

La principal ventaja de los métodos propuestos por Barry y Ver Hoef (1996) y Yao y Journel (1998) respecto a la aproximación de Shapiro y Botha (1991) es la obtención de modelos anisotrópicos (algo de especial interés en el caso espacio-temporal), aunque a costa de una mayor complejidad en su implementación en la práctica. Sin embargo puede extenderse fácilmente la metodología de Shapiro y Botha (1991) para este caso.

Según el teorema de Bochner, la expresión general de un covariograma bidimensional continuo en el origen es:

$$C(h_1, h_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos(\omega_1 h_1 + \omega_2 h_2) dF(\omega_1, \omega_2),$$

siendo dF una medida simétrica positiva y acotada. Aún podemos simplificar más esta expresión¹⁸, obteniendo que es equivalente a:

$$C(h_1, h_2) = \int_{-\infty}^{\infty} \int_0^{\infty} \cos(\omega_1 h_1) \cos(\omega_2 h_2) dF(\omega_1, \omega_2), \quad (4.48)$$

con dF positiva y acotada (considerada sobre $\mathbb{R} \times \mathbb{R}^{+,0}$).

Teniendo en cuenta la relación (2.20), $\gamma(h_1, h_2)$ es un semivariograma anisotrópico bidimensional acotado si y solo si es de la forma:

$$\gamma(h_1, h_2) = \nu_0 - \nu(h_1, h_2),$$

para $(h_1, h_2) \neq (0, 0)$, siendo ν_0 una constante positiva, $\nu(h_1, h_2)$ una función semidefinida positiva continua en el origen que admite la representación (4.48) y tales que:

$$c_0 = \nu_0 - \nu(0, 0) \geq 0.$$

Podemos considerar una simplificación análoga a la propuesta de Shapiro y Botha (1991) para facilitar el ajuste de este tipo de modelos a un conjunto de estimaciones piloto, suponiendo que dF en (4.48) es una medida atómica con un número finito de saltos positivos z_{ij} en puntos (x_i, y_j) ; es decir:

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} z_{ij}.$$

Los puntos de discretización (x_i, y_j) pueden tomarse por ejemplo equiespaciados:

$$x_i = i\phi_1, i = -I, \dots, I,$$

$$y_j = j\phi_2, j = 1, \dots, J,$$

siendo ϕ_1 y ϕ_2 dos números positivos (la idea de Genton y Gorschich, 2003, para el caso isotrópico se trasladaría aquí tomando $\phi_k = \pi / (2hmax_k)$ siendo $hmax_k$ el correspondiente máximo salto).

Los modelos obtenidos serán de la forma¹⁹:

¹⁸ Simplemente teniendo en cuenta que $\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ y utilizando de nuevo la simetría de estas funciones.

¹⁹ Esos modelos son equivalentes a algunos de los descritos por Rehman y Shapiro (1996) (donde consideraban en el caso multivariante), aunque realmente estos autores

$$\gamma(h_1, h_2) = \nu_0 - \sum_{i=-I}^I \sum_{j=1}^J \cos(x_i h_1) \cos(y_j h_2) z_{ij},$$

para $(h_1, h_2) \neq (0, 0)$, con:

$$z_{ij} \geq 0,$$

$$\nu_0 - \sum_{i=-I}^I \sum_{j=1}^J z_{ij} \geq 0.$$

El ajuste de este tipo de modelos a un conjunto de estimaciones piloto se realiza de forma análoga al caso isotrópico mediante programación cuadrática; en la sección 6.3 se tienen los detalles correspondientes a modelos similares y por tanto no va a ser tratado este tema aquí. Claramente esta aproximación es mucho más fácil de implementar en la práctica que las propuestas de Barry y Ver Hoef (1996) o Yao y Journel (1998). Aunque no están disponibles estudios sobre estos modelos (todavía no han sido ni siquiera utilizados), es de esperar que tengan un comportamiento y propiedades similares a otros modelos de este tipo.

Esta idea se puede generalizar para el caso anisotrópico en múltiples dimensiones; sin embargo, hay que tener en cuenta que el número de parámetros crece de forma significativa (exponencial) al aumentar el número de dimensiones por lo que pueden surgir problemas computacionales (aunque siempre menores que al considerar modelos basados en medias móviles), además de que el número de observaciones puede no ser muy grande. Un caso donde típicamente pueden aparecer estos problemas es el espacio-temporal donde el número total de dimensiones es como mínimo de dos (y normalmente mayor). Si el número de dimensiones es grande, se puede pensar en reducir el número de parámetros de estos modelos (completamente anisotrópicos) considerando que hay isotropía en algún componente del vector de salto. Por ejemplo, en el caso espacio-temporal habitualmente se supone que $\gamma(\mathbf{h}, t) \equiv \gamma(\|\mathbf{h}\|, t)$. La extensión de la aproximación Shapiro y Botha (1991) para este caso se tratará con detalle en la segunda parte de este trabajo.

sólo expusieron las expresiones generales sin simplificar (i.e. las correspondientes a utilizar directamente (2.11) y trabajar con números complejos en lugar de reales).

CAPÍTULO 5

GEOESTADÍSTICA MULTIVARIANTE

En algunas ocasiones, además de utilizar las observaciones $\{Z(\mathbf{s}_i) : i = 1, \dots, n\}$ para predecir el valor $Z(\mathbf{s}_0)$, también puede ser deseable tener en cuenta los valores de otras variables aleatorias (observadas en las mismas o en distintas posiciones espaciales). Este es uno de los motivos que promueven el estudio de los procesos espaciales multivariantes definidos en la sección 1.2.

En este capítulo se considerarán procesos espaciales multivariantes:

$$\{\mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_k(\mathbf{s}))' : \mathbf{s} \in D\} \quad (5.1)$$

donde $\{Z_j(\mathbf{s}) : j = 1, \dots, k; \mathbf{s} \in D\}$ son k procesos espaciales univariantes supuestamente interdependientes. La sección 5.1 se ocupa del estudio de las propiedades de este tipo de procesos. En la sección 5.2 describen brevemente algunos de los distintos métodos kriging de predicción lineal multivariante (denominados métodos *cokriging*) y la sección 5.3 se centra en las distintas aproximaciones para el modelado de la dependencia espacial multivariante.

Hay que recordar, como ya se comentó en la sección 1.2, que en ocasiones se utiliza esta metodología para el caso espacio-temporal (por ejemplo, considerando en (5.1) que $Z_i(\mathbf{s}) = Z(\mathbf{s}, t_i)$, $i = 1, \dots, k$); esta forma de proceder será considerada en el capítulo 6 y es el principal motivo de la inclusión de este capítulo. Por tanto este tema va a ser tratado aquí con menor profundidad que los anteriores; si se quiere obtener información adicional, algunas referencias que pueden ser de interés son por ejemplo Wackernagel (1998), Chilès y Delfiner (1999, capítulo 5) o Goovaerts (1997, capítulo 6).

5.1 INTRODUCCIÓN

Al igual que en el caso univariante (capítulo 2) en general únicamente se va a disponer de una realización parcial del proceso multivariante (5.1) y por tanto es necesario realizar también suposiciones acerca de $\mathbf{Z}(\mathbf{s})$. Por ejemplo, supondremos que existe:

$$E(\mathbf{Z}(\mathbf{s})) = \boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \dots, \mu_k(\mathbf{s}))', \forall \mathbf{s} \in D,$$

y en ocasiones también:

$$\text{Cov}(\mathbf{Z}(\mathbf{s}_1), \mathbf{Z}(\mathbf{s}_2)) = \mathbf{C}(\mathbf{s}_1, \mathbf{s}_2), \forall \mathbf{s}_1, \mathbf{s}_2 \in D.$$

En cuyo caso $\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2)$ será una matriz cuadrada de dimensión k no necesariamente simétrica, aunque verificando que:

$$\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2)' = \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1), \forall \mathbf{s}_1, \mathbf{s}_2 \in D.$$

Además, normalmente se considerará alguno de los siguientes tipos de estacionariedad.

5.1.1 Procesos espaciales multivariantes estacionarios

ESTACIONARIEDAD DE SEGUNDO ORDEN

Si el proceso $\mathbf{Z}(\mathbf{s})$ verifica:

$$E(\mathbf{Z}(\mathbf{s})) = \boldsymbol{\mu} = (\mu_1, \dots, \mu_k)', \forall \mathbf{s} \in D, \quad (5.2)$$

$$\text{Cov}(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)) = C_{ij}(\mathbf{s}_1 - \mathbf{s}_2), \forall \mathbf{s}_1, \mathbf{s}_2 \in D; i, j = 1, \dots, k,$$

se dice que es *estacionario de segundo orden*. Las funciones $C_{ij}(\cdot)$ se denominan *covariogramas cruzados* (naturalmente cuando $i = j$ será el covariograma de $Z_i(\mathbf{s})$, denominado también *autocovariograma* en el contexto multivariante). Si además $C_{ij}(\mathbf{h}) = C_{ij}^0(\|\mathbf{h}\|)$ se dice que el covariograma cruzado es isotrópico (y por lo tanto simétrico).

En algunos casos en lugar del covariograma se utilizará el *correlograma cruzado*:

$$\rho_{ij}(\mathbf{h}) = \frac{C_{ij}(\mathbf{h})}{\sqrt{C_{ii}(\mathbf{0})C_{jj}(\mathbf{0})}} \in [-1, +1].$$

ESTACIONARIEDAD INTRÍNSECA

Diremos que el proceso espacial multivariante $\mathbf{Z}(\mathbf{s})$ es *intrínsecamente estacionario* si su media es constante (i.e. se verifica también (5.2)) y además:

$$\text{Var}(Z_i(\mathbf{s}_1) - Z_j(\mathbf{s}_2)) = 2\gamma_{ij}(\mathbf{s}_1 - \mathbf{s}_2), \forall \mathbf{s}_1, \mathbf{s}_2 \in D; i, j = 1, \dots, k, \quad (5.3)$$

funciones que denominaremos *variogramas cruzados* o *variogramas cruzados basados en la varianza* (inicialmente propuestos¹ por Clark et al., 1989).

No obstante, existen diversas formas de generalizar los (auto)variogramas al caso multivariante (ver p.e. Künsch et al., 1997); por ejemplo, en numerosas ocasiones en la literatura se define el variograma cruzado como:

$$\begin{aligned} \text{Cov}(Z_i(\mathbf{s}_1) - Z_i(\mathbf{s}_2), Z_j(\mathbf{s}_1) - Z_j(\mathbf{s}_2)) &= 2v_{ij}(\mathbf{s}_1 - \mathbf{s}_2), \\ \forall \mathbf{s}_1, \mathbf{s}_2 \in D; i, j &= 1, \dots, k, \end{aligned} \quad (5.4)$$

(inicialmente propuesto por Journel y Huijbregts, 1978, p. 324) y se denomina a (5.3) *pseudo-variograma cruzado* (Myers, 1991). Sin embargo, teniendo en cuenta resultados recientes (e.g. Papritz et al., 1993; Cressie y Wikle, 1998; y comentarios siguientes), consideraremos que es preferible utilizar (5.3) en lugar de (5.4) y denominaremos a $v_{ij}(\cdot)$ *variograma cruzado basado en la covarianza* o *variograma cruzado tradicional*.

Análogamente a los casos anteriores, si los variogramas cruzados solamente son funciones de la distancia y no de la dirección del salto, se dirá que son isotrópicos.

RELACIÓN ENTRE LOS DISTINTOS TIPOS DE ESTACIONARIEDAD

Al igual que en el caso univariante, la clase de procesos multivariantes intrínsecamente estacionarios es más general que la clase de procesos estacionarios de segundo orden. A partir de los covariogramas cruzados $C_{ij}(\cdot)$ podemos obtener los correspondientes semivariogramas cruzados²:

¹ Aunque la definición utilizada por estos autores fue $\gamma_{ij}(\mathbf{h}) = \frac{1}{2} E \left(Z_i(\mathbf{s}) - Z_j(\mathbf{s} + \mathbf{h}) \right)^2$, distinta a la recomendada aquí por los mismos motivos que en el caso univariante.

² Por tanto los (pseudo) variogramas cruzados pueden modelar covarianzas negativas pese a lo que afirman algunos autores (e.g. Wakernagel, 1998, p. 150).

$$\gamma_{ij}(\mathbf{h}) = (C_{ii}(\mathbf{0}) + C_{jj}(\mathbf{0}))/2 - C_{ij}(\mathbf{h}),$$

y también los semivariogramas cruzados basados en la covarianza:

$$v_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - (C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h}))/2. \quad (5.5)$$

Además, a partir de los semivariogramas cruzados podemos obtener cualquier covarianza entre incrementos, ya que:

$$\begin{aligned} Cov(Z_i(\mathbf{s}_1) - Z_j(\mathbf{s}_2), Z_l(\mathbf{s}_3) - Z_m(\mathbf{s}_4)) = \\ -\gamma_{il}(\mathbf{s}_1 - \mathbf{s}_3) + \gamma_{im}(\mathbf{s}_1 - \mathbf{s}_4) + \gamma_{jl}(\mathbf{s}_2 - \mathbf{s}_3) - \gamma_{jm}(\mathbf{s}_2 - \mathbf{s}_4), \end{aligned}$$

y como caso particular:

$$v_{ij}(\mathbf{h}) = (\gamma_{ij}(\mathbf{h}) + \gamma_{ij}(-\mathbf{h}))/2 - \gamma_{ij}(\mathbf{0}).$$

por tanto siempre podemos expresar un variograma cruzado basado en la covarianza a partir de los variogramas cruzados. Sin embargo, el caso contrario sólo es posible si el variograma cruzado es una función par y se verifican algunas condiciones más restrictivas de estacionariedad (Papritz et al., 1993), en cuyo caso sólo se diferencian en una constante (Myers, 1991). Por tanto el variograma cruzado tradicional es algo más general en el sentido de que puede existir en casos en los que no existe (5.3).

Las condiciones para la existencia de los variogramas cruzados han sido estudiadas por Papritz et al. (1993). Entre otros resultados demostraron que las pendientes de los autovariogramas (y de ambos tipos de variogramas cruzados) deben coincidir cuando el salto tiende a infinito. Por tanto podríamos decir que los semivariogramas cruzados son poco más generales que los covariogramas cruzados, ya que sólo se diferencian en la inclusión de un componente lineal.

5.1.2 Algunas propiedades de los covariogramas y variogramas cruzados

Al contrario que en el caso univariante, en general los covariogramas cruzados no son simétricos:

$$C_{ij}(\mathbf{h}) \neq C_{ij}(-\mathbf{h}),$$

si $i \neq j$, aunque verifican que:

$$C_{ij}(\mathbf{h}) = C_{ji}(-\mathbf{h}),$$

y la desigualdad de Cauchy-Schwarz:

$$|C_{ij}(\mathbf{h})|^2 \leq C_{ii}(\mathbf{0})C_{jj}(\mathbf{0}),$$

y por tanto son también acotados. Además, como las matrices de covarianzas son semidefinidas positivas, se verifica también que³:

$$\sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^n \sum_{m=1}^n a_{il} a_{jm} C_{ij}(\mathbf{s}_l - \mathbf{s}_m) \geq 0$$

$$\forall n \geq 1, \forall \mathbf{s}_l \in D, \forall a_{il} \in \mathbb{R}; i = 1, \dots, k, l = 1, \dots, n.$$

Los semivariogramas cruzados verifican propiedades de simetría equivalentes:

$$\gamma_{ij}(\mathbf{h}) \neq \gamma_{ij}(-\mathbf{h}),$$

$$\gamma_{ij}(\mathbf{h}) = \gamma_{ji}(-\mathbf{h}),$$

y además son condicionalmente semidefinidos negativos:

$$\sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^n \sum_{m=1}^n a_{il} a_{jm} \gamma_{ij}(\mathbf{s}_l - \mathbf{s}_m) \leq 0 \tag{5.6}$$

$$\forall n \geq 1, \forall \mathbf{s}_l \in D, \forall a_{il} \in \mathbb{R}; i = 1, \dots, k, l = 1, \dots, n,$$

tales que $\sum_{l=1}^n a_{il} = 0, i = 1, \dots, k$.

Los semivariogramas cruzados basados en la covarianza sin embargo son simétricos:

$$v_{ij}(\mathbf{h}) = v_{ij}(-\mathbf{h}),$$

verifican la desigualdad de Cauchy-Schwarz:

$$|v_{ij}(\mathbf{h})|^2 \leq v_{ii}(\mathbf{h})v_{jj}(\mathbf{h}),$$

y son también condicionalmente semidefinidos negativos (i.e. verifican (5.6)).

³ Se deducen también otras propiedades elementales análogas al caso univariante mostrado en la sección 2.2.1, pero por motivos de extensión no se exponen aquí.

Además las formas de ambos tipos de variogramas cruzados pueden ser bastante diferentes de la de un autovariograma. Por ejemplo, $2\gamma_{ij}(\cdot)$ no tiene porque ser nulo en el origen; en general:

$$\gamma_{ij}(\mathbf{h}) \geq 0, \forall \mathbf{h}.$$

Los variogramas cruzados tradicionales si que toman el valor cero en el origen:

$$v_{ij}(\mathbf{0}) = 0,$$

aunque pueden tomar valores negativos en saltos $\mathbf{h} \neq \mathbf{0}$.

Uno de los inconvenientes de los variogramas cruzados tradicionales es su simetría. Por ejemplo, a partir de la relación (5.5) se deduce que, si descomponemos el covariograma cruzado en una función par y otra impar, el variograma cruzado tradicional solamente recoge el término par (ver p.e. Wakernagel, 1998, pp. 146-149; o Chilès y Delfiner, 1999, pp. 328-329; para más detalles). De hecho, para utilizar los variogramas cruzados tradicionales en lugar de los covariogramas cruzados en las ecuaciones cokriging debe suponerse simetría (algo que no ocurre con los semivariogramas cruzados basados en la varianza). Por tanto en algunos casos las ecuaciones de predicción lineal multivariante utilizando los variogramas cruzados tradicionales dan lugar a predictores no óptimos (ver p.e. Ver Hoef y Cressie, 1993, y comentarios en la sección 5.2).

Para comparar ambos tipos de variogramas cruzados son también de utilidad las siguientes expresiones (utilizadas en su estimación):

$$\begin{aligned} 2\gamma_{ij}(\mathbf{s}_1 - \mathbf{s}_2) &= E\left(\left(Z_i(\mathbf{s}_1) - \mu_i\right) - \left(Z_j(\mathbf{s}_2) - \mu_j\right)\right)^2 \\ &= E\left(Z_i(\mathbf{s}_1) - Z_j(\mathbf{s}_2)\right)^2 - \left(\mu_i - \mu_j\right)^2, \end{aligned} \quad (5.7)$$

$$2v_{ij}(\mathbf{s}_1 - \mathbf{s}_2) = E\left(\left(Z_i(\mathbf{s}_1) - Z_i(\mathbf{s}_2)\right)\left(Z_j(\mathbf{s}_1) - Z_j(\mathbf{s}_2)\right)\right).$$

A partir de estas expresiones se deduce que para la estimación (empírica o no paramétrica) de $2\gamma_{ij}(\cdot)$ no tienen porque coincidir las posiciones en las que se observan los procesos $Z_i(\mathbf{s})$ y $Z_j(\mathbf{s})$, mientras que $2v_{ij}(\cdot)$ sólo puede ser estimado a partir de pares de posiciones en los que ambos procesos son observados (ver p.e. Papritz et al., 1993). Este es quizás el principal inconveniente de los variogramas cruzados tradicionales.

La expresión (5.7) justificaría en principio los problemas achacados a los variogramas cruzados cuando las variables están medidas en escalas muy diferentes (e.g. Wakernagel, 1998, p. 150), aunque éstos afectarían igualmente al covariograma cruzado tradicional. Para evitar problemas numéricos, Cressie y Wikle (1998) propusieron estandarizar previamente los datos (además de demostrar que los predictores cokriging tradicionales, sección 5.2.1, utilizando $2\gamma_{ij}(\cdot)$ son invariantes frente a cambios en media y en escala).

Adicionalmente, la expresión (5.7) indica que para la estimación de los variogramas cruzados es necesario estimar también las medias; por lo que sería de esperar que la estimación de variogramas cruzados tenga pocas ventajas (o quizás ninguna) sobre la estimación de covariogramas cruzados (al contrario de lo que ocurre en el caso univariante, ver sección 4.1.1).

REPRESENTACIÓN ESPECTRAL

Al igual que en el caso univariante (sección 2.2.3) podemos caracterizar los covariogramas y variogramas cruzados continuos a partir de su representación espectral; sin embargo en este caso resulta más difícil utilizar las expresiones obtenidas en la construcción de modelos válidos. Como ejemplo se tratará el caso de covariogramas cruzados continuos.

Supongamos que $\mathbf{C}(\mathbf{h}) = [C_{lm}(\mathbf{h})]$ es la matriz $k \times k$ de covariogramas cruzados continuos de un proceso multivariante $\mathbf{Z}(\mathbf{s})$ estacionario de segundo orden (en este apartado se cambiará la notación habitual para evitar confusiones con $i = \sqrt{-1}$). Por simplicidad supongamos además que los covariogramas cruzados son integrables en valor absoluto (esto garantiza la existencia de las funciones de densidad espectrales, el caso general es análogo considerando integrales tipo Fourier-Stieljes). Por una generalización del teorema de Bochner debida a Cramer (1940), los covariogramas cruzados continuos $C_{lm}(\cdot)$ admiten la representación espectral:

$$C_{lm}(\mathbf{h}) = \int_{\mathbb{R}^d} e^{i(\boldsymbol{\omega} \cdot \mathbf{h})} f_{lm}(\boldsymbol{\omega}) d\boldsymbol{\omega}$$

donde las densidades espectrales:

$$f_{lm}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(\boldsymbol{\omega} \cdot \mathbf{h})} C_{lm}(\mathbf{h}) d\mathbf{h}$$

son en general funciones complejas (salvo cuando $l = m$) y tales que las matrices de densidades espectrales $\mathbf{F}(\boldsymbol{\omega}) = [f_{lm}(\boldsymbol{\omega})]$ son Hermitianas y semidefinidas positivas. Esto implica que:

$$f_{lm}(-\boldsymbol{\omega}) = \overline{f_{lm}(\boldsymbol{\omega})} = f_{ml}(\boldsymbol{\omega}), \quad (5.8)$$

$$|f_{lm}(\boldsymbol{\omega})|^2 \leq f_{ll}(\boldsymbol{\omega})f_{mm}(\boldsymbol{\omega}), \quad (5.9)$$

(aunque no son condiciones suficientes).

De forma análoga al caso univariante, se pueden obtener expresiones similares para el caso de los variogramas cruzados y también simplificaciones para el caso de isotropía (por ejemplo, la expresión correspondiente a variogramas cruzados tradicionales isotrópicos se tienen en Christakos, 1992, p. 98).

5.2 PREDICCIÓN ESPACIAL MULTIVARIANTE. COKRIGING

Supongamos que:

$$\{Z_j(\mathbf{s}_{ji}) : j = 1, \dots, k; i = 1, \dots, n_j\} \quad (5.10)$$

son los $n = n_1 + \dots + n_k$ valores observados a partir de los que se pretende predecir $Z_1(\mathbf{s}_0)$. Un problema más general es la predicción simultánea de un vector de variables $\mathbf{Z}(\mathbf{s}_0) = (Z_1(\mathbf{s}_0), \dots, Z_k(\mathbf{s}_0))'$ que se resuelve de modo análogo⁴ y por tanto no va a ser tratado aquí (este caso es tratado por ejemplo en Ver Hoef y Cressie, 1993).

Dependiendo de la configuración de los datos se suele distinguir entre dos situaciones:

- *Isotópica*: todas las variables son observadas en las mismas posiciones espaciales (i.e. $\mathbf{s}_{ji} = \mathbf{s}_i, \forall i, j$).
- *Heterotópica*: las variables son observadas en conjuntos de posiciones no necesariamente iguales (en el caso de no tener ninguna posición en común se dice que los datos son completamente heterotópicos).

⁴ La única posible dificultad está en la generalización del error en media cuadrática, para lo que se han considerado distintos criterios (e.g. Myers, 1982; Ver Hoef y Cressie, 1993; Cressie, 1993, p. 142).

En la literatura es habitual distinguir entre cokriging isotópico y cokriging heterotópico. Aquí se utilizará la notación (5.10) correspondiente a este último caso más general; además, denotaremos por $\mathbf{Z}_j = (Z_j(\mathbf{s}_{j1}), \dots, Z_j(\mathbf{s}_{jn_j}))'$ el vector de observaciones del proceso $Z_j(\cdot)$ y por $\mathbf{Z} = (\mathbf{Z}'_1, \dots, \mathbf{Z}'_k)'$ el vector formado por el total de observaciones.

Al igual que en el caso univariante (capítulo 3), dependiendo de las suposiciones acerca de la función de tendencia $E(\mathbf{Z}(\mathbf{s})) = \boldsymbol{\mu}(\mathbf{s})$, se distingue entre tres métodos de predicción lineal óptima multivariante:

4. Cokriging simple (CKS): se supone que la tendencia es conocida.
5. Cokriging ordinario (CKO): se supone que las medias son constantes (i.e. se verifica (5.2)) y desconocidas.
6. Cokriging universal (CKU): se supone que las tendencias $\mu_j(\cdot)$ (no nec. constantes y desconocidas) son combinaciones lineales de funciones o variables explicativas:

$$\mu_j(\mathbf{s}) = \mathbf{x}_j(\mathbf{s})' \boldsymbol{\beta}_j$$

donde $\mathbf{x}_j(\mathbf{s})$ es un vector $(p_j + 1) \times 1$ conocido y $\boldsymbol{\beta}_j$ es el correspondiente vector $(p_j + 1) \times 1$ de parámetros desconocidos.

Las ecuaciones de estos métodos son muy similares a las correspondientes del caso univariante (la principal diferencia son las restricciones de insesgadez sobre los pesos de las variables secundarias en el CKO y CKU), por este motivo sólo se mostrarán aquí las correspondientes al cokriging simple y al cokriging ordinario (sección 5.2.1). Las ecuaciones del cokriging universal están disponibles por ejemplo en Chilès y Delfiner (1999, sección 5.4.1) o Ver Hoef y Cressie (1993). Sin embargo, puede haber ciertos inconvenientes con estos métodos cokriging (que denominaremos tradicionales), por lo que se han propuesto algunas modificaciones o simplificaciones; en la sección 5.2.2 se muestran algunas de estas alternativas.

Al igual que en el caso univariante, la hipótesis de estacionariedad (e.g. $\mathbf{C}(\mathbf{u}, \mathbf{v}) = \mathbf{C}(\mathbf{u} - \mathbf{v})$) es conveniente en la práctica para el modelado de la dependencia espacial, pero no es necesaria para la predicción; por este motivo las ecuaciones serán expresadas en función de las matrices de covarianzas (y en los métodos en los que es posible, se indicará como son las expresiones resultantes considerando variogramas cruzados).

5.2.1 Métodos cokriging tradicionales

COKRIGING SIMPLE

Si la tendencia $E(\mathbf{Z}(\mathbf{s})) = \boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \dots, \mu_k(\mathbf{s}))'$ es conocida, al igual que en el caso univariante (sección 3.2), se deduce que el predictor kriging puede expresarse como el valor medio de la variable de interés más una combinación lineal de los residuos de los valores observados:

$$p_1(\mathbf{Z}, \mathbf{s}_0) = \mu_1(\mathbf{s}_0) + \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} (Z_j(\mathbf{s}_{ji}) - \mu_j(\mathbf{s}_{ji})),$$

(lo que garantiza que sea insesgado independientemente del conjunto de pesos).

El predictor óptimo se obtiene minimizando:

$$E \left(Z_1(\mathbf{s}_0) - \mu_1(\mathbf{s}_0) - \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} (Z_j(\mathbf{s}_{ji}) - \mu_j(\mathbf{s}_{ji})) \right)^2,$$

y de modo idéntico al caso del kriging simple, se obtienen las n ecuaciones cokriging en función de las matrices de covarianzas:

$$\sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} C_{jj'}(\mathbf{s}_{ji}, \mathbf{s}_{j'i'}) - C_{1j'}(\mathbf{s}_0, \mathbf{s}_{j'i'}) = 0,$$

$$j' = 1, \dots, k; i' = 1, \dots, n_{j'}.$$

Resolviendo este sistema se obtiene el predictor:

$$p_{CKS(1)}(\mathbf{Z}, \mathbf{s}_0) = \mu_1(\mathbf{s}_0) + \mathbf{c}'_1 \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}), \quad (5.11)$$

donde $\mathbf{c}_1 = (C_{11}(\mathbf{s}_0, \mathbf{s}_{11}), \dots, C_{1k}(\mathbf{s}_0, \mathbf{s}_{kn_k}))'$, $\boldsymbol{\mu} = E(\mathbf{Z})$ y $\boldsymbol{\Sigma} = Var(\mathbf{Z})$; y el mínimo e.m.c. de predicción (o varianza cokriging) viene dado por:

$$\sigma_{CKS(1)}^2(\mathbf{s}_0) = C_{11}(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}'_1 \boldsymbol{\Sigma}^{-1} \mathbf{c}_1,$$

obteniéndose por tanto las mismas expresiones que en el caso univariante.

En la práctica este predictor y la correspondiente varianza cokriging pueden calcularse utilizando el algoritmo propuesto al final de la sección 3.2; sin embargo, en este caso es más fácil que aparezcan inestabilidades numéricas en la práctica (normalmente debidas a una mayor densidad de muestreo de las variables secundarias) por lo que es recomendable realizar los cálculos en

máxima precisión (p.e. utilizando las rutinas DLFTDS y DLFIDS de la librería IMSL).

COKRIGING ORDINARIO

En el caso de medias constantes $E(\mathbf{Z}(\mathbf{s})) = (\mu_1, \dots, \mu_k)'$, $\forall \mathbf{s} \in D$, y desconocidas, el predictor kriging es una combinación lineal de los valores observados:

$$p_1(\mathbf{Z}, \mathbf{s}_0) = \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} Z_j(\mathbf{s}_{ji}), \tag{5.12}$$

verificando:

$$\begin{aligned} \sum_{i=1}^{n_1} \lambda_{1i} &= 1, \\ \sum_{i=1}^{n_j} \lambda_{ji} &= 0; \quad j = 2, \dots, k, \end{aligned} \tag{5.13}$$

(condición necesaria y suficiente para que el predictor sea uniformemente insesgado). El predictor óptimo se obtiene minimizando el correspondiente e.c.m. de predicción bajo las restricciones de insesgadez (5.13), siendo para ello necesario considerar k multiplicadores de Lagrange m_1, \dots, m_k (mientras que en el kriging ordinario sólo era necesario incluir uno). De modo análogo al caso univariante se obtienen las $n + k$ ecuaciones del cokriging ordinario:

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} C_{jj'}(\mathbf{s}_{ji}, \mathbf{s}_{j'i'}) - m_{j'} &= C_{1j'}(\mathbf{s}_0, \mathbf{s}_{j'i'}), \quad j' = 1, \dots, k; \quad i' = 1, \dots, n_{j'}, \\ \sum_{i=1}^{n_j} \lambda_{ji} &= \delta_{1j}, \quad j = 1, \dots, k, \end{aligned}$$

(siendo δ_{ij} la delta de Kronecker), y resolviendo este sistema se obtiene el predictor del cokriging ordinario; al igual que en el caso univariante, este predictor coincide con el obtenido sustituyendo en el predictor del cokriging simple (5.11) las medias teóricas por sus correspondientes estimadores lineales óptimos⁵ (lo cual

⁵ Para más detalles (formulación matricial) ver por ejemplo Myers (1982) o Ver Hoef y Cressie (1993).

es también válido para el cokriging universal). El mínimo error de predicción en media cuadrática o varianza cokriging resulta ser:

$$\sigma_{CKO(1)}^2 = C_{11}(\mathbf{s}_0, \mathbf{s}_0) - \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} C_{1j}(\mathbf{s}_0, \mathbf{s}_{ji}) + m_1.$$

Las ecuaciones se pueden formular en función de los variogramas cruzados sustituyendo simplemente $C_{ij}(\mathbf{u}, \mathbf{v})$ por $-\gamma_{ij}(\mathbf{u} - \mathbf{v})$ (ver p.e. Cressie, 1993, pp. 140-141). Si se pretende utilizar el variograma cruzado tradicional se debe verificar además la condición de simetría $C_{ij}(\mathbf{u}, \mathbf{v}) = C_{ij}(\mathbf{v}, \mathbf{u})$ (ver p.e. Journel y Huijbregts, 1978, p. 326; Myers, 1982, pp. 253-254). En el caso de asimetría, la predicción basada en el variograma cruzado tradicional da lugar a predictores no óptimos y en algunos casos considerablemente inferiores (e.g. Ver Hoef y Cressie, 1993, pp. 232-234).

El algoritmo descrito al final de la sección 3.3 puede ser utilizado para el cálculo del predictor y varianza cokriging; sin embargo pueden aparecer problemas computacionales incluso trabajando en máxima precisión (especialmente en el caso del cokriging universal). Para evitar estos problemas puede ser recomendable utilizar el algoritmo propuesto por Long y Myers (1997) cuando el número de procesos o de observaciones es grande.

INFLUENCIA DE LA INFORMACIÓN SECUNDARIA

La inclusión de información adicional da lugar a predicciones con un menor e.m.c. de predicción, y por tanto:

$$\sigma_{CK}^2(\mathbf{s}_0) \leq \sigma_K^2(\mathbf{s}_0), \forall \mathbf{s}_0 \in D,$$

aunque el beneficio obtenido puede no ser significativo en muchos casos; un claro ejemplo sería cuando los covariogramas cruzados entre la variable primaria y las variogramas secundarias son aproximadamente nulos:

$$C_{1i}(\mathbf{h}) \simeq 0, \forall \mathbf{h}, i = 2, \dots, k.$$

En el caso de que las variables secundarias estén incorreladas con la variable primaria ($C_{1i}(\cdot) \equiv 0, i = 2, \dots, k$) los predictores cokriging coincidirán con los correspondientes predictores kriging.

Además, los métodos cokriging tradicionales (simple, ordinario y universal con las restricciones de insesgadez habituales) asignan pesos nulos a las variables secundarias en el caso isotópico (todas las variables observadas en las mismas posiciones; ver sección 5.1) cuando los covariogramas cruzados entre las variables secundarias y la principal son proporcionales al auto-covariograma de la variable primaria:

$$C_{1i}(\mathbf{h}) = C_{i1}(\mathbf{h}) = \rho_{1i}C_{11}(\mathbf{h}), i = 2, \dots, k,$$

(por ejemplo, este es el caso del modelo intrínseco y del modelo de Markov MM1 mostrados en la sección 5.3.2). En estos casos, los predictores cokriging coinciden con los predictores kriging (ver p.e. Helterbrand y Cressie, 1994; Wackernagel, 1998, capítulo 25; Goovaerts, 1997, pp. 215-221). En general, se puede concluir que si la forma de los modelos de (co)variogramas cruzados es “más o menos similar” y no hay un submuestreo notable de la variable primaria, los métodos cokriging tradicionales no mejorarán mucho los resultados obtenidos con los métodos kriging univariantes⁶.

En la práctica se ha observado que los métodos cokriging (tradicionales) solamente mejoran de forma significativa los resultados obtenidos con la predicción kriging univariante cuando hay un submuestreo de la variable primaria en comparación con las secundarias y además las observaciones secundarias están bien correladas con $Z_1(\mathbf{s}_0)$ (e.g. Journel y Huijbregts, 1978, p. 326).

5.2.2 Variantes de los métodos cokriging tradicionales

COKRIGING ORDINARIO CENTRADO Y ESTANDARIZADO

Las restricciones del cokriging ordinario (y universal) sobre los pesos de las observaciones secundarias (suma nula en el caso del CKO), tienen principalmente dos inconvenientes:

- Algunos de los pesos secundarios son negativos, lo cual puede dar lugar a predicciones no válidas (e.g. Isaaks y Srivastava, 1989, pp. 411 y 416).

⁶ Otras referencias que pueden ser también de interés (además de las anteriores) son por ejemplo Goovaerts (1998) y Chilès y Delfiner (1999, sección 5.4.4).

- La mayoría de los pesos secundarios tienden a ser pequeños, reduciendo la influencia de los datos secundarios, lo que justifica también (ver comentarios al final de la sección anterior) que la eficiencia de los predictores cokriging no sea muy superior a la de los predictores kriging en muchos casos (lo que puede dar lugar a pensar que la estadística espacial multivariante puede ser una complicación innecesaria).

Por estos motivos se han propuesto algunas variantes del cokriging ordinario; por ejemplo, Isaaks y Srivastava (1989, pp. 408-416) propusieron utilizar, en lugar de (5.12) y (5.13), el predictor:

$$p_1(\mathbf{Z}, \mathbf{s}_0) = \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} (Z_j(\mathbf{s}_{ji}) - \hat{\mu}_j + \hat{\mu}_1),$$

verificando:

$$\sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} = 1. \quad (5.14)$$

A este predictor lo denominaremos predictor cokriging ordinario centrado (CKOC).

Otra alternativa, propuesta por Goovaerts (1998, p.33), consistiría en trabajar con los datos estandarizados:

$$\frac{(p_1(\mathbf{Z}, \mathbf{s}_0) - \hat{\mu}_1)}{C_{11}(\mathbf{0})} = \sum_{j=1}^k \sum_{i=1}^{n_j} \lambda_{ji} \frac{(Z_j(\mathbf{s}_{ji}) - \hat{\mu}_j)}{C_{jj}(\mathbf{0})},$$

considerando también la restricción (5.14). Este método, que denominaremos cokriging ordinario estandarizado⁷ (CKOE), incrementa aún más la influencia de las observaciones secundarias en la predicción (Goovaerts, 1998, pp. 35-39).

COKRIGING CO-POSICIONADO (COLOCALIZADO)

Cuando se conoce el valor de una variable secundaria en la posición de predicción, este valor oculta las observaciones secundarias más alejadas (produce un “efecto pantalla”) disminuyendo su influencia en el predictor (ver p.e.

⁷ Goovaerts (1998) lo denominó cokriging reescalado expresado en función del correlograma, ya que se puede obtener sustituyendo los covariogramas cruzados por correlogramas en las ecuaciones del cokriging ordinario centrado.

Goovaters, 1997, pp. 221-222), por lo que incluir únicamente la información secundaria en la posición de predicción (co-posicionada con $Z_1(\mathbf{s}_0)$) apenas disminuye la eficiencia del predictor (e.g. Goovaters, 1998, pp. 35-39). Esta forma de proceder, además de las ventajas numéricas (mayor rapidez evitando posibles inestabilidades de la matriz cokriging), facilita notablemente el modelado de la dependencia espacial multivariante (ver comentarios a continuación). Los predictores obtenidos de esta forma, propuestos inicialmente por Xu et al. (1992), los denominaremos predictores cokriging co-posicionados (o colocalizados; del inglés “collocated”).

Cokriging simple co-posicionado

En el caso de medias conocidas el método consiste en realizar cokriging simple de los valores considerados; el predictor es por tanto de la forma (Xu et al., 1992):

$$p_1(\mathbf{Z}, \mathbf{s}_0) = \mu_1(\mathbf{s}_0) + \sum_{i=1}^{n_1} \lambda_{1i} (Z_1(\mathbf{s}_{1i}) - \mu_1(\mathbf{s}_{1i})) + \sum_{j=2}^k \lambda_{j0} (Z_j(\mathbf{s}_0) - \mu_j(\mathbf{s}_0)),$$

y las correspondientes ecuaciones cokriging, en función de las matrices de covarianzas, son:

$$\sum_{i=1}^{n_1} \lambda_{1i} C_{11}(\mathbf{s}_{1i}, \mathbf{s}_{1i'}) + \sum_{j=2}^k \lambda_{j0} C_{j1}(\mathbf{s}_0, \mathbf{s}_{1i'}) = C_{11}(\mathbf{s}_0, \mathbf{s}_{1i'}), i' = 1, \dots, n_1,$$

$$\sum_{i=1}^{n_1} \lambda_{1i} C_{1j'}(\mathbf{s}_{1i}, \mathbf{s}_0) + \sum_{j=2}^k \lambda_{j0} C_{jj'}(\mathbf{s}_0, \mathbf{s}_0) = C_{1j'}(\mathbf{s}_0, \mathbf{s}_0), j' = 2, \dots, k.$$

Por tanto las auto-covarianzas de las variables secundarias sólo son evaluadas en $(\mathbf{s}_0, \mathbf{s}_0)$ (i.e. sólo se necesitan conocer sus varianzas en la posición de predicción), simplificándose su modelado en la práctica (ver modelos de Markov en la sección 5.3.2).

Cokriging ordinario co-posicionado

En el caso de medias desconocidas no se debería considerar la idea anterior utilizando el método de cokriging ordinario tradicional ya que se asignarían pesos nulos a las observaciones secundarias; en lugar de este método se puede utilizar por ejemplo el cokriging ordinario centrado o estandarizado

(e.g. Goovaters, 1997, pp. 236-237; Goovaters, 1998), obteniéndose con el CKOC el predictor:

$$p_1(\mathbf{Z}, \mathbf{s}_0) = \sum_{i=1}^{n_1} \lambda_{1i} Z_1(\mathbf{s}_{1i}) + \sum_{j=2}^k \lambda_{j0} (Z_j(\mathbf{s}_0) - \hat{\mu}_j + \hat{\mu}_1),$$

sujeito a:

$$\sum_{i=1}^{n_1} \lambda_{1i} + \sum_{j=2}^k \lambda_{j0} = 1.$$

Otra alternativa sería añadir también las observaciones secundarias “co-posicionadas” en las demás posiciones espaciales \mathbf{s}_{1i} y utilizar cokriging ordinario tradicional (e.g. Wackernagel, 1998, p.165).

5.3 MODELADO DE LA DEPENDENCIA ESPACIAL MULTIVARIANTE

En este caso el procedimiento habitual es también análogo al caso univariante mostrado en el capítulo 4, en la mayoría de los casos se realiza una estimación piloto de los (co)variogramas cruzados (sección 5.3.1) y posteriormente se ajustan modelos válidos al conjunto de estas estimaciones (secciones 5.3.2 y 5.3.3).

5.3.1 Estimación de los covariogramas y variogramas cruzados

Suponiendo que el proceso $\mathbf{Z}(\mathbf{s})$ es estacionario de segundo orden (sólo nos centraremos en el caso de media constante, en caso contrario habrá que proceder de forma análoga a la descrita en la sección 4.1.2), utilizando el método de los momentos obtenemos el estimador del covariograma cruzado:

$$\hat{C}_{ij}(\mathbf{h}) = \frac{1}{|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} (Z_i(\mathbf{s}_{il}) - \bar{Z}_i)(Z_j(\mathbf{s}_{jm}) - \bar{Z}_j), \quad \mathbf{h} \in \mathbb{R}^d,$$

siendo:

$$N_{ij}(\mathbf{h}) = \{(l, m) : \mathbf{s}_{il} - \mathbf{s}_{jm} = \mathbf{h}; l = 1, \dots, n_i, m = 1, \dots, n_j\}.$$

En la práctica, cuando los datos están irregularmente espaciados, se suelen considerar también regiones de tolerancia en torno a los saltos.

Si el proceso multivariante es intrínsecamente estacionario, utilizando también el método de los momentos (y teniendo en cuenta (5.7)), se obtiene el estimador del semivariograma cruzado (basado en la varianza):

$$\hat{\gamma}_{ij}(\mathbf{h}) = \frac{1}{2|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} (Z_i(\mathbf{s}_{il}) - Z_j(\mathbf{s}_{jm}))^2 - (\bar{Z}_i - \bar{Z}_j)^2, \quad \mathbf{h} \in \mathbb{R}^d;$$

aunque si las variables están medidas en escalas muy diferentes, es recomendable (e.g. Cressie y Wikle, 1998) calcular los variogramas cruzados utilizando los valores estandarizados:

$$\hat{\gamma}_{ij}^*(\mathbf{h}) = \frac{1}{2|N_{ij}(\mathbf{h})|} \sum_{N_{ij}(\mathbf{h})} (Z_i^*(\mathbf{s}_{il}) - Z_j^*(\mathbf{s}_{jm}))^2, \quad \mathbf{h} \in \mathbb{R}^d,$$

siendo $Z_i^*(\mathbf{s}) = (Z_i(\mathbf{s}) - \bar{Z}_i) / \hat{\sigma}_i$. Como ya se comentó en la sección 5.1.2, al ser necesario la estimación de las medias, es de esperar que la estimación del variograma cruzado no presente muchas ventajas sobre la estimación del covariograma cruzado.

Si se pretende utilizar los variogramas cruzados tradicionales (basados en la covarianza), en principio⁸ es necesario disponer de observaciones de ambas variables en las mismas posiciones espaciales:

$$\hat{v}_{ij}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} ((Z_i(\mathbf{s}_l) - Z_i(\mathbf{s}_m))(Z_j(\mathbf{s}_l) - Z_j(\mathbf{s}_m))), \quad \mathbf{h} \in \mathbb{R}^d,$$

con $N(\mathbf{h}) = \{(l, m) : \mathbf{s}_l - \mathbf{s}_m = \mathbf{h}; l, m = 1, \dots, n\}$.

De forma análoga al caso univariante se podría pensar en utilizar estimadores locales tipo núcleo; las expresiones de este tipo de estimadores se obtendrían también de forma natural a partir de las anteriores.

⁸ Papritz et al. (1993, p.1022) propusieron una alternativa para el ajuste del variograma cruzado tradicional en el caso de submuestreo de alguna de las variables, sin embargo esta aproximación resulta poco práctica.

5.3.2 Modelos paramétricos

Los estimadores mostrados en la sección anterior no pueden ser utilizados directamente en la predicción espacial multivariante (al igual que en el caso univariante, no verifican necesariamente las propiedades de los covariogramas o variogramas cruzados válidos). Como ya se comentó en la introducción de esta sección, se suele resolver este problema ajustando $k(k+1)/2$ modelos válidos de covariogramas o variogramas cruzados a un conjunto de estimaciones piloto $\hat{\mathbf{C}}(\mathbf{h}_l) = [\hat{C}_{ij}(\mathbf{h}_l)]$, $\hat{\mathbf{\Gamma}}(\mathbf{h}_l) = [\hat{\gamma}_{ij}(\mathbf{h}_l)]$ o $\hat{\mathbf{\Upsilon}}(\mathbf{h}_l) = [\hat{v}_{ij}(\mathbf{h}_l)]$; $l = 1, \dots, M$. Sin embargo, esto presenta muchas más dificultades que en el caso univariante, principalmente debidas a las complicaciones en la obtención de modelos válidos. A continuación se presenta una revisión con algunos de los modelos paramétricos de covariogramas y variogramas cruzados (basados en la varianza y en la covarianza) utilizados en la literatura.

MODELO INTRÍNSECO

El modelo de dependencia espacial multivariante más simple se obtiene suponiendo que los covariogramas son proporcionales a un modelo básico:

$$C_{ij}(\mathbf{h}) = \rho_{ij} C_0(\mathbf{h}),$$

que se puede expresar en forma matricial como:

$$\mathbf{C}(\mathbf{h}) = \mathbf{C}(\mathbf{0})\rho_0(\mathbf{h}).$$

En el caso de los semivariogramas cruzados tradicionales, denotando por $\mathbf{\Upsilon}(\mathbf{h}) = [v_{ij}(\mathbf{h})]$, la expresión matricial de este tipo de modelos es:

$$\mathbf{\Upsilon}(\mathbf{h}) = \mathbf{B}\gamma_0(\mathbf{h}),$$

siendo \mathbf{B} una matriz $k \times k$ semidefinida positiva. La expresión anterior es también válida para el caso de los variogramas cruzados basados en la varianza siempre que el variograma $\gamma_0(\cdot)$ sea acotado (estacionariedad de segundo orden); en caso contrario, teniendo en cuenta que las pendientes de los auto-variogramas y variogramas cruzados deben coincidir (Papritz et al., 1993; ver también sección 5.1.1), los modelos intrínsecos válidos serán de la forma:

$$\gamma_{ij}(\mathbf{h}) = b\gamma_0(\mathbf{h}),$$

siendo $b \geq 0$.

Es importante recordar que al utilizar este tipo de modelos en los métodos cokriging tradicionales (simple, ordinario y universal con las restricciones de insesgadez habituales) los pesos correspondientes a las variables secundarias son nulos en el caso isotópico (ver sección 5.2.1).

Estos modelos pueden verse también como un caso particular del modelo lineal descrito a continuación.

MODELO LINEAL DE CO-REGIONALIZACIÓN

La misma idea considerada en el caso univariante (sección 4.2.3) puede ser utilizada también para el caso multivariante. Por ejemplo, si suponemos estacionariedad de segundo orden y que los procesos verifican⁹:

$$Z_i(\mathbf{s}) = \sum_{l=0}^L \sum_{m=1}^{M_l} a_{im}^l Y_m^l(\mathbf{s}) + \mu_i,$$

donde $Y_m^l(\mathbf{s})$ son variables estacionarias de segundo orden, de media cero y covariograma $C_l(\mathbf{h})$ (se suelen considerar también modelos básicos), y mutuamente independientes¹⁰; de esta forma se obtiene el denominado *modelo lineal de co-regionalización* (Journel y Huijbregts, 1978, p. 172) expresado en función de los covariogramas cruzados:

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^L \sum_{m=1}^{M_l} a_{im}^l a_{jm}^l C_l(\mathbf{h}) = \sum_{l=0}^L b_{ij}^l C_l(\mathbf{h}),$$

o en forma matricial:

$$\mathbf{C}(\mathbf{h}) = \sum_{l=0}^L \mathbf{B}_l C_l(\mathbf{h})$$

donde las $L + 1$ matrices \mathbf{B}_l son semidefinidas positivas (condición suficiente para obtener modelos válidos).

Es importante destacar que por construcción $b_{ij}^l = b_{ji}^l$ y por tanto los modelos obtenidos de esta forma son simétricos:

⁹ Esta hipótesis es también la base del kriging factorial multivariante (e.g. Goovaters, 1997, sección 6.2.9).

¹⁰ Se han generalizado también este tipo de modelos para el caso de dependencia entre los componentes (Vargas-Guzmán et al., 2002)

$$C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h}) = C_{ij}(-\mathbf{h}).$$

El modelado de asimetrías se suele realizar de la forma descrita en la sección 5.3.3, considerando parámetros de desplazamiento espacial (Journel y Huijbregts, 1978, p. 173); aunque también se han considerado para este caso modelos bilineales de covarianzas obtenidos a partir de modelos lineales complejos (e.g. Wackernagel, 1998, cap. 29).

Suponiendo estacionariedad intrínseca, obtendríamos el correspondiente modelo expresado en función de los semivariogramas cruzados tradicionales:

$$\Upsilon(\mathbf{h}) = \sum_{l=0}^L \mathbf{B}_l \gamma_l(\mathbf{h})$$

donde las $L + 1$ matrices \mathbf{B}_l deben ser semidefinidas positivas y los (auto) semivariogramas $\gamma_l(\mathbf{h})$ deben ser modelo válidos. Este modelo sin embargo no es en general válido para los semivariogramas cruzados (basados en la varianza), ya que los componentes no acotados deben ser los mismos en todos los semivariogramas. Por ejemplo, en el caso bivalente, si los L_0 primeros componentes se corresponden a procesos no estacionarios de segundo orden y los restantes a procesos estacionarios, obtendríamos el siguiente modelo (simétrico) para los variogramas cruzados (Papritz et al., 1993, p.1021):

$$\gamma_{12}(\mathbf{h}) = \sum_{l=1}^{L_0} b^l \gamma_l(\mathbf{h}) + \sum_{l=L_0+1}^L \left(\frac{1}{2}(b_{11}^l + b_{22}^l) C_l(\mathbf{0}) - b_{12}^l C_l(\mathbf{h}) \right)$$

Para la construcción de modelos válidos (matrices semidefinidas positivas) suelen tener en cuenta algunas reglas prácticas (e.g. Goovaerts, 1997, p. 114); por ejemplo, aunque los semivariogramas no tienen porque incluir todas las $L + 1$ estructuras básicas, si se incluye alguna en un semivariograma cruzado debe estar presente también en los auto-semivariogramas correspondientes.

En el ajuste de modelos de este tipo en la práctica se suele dar prioridad a los auto-semivariogramas (e.g. Goovaerts, 1997, p. 117-122), determinando en primer lugar las estructuras que permiten ajustar los auto-semivariogramas (de forma independiente) y finalmente ajustando de forma conjunta todos los semivariogramas. Para realizar el ajuste conjunto se suele recurrir al algoritmo m.c.p. propuesto por Goulard (1989) (ver también Goulard y Voltz, 1992; Wackernagel, 1998, p. 175 y Goovaters, 1997, apéndice A), aunque otra

alternativa que puede ser más interesante es la reparametrización del modelo considerando las factorizaciones de Cholesky de las matrices \mathbf{B}_l (de forma análoga a la descrita por Rehman y Shapiro, 1996, p. 220, para otro tipo de modelos; ver siguiente sección).

MODELOS DE MARKOV

En el cokriging co-posicionado el covariograma de las variables secundarias sólo es evaluado en el origen, lo cual puede simplificar notablemente el modelado de la dependencia espacial multivariante. Por ejemplo, si sólo se considera una variable secundaria sólo es necesario modelar $C_{11}(\mathbf{h})$, $C_{12}(\mathbf{h})$ y $C_{22}(\mathbf{0})$ para poder aplicar estos métodos y no es necesario realizar ninguna suposición sobre $C_{22}(\mathbf{h})$ para $\mathbf{h} \neq \mathbf{0}$. Por este motivo Almeida y Journel (1994) propusieron el siguiente modelo de Markov denominado MM1:

$$\rho_{12}(\mathbf{h}) = \rho_{12}(\mathbf{0})\rho_{11}(\mathbf{h}).$$

Este modelo se basa en la siguiente hipótesis de Markov:

$$E\{Z_2(\mathbf{u})|Z_1(\mathbf{u}) = z_1, Z_1(\mathbf{u}') = z_1'\} = E\{Z_2(\mathbf{u})|Z_1(\mathbf{u}) = z_1\}, \forall \mathbf{u}, \mathbf{u}', z_1, z_1'.$$

Sin embargo en ciertas ocasiones puede ser más razonable la hipótesis:

$$E\{Z_1(\mathbf{u})|Z_2(\mathbf{u}) = z_2, Z_2(\mathbf{u}') = z_2'\} = E\{Z_1(\mathbf{u})|Z_2(\mathbf{u}) = z_2\}, \forall \mathbf{u}, \mathbf{u}', z_2, z_2',$$

relacionada con la idea del cokriging co-posicionado de que el valor de variable secundaria en la posición de predicción oculta las observaciones secundarias más alejadas. A partir de esta hipótesis se obtiene el denominado modelo MM2 (Journel, 1999):

$$\rho_{12}(\mathbf{h}) = \rho_{12}(\mathbf{0})\rho_{22}(\mathbf{h}).$$

OTROS MODELOS PARAMÉTRICOS

Aunque el modelo lineal de co-regionalización ha sido el más utilizado en la literatura, existen numerosos ejemplos donde se muestra que puede ser bastante restrictivo (justificando también el “pesimismo” con los métodos cokriging). Con la idea de disponer de mayor flexibilidad, Myers (1982) propuso un criterio para la obtención de modelos (simétricos) para covariogramas y

variogramas cruzados a partir de modelos univariantes. Considerando el proceso espacial:

$$Z_{ij}^+(\mathbf{s}) = Z_i(\mathbf{s}) + Z_j(\mathbf{s}),$$

suponiendo estacionariedad de segundo orden y que el covariograma cruzado es simétrico, se deduce que:

$$C_{ij}^+(\mathbf{h}) = C_{ii}(\mathbf{h}) + C_{jj}(\mathbf{h}) + 2C_{ij}(\mathbf{h}),$$

y por tanto se pueden obtener modelos de covariogramas cruzados a partir de los auto-covariogramas de los correspondientes procesos y del proceso suma. En el caso de procesos intrínsecamente estacionarios, considerando el semivariograma cruzado tradicional, se obtiene que:

$$v_{ij}(\mathbf{h}) = \frac{1}{2}(v_{ij}^+(\mathbf{h}) - v_{ii}(\mathbf{h}) - v_{jj}(\mathbf{h})),$$

(y una expresión algo más complicada para el caso del semivariograma cruzado basado en la varianza). El principal problema de esta aproximación es que, al ajustar de forma independiente los modelos univariantes, el semivariograma (o covariograma) cruzado obtenido no es necesariamente válido. En general habrá que verificar su validez en todos los saltos \mathbf{h} que se vayan a considerar; por ejemplo, en el caso del semivariograma cruzado tradicional, se suele comprobar que se verifica la desigualdad de Cauchy-Schwarz en todos los saltos:

$$|v_{ij}(\mathbf{h})|^2 \leq v_{ii}(\mathbf{h})v_{jj}(\mathbf{h}), \forall \mathbf{h},$$

aunque esto no es una condición suficiente.

La representación en medias móviles descrita en la sección 4.2.1 puede ser utilizada también para la construcción de modelos válidos de covariogramas o variogramas cruzados, teniendo en cuenta que (Yaglom, 1986, p. 418):

$$Cov\left(\int_{\mathbb{R}^d} f_1(\mathbf{x})W(\mathbf{x})d\mathbf{x}, \int_{\mathbb{R}^d} f_2(\mathbf{x} - \mathbf{s})W(\mathbf{x})d\mathbf{x}\right) = \int_{\mathbb{R}^d} f_1(\mathbf{x})f_2(\mathbf{x} - \mathbf{s})W(\mathbf{x})d\mathbf{x},$$

para f_1 y f_2 funciones reales con cuadrado integrable en \mathbb{R}^d y $W(\cdot)$ proceso espacial ruido blanco. Utilizando este resultado Ver Hoef y Barry (1998) propusieron un criterio para la obtención de modelos válidos (incluyendo además de los parámetros correspondientes a los auto-variogramas, parámetros de desplazamiento espacial para modelar posibles asimetrías y parámetros para

controlar el grado de dependencia). Sin embargo las expresiones de los modelos obtenidos (suponiendo que se puede encontrar la solución explícita de las integrales involucradas) pueden ser bastante complicadas, dificultando el uso en la práctica de este tipo de modelos. La ventaja de esta aproximación es que permite obtener modelos válidos por construcción, evitando las dificultades que surgen al utilizar otros modelos debidas a las restricciones sobre los parámetros para garantizar su validez.

5.3.3 Modelado no paramétrico

La mayoría de las aproximaciones flexibles para el modelado de la dependencia espacial univariante descritas en la sección 4.5 se han generalizado también para el caso multivariante. En esta sección se describen brevemente los distintos métodos, tratando con mayor detalle la extensión de la metodología propuesta para el caso univariante por Shapiro y Botha (1991) (mostrada en la sección 4.5.1).

GENERALIZACIÓN DE LA APROXIMACIÓN DE SHAPIRO-BOTHA

Rehman y Shapiro (1996) extendieron la metodología propuesta por Shapiro y Botha (1991) para el modelado de semivariogramas cruzados. La base de esta aproximación es la misma que la del caso univariante y consiste en la discretización las funciones de densidad espectrales. Por ejemplo, en el caso unidimensional ($d = 1$) si se consideran saltos $z_{lm}^j = a_{lm}^j + ib_{lm}^j$ (siendo $i = \sqrt{-1}$) en puntos de discretización $x_j, j = 1, \dots, J$, los modelos asimétricos obtenidos son de la forma:

$$\gamma_{lm}(r) = \nu_{lm} - \sum_{j=1}^J (\cos(x_j r) a_{lm}^j - \text{sen}(x_j r) b_{lm}^j),$$

con $\mathbf{Z}^j = [z_{lm}^j]$ matrices Hermitianas semidefinidas positivas (por tanto $b_{ll}^j = 0$), $j = 1, \dots, J$, y ν_{lm} números reales positivos verificando:

$$\nu_{lm} - \sum_{j=1}^J |z_{lm}^j| \geq 0, \tag{5.15}$$

para $l, m = 1, \dots, k$.

La principal complicación respecto al caso univariante reside en las restricciones que deben verificar los parámetros. En el caso bivalente ($k = 2$) la condición de matrices semidefinidas positivas es equivalente a las restricciones:

$$a_{11}^j \geq 0, a_{22}^j \geq 0, \text{ y } a_{11}^j a_{22}^j \geq (a_{12}^j)^2 + (b_{12}^j)^2, j = 1, \dots, J, \quad (5.16)$$

sin embargo en el caso general multivariante es bastante más difícil chequear dicha condición. Por este motivo cuando $k > 2$ Rehman y Shapiro (1996) propusieron reparametrizar el modelo a partir de las factorizaciones de Cholesky de las matrices \mathbf{Z}^j , aunque los modelos resultantes no son funciones lineales de los parámetros (fijados los ptos. de discretización) y por tanto el ajuste de este tipo de modelos es mucho más complicado que en el caso univariante. Hay que destacar también que incluso en el caso de $k = 2$, el ajuste no puede realizarse mediante programación cuadrática ya que tanto (5.15) como (5.16) son restricciones no lineales.

Estos modelos pueden ser generalizados para el caso multidimensional $d > 1$ aunque las expresiones obtenidas resultan algo más complicadas (de hecho Rehman y Shapiro, 1996, únicamente mostraron las expresiones sin simplificar para el caso $d = 2$ y $d = 3$). Por ejemplo, en el caso de los auto-semivariogramas (anisotrópicos), se obtendrían expresiones equivalentes a las mostradas en la sección 4.5.4. Por este motivo Rehman y Shapiro (1996) consideraron modelos isotrópicos (por tanto simétricos) cuando $d > 1$ (con expresiones casi idénticas a las del caso isotrópico univariante descrito en la sección 4.5.1); para conseguir mayor flexibilidad de estos modelos consideraron además anisotropía geométrica (ver sección 2.2.2) y parámetros de desplazamiento espacial para modelar semivariogramas cruzados asimétricos (sección 5.3.4). Sin embargo, incluso utilizando los modelos isotrópicos más sencillos, surgen el mismo tipo de problemas que en el caso unidimensional (algo menores ya que se elimina la componente imaginaria) debidos a las restricciones sobre los parámetros para garantizar modelos válidos.

OTROS MÉTODOS

En la sección 4.5.2 se describió la aproximación de Yao y Journel (1998) en el caso univariante, aunque de hecho estos autores se centraron en el caso

multivariante. La principal diferencia en el caso general respecto a la metodología descrita en dicha sección, reside a la hora de suavizar la tabla (compleja) de pseudo-densidad espectral correspondiente al covariograma cruzado (obtenida también utilizando la FFT); el suavizado se debe realizar bajo la restricción (5.8) de simetría respecto a la frecuencia cero (para obtener covarianzas no imaginarias) y las restricciones (5.9) para garantizar modelos válidos (aunque sólo en el caso bivariante). Volviendo a transformar a la escala original la tabla de densidad espectral resultante, se obtienen tablas de covarianzas cruzadas (en general asimétricas) que pueden ser utilizadas para la predicción espacial multivariante. Teniendo en cuenta esto, las observaciones realizadas al final de la sección 4.5.2 sobre esta metodología son aplicables también al caso multivariante.

Se pueden obtener también modelos flexibles de semivariogramas cruzados utilizando la representación en medias móviles. Ver Hoef et al. (2003) propusieron utilizar las funciones de medias móviles consideradas por Barry y Ver Hoef (1996) para el caso univariante (ver sección 4.5.3) para obtener modelos de semivariogramas cruzados empleando la metodología sugerida por Ver Hoef y Barry¹¹ (1998) (ver sección anterior). Naturalmente los problemas computacionales descritos en la sección 4.5.3 para el caso univariante, se incrementan al considerar múltiples variables.

5.3.4 Modelado de asimetría

Utilizando la mayoría de las aproximaciones descritas anteriormente se obtienen inicialmente modelos simétricos de semivariogramas o covariogramas cruzados (salvo con los métodos propuestos por Rehman y Shapiro, 1996, y Yao y Journel, 1998). Sin embargo a partir de modelos simétricos se pueden obtener otros asimétricos introduciendo un desfase espacial (“lag effect”); por ejemplo Journel y Huijbregts (1978, p. 173) propusieron, para el caso del modelo lineal de co-regionalización, desplazar los procesos $Z_i(\mathbf{s})$ considerando:

¹¹ Aunque en lugar de utilizar el criterio de ajuste m.c.p., Ver Hoef et al. (2003) propusieron la estimación de los parámetros utilizando máxima verosimilitud restringida (ver sección 4.3.2).

$$Z_i^*(\mathbf{s}) = Z_i(\mathbf{s} - \mathbf{\Delta}_i) = \sum_{l=0}^L \sum_{m=1}^{M_l} a_{im}^l Y_m^l(\mathbf{s} - \mathbf{\Delta}_i) + \mu_i,$$

donde $\mathbf{\Delta}_i \in \mathbb{R}^d$, $i = 1, \dots, k$, son los parámetros de desplazamiento espacial (normalmente se suele fijar $\mathbf{\Delta}_1 = \mathbf{0}$). De esta forma se obtienen los modelos lineales de co-regionalización asimétricos:

$$C_{ij}^*(\mathbf{h}) = C_{ij}(\mathbf{h} + \mathbf{\Delta}_i - \mathbf{\Delta}_j) = \sum_{l=0}^L \sum_{m=1}^{M_l} a_{im}^l a_{jm}^l C_l(\mathbf{h} + \mathbf{\Delta}_i - \mathbf{\Delta}_j).$$

Esta aproximación se ha considerado también para el caso de modelos de semivariogramas cruzados; si $\gamma_{ij}(\mathbf{h})$, $i, j = 1, \dots, k$, son modelos válidos de semivariogramas cruzados simétricos, entonces:

$$\gamma_{ij}^*(\mathbf{h}) = \gamma_{ij}(\mathbf{h} + \mathbf{\Delta}_i - \mathbf{\Delta}_j), \quad i, j = 1, \dots, k$$

son semivariogramas cruzados asimétricos también válidos.

5.3.5 Observaciones

El modelado de la dependencia espacial en el caso multivariante presenta muchas más dificultades que en el caso univariante. Es necesario el modelado conjunto de $k(k+1)/2$ semivariogramas o covariogramas cruzados y cuando $k > 2$ se complica significativamente el problema debido a las condiciones necesarias para garantizar modelos válidos.

En el caso general de modelos asimétricos aparecen complicaciones relacionadas con el número de dimensiones similares a considerar modelos no isotrópicos en el caso univariante (ver p.e. la sección 4.5.4). Se pueden evitar estos inconvenientes considerando modelos isotrópicos y parámetros de desplazamiento espacial (aunque no resulta claro el papel de estos parámetros en el modelado de la asimetría), sin embargo persisten las mismas dificultades para garantizar modelos válidos cuando el número de variables es grande (mayor que dos).

El ajuste de forma conjunta o por separado de los semivariogramas cruzados es también un tema abierto. Si se da preferencia al ajuste de los auto-semivariogramas como proponen algunos autores y son ajustados por separado en un primer paso, las restricciones para garantizar modelos válidos pueden

producir que el ajuste posterior de los semivariogramas cruzados sea poco satisfactorio (incluso utilizando modelos flexibles). Teniendo en cuenta las observaciones realizadas en la sección 5.2, como es de esperar una menor influencia de la información secundaria, debería ser prioritario obtener un buen ajuste del auto-variograma de la variable principal.

Como conclusión podríamos decir que uno de los temas más interesantes de la geoestadística espacial multivariante que requieren de mayor estudio, es justamente la construcción de modelos flexibles de (co)variogramas cruzados y de fácil implementación (ajuste) en la práctica.

PARTE II

GEOESTADÍSTICA

ESPACIO-TEMPORAL

CAPÍTULO 6

MODELOS ESPACIO-TEMPORALES

El modelado de procesos espacio-temporales es un problema crucial en muchas disciplinas (ver sección 1.1). En estos campos, el uso de modelos geoestadísticos espacio-temporales se ha generalizado en los últimos años (ver p.e. Kyriakidis y Journel, 1999). Sin embargo, muchas de las aproximaciones empleadas inicialmente se basan en modelos originalmente desarrollados para el caso puramente espacial o temporal, y consideran la dependencia espacio-temporal de forma separada o no la explotan adecuadamente. Por este motivo se han realizado recientemente múltiples propuestas para el modelado de la dependencia espacio-temporal. La demanda de este tipo de modelos continúa siendo creciente, por lo que es de esperar que también se produzca un desarrollo importante en este campo durante los próximos años.

En este capítulo nos centraremos en la extensión de los modelos geoestadísticos espaciales para el caso de la inclusión de una dimensión temporal adicional. En la sección 6.1 se hace una pequeña revisión de las aproximaciones disponibles actualmente. En la sección 6.2 se presentan algunos de los modelos paramétricos de (co)variogramas no separables más conocidos. Al final de este capítulo, en la sección 6.3, se generaliza la aproximación de Shapiro y Botha (1991) para la obtención de familias flexibles de modelos válidos de variogramas estacionarios espacio-temporales.

6.1 INTRODUCCIÓN

En el pasado se han considerado numerosas alternativas para el modelado de procesos espacio-temporales (ver p.e. Kyriakidis y Journel, 1999). Podríamos clasificar las distintas formas de proceder en dos grandes grupos dependiendo del

punto de vista inicial: modelos considerando procesos multivariantes espaciales o temporales (sección 6.1.1) y modelos considerando un proceso espacio-temporal univariante (sección 6.1.2). Recientemente parece ser que se ha optado por la segunda de las opciones (y es la única que va a ser considerada en los siguientes capítulos), principalmente debido al mayor número de modelos de variogramas espacio-temporales disponibles gracias a los últimos trabajos realizados en este campo (ver secciones 6.2 y 6.3). Aunque en teoría ambos puntos de vista tienen bastante en común, existen importantes diferencias a la hora de aplicarlos en la práctica; en la sección 6.1.3 se comparan ambas aproximaciones.

Supongamos por el momento que el proceso espacio-temporal $Z(\mathbf{s}, t)$ admite una descomposición de la forma:

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \delta(\mathbf{s}, t), \quad (6.1)$$

siendo $\mu(\mathbf{s}, t)$ la función determinística¹ de tendencia y $\delta(\mathbf{s}, t)$ un proceso espacial de media cero y con variograma no necesariamente estacionario:

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

y en ocasiones además se supondrá que este proceso tiene también covariograma:

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

En este trabajo nos centraremos sin embargo en el caso de variogramas y covariogramas estacionarios (i.e. funciones de $\mathbf{s}_1 - \mathbf{s}_2$ y $t_1 - t_2$); ya que en la práctica normalmente sólo se dispondrá de una única realización parcial del proceso dificultando la inferencia sobre variogramas o covariogramas no estacionarios. Adicionalmente, ver comentarios al final de la sección 6.1.2, los modelos estacionarios pueden ser utilizados como punto de partida para el modelado de procesos no estacionarios.

¹ También han sido consideradas, especialmente en el campo espacio-temporal, tendencias aleatorias (ver p.e. Kyriakidis y Journel, 1999, pp. 661-663; Chilès y Delfiner, 1999, sección 5.8.4); este caso sin embargo no va a ser considerado aquí, aunque la forma de proceder es similar (la única diferencia está a la hora de eliminar la tendencia, utilizándose un método kriging en el caso aleatorio; ver p.e. Wackernagel, 1998, pp. 212-213).

6.1.1 Aproximación espacial multivariante

Como ya se comentó anteriormente, en algunas ocasiones se ha utilizado la metodología espacial multivariante descrita en el capítulo 5 para el modelado de la dependencia espacio-temporal. Este tipo de aproximaciones han sido consideradas principalmente en los primeros trabajos en los que comenzó a ser necesario el modelado de procesos espacio-temporales y resultaban poco satisfactorios o adecuados los modelos de (co)variogramas espacio-temporales disponibles en aquel momento.

Dependiendo de la configuración de los datos se suelen considerar dos alternativas. Si el muestreo espacial es más denso que el temporal y se observa el proceso en t_1, \dots, t_T instantes temporales (y para cada instante t_i se dispone de suficientes observaciones espaciales), se suele considerar el proceso espacial multivariante:

$$\mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_T(\mathbf{s}))', \quad (6.2)$$

siendo $Z_i(\mathbf{s}) = Z(\mathbf{s}, t_i)$, $i = 1, \dots, T$ (e.g. Egbert y Lettenmaier, 1986). Mientras que si el muestreo de las observaciones es más denso en el tiempo, incluyendo un número menor de posiciones espaciales $\mathbf{s}_1, \dots, \mathbf{s}_N$, se suele utilizar el proceso temporal (espacial unidimensional) multivariante:

$$\mathbf{Z}(t) = (Z_1(t), \dots, Z_N(t))', \quad (6.3)$$

con $Z_j(t) = Z(\mathbf{s}_j, t)$, $j = 1, \dots, N$ (e.g. Solow y Gorelick, 1986). Por tanto, en cualquiera de los casos, debe haber cierta clase de regularidad (temporal o espacial) en el conjunto de observaciones; por ejemplo, si el proceso es observado siempre en distintas posiciones espaciales y temporales no podríamos utilizar ninguna de estas alternativas.

La representación elegida limita además las posiciones espacio-temporales en las que se pueden obtener predicciones sin modelado adicional² (utilizando la metodología espacial multivariante habitual descrita en el capítulo 5). Considerando el proceso espacial multivariante (6.2) y utilizando alguno de los

² En ocasiones, modelos discretos son extendidos al caso continuo interpolando (mediante algún método kriging por ejemplo) los valores obtenidos de los parámetros en los puntos de discretización.

métodos cokriging, sólo se pueden obtener predicciones en posiciones (\mathbf{s}, t) con $t \in \{t_1, \dots, t_T\}$, mientras que si se consideran modelos del tipo (6.3) solo se puede predecir en (\mathbf{s}, t) con $\mathbf{s} \in \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$.

Si se consideran aproximaciones del tipo (6.2) es necesario el modelado de $T(T+1)/2$ auto-variogramas y variogramas cruzados (o covariogramas); lo cual puede resultar bastante complicado teniendo en cuenta las observaciones realizadas en el capítulo 5 (ver p.e. sección 5.3.5). A partir de la descomposición (6.1), se obtiene que:

$$\gamma_{ij}(\mathbf{h}) = \gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}, t_i, t_j) \equiv \gamma(\mathbf{h}, t_i, t_j),$$

y por tanto equivale a suponer que el semivariograma del proceso espacio-temporal es estacionario en el espacio (en cuyo caso diremos que el semivariograma es espacial-estacionario), aunque no es necesario que haya estacionariedad temporal; esta será la principal ventaja de este tipo de modelos respecto a las aproximaciones espacio-temporales univariantes habitualmente consideradas. Las mismas observaciones sirven también para el caso de considerar procesos temporales multivariantes, donde habrá que modelar $N(N+1)/2$ variogramas cruzados y no es necesario que exista estacionariedad espacial.

6.1.2 Aproximación espacio-temporal univariante

En este caso normalmente se considera el proceso espacio-temporal $Z(\mathbf{s}, t)$ admite una descomposición del tipo (6.1), aunque suponiendo que el proceso $\delta(\mathbf{s}, t)$ verifica algún tipo de estacionariedad. Usualmente se considera que este proceso es intrínsecamente estacionario con semivariograma:

$$\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}, t, t + u) \equiv \gamma(\mathbf{h}, u),$$

o estacionario de segundo orden con covariograma:

$$C(\mathbf{s}, \mathbf{s} + \mathbf{h}, t, t + u) \equiv C(\mathbf{h}, u).$$

La hipótesis de isotropía, tradicionalmente considerada en el caso puramente espacial, es claramente poco adecuada para el caso espacio-temporal (implicaría suponer que un salto espacial es igual a un salto temporal); por este

motivo se han considerado numerosas alternativas para la obtención de modelos válidos de semivariogramas espacio-temporales:

1. Modelos de variogramas geoméricamente anisotrópicos (e.g. Dimitrakopoulos y Lou, 1994). En ocasiones se supone que:

$$\gamma(\mathbf{h}, u) = \gamma^0\left(\sqrt{a_1 \|\mathbf{h}\|^2 + a_2 u^2}\right),$$

siendo $\gamma^0(\cdot)$ un semivariograma isotrópico (p.e. alguno de los mostrados en la sección 4.2); el caso general se tiene en la sección 2.2.2. Sin embargo esta hipótesis puede ser también poco adecuada para el caso espacio-temporal, principalmente debido a que implica asumir que la forma del semivariograma es la misma en el espacio y en el tiempo. Un ejemplo de semivariograma de este tipo, (que será utilizado en capítulos siguientes), es el siguiente modelo de semivariograma esférico anisotrópico (SVESFA):

$$\gamma(\mathbf{h}, u | \boldsymbol{\theta}) = \gamma^0(\sqrt{\|\mathbf{h}\|^2 + bu^2} | \tilde{\boldsymbol{\theta}}),$$

$$\gamma^0(h | \tilde{\boldsymbol{\theta}}) = \begin{cases} 0 & \text{si } h = 0 \\ c_0 + \sigma^2 \left\{ \frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\} & \text{si } 0 < h \leq a \\ c_0 + \sigma^2 & \text{si } h > a \end{cases} \quad (6.4)$$

siendo $\boldsymbol{\theta} = (c_0, a, b, \sigma^2)'$, con $c_0 \geq 0$ (efecto nugget), $a \geq 0$ (rango semivariograma esférico), $b \geq 0$ (parámetro de interacción espacio-temporal) y $\sigma^2 > 0$ (umbral parcial). En la figura 6.1(a) se muestra un ejemplo de la forma de este modelo.

2. Modelos de covariogramas separables (e.g. Guttorp et al., 1992; sección 2.2.2), también denominados covariogramas producto o factorizables:

$$C(\mathbf{h}, u) = C_s(\mathbf{h})C_t(u), \quad (6.5)$$

considerando por tanto que la dependencia espacial es independiente de la dependencia temporal³. Lo que equivale a suponer que el proceso $\delta(\mathbf{s}, t)$ admite la descomposición:

³ Nótese que (6.5) implica que, fijados \mathbf{h}_1 y \mathbf{h}_2 , $C(\mathbf{h}_1, \cdot)$ es proporcional a $C(\mathbf{h}_2, \cdot)$ (i.e. las covarianzas cruzadas de las series temporales correspondientes a cualquier par de posiciones espaciales siempre tienen la misma forma). Análogamente, si se fijan dos

$$\delta(\mathbf{s}, t) = \delta_1(\mathbf{s})\delta_2(t)$$

siendo $\delta_1(\mathbf{s})$ y $\delta_2(t)$ procesos independientes⁴. A partir de (6.5), teniendo en cuenta la relación entre el covariograma y el variograma, se deduce que:

$$\gamma(\mathbf{h}, u) = \gamma_s(\mathbf{h})C_t(0) + C_s(\mathbf{0})\gamma_t(u) - \gamma_s(\mathbf{h})\gamma_t(u).$$

Un semivariograma de este tipo (que será utilizado también en capítulos siguientes) es el modelo de semivariograma correspondiente a un covariograma exponencial separable (SVEXPS):

$$\gamma(\mathbf{h}, u | \boldsymbol{\theta}) = \begin{cases} c_0 + \sigma^2 \left(1 - \exp \left\{ -\frac{u}{a} - \frac{\|\mathbf{h}\|}{b} \right\} \right) & \text{si } \mathbf{h} \neq \mathbf{0} \text{ o } u \neq 0 \\ 0 & \text{si } \mathbf{h} = \mathbf{0} \text{ y } u = 0 \end{cases} \quad (6.6)$$

siendo $\boldsymbol{\theta} = (c_0, a, b, \sigma^2)'$, con $c_0 \geq 0$ (efecto nugget), $a \geq 0$ (parámetro de escala temporal), $b \geq 0$ (parámetro de escala espacial) y $\sigma^2 > 0$ (umbral parcial). En la figura 6.1(b) se tiene un ejemplo del aspecto que presentan este tipo de modelos.

3. Modelos de (co)variogramas anidados (e.g. Rouhani y Hall, 1989), también denominados variogramas suma o con anisotropía zonal (sección 2.2.2):

$$\gamma(\mathbf{h}, u) = \gamma_s(\mathbf{h}) + \gamma_t(u),$$

considerando también de forma independiente la dependencia espacial y temporal. En el caso de estacionariedad de segundo orden se tendría que:

$$C(\mathbf{h}, u) = C_s(\mathbf{h}) + C_t(u),$$

lo que equivale a suponer que el proceso $\delta(\mathbf{s}, t)$ admite la descomposición:

$$\delta(\mathbf{s}, t) = \delta_1(\mathbf{s}) + \delta_2(t).$$

siendo $\delta_1(\mathbf{s})$ y $\delta_2(t)$ procesos independientes (por tanto también podríamos clasificar estos modelos como separables).

saltos temporales, la forma del covariograma (como función del salto espacial) es la misma.

⁴ También se han considerados algunas variaciones sobre los modelos de este tipo, por ejemplo considerando procesos tipificados (Haas, 1995).

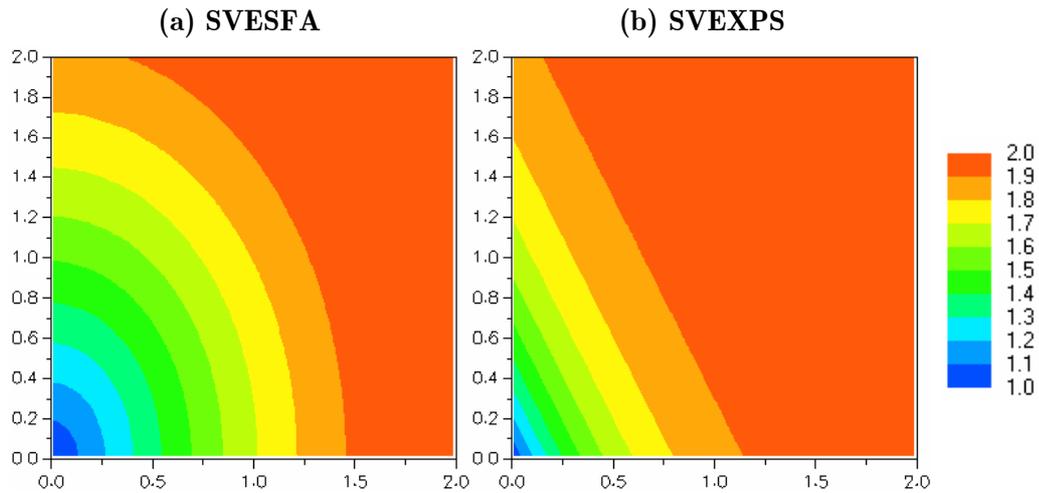


Figura 6.1. Gráfico de contorno de los modelos: (a) SVESFA (ec. (6.4) con $c_0 = 1$, $a = 2$, $b = 0.5$ y $\sigma^2 = 1$) y (b) SVEXPS (ec. (6.6) con $c_0 = 1$, $a = 0.5$, $b = 0.25$ y $\sigma^2 = 1$).

4. Modelos de variogramas espaciales no estacionarios (e.g. Sampson y Guttorp, 1992; ver también sección 2.2.2):

$$\gamma(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) \equiv \gamma(\mathbf{s}_1, \mathbf{s}_2).$$

En este caso se supone que la tendencia captura totalmente la variabilidad temporal y que el proceso $\delta(\mathbf{s}, t) \equiv \delta(\mathbf{s})$ es puramente espacial; de esta forma las observaciones temporales se corresponden con T replicas independientes de este proceso y es más fácil realizar inferencias sobre variogramas no estacionarios. Una revisión bastante detallada de este tipo de modelos se tiene en Guttorp y Sampson (1994).

5. Modelos de (co)variogramas espacio-temporales estacionarios no separables (e.g. Jones y Zhang, 1997; Cressie y Huang, 1999; Christakos, 2000; De Cesare et al., 2001; De Iaco et al., 2001, 2002a; Gneiting, 2002; Ma, 2002; Stein, 2003). Como los modelos anteriores (basados en la metodología espacial tradicional) pueden ser en ciertas ocasiones poco adecuados para el caso espacio-temporal (ver p.e. Journel, 1986), se ha realizado recientemente un esfuerzo considerable para la obtención de nuevos modelos de dependencia que soporten una posible interacción espacio-temporal (en la sección 6.2 se hace una pequeña revisión de los

modelos paramétricos más conocidos). La mayoría de estos modelos suponen isotropía en la componente espacial y temporal, i.e.:

$$\gamma(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) \equiv \gamma(\|\mathbf{s}_1 - \mathbf{s}_2\|, |t_1 - t_2|),$$

aunque en ocasiones pueden ser deseables modelos no simétricos (ver comentarios al final de la sección 6.2).

6. Modelos de (co)variogramas espacio-temporales no estacionarios (e.g. Christakos, 2000; Fuentes y Smith, 2001; Fuentes, 2002; Ma, 2002, 2003). Muy recientemente también se han comenzado a desarrollar modelos no estacionarios que pueden ser utilizados en el caso espacio-temporal. Aunque todavía es necesario un mayor estudio que permita la obtención de nuevos modelos y procedimientos para su uso en la práctica, es de esperar un desarrollo creciente de este tipo de modelos en los próximos años.

Como ya se manifestó anteriormente, en este trabajo nos centraremos únicamente en el caso de procesos estacionarios. Los modelos estacionarios pueden ser utilizados también para el caso de procesos no estacionarios, por ejemplo procediendo de forma local. Estas aproximaciones, que denominaremos *tipo ventana móvil* (e.g. Haas, 1990, 1995), se basan en la idea de que es razonable suponer un modelo homogéneo en el entorno (ventana) de una posición espacial. Además, los modelos de variogramas espacio-temporales estacionarios pueden ser utilizados para la obtención de modelos no estacionarios. De hecho es así como han sido obtenidos la mayoría de los modelos disponibles en la actualidad. Por ejemplo Fuentes y Smith (2001) y Fuentes (2002), basándose en la idea anterior de estacionariedad local, obtuvieron modelos no estacionarios a partir de convoluciones de procesos localmente estacionarios. Un resultado interesante, debido a Ma (2003), es que si $\gamma(\mathbf{h}, u)$ es un variograma espacio-temporal estacionario entonces:

$$C_1(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = \gamma(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2),$$

$$C_2(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = \gamma(\mathbf{s}_1, t_1) + \gamma(\mathbf{s}_2, t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2),$$

$$C_3(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) = 2\{\gamma(\mathbf{s}_1, t_1) + \gamma(\mathbf{s}_2, t_2)\} - \gamma(\mathbf{s}_1 + \mathbf{s}_2, t_1 + t_2) - \gamma(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2),$$

son covariogramas espacio-temporales no estacionarios válidos.

6.1.3 Observaciones

Como cabía esperar existen bastantes puntos en común entre las aproximaciones espaciales multivariantes (sección 6.1.1) y la aproximación espacio-temporal univariante (sección 6.1.2). Si el objetivo final es la predicción, puede verse que los métodos kriging en el contexto espacio-temporal son equivalentes a utilizar métodos cokriging con cualquiera de las aproximaciones multivariantes. Por ejemplo, si suponemos en la descomposición (6.1) que la tendencia es conocida (y que el covariograma existe), puede probarse fácilmente (e.g. Kyriakidis y Journel, 1999, pp. 673-675) que el predictor del kriging simple espacio-temporal (ver sección 3.2) coincide con el predictor cokriging simple (sección 5.2.1) obtenido considerando el proceso espacial multivariante (6.2); suponiendo estacionariedad, simplemente hay que tener en cuenta la relación:

$$C_{ij}(\mathbf{h}) = C(\mathbf{h}, t_i - t_j).$$

Si se consideran vectores de procesos temporales del tipo (6.3) se obtiene también mismo resultado, ya que en este caso:

$$C_{ij}(u) = C(\mathbf{s}_i - \mathbf{s}_j, u).$$

En el caso de media desconocida hay que tener cuidado sin embargo con las hipótesis que asumen sobre ella los métodos utilizados. Por ejemplo, en el caso de media constante, el predictor del kriging ordinario (espacio-temporal) no va a coincidir con el predictor cokriging ordinario tradicional ya que al utilizar este último se asume que las medias de los componentes del proceso multivariante son distintas (dando lugar por tanto a un predictor no óptimo).

En principal inconveniente de la aproximación espacial multivariante, además de los descritos en la sección 6.1.1, es la dificultad para la obtención de modelos válidos de dependencia. Teniendo en cuenta las observaciones realizadas en el capítulo 5 (ver p.e. la sección 5.3.5), el problema de incluir una dimensión adicional es menor que los debidos a considerar múltiples variables. Este es posiblemente el principal motivo que justifica que en los últimos años se haya optado aparentemente por la aproximación espacio-temporal (única considerada en el resto de este trabajo). Adicionalmente, en ocasiones puede ser de interés el modelado de procesos espacio-temporales multivariantes. Para ello podemos recurrir a modelos de variogramas espacio-temporales junto con alguna

de las herramientas descritas en la sección 5.3, utilizando por ejemplo el modelo lineal de co-regionalización (e.g. De Iaco et al., 2003). Está claro que utilizar en este caso la aproximación espacial multivariante presentará muchas más complicaciones.

6.2 MODELOS PARAMÉTRICOS DE SEMIVARIOGRAMAS NO SEPARABLES

Se han utilizado una gran variedad de métodos para la obtención de modelos de variogramas espacio-temporales no separables, por ejemplo a partir de ecuaciones diferenciales estocásticas (e.g. Jones y Zhang, 1997 y Christakos, 2000), utilizando como base modelos separables (e.g. De Cesare et al., 2001; De Iaco et al., 2001, 2002a y Ma, 2002) o a partir de la representación espectral (e.g. Cressie y Huang, 1999; Gneiting, 2002 y Stein, 2003).

El principal inconveniente de los métodos basados en ecuaciones diferenciales estocásticas es que en la mayoría de las ocasiones no se pueden obtener expresiones explícitas de los modelos de covariogramas, por lo que su uso en la práctica presenta más dificultades.

Una extensión de los modelos separables, propuesta inicialmente por De Cesare et al. (2001), es el denominado *modelo de covariograma suma-producto*:

$$C(\mathbf{h}, u) = k_1 C_s(\mathbf{h}) C_t(u) + k_2 C_s(\mathbf{h}) + k_3 C_t(u),$$

donde $C_s(\cdot)$ y $C_t(\cdot)$ son covariogramas válidos y $k_1 > 0$, $k_2 \geq 0$, $k_3 \geq 0$ para garantizar su validez. De Iaco et al. (2001) propusieron la siguiente generalización de este modelo:

$$\gamma(\mathbf{h}, u) = \gamma_s(\mathbf{h}) + \gamma_t(u) - k \gamma_s(\mathbf{h}) \gamma_t(u),$$

que denominaron *modelo suma-producto generalizado*, donde $\gamma_s(\cdot)$ y $\gamma_t(\cdot)$ son variogramas acotados válidos y:

$$0 < k \leq 1 / \max \{ \text{umbral}(\gamma_s(\mathbf{h})), \text{umbral}(\gamma_t(u)) \}.$$

Una ventaja sugerida por los autores es que el ajuste de estos modelos puede llevarse a cabo en la práctica a partir de los semivariogramas marginales $\gamma(\mathbf{h}, 0)$ y $\gamma(\mathbf{0}, u)$ (procediendo de forma totalmente análoga al caso puramente espacial). Sin

embargo esto resulta en una severa restricción a la hora de modelar posibles interacciones espacio-temporales, por lo que resulta poco recomendable esta forma proceder en la práctica (de hecho es de esperar que estos modelos no sean muy adecuados para modelar interacciones de este tipo). Otros modelos paramétricos no separables obtenidos a partir de covariogramas separables, son por ejemplo los propuestos por Ma (2002):

$$\rho(\mathbf{h}, u) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \rho_s^i(\mathbf{h}) \rho_t^j(u) p_{ij},$$

siendo $\rho_s^i(\cdot)$ y $\rho_t^j(\cdot)$ correlogramas válidos para $i, j = 1, 2, \dots$ y $p_{ij} \geq 0$ tales que:

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_{ij} = 1.$$

También se han derivado modelos no separables aplicando la propiedad (2.8) a modelos separables (e.g. De Iaco et al., 2002a y Ma, 2002) o al modelo suma-producto (De Iaco et al., 2002a).

Cressie y Huang (1999) obtuvieron varias familias de covariogramas no separables a partir de la representación espectral:

$$C(\mathbf{h}, u) = \int_{\mathbb{R}^d} e^{i(\boldsymbol{\omega} \cdot \mathbf{h})} \rho(\boldsymbol{\omega}, u) k(\boldsymbol{\omega}) d\boldsymbol{\omega},$$

siendo $\rho(\boldsymbol{\omega}, \cdot)$ un correlograma continuo $\forall \boldsymbol{\omega} \in \mathbb{R}^d$, verificando:

$$\int \rho(\boldsymbol{\omega}, u) du < \infty,$$

$$\text{y } k(\boldsymbol{\omega}) > 0 \text{ con } \int_{\mathbb{R}^d} k(\boldsymbol{\omega}) d\boldsymbol{\omega} < \infty.$$

Esta aproximación fue generalizada posteriormente por Gneiting (2002), obteniendo modelos de la forma:

$$C(\mathbf{h}, u) = \frac{\sigma^2}{\psi(|u|^2)^{d/2}} \varphi\left(\frac{\|\mathbf{h}\|^2}{\psi(|u|^2)}\right),$$

siendo $\varphi(t), t \geq 0$, una función completamente monótona y $\psi(t), t \geq 0$, una función positiva con derivada completamente monótona⁵ (en Gneiting, 2002, p. 591, se tienen algunos ejemplos de este tipo de funciones). Recordando que una

⁵ Aunque estos modelos son no separables, la función $\varphi(\cdot)$ está asociada a la componente espacial y $\psi(\cdot)$ a la temporal.

función $\varphi(t)$, $t \geq 0$, es completamente monótona si y sólo si es continua en 0 y verifica:

$$(-1)^n \varphi^{(n)}(t) \geq 0, \forall t > 0, (n = 0, 1, 2, \dots).$$

Por ejemplo, algunos de los modelos obtenidos utilizando esta aproximación (y que serán utilizados en los próximos capítulos) son los siguientes:

- Modelo de semivariograma (*SVCH2*) correspondiente al ejemplo 2 de Cressie y Huang (1999, p. 1333):

$$\gamma(\mathbf{h}, u | \boldsymbol{\theta}) = \begin{cases} c_0 + \sigma^2 \left(1 - \frac{1}{(a|u| + 1)^{d/2}} \exp \left\{ -\frac{b^2 \|\mathbf{h}\|^2}{a|u| + 1} \right\} \right) & \text{si } \mathbf{h} \neq \mathbf{0} \text{ o } u \neq 0 \\ 0 & \text{si } \mathbf{h} = \mathbf{0} \text{ y } u = 0. \end{cases} \quad (6.7)$$

- Modelo de semivariograma (*SVCH4*) correspondiente al ejemplo 4 de Cressie y Huang (1999, p. 1333):

$$\gamma(\mathbf{h}, u | \boldsymbol{\theta}) = \begin{cases} c_0 + \sigma^2 \left(1 - \frac{a|u| + 1}{\{(a|u| + 1)^2 + b^2 \|\mathbf{h}\|^2\}^{(d+1)/2}} \right) & \text{si } \mathbf{h} \neq \mathbf{0} \text{ o } u \neq 0 \\ 0 & \text{si } \mathbf{h} = \mathbf{0} \text{ y } u = 0. \end{cases} \quad (6.8)$$

Donde, en las expresiones anteriores, d es la dimensión espacial y $\boldsymbol{\theta} = (c_0, a, b, \sigma^2)$, con $c_0 \geq 0$ (efecto nugget), $a \geq 0$ (parámetro de escala temporal), $b \geq 0$ (parámetro de escala espacial) y $\sigma^2 > 0$ (umbral parcial). En la figura 6.2 se muestran gráficos de contorno de estos modelos.

Los modelos anteriores de semivariogramas, al igual que la casi totalidad de los modelos paramétricos disponibles en la actualidad, son funciones del salto espacial \mathbf{h} únicamente a través de la distancia $\|\mathbf{h}\|$. La suposición de isotropía en la componente espacial puede ser una propiedad interesante; sin embargo, si un semivariograma espacio-temporal es simétrico en el salto espacial entonces necesariamente también es simétrico en el salto temporal, ya que:

$$\gamma(\mathbf{h}, u) = \gamma(-\mathbf{h}, -u) = \gamma(\mathbf{h}, -u), \quad (6.9)$$

(por tanto un semivariograma espacial-isotrópico es también temporal-isotrópico); Gneiting (2002) denominó a la propiedad (6.9) *simetría espacio-temporal completa*. En algunos casos esta suposición puede ser poco apropiada, por ejemplo cuando hay un flujo temporal en una dirección (ver p.e. Gneiting, 2002, sección 4 ó Stein, 2003, sección 7), y por tanto puede ser deseable disponer

de modelos de semivariogramas espacio-temporales que no posean esta propiedad. Actualmente se han derivado algunos modelos asimétricos a partir de modelos completamente simétricos. Por ejemplo, Stein (2003, sección 5) propuso un método para la obtención de semivariogramas no completamente simétricos utilizando las derivadas de modelos completamente simétricos. Sin embargo, en la mayoría de los casos (e.g. Gneiting, 2002, p. 598; Stein, 2003, sección 7) estos modelos se han obtenido perturbando un modelo espacial-isotrópico, reemplazando en su expresión $\|\mathbf{h}\|$ por $\|\mathbf{h} - \mathbf{v}t\|$, con $\mathbf{v} \in \mathbb{R}^d$ (parámetro que se podría interpretar como el vector de velocidad media del campo). Otra alternativa sería utilizar modelos completamente anisotrópicos (más generales), como por ejemplo los mostrados en la sección 4.5.4. Aunque todavía es un problema abierto el encontrar modelos flexibles más simples que permitan modelar asimetrías de este tipo.

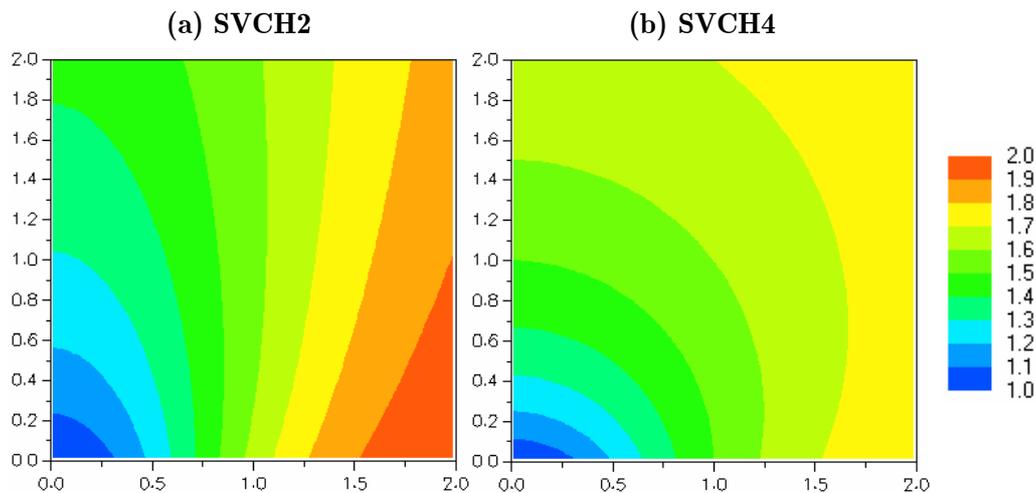


Figura 6.2. Gráfico de contorno de los modelos: (a) SVCH2 (ec. (6.7)) y (b) SVCH4 (ec.(6.8)) con parámetros $d = 1$, $c_0 = 1$, $a = 1$, $b = 1$ y $\sigma^2 = 1$.

6.3 MODELOS FLEXIBLES DE SEMIVARIOGRAMAS ESPACIO-TEMPORALES

Como se mostró en las secciones anteriores, recientemente se han obtenido algunos modelos paramétricos de variogramas estacionarios espacio-temporales no separables. Sin embargo, aunque se incremente el número de

modelos de variogramas espacio-temporales disponibles (algo de esperar en los próximos años), al utilizar en la práctica modelos paramétricos persistirá el problema de una posible mala especificación del modelo. Por tanto, al igual que en el caso espacial (ver sección 4.5), es también deseable disponer de modelos de variogramas espacio-temporales suficientemente flexibles que se puedan adaptar a cualquier situación. Los modelos espaciales completamente anisotrópicos propuestos en la sección 4.5.4 son también adecuados para el caso espacio-temporal; sin embargo, teniendo en cuenta las observaciones realizadas al final de dicha sección, pueden surgir problemas con estos modelos cuando el número de dimensiones es grande. Una forma natural de evitar estos problemas es suponer isotropía en la componente espacial (lo que en cierta forma equivale a reducir la dimensión espacial a uno).

Fernández-Casal et al. (2001, 2003a) generalizaron la aproximación de Shapiro y Botha (1991) para la obtención de familias flexibles de modelos válidos de variogramas estacionarios anisotrópicos en dos componentes. Estos modelos, que pueden ser utilizados para el caso espacio-temporal, se presentan a continuación.

6.3.1 Modelos anisotrópicos en dos componentes

Podemos obtener modelos válidos de semivariogramas espacio-temporales como casos particulares de modelos espaciales anisotrópicos en dos componentes; lo que adicionalmente proporciona mayor variedad de modelos aplicables también al caso espacio-temporal.

Supongamos que $Z(\mathbf{s})$ es un proceso espacial en \mathbb{R}^d estacionario de segundo orden, es decir:

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

$$\text{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{h}),$$

siendo $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$. Supongamos además que, sin ser necesariamente el covariograma isotrópico, hay isotropía dentro de dos componentes del vector de salto:

$$C(\mathbf{h}) = C(\|\mathbf{h}_1\|, \|\mathbf{h}_2\|) = C(r, u), \quad (6.10)$$

siendo $\mathbf{h}_1 \in \mathbb{R}^{d_1}$ un vector formado por d_1 componentes del salto \mathbf{h} y $\mathbf{h}_2 \in \mathbb{R}^{d_2}$ el vector formado por las $d_2 = d - d_1$ componentes restantes. Nótese que este es por ejemplo el caso de un proceso espacio-temporal estacionario con un covariograma espacial-isotrópico/temporal-isotrópico $C(\mathbf{h}, t) \equiv C(\|\mathbf{h}\|, |t|)$.

Si además el covariograma es continuo en el origen, por el teorema de Bochner (Bochner, 1955) el covariograma admite una representación de la forma (2.11), i.e.:

$$C(\mathbf{h}_1, \mathbf{h}_2) = \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_2}} e^{i(\boldsymbol{\omega} \cdot \mathbf{h}_1 + \boldsymbol{\tau} \cdot \mathbf{h}_2)} dF(\boldsymbol{\omega}, \boldsymbol{\tau}) \quad (6.11)$$

donde dF es una medida positiva finita, y siendo el recíproco también cierto. Naturalmente si la función de distribución espectral F es diferenciable, podemos expresar la ecuación anterior en términos de la función de densidad espectral⁶ $f(\boldsymbol{\omega}, \boldsymbol{\tau}) \geq 0$.

De forma análoga al caso de isotropía espacial (sección 2.2.3), podemos obtener la siguiente expresión de la ecuación (6.11) anterior correspondiente a un covariograma anisotrópico en dos componentes del tipo (6.10):

$$C(r, u) = \int_0^\infty \int_0^\infty \kappa_{d_1}(\lambda r) \kappa_{d_2}(vu) dG_d(\lambda, v), \quad (6.12)$$

donde $\kappa_d(\cdot)$ está definida por (2.16) y G_d es una función positiva acotada en $[0, \infty) \times [0, \infty)$ de propiedades similares a las de una función de distribución bidimensional. Esta función es de la forma:

$$G_d(\lambda, v) = \int_{\|\boldsymbol{\omega}\| < \lambda} \int_{\|\boldsymbol{\tau}\| < v} dF(\boldsymbol{\omega}, \boldsymbol{\tau}),$$

siendo dF una medida positiva simétrica. Además, al igual que en el caso espacial isotrópico, se verifica que:

$$C(0, 0) = \int_0^\infty \int_0^\infty dG_d(\lambda, v).$$

Hay que destacar que la gran mayoría de los modelos espacio-temporales disponibles en la actualidad (ver secciones anteriores) son casos particulares de (6.12). Por ejemplo, la suposición de covariogramas separables del tipo

⁶ Además, como en el caso de procesos reales el covariograma es una función par, podemos sustituir en las expresiones anteriores los factores exponenciales por cosenos y la función de densidad espectral es también una función par

$C(\mathbf{h}, t) = C_1(\|\mathbf{h}\|)C_2(|t|)$, frecuentemente utilizados en la práctica para simplificar el problema del modelado de la dependencia espacio-temporal, se corresponden con el caso de $f(\boldsymbol{\omega}, \boldsymbol{\tau}) \equiv f_1(\|\boldsymbol{\omega}\|)f_2(\|\boldsymbol{\tau}\|)$.

Al igual que en el caso espacial, en general las expresiones de la ecuación (6.12) incluirán funciones trigonométricas cuando d_i es impar y funciones de Bessel de orden entero cuando d_i es par. Por ejemplo, para $d_1 = 2$, $d_2 = 1$ y $d_1 = 3$, $d_2 = 1$, obtendríamos respectivamente expresiones de la forma:

$$C(r, u) = \int_0^\infty \int_0^\infty J_0(\lambda r) \cos(vu) dG(\lambda, v),$$

$$C(r, u) = \int_0^\infty \int_0^\infty \frac{\text{sen } \lambda r}{\lambda r} \cos(vu) dG(\lambda, v).$$

Es importante destacar también que un covariograma que admite la representación (6.12) es un covariograma válido en $\mathbb{R}^{d'_1} \times \mathbb{R}^{d'_2}$, $\forall d'_1 \leq d_1$ y $\forall d'_2 \leq d_2$. Además, de modo análogo al caso espacial isotrópico (ver sección 2.2.3), puede verse fácilmente que:

$$C(r, u) = \int_0^\infty \int_0^\infty \kappa_\infty(\lambda r) \kappa_{d_2}(vu) dG(\lambda, v) = \int_0^\infty \int_0^\infty e^{-\lambda^2 r^2} \kappa_{d_2}(vu) dG(\lambda, v),$$

es un covariograma válido para cualquier d_1 , y que:

$$C(r, u) = \int_0^\infty \int_0^\infty \kappa_\infty(\lambda r) \kappa_\infty(vu) dG(\lambda, v) = \int_0^\infty \int_0^\infty e^{-\lambda^2 r^2 - v^2 u^2} dG(\lambda, v),$$

es un covariograma válido para cualesquiera d_1 y d_2 .

A partir de las observaciones anteriores, y teniendo en cuenta la relación (2.20), se deduce que una función es el semivariograma de un proceso espacial anisotrópico en dos componentes si y sólo si admite una representación de la forma:

$$\gamma(r, u) = \begin{cases} \nu_0 - \nu(r, u) & \text{si } r > 0 \text{ o } u > 0 \\ 0 & \text{si } r = u = 0 \end{cases} \quad (6.13)$$

siendo ν_0 una constante positiva y $\nu(r, u)$ una función semidefinida positiva continua en el origen que admite la representación (6.12), y tales que:

$$c_0 = \nu(0, 0) = \nu_0 - \int_0^\infty \int_0^\infty dG(\lambda, v) \geq 0.$$

Utilizando modelos del tipo (6.13), el problema del modelado de la dependencia espacio-temporal de un conjunto de datos se reduce a encontrar

una constante positiva ν_0 y una función G adecuadas. Para facilitar la resolución de este problema podemos considerar una simplificación análoga a la propuesta por Shapiro y Botha (1991) para el caso espacial isotrópico (sección 4.5.1), o a la del caso completamente anisotrópico descrito en la sección 4.5.4, y considerar únicamente medidas atómicas. Es decir, suponer que G es una función con un número finito de saltos positivos z_{ij} en puntos (x_i, y_j) , de la forma:

$$G(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} z_{ij}.$$

Los modelos de semivariogramas (acotados) obtenidos, que denominaremos *modelos de Shapiro-Botha extendidos* (SVSBE), son de la forma:

$$\gamma(r, u) = \nu_0 - \nu(r, u) = \nu_0 - \sum_{i=1}^I \sum_{j=1}^J \kappa_{d_1}(x_i r) \kappa_{d_2}(y_j u) z_{ij}, \quad (6.14)$$

para $(r, u) \neq (0, 0)$, verificando:

$$z_{ij} \geq 0, \quad (6.15)$$

$$\nu_0 - \sum_{i=1}^I \sum_{j=1}^J z_{ij} \geq 0. \quad (6.16)$$

El ajuste de este tipo de modelos se trata en la siguiente sección; donde se muestra que, si se emplea el criterio m.c.p. (ver sección 4.3.1), este problema se puede resolver también fácilmente mediante programación cuadrática. La selección de los puntos de discretización (x_i, y_j) se discutirá en la sección 6.2.3.

6.3.2 Ajuste

Entre los criterios de bondad de ajuste para encontrar un variograma válido que describa adecuadamente la dependencia presente en los datos (sección 4.3), el criterio de mínimos cuadrados ponderados podría ser una buena opción en este caso. La elección de este método estaría justificada principalmente por dos razones: requiere pocas suposiciones acerca de la distribución de los datos (algo en concordancia con la idea de la estimación no paramétrica) y puede llevarse a cabo fácilmente en la práctica.

Utilizando el criterio m.c.p., si:

$$\{\hat{\gamma}_{kl} = \hat{\gamma}(r_k, u_l) : (k, l) \in M\},$$

es el conjunto de estimaciones piloto del semivariograma⁷, donde M representa el conjunto de índices correspondiente a los saltos considerados, el problema del ajuste de un modelo válido se reduce a encontrar el vector $\boldsymbol{\theta} = (z_{11}, \dots, z_{IJ}, \nu_0)'$ de dimensión $I \times J + 1$ que minimiza la función:

$$Q(\boldsymbol{\theta}) = \sum_{(k,l) \in M} w_{kl} \left(\hat{\gamma}_{kl} - \nu_0 + \sum_{i=1}^I \sum_{j=1}^J \kappa_{d_1}(x_i r_k) \kappa_{d_2}(y_j u_l) z_{ij} \right)^2,$$

sujeto a las restricciones lineales (6.15) y (6.16).

Al igual que en el caso isotrópico de Shapiro y Botha (1991), se trata de un problema de programación cuadrática que se resuelve de modo idéntico (p.e. utilizando la rutina QPROG de la librería IMSL). Naturalmente, para este problema de minimización tenga solución única se debe verificar necesariamente que $|M| \geq I \times J + 1$ (donde $|M|$ es el número de saltos utilizados en el ajuste). Si en la elección de los pesos se sigue el criterio propuesto por Cressie (1985) de tomar $w_{kl} = |N(r_k, u_l)| / \gamma(r_k, u_l)^2$, siendo $|N(r_k, u_l)|$ el número de pares utilizados en la estimación $\hat{\gamma}_{kl} = \hat{\gamma}(r_k, u_l)$, habrá que proceder también de forma iterativa, tomando $w_{kl} = 1$ en el primer paso (m.c.o.) y recalculando los pesos en cada iteración hasta convergencia.

Después de resolver este problema, si $\bar{\boldsymbol{\theta}} = (\bar{z}_{11}, \dots, \bar{z}_{IJ}, \bar{\nu}_0)'$ es la solución óptima obtenida, el modelo de semivariograma ajustado vendrá dado por:

$$\bar{\gamma}(r, u) = \bar{\nu}_0 - \sum_{i=1}^I \sum_{j=1}^J \kappa_{d_1}(x_i r) \kappa_{d_2}(y_j u) \bar{z}_{ij},$$

para $(r, u) \neq (0, 0)$; y la correspondiente suma de cuadrados ponderados (MCP) será:

$$\text{MCP} = \sum_{(k,l) \in M} |N(r_k, u_l)| \left(\frac{\hat{\gamma}(r_k, u_l)}{\bar{\gamma}(r_k, u_l)} - 1 \right)^2 / \sum_{(k,l) \in M} |N(r_k, u_l)|, \quad (6.17)$$

(se define de esta forma para poder comparar ajustes a diferentes estimadores piloto).

⁷ El problema de la obtención de un conjunto de estimaciones piloto adecuado se trata en el siguiente capítulo.

Los modelos obtenidos de esta manera son también sumamente flexibles y casi siempre producen buenas aproximaciones a las estimaciones piloto. Esto puede dar lugar a problemas de “sobreajuste” y también sugiere la importancia de utilizar estimaciones piloto eficientes. Para evitar estos problemas, siguiendo la idea de Febrero-Bande et al. (1998), propondremos el uso de un estimador no paramétrico del semivariograma (capítulo 7). Una aproximación alternativa⁸ sería la inclusión de restricciones adicionales (lineales) de suavidad, monotonía o convexidad, aunque esto sería más complicado que en el caso unidimensional (isotrópico) considerado por Shapiro y Botha (1991, pp. 91-94).

Podrían considerarse otros criterios para la estimación de los parámetros de estos modelos. Por ejemplo, el método m.c.g. puede implementarse del mismo modo que el método anterior y podría mejorar significativamente el comportamiento de los modelos ajustados. Sin embargo, para poder utilizar el criterio m.c.g. serían necesarias expresiones que permitirían el cálculo de $Var(\hat{\gamma}(r_k, u_l))$ y $Cov(\hat{\gamma}(r_k, u_l), \hat{\gamma}(r_i, u_j))$. Estas expresiones están disponibles para el estimador empírico (Cressie, 1985), aunque no actualmente para el estimador no paramétrico propuesto en el capítulo siguiente. Otros métodos que requieren de una minimización no lineal multidimensional deberían descartarse por el momento para este tipo de modelos, ya que complicarían significativamente la obtención de una solución óptima. Éste es el caso por ejemplo de los métodos basados en máxima verosimilitud, dónde además es en principio necesaria la suposición de normalidad (difícil de chequear en la práctica).

También es importante señalar, especialmente al utilizar una aproximación no paramétrica, que siempre se debe tener cuidado con las extrapolaciones. Los modelos ajustados no deberían ser evaluados en saltos mucho mayores que los máximos saltos considerados en el ajuste⁹. Por este motivo, se recomienda utilizar kriging local (tipo ventana móvil) para evitar los posibles inconvenientes de la extrapolación.

⁸ También se podría pensar en obtener un ajuste más suave utilizando modelos correspondientes a dimensiones mayores menos flexibles.

⁹ En caso contrario puede ocurrir, por ejemplo, que observaciones muy alejadas de la posición de predicción reciban mayor peso que observaciones cercanas.

6.3.3 Selección de los puntos de discretización

Para la selección de los puntos de discretización (x_i, y_j) podemos seguir el criterio inicial de Shapiro y Botha (1991) y tomar estos puntos regularmente espaciados:

$$\begin{aligned} x_i &= i\phi_1, i = 1, \dots, I, \\ y_j &= j\phi_2, j = 1, \dots, J, \end{aligned} \tag{6.18}$$

siendo ϕ_1 y ϕ_2 dos números positivos. En cuyo caso el problema crucial es la selección de valores apropiados para ϕ_1 y ϕ_2 . Estos pueden tomarse por ejemplo de forma que se minimice la suma de cuadrados ponderados del ajuste (e.g. Fernández-Casal et al., 2001, 2003a; sección 7.1.2), aunque procediendo de esta manera aparecen problemas computacionales cuando el número de puntos de discretización es grande. Por ejemplo, en un ajuste con $I = 13$ y $J = 14$ (183 parámetros), el tiempo de computación del algoritmo iterativo de programación cuadrática resultó de 14.5 segundos en un PIII a 750MHz (con 6 iteraciones m.c.p. y fijando los puntos de discretización (x_i, y_j)). En este caso, si se realizara una búsqueda de ϕ_1 y ϕ_2 en una rejilla 10×10 el tiempo de computación sería de aproximadamente 24 minutos. Además, teniendo en cuenta los comentarios realizados al final de la sección 4.5.1 para el caso espacial isotrópico, la selección de nodos regularmente espaciados podría producir en algunos casos que el semivariograma ajustado presente oscilaciones anómalas¹⁰.

Otra alternativa sería seguir la aproximación sugerida por Genton y Gorsich (2003) para el caso espacial isotrópico (ver sección 4.5.1) y seleccionar los nodos a partir de los ceros de las correspondientes funciones de Bessel. Podemos extender esta idea para el caso espacio-temporal seleccionando los puntos de discretización de la siguiente forma:

- Si $x_i, i = 1, \dots, I$ son los nodos correspondientes a $\kappa_d(\cdot)$, entonces podemos fijar $x_1 = 0$ y $x_i = q_{i-1}/r_{\max}$, $i = 2, \dots, I$, siendo q_i las $I - 1$ primeras raíces de $J_{(d-2)/2}(\cdot)$ y r_{\max} el correspondiente salto máximo. ·

¹⁰ Aunque en el análisis de múltiples ajustes utilizando el criterio anterior (de minimizar el valor m.c.p.) no se observaron oscilaciones de este tipo.

- Si $d = \infty$ podemos tomar $x_i = \phi(i-1)/r_{\max}$, $i = 1, \dots, I$, siendo ϕ un número positivo. Adicionalmente, se observó que en este caso la elección del parámetro ϕ es menos importante¹¹.

Además de las ventajas sugeridas por Genton y Gorschich (2003), este método puede llevarse a cabo fácilmente en la práctica, reduciendo considerablemente el tiempo de computación necesario para realizar el ajuste de un modelo de este tipo a un conjunto de estimaciones piloto (utilizando p.e. el algoritmo desarrollado por Ball (2000) para calcular rápidamente los ceros de las funciones de Bessel). Sin embargo, aunque utilizando este criterio se obtienen en general muy buenos ajustes a las estimaciones piloto (Fernández-Casal et al., 2003b), se observó que el ajuste cerca del origen resulta en ocasiones menos satisfactorio que el obtenido utilizando el criterio anterior (aparentemente este criterio tiende a producir peores estimaciones del efecto nugget). Esta claro que la selección de los puntos de discretización es todavía un tema abierto que requiere de mayor estudio.

¹¹ En todos los casos considerados en capítulos siguientes (en los que se utilizó este criterio) se tomó $\phi = 0.3$.

CAPÍTULO 7

ESTIMACIÓN DE UN SEMIVARIOGRAMA ESPACIO-TEMPORAL

Este capítulo se centra en la estimación del semivariograma de un proceso espacio-temporal estacionario $Z(\mathbf{s}, t)$. El objetivo será por tanto la obtención de un conjunto adecuado de estimaciones piloto:

$$\left\{ \hat{\gamma}_{kl} = \hat{\gamma}(r_k, u_l) : (k, l) \in M \subset \mathbb{Z} \times \mathbb{Z} \right\},$$

a partir de una realización parcial $\{Z(\mathbf{s}_1, t_1), \dots, Z(\mathbf{s}_n, t_n)\}$ del proceso $Z(\mathbf{s}, t)$. Nótese que las posiciones de observación pueden estar regularmente o irregularmente espaciadas (aunque en la práctica suele haber regularidad en las posiciones temporales). El caso de media no constante no va a ser tratado aquí, el procedimiento a seguir sería el mismo que el descrito en la sección 4.1.2. Supondremos además por comodidad que hay isotropía en la componente espacial (y por tanto también en la temporal, ver sección 6.2), i.e.:

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

(el caso general completamente anisotrópico es análogo).

En la sección 7.1 se muestra la expresión del estimador empírico del semivariograma para este caso y se incluye un pequeño estudio de simulación como ejemplo de los problemas que puede aparecer al utilizar este estimador. En la sección 7.2 se presenta, entre otros estimadores no paramétricos, el estimador lineal local del semivariograma que será utilizado en capítulos siguientes.

7.1 ESTIMADOR EMPÍRICO

Como ya se comentó en la sección 4.1.1, utilizando el método de los momentos se obtiene el estimador empírico (o clásico) del semivariograma (4.3) (Matheron, 1962). En el caso de datos regularmente espaciados, utilizando una notación más explícita, este estimador tendría la siguiente expresión:

$$\hat{\gamma}(r_k, u_l) = \frac{1}{2|N(r_k, u_l)|} \sum_{N(r_k, u_l)} (Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2, \quad (7.1)$$

siendo:

$$N(r_k, u_l) = \{(i, j) : \|\mathbf{s}_i - \mathbf{s}_j\| = r_k, |t_i - t_j| = u_l\}. \quad (7.2)$$

En este caso además, se suele considerar un conjunto de índices de la forma:

$$M = \{(k, l) : 0 \leq k \leq K, 0 \leq l \leq L, (k, l) \neq (0, 0)\}, \quad (7.3)$$

(ver p.e. Fernández-Casal et al., 2003a). Nótese que el semivariograma en el origen es siempre nulo, por tanto no tiene sentido estimarlo en dicho salto.

Si las posiciones espaciales de los datos no son regulares se utiliza una versión suavizada de este estimador, utilizando en lugar de (7.2):

$$N(\text{Tol}(r_k), u_l) = \{(i, j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in \text{Tol}(r_k), |t_i - t_j| = u_l\},$$

siendo $\text{Tol}(r_k)$ una región de tolerancia en torno a r_k (realmente se suele proceder de esta forma incluso cuando hay regularidad en las posiciones espaciales). Si las coordenadas temporales de las observaciones tampoco están regularmente espaciadas, se procede de forma análoga considerando regiones de tolerancia en torno a los saltos temporales. En estos casos, el conjunto de índices suele ser de la forma:

$$M = \{(k, l) : 1 \leq k \leq K, 1 \leq l \leq L\},$$

(e.g. Fernández-Casal et al., 2003b).

De forma totalmente análoga se obtienen las expresiones de cualquiera de las versiones del estimador (7.1); por ejemplo, el estimador robusto propuesto por Hawkins y Cressie (1984):

$$2\bar{\gamma}(r_k, u_l) = \left(\frac{1}{|N(r_k, u_l)|} \sum_{N(r_k, u_l)} |Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)|^{\frac{1}{2}} \right)^4 \left/ \left(0.457 + \frac{0.494}{|N(r_k, u_l)|} + \frac{0.045}{|N(r_k, u_l)|^2} \right) \right.$$

7.1.1 Selección de los saltos y las regiones de tolerancia

Un problema todavía abierto es la selección de los saltos y las regiones de tolerancia adecuadas al utilizar este tipo de estimadores. Normalmente los saltos se suelen tomar regularmente espaciados:

$$\begin{aligned} r_k &= k\Delta_S, \\ u_l &= l\Delta_T, \end{aligned} \tag{7.4}$$

siendo Δ_S y Δ_T dos números positivos (con $\Delta_T = 1$ en la mayoría de los casos), y considerando regiones de tolerancia de la forma:

$$Tol(r_k) = [(k - \frac{1}{2})\Delta_S, (k + \frac{1}{2})\Delta_S).$$

Hay que tener en cuenta que cuanto más grande es el valor del variograma teórico mayor varianza tiene el estimador empírico (ver ec. (4.20)). Esto suele producir que las estimaciones en saltos alejados del origen tengan mayor varianza (pudiendo llegar a ser extremadamente grande). La recomendación de Journel y Huijbregts (1978, p. 194) es realizar el ajuste a lo sumo hasta la mitad del salto máximo posible y de forma que el número de aportaciones a la estimación en cada salto sea por lo menos de 30. Según la experiencia personal (que coincide en parte con las observaciones de Hass, 1990, p. 954), aunque este valor puede ser adecuado para saltos próximos al origen (incluso uno menor, p.e. 20), la estimación eficiente en saltos grandes suele requerir de un número bastante mayor de aportaciones. Por este motivo un buen criterio puede ser por ejemplo seleccionar los saltos de estimación regularmente espaciados, tomando Δ_S y Δ_T de forma que el número de aportaciones a las estimaciones en los saltos más próximos al origen sea por lo menos de 30 (aunque un valor menor podría ser también adecuado) y únicamente considerar saltos hasta la mitad del correspondiente salto máximo. Si se utiliza este criterio, el número de aportaciones en todos los saltos será normalmente superior a 30 y muy superior en los saltos más grandes (ya que al tomar los saltos regularmente espaciados, el número de aportaciones típicamente aumenta rápidamente hasta la mitad del salto máximo disminuyendo posteriormente).

Siguiendo más de cerca la recomendación de Journel y Huijbregts (1978), otro posible criterio sería considerar saltos irregularmente espaciados de forma

que el número de aportaciones a la estimación en cada salto fuera como mínimo 30. Sin embargo, después de realizar un pequeño estudio de simulación (en el caso espacial isotrópico), se observó que con el criterio anterior se obtenían resultados mucho mejores. Se podría pensar también en aumentar el número mínimo de aportaciones; esto produciría sin embargo estimaciones más alejadas del origen, resultando contraproducente si el objetivo final es la predicción (ver p.e. sección 3.4.2).

7.1.2 Un ejemplo de la eficiencia del estimador empírico

Como ya se comentó en las secciones 4.5.1 y 6.2.2, si se utiliza el estimador empírico del semivariograma con modelos flexibles pueden aparecer problemas de “sobreajuste”, ya que las estimaciones piloto pueden tener mucha variabilidad produciendo que el semivariograma ajustado tenga también una forma muy irregular. Algunos autores (normalmente aquellos que apoyan la estimación por máxima verosimilitud y están en contra de los métodos de mínimos cuadrados) cuestionan la utilidad del estimador empírico como herramienta para seleccionar modelos de semivariograma y estimar los parámetros de dichos modelos (e.g. Stein, 1999, sección 6.9). Sin embargo, en la mayoría de los casos los ejemplos que manejan en su argumentación fueron obtenidos a partir de una única simulación de un proceso espacial y por tanto habría que tener cuidado al extraer conclusiones de ellos. Fernández-Casal et al. (2003a) realizaron un estudio de simulación más completo empleando el estimador empírico (7.1) junto a distintos modelos de semivariogramas espacio-temporales (entre ellos los modelos flexibles anisotrópicos en dos componentes presentados en la sección 6.3). A continuación se presentan algunos de los resultados obtenidos en dicho estudio de simulación, aunque solamente los relativos al ajuste de los distintos modelos. De estos resultados se deduce, entre otras cosas, que obtener buenos ajustes al estimador empírico no garantiza necesariamente mejores resultados. Más adelante, en el capítulo 9, se mostrarán los resultados de un estudio de simulación más completo donde se compara este estimador con un estimador no paramétrico propuesto en la siguiente sección.

En el estudio de simulación se consideró una rejilla (diseño fijo) de dimensiones 10×10 ($n = 100$) en el dominio $D = [0,2] \times [0,2] \subset \mathbb{R} \times \mathbb{R}$

(dimensión espacial $d = 1$). Utilizando el modelo de kriging ordinario, se generaron 1000 simulaciones en esas posiciones espacio-temporales (ver figura 7.1). Los valores fueron generados a partir de procesos normales utilizando los siguientes variogramas teóricos: los modelos SVCH2 (ec. (6.7)) y SVCH4 (ec. (6.8)) de Cressie y Huang (1999), con parámetros $d = 1$, $c_0 = 1$, $a = 1$, $b = 1$ y $\sigma^2 = 1$; el modelo esférico anisotrópico SVESFA (ec. (6.4)) con $c_0 = 1$, $a = 0.5$, $b = 0.25$ y $\sigma^2 = 1$; y el modelo exponencial separable SVEXPS (ec. (6.6)) con $c_0 = 1$, $a = 0.5$, $b = 1$ y $\sigma^2 = 1$ (ver figuras 6.1 y 6.2).

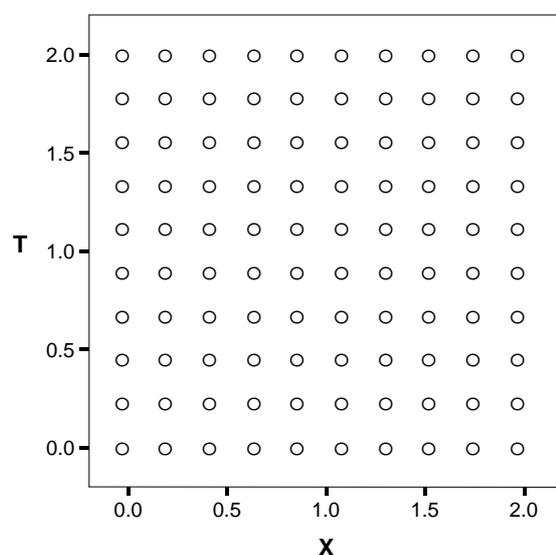


Figura 7.1: Posiciones espacio-temporales de los datos.

En cada simulación, utilizando el estimador empírico (7.1), se obtuvieron estimaciones piloto del semivariograma en saltos equiespaciados de la forma (7.4) con $\Delta_S = \Delta_T = 2/9$ y considerando un conjunto de índices de la forma (7.3) con $K = L = 5$ (de esta forma se sigue el criterio de Journel y Huijbregts 1978, p. 194; ver sección 4.1.1). Al conjunto de estimaciones piloto obtenidas se le ajustaron los modelos paramétricos anteriores y los siguientes modelos de Shapiro-Botha extendidos: SVSBE(1,1) modelo anisotrópico en dos componentes (6.14) con $d_1 = d_2 = 1$, SVSBE(∞ ,1) modelo (6.14) con $d_1 = \infty$ y $d_2 = 1$, y SVSBE(∞ , ∞) modelo (6.14) con $d_1 = \infty$ y $d_2 = \infty$.

En el ajuste de los modelos de Shapiro-Botha extendidos se utilizó la rutina QPROG de la librería IMSL para resolver los problemas de programación cuadrática (se procedió de forma iterativa; utilizando m.c.o., i.e. $w_{kl} = 1$, en el

primer paso y recalculando los pesos en cada iteración hasta convergencia). Los valores de I y J en (6.14) se fijaron igual al correspondiente número de saltos utilizados en el ajuste menos uno, es decir $I = J = 5$ (de esta forma el número de parámetros es menor que el número de estimaciones). Los puntos de discretización se tomaron equiespaciados (ec. (6.18)), tomando los valores de ϕ_1 y ϕ_2 de forma que se minimizara la suma de cuadrados ponderados del ajuste (se consideraron como dos parámetros adicionales; para su estimación, se evaluó la función objetivo en una rejilla bidimensional y se seleccionaron los valores que minimizaban la suma de cuadrados ponderados). En el ajuste de los demás modelos se utilizó el algoritmo iterativo descrito al final de la sección 4.3.1 (combinando regresión lineal m.c.p. con un algoritmo Levenberg-Marquardt modificado).

Para comparar los distintos modelos, además de la suma de cuadrados ponderados del ajuste al semivariograma empírico:

$$\text{SCP} = \sum_{(k,l) \in M} |N(r_k, u_l)| \left(\frac{\hat{\gamma}(r_k, u_l)}{\bar{\gamma}(r_k, u_l)} - 1 \right)^2,$$

se calculó también el error cuadrático relativo para medir el ajuste al variograma teórico:

$$\text{ECRT} = \sum_{(k,l) \in M} \left(\frac{\bar{\gamma}(r_k, u_l)}{\gamma(r_k, u_l)} - 1 \right)^2,$$

donde $\hat{\gamma}(\cdot, \cdot)$ es el estimador empírico del semivariograma, $\bar{\gamma}(\cdot, \cdot)$ el semivariograma ajustado, $\gamma(\cdot, \cdot)$ el semivariograma teórico y M el conjunto de índices correspondiente a los saltos utilizados en el ajuste. Además se calcularon los valores correspondientes al variograma teórico (TEOR) empleado para generar las observaciones (utilizándolo en lugar del modelo ajustado).

Los valores obtenidos se muestran en la tabla 7.1; si comparamos el ajuste al semivariograma empírico (SCP) observamos que los modelos de Shapiro-Botha extendidos son mucho más flexibles y se adaptan mejor a las estimaciones del semivariograma en todos los casos. Dentro de éstos, los modelos más específicos son los que obtienen mejores ajustes; algo que concuerda con observaciones anteriores (al considerar modelos válidos en dimensiones mayores se pierde flexibilidad). Si comparamos los valores de ECRT para medir la

proximidad al semivariograma teórico (en los saltos utilizados en el ajuste), en general ocurre lo contrario y los modelos de Shapiro-Botha extendidos se ajustan peor. En la mayoría de los casos el modelo correspondiente al semivariograma teórico es el que aparentemente obtiene mejores resultados. Esto es lo que nos induce a pensar que el estimador clásico del semivariograma es poco eficiente, ya que los modelos más ajustados al estimador empírico son los que están a mayor distancia del semivariograma teórico (lo que se corresponde con los grandes valores de SCP obtenidos utilizando los semivariogramas teóricos).

Como ejemplo adicional en la figura 7.2 se comparan las estimaciones del efecto nugget obtenidas con los distintos modelos, las estimaciones logradas con los modelos de Shapiro-Botha extendidos son las que en general se aproximan mejor al verdadero valor (especialmente los menos flexibles), por lo que podemos decir que el comportamiento de éstos modelos cerca del origen es bastante bueno (y por tanto el del estimador empírico en saltos pequeños no es muy malo). Destaca también el mal comportamiento del modelo SVCH4 (produciendo subestimaciones del efecto nugget), incluso cuando es el utilizado para generar las simulaciones (figura 7.1-b). Por tanto, incluso utilizando el modelo paramétrico correcto con el estimador empírico, las estimaciones finales de los parámetros pueden ser poco eficientes (naturalmente si el modelo paramétrico está mal especificado los resultados pueden ser mucho peores).

Las observaciones anteriores confirman la idea de que al utilizar modelos flexibles es importante utilizar estimaciones piloto eficientes. Más adelante (en el capítulo 9) se corroborará esta afirmación al comparar, también mediante simulación, el estimador empírico con el estimador lineal local propuesto en la siguiente sección.

Semivariograma		SCP			ECRT		
Teórico	Ajustado	Media	Mediana	Des.Std.	Media	Mediana	Des.Std.
SVCH2	SVSBE(1,1)	19.48	18.28	7.75	1.998	.951	3.427
	SVSBE(∞ ,1)	24.14	23.00	8.61	1.974	.933	3.436
	SVSBE(∞ , ∞)	26.99	25.68	9.58	1.899	.856	3.392
	SVCH2	30.54	29.40	10.92	1.831	.819	3.190
	SVCH4	33.36	31.54	12.65	1.783	.860	3.001
	SVESFA	34.24	31.13	15.39	1.755	.846	2.897
	SVEXPS	51.97	39.61	44.46	1.702	.845	2.607
	TEOR	186.85	98.18	280.46	.000	.000	.000
SVCH4	SVSBE(1,1)	17.73	16.40	7.69	1.601	.881	2.350
	SVSBE(∞ ,1)	21.94	20.47	8.76	1.553	.846	2.354
	SVSBE(∞ , ∞)	26.66	25.33	10.49	1.486	.773	2.363
	SVCH2	36.60	31.90	20.45	1.433	.727	2.241
	SVCH4	37.33	33.48	18.27	1.386	.710	2.144
	SVESFA	32.11	29.83	12.45	1.414	.722	2.225
	SVEXPS	38.58	33.86	21.11	1.410	.719	2.235
	TEOR	156.66	94.60	200.53	.000	.000	.000
SVESFA	SVSBE(1,1)	22.73	19.91	12.30	1.470	.804	2.157
	SVSBE(∞ ,1)	27.53	25.14	12.92	1.349	.698	2.128
	SVSBE(∞ , ∞)	31.33	28.98	13.73	1.310	.658	2.130
	SVCH2	51.82	46.02	24.74	1.551	.824	2.416
	SVCH4	46.85	41.53	21.67	1.460	.729	2.341
	SVESFA	35.59	32.82	15.01	1.368	.689	2.259
	SVEXPS	39.31	35.40	17.83	1.274	.613	2.069
	TEOR	148.49	92.59	188.64	.000	.000	.000
SVEXPS	SVSBE(1,1)	22.48	19.70	12.44	1.882	.967	3.218
	SVSBE(∞ ,1)	24.91	22.22	12.29	1.775	.866	3.219
	SVSBE(∞ , ∞)	28.01	25.50	13.38	1.725	.812	3.212
	SVCH2	48.75	41.15	28.39	1.987	.929	3.679
	SVCH4	44.87	40.23	23.90	1.833	.886	3.324
	SVESFA	34.83	31.30	15.92	1.762	.811	3.367
	SVEXPS	43.96	37.03	25.89	1.590	.779	2.823
	TEOR	180.35	105.13	272.00	.000	.000	.000

Tabla 7.1: Medidas del ajuste obtenido con los distintos modelos.

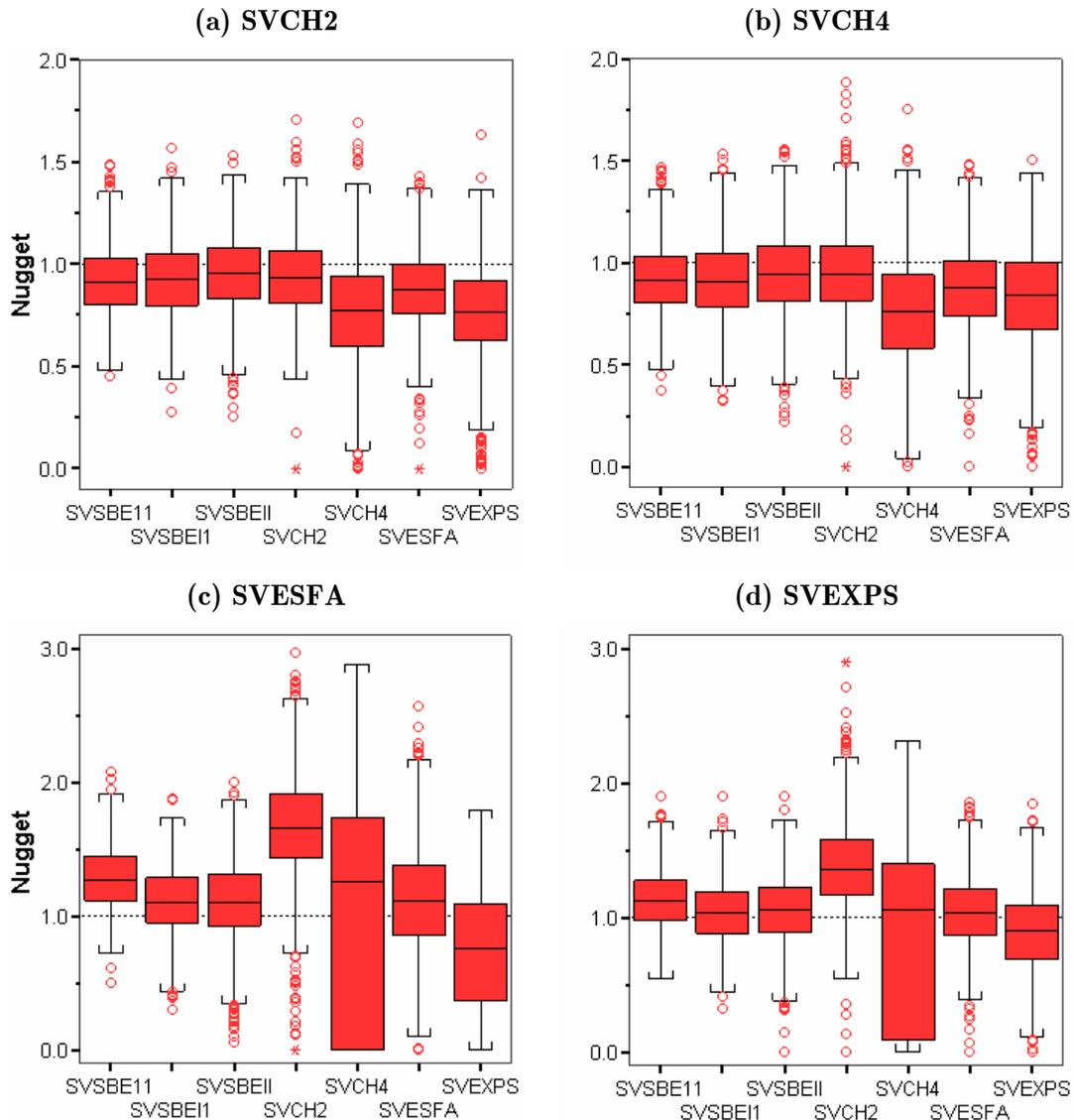


Figura 7.2. Estimaciones del efecto nugget obtenidas al ajustar los distintos modelos (las líneas de puntos se corresponden con el valor teórico).

7.2 ESTIMADORES NO PARAMÉTRICOS

Para la obtención de estimaciones más eficientes podemos pensar en utilizar también las ideas de suavizado tipo núcleo. Los estimadores no paramétricos mostrados en la sección 4.1.1 para el caso espacial isotrópico se extienden fácilmente para el caso espacial-isotrópico/temporal-isotrópico. Por ejemplo, podemos obtener estimadores de este tipo simplemente aplicando

directamente la metodología de regresión no paramétrica multivariante al conjunto de datos¹:

$$\{(\mathbf{X}_{ij}, Y_{ij}) : 1 \leq i < j \leq n\}$$

con:

$$\begin{aligned}\mathbf{X}_{ij} &= (\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|)' , \\ Y_{ij} &= (Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2 .\end{aligned}$$

Otras aproximaciones utilizadas en el caso espacial isotrópico para la construcción de estimadores piloto se podrían extender también para este caso. Por ejemplo, siguiendo la idea de Hawkins y Cressie (1984) (ser sección 4.1.1), podríamos obtener estimadores más robustos considerando como respuesta, en lugar de Y_{ij} , la raíz cuadrada de las diferencias en valor absoluto:

$$Y_{ij}^{1/4} = |Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)|^{1/2} .$$

Aunque las estimaciones obtenidas aplicando directamente los métodos de regresión no paramétrica multivariante a este conjunto de datos también habría que transformarlas adecuadamente a la escala original.

A continuación se presentan algunos de los estimadores pilotos no paramétricos que se pueden obtener de esta forma.

ESTIMADOR NADARAYA-WATSON

La expresión de un estimador de este tipo sería la siguiente:

$$\hat{\gamma}(r, u) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n K_{\mathbf{H}} \left(\left(\begin{array}{c} \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{array} \right) \right) (Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2}{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n K_{\mathbf{H}} \left(\left(\begin{array}{c} \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{array} \right) \right)} ,$$

con $K_{\mathbf{H}}(\mathbf{v}) = \frac{1}{|\mathbf{H}|} K(\mathbf{H}^{-1}\mathbf{v})$, siendo $K(\cdot)$ una función núcleo bidimensional y \mathbf{H} la correspondiente matriz de ventanas.

¹ Al igual que en el caso isotrópico mostrado al final de la sección 4.1.1, en los estimadores considerados aquí no se incluirán los datos correspondientes a los índices $i = j$ para cubrir la posible falta de continuidad del semivariograma en el origen.

Un problema con estos estimadores, bastante conocido en la estimación no paramétrica de curvas y superficies, es el denominado “efecto frontera”. Si se utiliza este tipo de estimadores con núcleos simétricos, las estimaciones obtenidas en los extremos de las observaciones pueden ser muy poco eficientes. Como ya se comentó anteriormente, en la estimación del semivariograma es especialmente importante capturar la conducta cerca del origen (ver p.e. sección 3.4.2); por lo que este problema resulta aún más importante. Para solventar eficientemente este inconveniente García-Soidán et al. (2003a) propusieron, para el caso espacial isotrópico, utilizar una combinación de núcleos asimétricos. Se podría extender igualmente esta aproximación para el caso espacial-isotrópico/temporal-isotrópico; sin embargo, siguiendo la alternativa sugerida por García-Soidán et al. (2003b), se propondrá el uso del estimador lineal local mostrado a continuación.

ESTIMADOR LINEAL LOCAL

Como ya se comentó anteriormente, otros estimadores locales tipo núcleo muy conocidos son los estimadores polinómicos locales (ver p.e. Fan y Gijbels, 1996). Una de las principales ventajas de estos estimadores frente a los estimadores tipo Nadaraya-Watson es la ausencia de efectos frontera.

Podemos obtener una estimación piloto de $\gamma(r, u)$ mediante regresión lineal local multivariante minimizando:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j) \right)^2 - (\beta_0, \beta_{10}, \beta_{01}) \begin{pmatrix} 1 \\ \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{pmatrix} \times K_{\mathbf{H}} \left(\begin{pmatrix} \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{pmatrix} \right), \quad (7.5)$$

(utilizando la notación anterior). Si $(\hat{\beta}_0, \hat{\beta}_{10}, \hat{\beta}_{01})$ es la solución del problema de mínimos cuadrados ponderados (7.5), entonces la estimación piloto de $\gamma(r, u)$ será:

$$\hat{\gamma}(r, u) = \frac{1}{2} \hat{\beta}_0. \quad (7.6)$$

ESTIMADOR LINEAL LOCAL ROBUSTO

Siguiendo la idea descrita al principio de esta sección podríamos obtener versiones robustas de los estimadores anteriores; por ejemplo, en el caso del estimador lineal local, minimizando:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(|Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)|^{\frac{1}{2}} - (\beta_0, \beta_{10}, \beta_{01}) \begin{pmatrix} 1 \\ \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{pmatrix} \right)^2 \times \\ K_{\mathbf{H}} \left(\begin{pmatrix} \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{pmatrix} \right).$$

Si $(\bar{\beta}_0, \bar{\beta}_{10}, \bar{\beta}_{01})$ es la solución del problema anterior, procediendo de forma análoga a la propuesta de Hawkins y Cressie (1984) para el caso espacial isotrópico, se obtiene² la estimación robusta de $\gamma(r, u)$:

$$\bar{\gamma}(r, u) = \frac{1}{2A(h)} (\bar{\beta}_0)^4,$$

con:

$$A(h) = 0.457 + 0.494 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij}^2,$$

siendo w_{ij} los pesos de la estimación local lineal de $\bar{\beta}_0$, es decir:

$$\bar{\beta}_0 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} |Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)|^{\frac{1}{2}},$$

² Simplemente teniendo en cuenta que si:

$$W = \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} \frac{|Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)|^{\frac{1}{2}}}{(2\gamma(\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|))^{\frac{1}{4}}},$$

entonces $E(W) \simeq 0.82216$ y $Var(W) \simeq 0.12192 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij}^2$ suponiendo independencia (algo más apropiado en este caso, ya que la raíz cuadrada de las diferencias en valor absoluto están menos correladas que las diferencias al cuadrado; ver p.e. Cressie, 1993, p. 76) y aplicando posteriormente el método delta.

(realmente esta corrección serviría para cualquiera de los estimador no paramétricos que verifiquen $\sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} = 1$).

DETALLES COMPUTACIONALES

Hay que tener en cuenta que el numero de datos utilizados en la estimación ($n(n-1)/2$) puede llegar a ser extremadamente grande, por lo que es recomendable utilizar algún método que permita acelerar el cálculo de las estimaciones. En los casos mostrados en este trabajo se decidió emplear la técnica conocida como binning lineal por sus ventajas en la precisión de las aproximaciones obtenidas (ver p.e. Wand, 1994).

Adicionalmente, en la estimación no paramétrica multivariante normalmente se consideran formas más simples para la matriz de ventanas \mathbf{H} , como por ejemplo:

$$\mathbf{H} = \text{diag}(h_1^2, h_2^2),$$

o incluso:

$$\mathbf{H} = h^2 \mathbf{S}, \quad (7.7)$$

siendo \mathbf{S} la matriz de covarianzas muestral:

$$\mathbf{S} = \text{Cov}\left(\left\{(\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|)' : 1 \leq i < j \leq n\right\}\right),$$

(ver p.e. Wand y Jones, 1995, pp. 105-108). Si se escoge una matriz de ventanas más compleja aumentará el tiempo de computación y se pueden agravar los problemas relacionados con la minimización multidimensional (aunque éstos pueden solventarse, por lo menos en parte, utilizando algoritmos genéticos de optimización; ver por ejemplo Goldberg, 1989). En los casos considerados en este trabajo se obtuvieron resultados casi idénticos con las distintas opciones para la matriz \mathbf{H} (incluyendo el caso general).

SELECCIÓN DE LA VENTANA ÓPTIMA

Las estimaciones obtenidas dependen en gran medida de la matriz de ventanas \mathbf{H} empleada ($\hat{\gamma}(r, u) \equiv \hat{\gamma}(r, u; \mathbf{H})$), por lo que la principal cuestión a la hora de utilizar este estimador es la selección de una matriz de ventanas

adecuada. Para solventar este problema podemos considerar dos opciones de partida: basarse en la idea de que el objetivo final es la obtención de predicciones kriging o en alguna de las aproximaciones tradicionalmente utilizadas en la estimación funcional no paramétrica. Por ejemplo, como aproximación inicial, podríamos pensar en utilizar alguno de los siguientes criterios para la obtención de la matriz de ventanas óptima:

1. Validación cruzada en la predicción kriging (VCK):

Seleccionar \mathbf{H} de forma que se minimice el error cuadrático medio de validación cruzada:

$$\text{ECMVC} = \frac{1}{n} \sum_{i=1}^n \left(\hat{Z}_{-i}(\mathbf{s}_i, t_i) - Z(\mathbf{s}_i, t_i) \right)^2 \quad (7.8)$$

donde $\hat{Z}_{-i}(\mathbf{s}_i, t_i)$ es el predictor kriging de $Z(\mathbf{s}_i, t_i)$ calculado a partir de $\{Z(\mathbf{s}_j, t_j) : j \neq i\}$ empleando las estimaciones piloto $\hat{\gamma}(r, u; \mathbf{H})$ (obtenidas utilizando todas las observaciones). El problema principal de esta aproximación es que puede requerir de mucho tiempo de computación, sobre todo si el número de observaciones es grande y la forma de la matriz de ventanas es compleja. Para calcular esta medida para una matriz fija \mathbf{H} , se deben obtener las correspondientes estimaciones piloto del semivariograma, ajustar un modelo válido³ y finalmente calcular las predicciones kriging.

2. Validación cruzada en la estimación del variograma (VCE):

Buscar \mathbf{H} de forma que se minimice alguno de los criterios de error empleados en la estimación no paramétrica tradicional. Por ejemplo el error cuadrático de validación cruzada en la estimación del variograma:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n \left((Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2 - 2\hat{\gamma}_{-(i,j)}(\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|) \right)^2,$$

o el error cuadrático relativo:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{(Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2}{2\hat{\gamma}_{-(i,j)}(\|\mathbf{s}_i - \mathbf{s}_j\|, |t_i - t_j|)} - 1 \right)^2, \quad (7.9)$$

³ Si además se ajusta un modelo SVSBE utilizando el primero de los criterios mostrados en la sección 6.2.3 para la selección de los puntos de discretización, este método resulta ser computacionalmente prohibitivo.

siendo $\hat{\gamma}_{-(i,j)}(\cdot, \cdot)$ la estimación piloto del semivariograma obtenida sin considerar $(Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2$. La ventaja de utilizar alguno de estos criterios es que han sido estudiados extensivamente en el contexto de la regresión local multivariante y están disponibles algoritmos rápidos para el cálculo de estas medidas (e.g. Turlach y Wand, 1996). En los resultados mostrados en próximos capítulos se consideró únicamente el error cuadrático relativo con este criterio.

La selección del parámetro \mathbf{H} es un tema también abierto que debe ser estudiado con mayor profundidad. Por ejemplo, la idea de utilizar una validación cruzada eliminando más de una observación (descrita al final de la sección 4.4) puede ser más apropiada para este caso⁴. Adicionalmente, si se deducen las expresiones del sesgo y varianza condicionales del estimador lineal local (de forma análoga a la estimación núcleo multivariante, e.g. Wand y Jones, 1994; o a la regresión local multivariante, e.g. Ruppert y Wand, 1994), se podrían obtener la ventana local o global asintóticamente óptima utilizando por ejemplo el criterio MSE o MSIE (ver p.e. García-Soidán et al., 2003b, para el caso espacial isotrópico). Todavía no se ha investigado ninguno de estos puntos, pero teniendo en cuenta el comportamiento más estable en otras situaciones, los métodos de selección de la ventana local óptima tipo plug-in podrían mejorar significativamente la conducta del estimador lineal local del semivariograma (aunque pueden ser mucho más difíciles de obtener que en el contexto de independencia).

⁴ Realmente, si se emplea alguna técnica binning para el cálculo de las estimaciones y se emplea el criterio VCE sobre los datos discretizados como si fuesen los originales, al eliminar una de estas observaciones se eliminarán en general varias de las observaciones iniciales.

CAPÍTULO 8

APLICACIÓN A DATOS REALES

En este capítulo se muestra un ejemplo de la aplicación a un conjunto de datos reales de la aproximación propuesta para el modelado de la dependencia espacio-temporal (combinando el estimador lineal local del semivariograma, mostrado en el capítulo anterior, y los modelos de Shapiro-Botha extendidos descritos en la sección 6.3). Las observaciones son medidas tomadas cada seis horas (unidad de tiempo) de la componente este-oeste de la velocidad de viento (en m/seg) a 10 metros de la superficie de una rejilla de 17×17 posiciones espaciales (regularmente espaciadas, aprox. 210 km.) localizada en la parte occidental del Océano Pacífico tropical (145°E - 175°E , 14°S - 16°N ; al noreste de Australia y Nueva Guinea). Este mismo conjunto de datos ha sido analizado anteriormente por distintos autores, como por ejemplo Cressie y Huang (1999) o Wikle y Cressie (1999). En lugar de emplear todas las observaciones temporales (480 intervalos de tiempo que corresponden al período comprendido entre noviembre de 1992 y febrero de 1993), se decidió considerar únicamente los primeros 56 instantes temporales para reducir el tiempo de computación; por tanto los resultados mostrados a continuación se obtuvieron a partir de una muestra espacio-temporal de 16184 observaciones. En la figura 8.1 se muestran las observaciones correspondientes a 9 instantes temporales (nótese que valores negativos se corresponden con viento del este y valores positivos con viento del oeste).

Después de un análisis exploratorio de los datos¹, puede verse (e.g. Cressie y Huang, 1999, pp. 1336-1338) que la suposición de estacionariedad

¹ Por motivos de extensión no se incluyeron en este trabajo las técnicas disponibles para la exploración de datos espaciales. Para información sobre este tema ver por ejemplo

espacio-temporal es razonable para este conjunto de datos; lo que nos permite emplear el modelo del kriging ordinario y estimar el semivariograma directamente de los datos originales (sin necesidad de eliminar la tendencia).

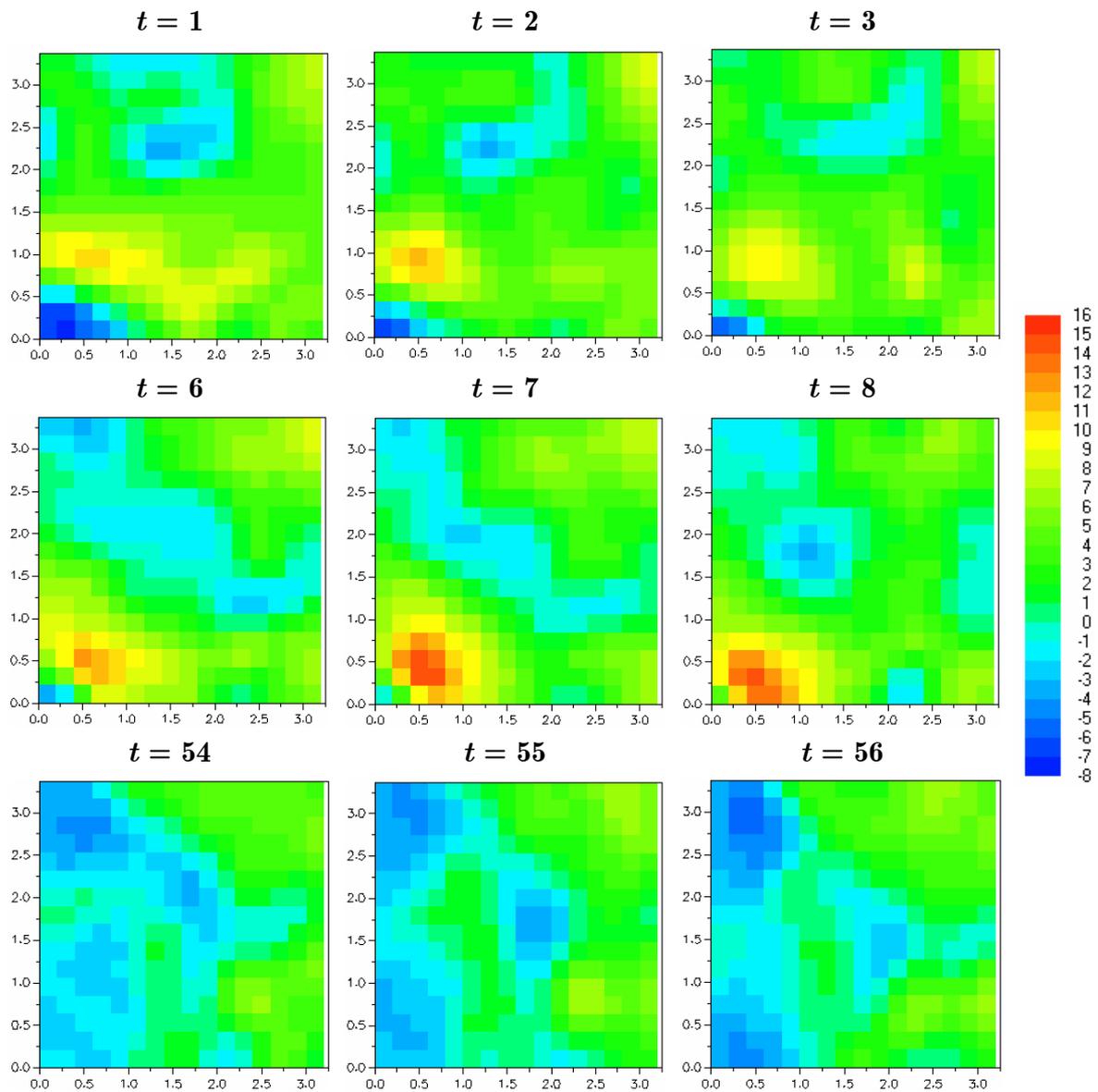


Figura 8.1: Datos de la componente este-oeste de la velocidad de viento (m/seg) en las 17×17 posiciones de observación para distintos instantes temporales (las escalas de los ejes X e Y en miles de kilómetros).

Como en ocasiones el objetivo es la obtención de predicciones futuras a partir de observaciones anteriores, se decidió utilizar esta idea para estudiar la

conducta de los distintos modelos². Para cada uno de los instantes temporales $t_0 = 9, \dots, 56$ (dos semanas) y para una de las 289 posiciones espaciales de observación \mathbf{s}_0 , se calculó el predictor (local) del kriging ordinario $\hat{Z}(\mathbf{s}_0, t_0)$ y la correspondiente varianza kriging $\sigma^2(\mathbf{s}_0, t_0)$ a partir de los datos de los 8 instantes temporales anteriores $\{Z(\mathbf{s}_i, t_i) : t_0 - 8 \leq t_i \leq t_0 - 1\}$. Para la obtención de estos valores se empleó el estimador lineal local (7.6) y los siguientes modelos de Shapiro-Botha extendidos: SVSBE(2,1) (modelo (6.14) con $d_1 = 2$ y $d_2 = 1$), SVSBE(∞ ,1) (modelo (6.14) con $d_1 = \infty$ y $d_2 = 1$) y SVSBE(∞ , ∞) (modelo (6.14) con $d_1 = \infty$ y $d_2 = \infty$).

8.1 ESTIMACIÓN Y AJUSTE DE LOS MODELOS

En cada subconjunto de datos se calcularon las estimaciones lineales locales del semivariograma empleando binning lineal (nótese que, aunque el número de observaciones no es muy grande 2312, el número total de pares es de 2671516). La matriz de ventanas se escogió de la forma $\mathbf{H} = h^2\mathbf{S}$ (ec. (7.7)) y se empleó el criterio de validación cruzada en la estimación del variograma (VCE), minimizando el error cuadrático relativo (7.9) para seleccionar la matriz óptima. De esta forma se obtuvieron estimaciones piloto $\hat{\gamma}(r_i, u_j)$ en los saltos $r_i = 0.1586 \cdot i$ (en miles de kilómetros), $u_j = j$ (1 unidad = 6 horas), $i = 0, \dots, 29$, $j = 0, \dots, 8$, $(i, j) \neq (0, 0)$.

Siguiendo el criterio de Journel y Huijbregts (1978, p. 194), el ajuste de los modelos SVSBE únicamente se realizó hasta la mitad de los correspondientes saltos máximos. Para el ajuste de estos modelos se siguió el mismo procedimiento descrito en la sección 7.1.2 (de forma iterativa, empezando con m.c.o. y utilizando la rutina QPROG de la IMSL). Los valores de I y J en (6.14) se fijaron igual al correspondiente número de saltos utilizados en el ajuste menos uno, es decir $I = 15$ y $J = 4$. Los puntos de discretización sin embargo se tomaron siguiendo el criterio descrito al final de la sección 6.2.3 (basado en la propuesta de Genton y Gorsich, 2003), utilizando el algoritmo desarrollado por

² En Fernández-Casal et al (2003a) se tiene un ejemplo sobre este mismo conjunto de datos, empleando el estimador empírico y los modelos SVSBE, pero considerando interpolación (validación cruzada tradicional) en lugar de extrapolación.

Ball (2000) para el cálculo de los ceros de las funciones de Bessel. En la figura 8.2 se muestra un ejemplo, correspondiente al primer subconjunto de datos, con las estimaciones piloto obtenidas de esta forma y los ajustes de los distintos modelos.

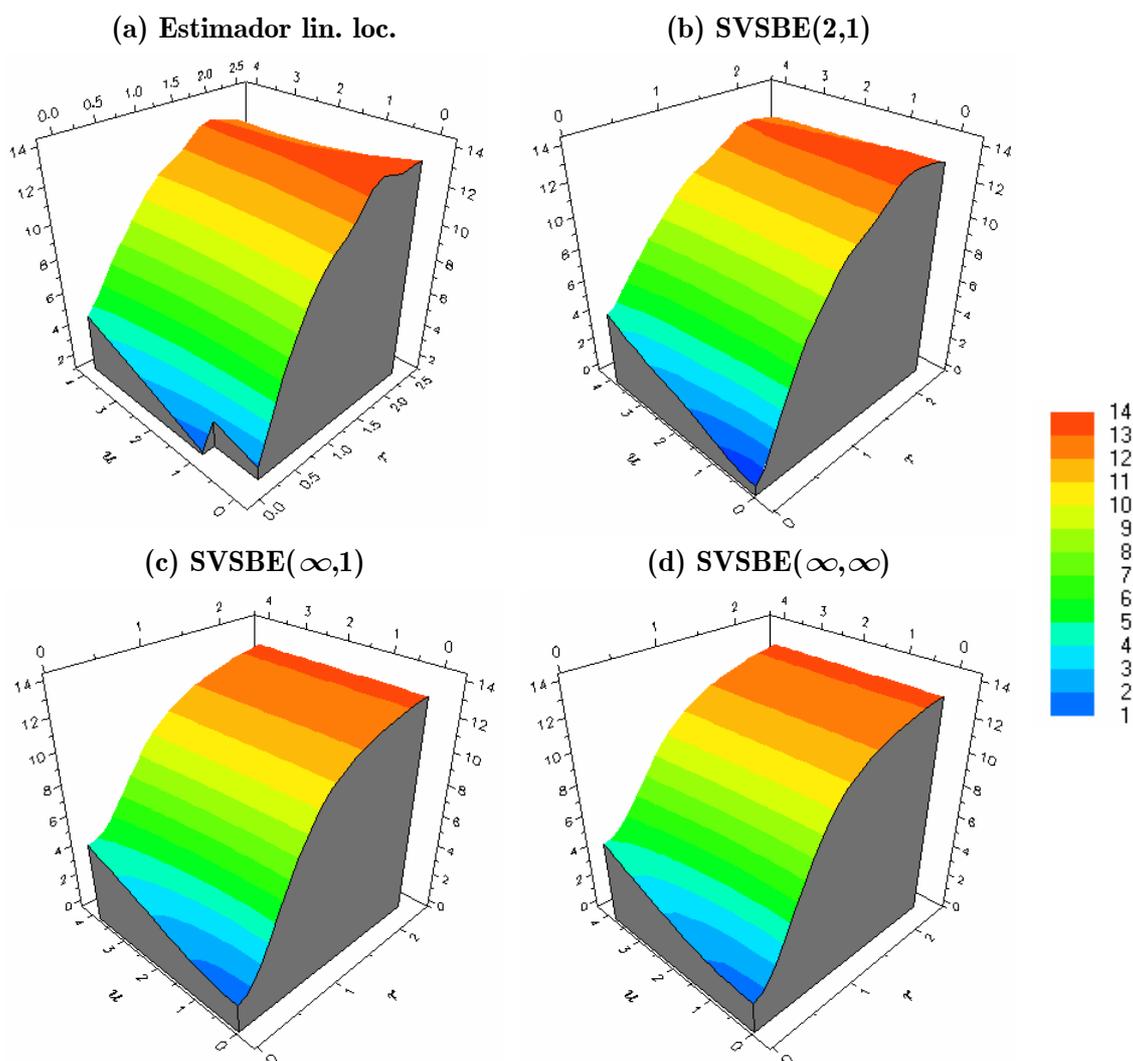


Figura 8.2: Gráficos 3D, correspondientes a los 8 primeros instantes temporales, con las estimaciones lineales locales del semivariograma (en los saltos considerados en el ajuste) y los distintos modelos ajustados.

En la tabla 8.1 se tiene un resumen de las estimaciones del efecto nugget y de los valores MCP (ec. (6.17)) obtenidos en los ajustes de los modelos SVSBE. Como cabía esperar, se observa que el modelo más flexible es el específico a las dimensiones de los datos y que al considerar modelos más

generales se obtienen peores ajustes. También se observa en este caso que al considerar modelos menos flexibles se obtienen estimaciones mayores del efecto nugget, algo que se puede ver también en el ejemplo mostrado en la figura 8.2.

Modelo ajustado	MCP			c_0		
	Media	Mediana	Des.Std.	Media	Mediana	Des.Std.
SVSBE(2,1)	4.07E-05	3.38E-05	2.12E-05	.354	.363	.102
SVSBE(∞ ,1)	.0010	.0009	.0004	.897	.832	.242
SVSBE(∞ , ∞)	.0012	.0012	.0003	1.136	1.086	.238

Tabla 8.1: Medidas resumen de los valores MCP y estimaciones del efecto nugget obtenidos en el ajuste los distintos modelos de Shapiro-Botha extendidos.

8.2 PREDICCIONES

Como se comentó al principio de este capítulo, en cada uno de los instantes temporales $t_0 = 9, \dots, 56$, se utilizaron los modelos SVSBE ajustados para calcular las predicciones $\hat{Z}(\mathbf{s}_0, t_0)$ y varianzas kriging $\sigma^2(\mathbf{s}_0, t_0)$ en las posiciones de observación \mathbf{s}_0 . Por ejemplo, en la figura 8.3 se tienen las superficies de predicción correspondientes al modelo SVSBE(2,1) en los tres últimos instantes temporales (estas superficies pueden ser comparadas con las mostradas en la figura 8.1).

Para comparar las predicciones obtenidas utilizando los diferentes modelos SVSBE, se calcularon las siguientes medidas para cada posición de predicción (\mathbf{s}_0, t_0) :

- El error cuadrático de predicción:

$$ECP = (\hat{Z}(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0))^2.$$

- El error cuadrático de predicción adimensional:

$$ECPA = \left((\hat{Z}(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0)) / \sigma(\mathbf{s}_0, t_0) \right)^2,$$

(que debería ser próximo a 1 si hay concordancia entre las varianzas

kriging y las varianzas observadas).

- La cobertura (COP) del intervalo de predicción al 95% suponiendo normalidad.

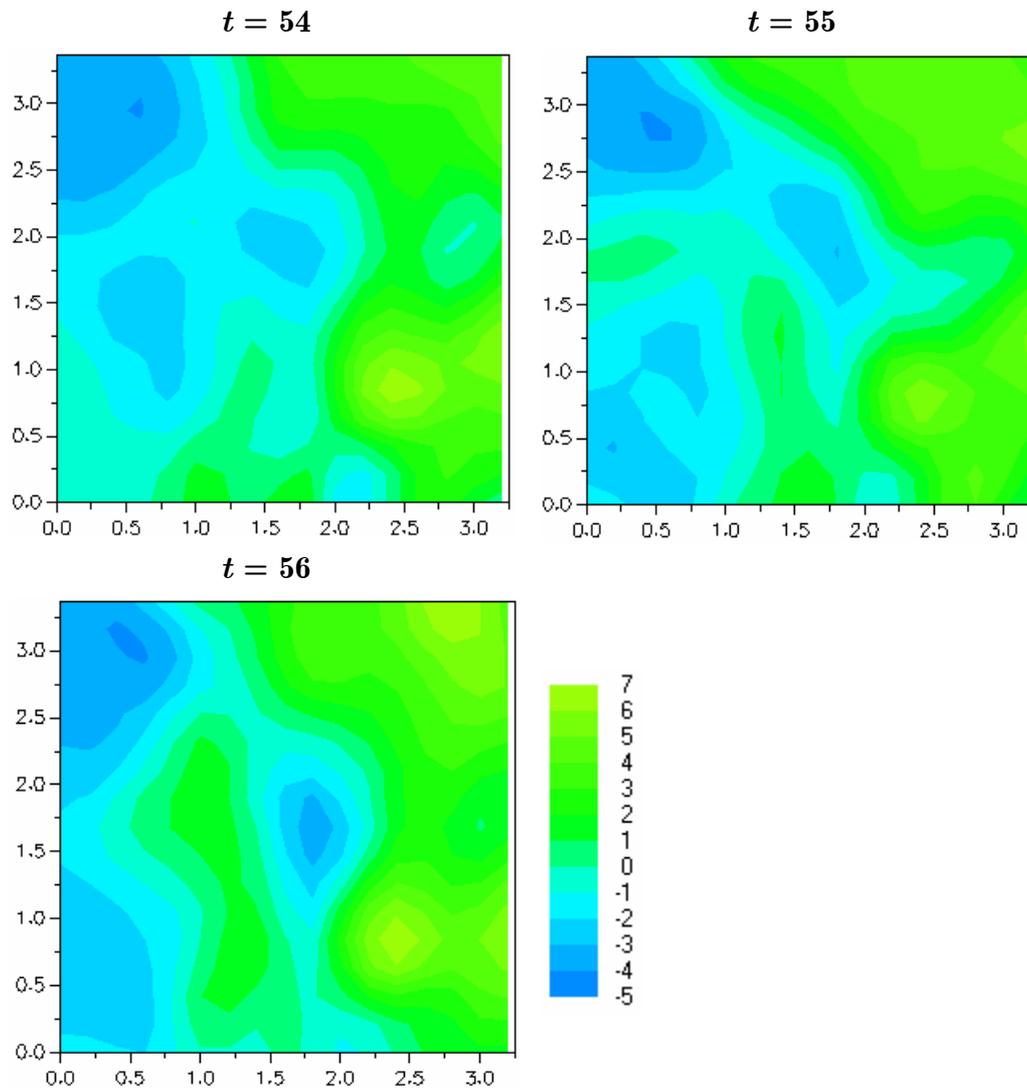


Figura 8.3: Predicciones de la componente este-oeste de la velocidad de viento en torno a las posiciones de observación para los tres últimos instantes temporales (calculadas a partir de los correspondientes 8 anteriores) empleando el modelo SVSBE(2,1) (con escalas iguales a las de la figura 8.1).

En la tabla 8.2 se muestra un resumen de los valores anteriores y adicionalmente en figura 8.4 se muestran las superficies de error obtenidas

promediando los valores de ECP en los 48 instantes temporales considerados, i.e.:

$$\text{MECP}(\mathbf{s}_0) = \frac{1}{48} \sum_{t_0=9}^{56} (\hat{Z}(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0))^2. \quad (8.1)$$

Teniendo en cuenta el error cuadrático de predicción (ECP), observamos en este caso que los modelos más flexibles con mejores ajustes al estimador piloto (ver tabla 8.1) son los que obtienen mejores resultados. Si se comparan los valores del error cuadrático de predicción adimensional (ECPA), se observa que el modelo SVSBE($\infty,1$) es el que obtiene mejores estimaciones de la varianza de predicción. Los valores correspondientes al modelo SVSBE(2,1) son también bastante buenos, mientras que con el modelo SVSBE(∞,∞) se obtuvieron estimaciones muy optimistas de la varianza kriging³. Naturalmente los valores de ECPA están en concordancia con las coberturas de los intervalos de predicción; cuanto mayor es el valor del ECPA menor es la cobertura del intervalo de confianza.

		SVSBE(2,1)	SVSBE($\infty,1$)	SVSBE(∞,∞)
ECP	Media	2.075	2.130	2.337
	Mediana	.911	.915	.983
	Desv. Std.	3.180	3.290	3.598
ECPA	Media	1.223	1.126	1.429
	Mediana	.528	.483	.598
	Desv. Std.	1.867	1.729	2.203
COP	Media	.927	.936	.900
	Desv. Std.	.260	.246	.300

Tabla 8.2: Medidas de resumen de las predicciones obtenidas con los diferentes modelos ajustados.

³ Nótese que es de esperar que los valores del ECPA sean mayores que 1, ya que en la estimación de la varianza kriging se supone que el variograma es conocido (ver sección 3.4.2).

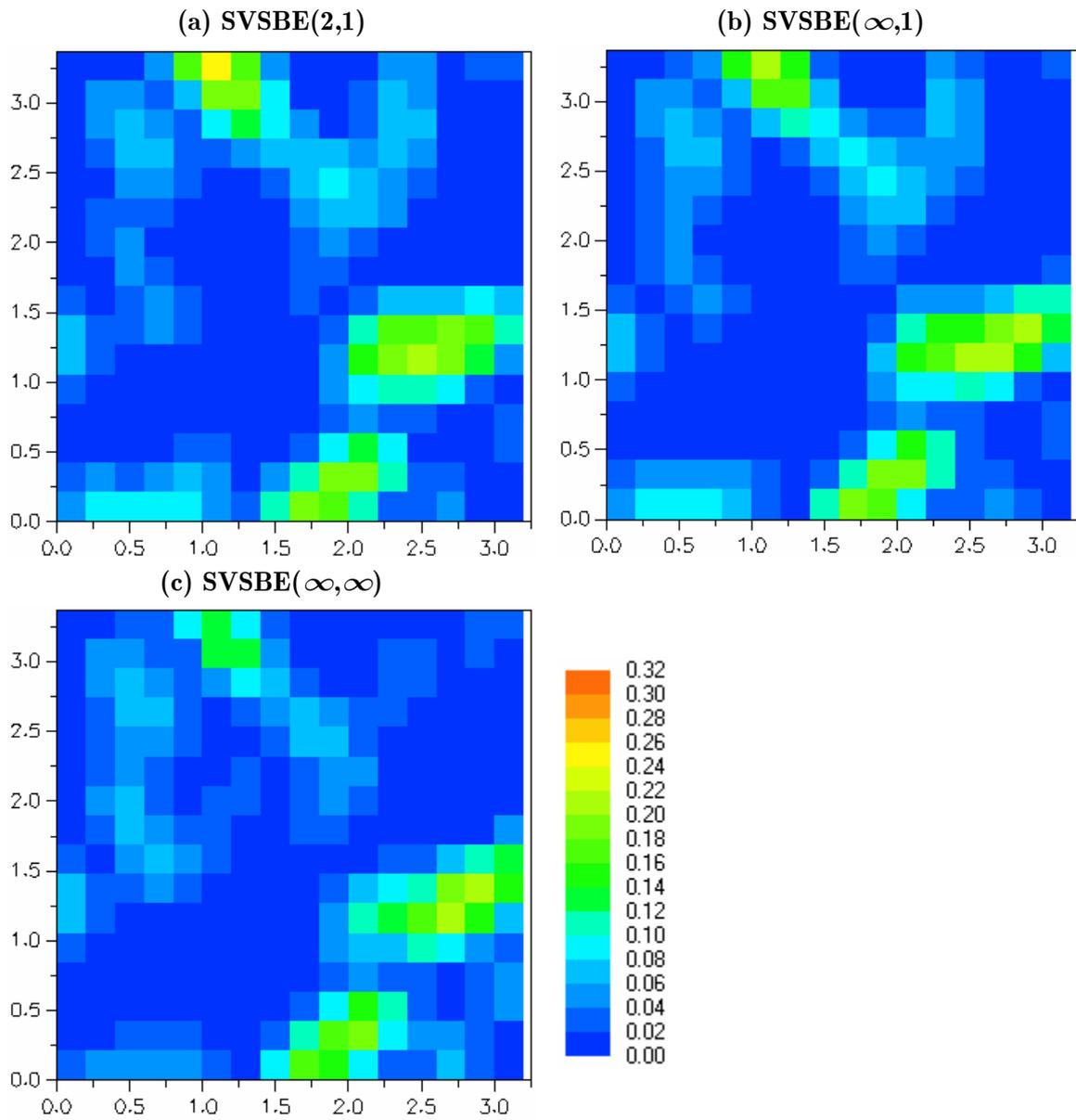


Figura 8.4: Gráficos de contorno de $MECP(s_0)$ (ec. (8.1)) correspondientes a los distintos modelos ajustados (las escalas de los ejes X e Y en miles de kilómetros).

CAPÍTULO 9

ESTUDIO DE SIMULACIÓN

En este capítulo se muestran los resultados obtenidos en un estudio de simulación comparando la aproximación propuesta con los métodos tradicionalmente utilizados en el modelado de la dependencia espacial (descritos en la sección 4.3). Se consideró el ajuste por m.c.p. de un modelo paramétrico de semivariograma a un estimador piloto y los métodos basados en máxima verosimilitud: estimación por máxima verosimilitud (MV) y máxima verosimilitud restringida (MVR). Adicionalmente, se estudió el comportamiento en la estimación m.c.p. de los estimadores empírico y lineal local mostrados en el capítulo anterior.

En el estudio se consideraron dos tamaños muestrales: $n = 100$ y $n = 250$, y dos modelos paramétricos como semivariogramas teóricos: el modelo SVCH2 (ec. (6.7)) de Cressie y Huang (1999) con parámetros $c_0 = 0.5$, $a = 1$, $b = 1$ y $\sigma^2 = 1.5$, y el modelo exponencial separable SVEXPS (ec. (6.6)) con $c_0 = 0.5$, $a = 0.5$, $b = 1$ y $\sigma^2 = 1.5$. Para cada tamaño muestral y para cada semivariograma teórico se generaron 1000 muestras en posiciones aleatorias en el dominio $D = [0,2] \times [0,2] \subset \mathbb{R} \times \mathbb{R}$ (dimensión espacial $d = 1$). En cada simulación, considerando una distribución uniforme en el dominio D , se generó un conjunto de posiciones espacio-temporales de tamaño n ; en este conjunto de posiciones y en tres posiciones de predicción (ver figura 9.1) se generaron valores de un proceso normal estacionario (empleando el correspondiente semivariograma teórico). A partir de estas observaciones se obtuvieron estimaciones piloto del semivariograma utilizando el estimador empírico (7.1) y el estimador lineal local (7.6).

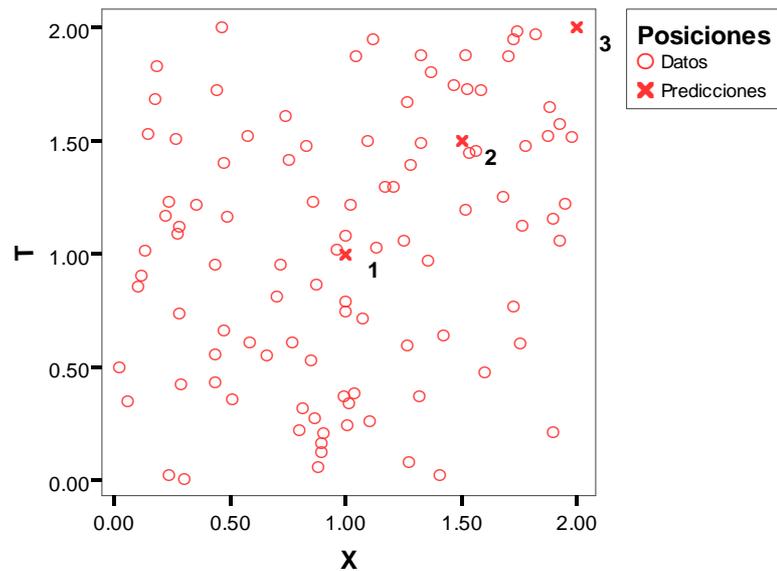


Figura 9.1: Posiciones de predicción y una generación de las posiciones aleatorias de las observaciones ($n = 100$).

Empleando el estimador empírico del semivariograma se calcularon estimaciones piloto en saltos $\{(r_k, u_l) : r_k = \Delta \cdot k, u_l = \Delta \cdot l, 1 \leq k, l \leq L\}$ con $\Delta = 2/L$, $L = 17$ en el caso de muestras de tamaño 100 y $L = 29$ en el caso de muestras de tamaño 250 (estos valores se seleccionaron de forma que el número de aportaciones para la estimación en el primer salto (r_1, u_1) fuese por lo menos de 30 en la mayoría de los casos). Con el estimador lineal local se obtuvieron estimaciones piloto, empleando binning lineal, en conjuntos de saltos de la forma $\{(r_k, u_l) : r_k = \Delta \cdot k, u_l = \Delta \cdot l, 0 \leq k, l \leq L, (k, l) \neq (0, 0)\}$ con $\Delta = 2/(L - 1)$, $L = 35$ en el caso de $n = 100$ y $L = 39$ para $n = 250$ (estos valores fueron seleccionados siguiendo el criterio propuesto por Wand, 1994). A los conjuntos de estimaciones piloto obtenidos se ajustaron, empleando el criterio m.c.p, los dos modelos paramétricos anteriores y los siguientes modelos de Shapiro-Botha extendidos: SVSBE(1,1) modelo anisotrópico en dos componentes (6.14) con $d_1 = d_2 = 1$, SVSBE($\infty, 1$) modelo (6.14) con $d_1 = \infty$ y $d_2 = 1$, y SVSBE(∞, ∞) modelo (6.14) con $d_1 = \infty$ y $d_2 = \infty$. Para el ajuste de los modelos SVSBE se siguió un procedimiento análogo al descrito en el capítulo anterior (los valores de I y J en (6.14) se fijaron igual al correspondiente número de saltos utilizados en el ajuste menos uno, y los puntos de discretización se tomaron siguiendo el segundo de los criterios descritos en la

sección 6.2.3 para disminuir el tiempo de computación). En el ajuste de los modelos paramétricos se utilizó el algoritmo iterativo descrito al final de la sección 4.3.1 (combinando regresión lineal m.c.p. con un algoritmo Levenberg-Marquardt modificado con restricciones en los parámetros).

Adicionalmente a la estimación m.c.p., se obtuvieron estimaciones por MV y MVR de los parámetros de los modelos SVCH2 y SVEXPS (por las razones dadas al final de la sección 6.2.2, no se utilizó la estimación máximo verosímil con los modelos SVSBE). Para evitar el problema de la posible multimodalidad de la superficie de verosimilitud (ver p.e. Mardia y Watkins, 1989) se siguió el procedimiento descrito al final de la sección 4.3.2 (empleando un algoritmo micro-genético de optimización) para resolver el problema de minimización multidimensional no lineal.

Finalmente, empleando las estimaciones definitivas del semivariograma, se obtuvieron predicciones de validación cruzada en las posiciones de observación y en las tres posiciones de predicción utilizando el método del kriging ordinario descrito en la sección 3.2. Estas predicciones se calcularon de forma local, empleando un radio máximo de búsqueda de 0.8 en el caso de $n = 100$ y de 0.6 en el caso de $n = 250$; y de forma que el número de observaciones en el vecindario estuviera comprendido entre 10 y 30 (para acelerar los cálculos se empleó la técnica de búsqueda por súper-bloques descrita en la sección 3.4.3 y el método mostrado en la sección 4.4 para la obtención de las predicciones de validación cruzada).

Un resumen de los resultados obtenidos en este estudio se muestra a continuación.

9.1 ESTIMACIONES PILOTO DEL SEMIVARIOGRAMA

Para comparar los dos estimadores piloto considerados, además de calcular valores MCP utilizando el semivariograma teórico en lugar de un modelo ajustado:

$$\text{MCP}^* = \frac{\sum_{(k,l) \in M} |N(r_k, u_l)| \left(\frac{\hat{\gamma}(r_k, u_l)}{\gamma(r_k, u_l)} - 1 \right)^2}{\sum_{(k,l) \in M} |N(r_k, u_l)|},$$

se calculó el promedio del error cuadrático de la estimación del semivariograma:

$$PEC = \frac{1}{|M|} \sum_{(k,l) \in M} (\hat{\gamma}(r_k, u_l) - \gamma(r_k, u_l))^2,$$

siendo $\hat{\gamma}(\cdot, \cdot)$ la estimación piloto, $\gamma(\cdot, \cdot)$ el semivariograma teórico y M el conjunto de índices correspondiente a los saltos utilizados en el ajuste (nótese que es de esperar que en la primera medida tengan mayor peso los errores en saltos próximos al origen, mientras que en la segunda tendrán mayor peso los errores en saltos mayores).

Semivar. y tamaño muestral			MCP*			PEC		
Teórico	n	Estimador	Media	Mediana	Des.Std.	Media	Mediana	Des.Std.
SVCH2	100	Empírico	.119	.078	.221	.129	.077	.304
		Lin. loc. (VCE)	.088	.051	.200	.090	.047	.244
		Lin. loc. (VCK)	.086	.049	.201	.088	.046	.244
	250	Empírico	.054	.034	.087	.043	.025	.081
		Lin. loc. (VCE)	.038	.020	.081	.029	.015	.069
		Lin. loc. (VCK)	.038	.019	.085	.029	.014	.071
SVEXPS	100	Empírico	.115	.083	.137	.315	.210	.448
		Lin. loc. (VCE)	.085	.056	.125	.217	.132	.374
		Lin. loc. (VCK)	.080	.051	.121	.205	.121	.352
	250	Empírico	.060	.041	.060	.135	.087	.150
		Lin. loc. (VCE)	.043	.025	.055	.093	.054	.126
		Lin. loc. (VCK)	.042	.025	.054	.090	.051	.123

Tabla 9.1: Comparación de las estimaciones piloto obtenidas con el estimador empírico y lineal local con los dos criterios para la selección de la ventana descritos al final de la sección 7.2 (utilizando el semivariograma teórico para el cálculo de los valores MCP).

Un resumen de los valores obtenidos se muestra en la tabla 9.1; donde podemos observar que el estimador lineal local proporciona mejores resultados en todos los casos. Además, no se observan grandes diferencias entre los dos criterios de selección de la ventana (sobre todo para $n = 250$); por tanto, a la vista de estos resultados, el criterio VCE sería aparentemente el preferible en la

práctica cuando el número de observaciones es suficientemente grande, ya que requiere de mucho menos tiempo de computación.

Además de las medidas anteriores, se calculó también el error cuadrático medio en los saltos (r_k, u_l) :

$$ECM(r_k, u_l) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\gamma}_i(r_k, u_l) - \gamma(r_k, u_l))^2,$$

denotando por $\hat{\gamma}_i(\cdot, \cdot)$ las estimaciones piloto obtenidas en la i -ésima simulación.

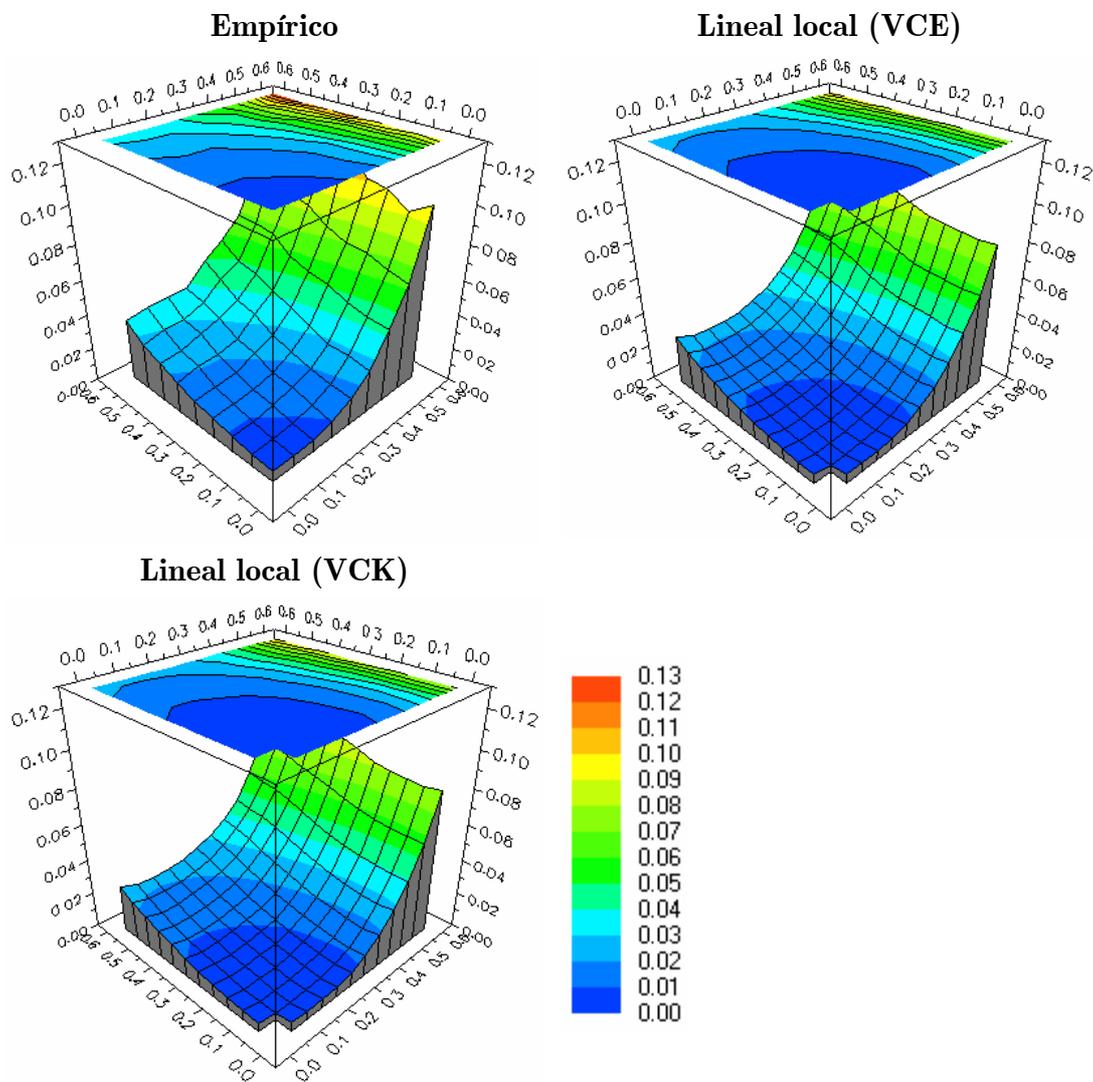


Figura 9.2: Superficies del error cuadrático medio de las estimaciones piloto del semivariograma, para $n = 250$ y utilizando el modelo SVCH2 como semivariograma teórico.

Por ejemplo, las superficies de $ECM(r_k, u_l)$ correspondientes a los modelos teóricos SVCH2 y SVEXPS para el caso de $n = 250$ se muestran en las figuras 9.2 y 9.3 respectivamente. Se observa también en todos los casos que con el estimador lineal local se obtienen mejores resultados, especialmente en torno al origen (lo cual es importante si el objetivo final es la predicción; ver sección 3.4.2).

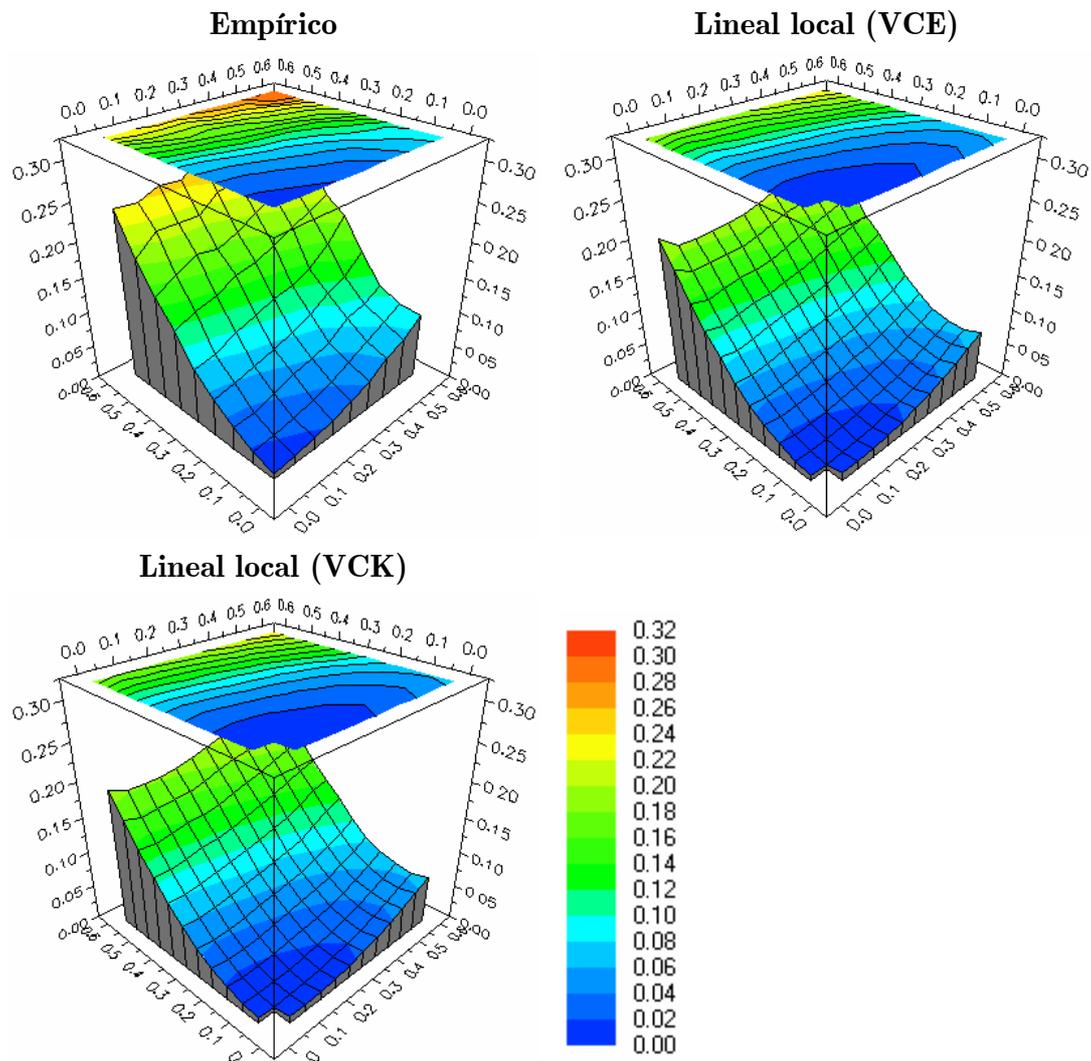


Figura 9.3: Superficies del error cuadrático medio de las estimaciones piloto del semivariograma, para $n = 250$ y utilizando el modelo SVEXPS como semivariograma teórico.

9.2 MODELOS AJUSTADOS

Para comparar el ajuste con los distintos modelos, además de los valores MCP obtenidos con los criterios m.c.p:

$$\text{MCP} = \frac{\sum_{(k,l) \in M} |N(r_k, u_l)| \left(\frac{\hat{\gamma}(r_k, u_l)}{\bar{\gamma}(r_k, u_l)} - 1 \right)^2}{\sum_{(k,l) \in M} |N(r_k, u_l)|},$$

se calculó también el error cuadrático relativo medio para medir el ajuste al variograma teórico:

$$\text{ECRM} = \frac{1}{|M|} \sum_{(k,l) \in M} \left(\frac{\bar{\gamma}(r_k, u_l)}{\gamma(r_k, u_l)} - 1 \right)^2,$$

donde $\hat{\gamma}(\cdot, \cdot)$ es el correspondiente estimador piloto del semivariograma, $\bar{\gamma}(\cdot, \cdot)$ el semivariograma ajustado, $\gamma(\cdot, \cdot)$ el semivariograma teórico y M el conjunto de índices correspondiente a los saltos utilizados en el ajuste.

Como ejemplo de los resultados obtenidos, en las tablas 9.2 y 9.3 se muestran resúmenes de las medidas calculadas utilizando los modelos SVCH2 y SVEXPS como semivariogramas teóricos para el caso de $n = 250$ (en los otros casos se observó un comportamiento similar de los distintos modelos). Si comparamos el ajuste al estimador piloto (valores MCP), como cabía esperar, los modelos de Shapiro-Botha son siempre los que obtienen mejores resultados (y al igual que en ejemplos anteriores, los modelos más específicos son los que se ajustan mejor al semivariograma piloto). También observamos pocas diferencias entre los dos criterios de selección de la ventana.

Si comparamos los valores ECRM, como regla general observamos que el modelo SVSBE(1,1) es el que presenta un peor comportamiento, mejorando considerablemente cuando se utiliza el estimador lineal local; obteniéndose resultados bastante parecidos con los demás modelos. En general se observó que con el modelo SVSBE(1,1) no se obtienen buenas estimaciones del semivariograma en los saltos pequeños especialmente al utilizar el semivariograma empírico (después de realizar algunas pruebas, parece ser que la opción escogida para seleccionar los puntos de discretización produce que los modelos SVSBE tengan menor flexibilidad en torno al origen en comparación con los ajustes obtenidos utilizando la primera de las opciones descritas en la

sección 6.2.3). Respecto a los demás resultados obtenidos, en general se observó que los métodos de estimación basados en máxima verosimilitud obtenían los mejores resultados cuando el modelo paramétrico estaba especificado correctamente, aunque empeorando considerablemente cuando el modelo no está correctamente especificado. También se observó que en la mayoría de los casos la estimación por MVR mejoraba los resultados obtenidos con la estimación MV (y en ocasiones de forma significativa).

Estimación		MCP			ECRM		
Método	Modelo	Media	Mediana	Des.Std.	Media	Mediana	Des.Std.
MCP Empírico	SVSBE(1,1)	.013	.013	.003	.041	.022	.088
	SVSBE(∞ ,1)	.016	.016	.003	.039	.020	.090
	SVSBE(∞ , ∞)	.018	.017	.003	.038	.018	.090
	SVCH2	.019	.018	.006	.038	.017	.090
	SVEXPS	.021	.020	.007	.037	.018	.077
MCP Lin. loc. (VCE)	SVSBE(1,1)	.001	.001	.001	.037	.019	.082
	SVSBE(∞ ,1)	.002	.002	.001	.036	.018	.083
	SVSBE(∞ , ∞)	.003	.003	.001	.036	.018	.083
	SVCH2	.005	.003	.014	.034	.017	.078
	SVEXPS	.006	.005	.006	.034	.018	.069
MCP Lin. loc. (VCK)	SVSBE(1,1)	.001	.001	.001	.037	.019	.085
	SVSBE(∞ ,1)	.002	.002	.001	.036	.018	.086
	SVSBE(∞ , ∞)	.002	.002	.001	.036	.018	.086
	SVCH2	.004	.003	.011	.035	.017	.084
	SVEXPS	.006	.004	.006	.034	.018	.070
MV	SVCH2				.032	.021	.032
	SVEXPS				.041	.032	.031
MVR	SVCH2				.028	.018	.029
	SVEXPS				.038	.029	.030

Tabla 9.2: Medidas del ajuste al estimador piloto y al semivariograma teórico correspondientes a muestras de tamaño $n = 250$ generadas utilizando el modelo SVCH2 como semivariograma teórico.

Estimación		MCP			ECRM		
Método	Modelo	Media	Mediana	Des.Std.	Media	Mediana	Des.Std.
MCP Empírico	SVSBE(1,1)	.013	.012	.003	.047	.028	.059
	SVSBE(∞ ,1)	.017	.016	.004	.044	.025	.060
	SVSBE(∞ , ∞)	.020	.019	.005	.043	.023	.060
	SVCH2	.022	.021	.007	.043	.023	.062
	SVEXPS	.022	.021	.006	.041	.022	.059
MCP Lin. loc. (VCE)	SVSBE(1,1)	.001	.001	.001	.042	.024	.055
	SVSBE(∞ ,1)	.002	.002	.001	.041	.023	.056
	SVSBE(∞ , ∞)	.003	.003	.002	.041	.023	.056
	SVCH2	.007	.006	.007	.041	.023	.055
	SVEXPS	.006	.006	.004	.038	.022	.050
MCP Lin. loc. (VCK)	SVSBE(1,1)	.001	.001	.001	.041	.024	.054
	SVSBE(∞ ,1)	.002	.001	.001	.041	.023	.055
	SVSBE(∞ , ∞)	.003	.002	.002	.040	.023	.055
	SVCH2	.006	.005	.005	.040	.023	.053
	SVEXPS	.005	.005	.006	.037	.022	.050
MV	SVCH2				.039	.031	.030
	SVEXPS				.025	.013	.030
MVR	SVCH2				.037	.028	.030
	SVEXPS				.021	.010	.028

Tabla 9.3: Medidas del ajuste al estimador piloto y al semivariograma teórico correspondientes a muestras de tamaño $n = 250$ generadas utilizando el modelo SVEXP como semivariograma teórico.

Como ejemplo adicional en la figura 9.4 se comparan las estimaciones del efecto nugget obtenidas con los diferentes métodos de estimación para el caso de muestras de tamaño 250 (aunque considerando únicamente el criterio VCE para la selección de la ventana en la estimación lineal local). En los métodos m.c.p se observa que al utilizar el estimador lineal local se obtienen estimaciones casi siempre las más cercanas al verdadero valor, especialmente al utilizar el modelo SVSBE(1,1) más flexible (algo esperado teniendo en cuenta los resultados mostrados en la sección anterior). En general se observó que los modelos SVSBE

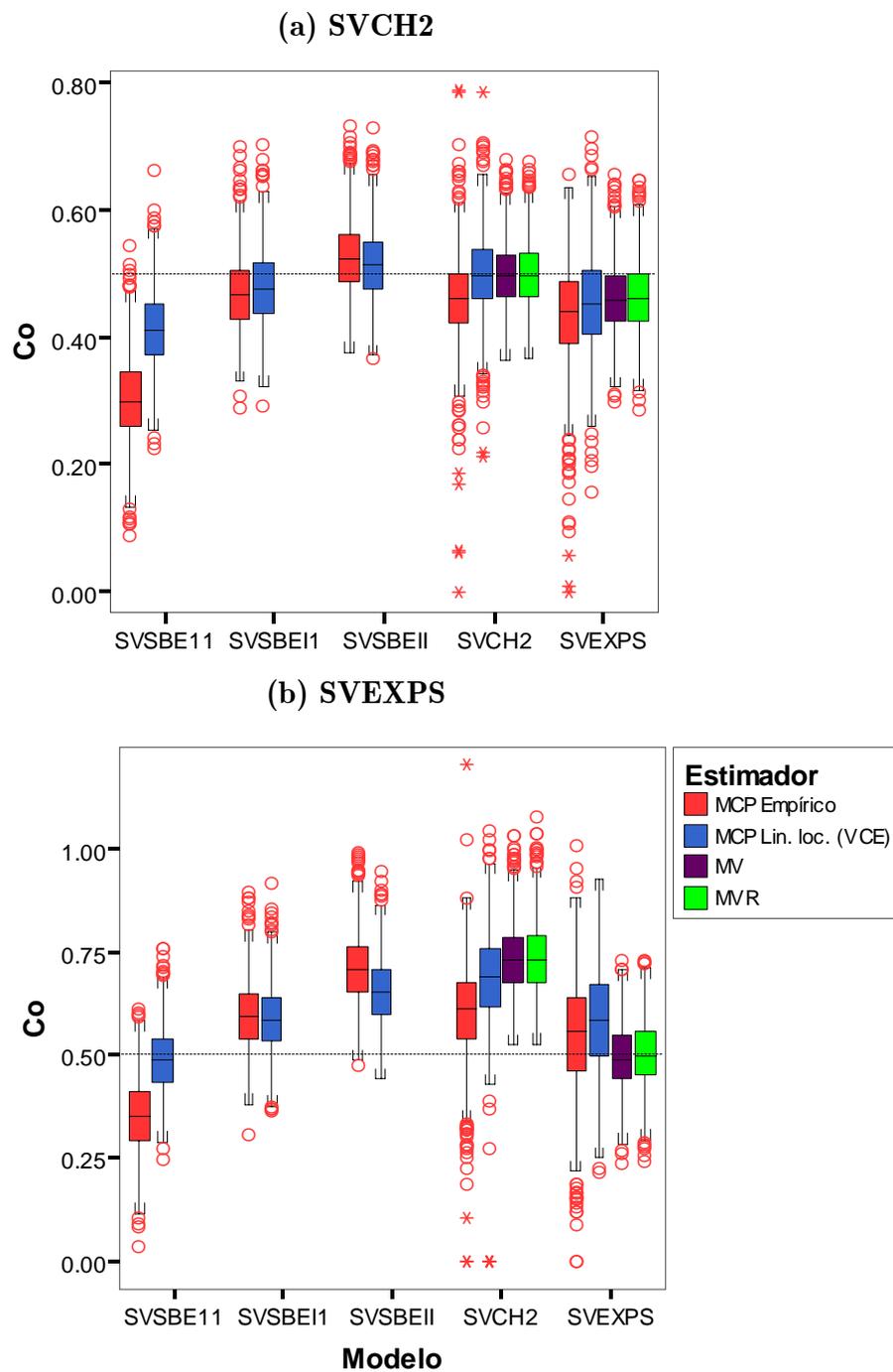


Figura 9.4: Estimaciones del efecto nugget utilizando los modelos SVCH2 (a) y SVEXPS (b) como semivariogramas teóricos para el caso de $n = 250$.

más flexibles obtienen estimaciones del efecto nugget menores que los más generales¹. En cuanto a los métodos basados en máxima verosimilitud, se observó también que en general obtenían los mejores resultados cuando se empleaba el modelo teórico en la estimación; empeorando igualmente cuando el modelo no estaba correctamente especificado.

Para evaluar si los modelos de variograma ajustados representaban adecuadamente la dependencia espacio-temporal de los datos se decidió emplear la técnica tradicional de validación cruzada descrita en la sección 4.4. Además del error cuadrático medio de validación cruzada:

$$\text{ECMVC} = \frac{1}{n} \sum_{i=1}^n \left(\hat{Z}_{-i}(\mathbf{s}_i, t_i) - Z(\mathbf{s}_i, t_i) \right)^2,$$

donde $\hat{Z}_{-i}(\mathbf{s}_i, t_i)$ es el predictor kriging de $Z(\mathbf{s}_i, t_i)$ calculado a partir del resto de observaciones, se calculó también la cobertura (COVC) de los correspondientes intervalos de predicción al 95% (suponiendo normalidad).

En las tablas 9.4 y 9.5 se muestran resúmenes de las medidas calculadas, para el caso de $n = 250$, utilizando los modelos teóricos SVCH2 y SVEXPS respectivamente. Aquí observamos también que al utilizar el estimador lineal local en la estimación m.c.p. se obtienen menores valores del ECMVC; sobre todo en el caso del modelo SVSBE(1,1) que es el peor con el estimador empírico y el mejor con el estimador lineal local. Además, podemos observar que con el criterio VCK para la selección de la ventana en la estimación lineal local, se obtienen resultados ligeramente mejores que con el criterio VCE, especialmente en el caso de los modelos más flexibles. Si comparamos las coberturas de los intervalos de predicción (COVC) podemos ver que en estos casos son todos bastante buenos, exceptuando el método m.c.p con el modelo SVSBE(1,1) y el estimador empírico. En general se observó también que cuando al utilizar el método de m.c.p. con el estimador lineal local se obtenían mejores valores de CVCOV en todos los casos.

¹ Esto aparentemente es también debido al empleo del criterio de selección de los nodos basado en los ceros de las funciones de Bessel. Por ejemplo, en la figura 7.1 se mostraban estimaciones del efecto nugget utilizando el primer criterio descrito en la sección 6.2.3, mostrando en aquel caso un mejor comportamiento de los modelos SVSBE en la estimación del efecto nugget al emplear el estimador empírico.

Respecto a los métodos de máxima verosimilitud, se obtuvieron también en general buenos resultados cuando el modelo estaba especificado correctamente (aunque inferiores en muchos casos a los obtenidos utilizando el método m.c.p. con el estimador lineal local); empeorando considerablemente cuando se empleaba un modelo distinto al semivariograma teórico.

Estimación		ECMVC			COVC	
Método	Modelo	Media	Mediana	Des.Std.	Media	Des.Std.
MCP Empírico	SVSBE(1,1)	.610	.607	.062	.927	.017
	SVSBE(∞ ,1)	.588	.585	.054	.946	.012
	SVSBE(∞ , ∞)	.592	.588	.055	.949	.010
	SVCH2	.639	.590	1.461	.947	.020
	SVEXPS	.605	.602	.057	.944	.016
MCP Lin. loc. (VCE)	SVSBE(1,1)	.583	.579	.055	.947	.011
	SVSBE(∞ ,1)	.585	.580	.054	.949	.011
	SVSBE(∞ , ∞)	.589	.585	.054	.950	.010
	SVCH2	.591	.589	.055	.951	.012
	SVEXPS	.603	.601	.056	.946	.013
MCP Lin. loc. (VCK)	SVSBE(1,1)	.580	.577	.054	.948	.011
	SVSBE(∞ ,1)	.584	.580	.054	.949	.011
	SVSBE(∞ , ∞)	.589	.585	.054	.950	.010
	SVCH2	.591	.588	.055	.951	.012
	SVEXPS	.603	.599	.056	.946	.013
MV	SVCH2	.587	.583	.053	.951	.009
	SVEXPS	.593	.591	.054	.952	.009
MVR	SVCH2	.586	.584	.053	.951	.009
	SVEXPS	.593	.591	.054	.952	.009

Tabla 9.4: Medidas de validación cruzada correspondientes a muestras de tamaño $n = 250$ generadas utilizando el modelo SVCH2 como semivariograma teórico.

Estimación		ECMVC			COVC	
Método	Modelo	Media	Mediana	Des.Std.	Media	Des.Std.
MCP Empírico	SVSBE(1,1)	.835	.831	.088	.922	.021
	SVSBE(∞ ,1)	.802	.799	.077	.945	.013
	SVSBE(∞ , ∞)	.824	.820	.080	.952	.011
	SVCH2	.831	.827	.085	.948	.017
	SVEXPS	.809	.801	.084	.949	.018
MCP Lin. loc. (VCE)	SVSBE(1,1)	.785	.781	.077	.948	.012
	SVSBE(∞ ,1)	.791	.788	.075	.950	.011
	SVSBE(∞ , ∞)	.806	.800	.076	.951	.011
	SVCH2	.835	.829	.087	.955	.012
	SVEXPS	.805	.798	.079	.954	.013
MCP Lin. loc. (VCK)	SVSBE(1,1)	.778	.775	.075	.949	.011
	SVSBE(∞ ,1)	.789	.786	.075	.950	.011
	SVSBE(∞ , ∞)	.805	.800	.076	.951	.011
	SVCH2	.834	.830	.082	.955	.012
	SVEXPS	.805	.799	.080	.954	.013
MV	SVCH2	.832	.828	.082	.960	.010
	SVEXPS	.783	.778	.073	.951	.010
MVR	SVCH2	.831	.827	.082	.960	.010
	SVEXPS	.783	.779	.073	.951	.010

Tabla 9.5: Medidas de validación cruzada correspondientes a muestras de tamaño $n = 250$ generadas utilizando el modelo SVEXPS como semivariograma teórico.

9.3 PREDICCIONES

Para comparar las predicciones obtenidas utilizando los diferentes métodos, para cada posición de predicción (\mathbf{s}_0, t_0) se calcularon los errores cuadráticos de predicción:

$$ECP = (\hat{Z}(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0))^2,$$

y las coberturas (COP) de los intervalos de predicción al 95% suponiendo

normalidad. En las tablas 9.6 y 9.7 se muestran un resumen² de los valores obtenidos para el caso de $n = 250$.

Estimación		ECP			COP	
Método	Modelo	Media	Mediana	Des.Std.	Media	Des.Std.
MCP Empírico	SVSBE(1,1)	.698	.309	.974	.916	.278
	SVSBE(∞ ,1)	.650	.304	.916	.945	.229
	SVSBE(∞ , ∞)	.647	.313	.911	.952	.214
	SVCH2	.645	.299	.908	.949	.221
	SVEXPS	.653	.305	.920	.950	.218
MCP Lin. loc. (VCE)	SVSBE(1,1)	.651	.304	.906	.947	.225
	SVSBE(∞ ,1)	.643	.305	.907	.948	.223
	SVSBE(∞ , ∞)	.644	.305	.908	.954	.210
	SVCH2	.640	.303	.901	.955	.207
	SVEXPS	.653	.303	.919	.950	.217
MCP Lin. loc. (VCK)	SVSBE(1,1)	.649	.304	.907	.946	.226
	SVSBE(∞ ,1)	.643	.304	.907	.948	.222
	SVSBE(∞ , ∞)	.644	.305	.907	.954	.210
	SVCH2	.640	.303	.902	.956	.204
	SVEXPS	.653	.308	.917	.950	.217
MV	SVCH2	.632	.302	.888	.953	.211
	SVEXPS	.647	.307	.916	.955	.207
MVR	SVCH2	.631	.303	.885	.955	.208
	SVEXPS	.646	.307	.915	.956	.205

Tabla 9.6: Medidas de eficiencia de las predicciones³ correspondientes a muestras de tamaño $n = 250$ y al semivariograma teórico SVCH2.

² Hay que tener en cuenta que, al considerar de forma conjunta los valores de las tres posiciones de predicción, los valores del ECP correspondientes a la posición 3 tienen un efecto dominante sobre la media global.

³ En el cálculo de estas medidas se eliminó un valor atípico correspondiente al modelo SVCH2 con el estimador empírico.

Estimación		ECP			COP	
Método	Modelo	Media	Mediana	Des.Std.	Media	Des.Std.
MCP Empírico	SVSBE(1,1)	.994	.436	1.432	.914	.280
	SVSBE(∞ ,1)	.932	.428	1.362	.942	.234
	SVSBE(∞ , ∞)	.940	.441	1.354	.950	.219
	SVCH2	.928	.419	1.330	.947	.225
	SVEXPS	.914	.424	1.314	.952	.214
MCP Lin. loc. (VCE)	SVSBE(1,1)	.927	.436	1.334	.939	.240
	SVSBE(∞ ,1)	.923	.419	1.343	.943	.231
	SVSBE(∞ , ∞)	.920	.421	1.330	.946	.227
	SVCH2	.927	.418	1.317	.954	.210
	SVEXPS	.912	.430	1.314	.956	.204
MCP Lin. loc. (VCK)	SVSBE(1,1)	.924	.425	1.337	.940	.237
	SVSBE(∞ ,1)	.921	.419	1.344	.946	.227
	SVSBE(∞ , ∞)	.920	.419	1.332	.947	.224
	SVCH2	.930	.422	1.324	.953	.212
	SVEXPS	.909	.423	1.301	.957	.203
MV	SVCH2	.927	.418	1.319	.957	.203
	SVEXPS	.889	.411	1.277	.950	.217
MVR	SVCH2	.926	.419	1.317	.957	.204
	SVEXPS	.888	.406	1.275	.953	.212

Tabla 9.7: Medidas de eficiencia de las predicciones correspondientes a muestras de tamaño $n = 250$ y al semivariograma teórico SVEXPS.

Respecto al error cuadrático de predicción vemos también que en la estimación m.c.p. se obtienen casi siempre mejores valores con el estimador lineal local que con el estimador empírico (principalmente en el caso del modelo SVSBE(1,1)). Entre los modelos SVSBE el que aparentemente obtiene mejores resultados en estos casos⁴ es el modelo SVSBE(∞ ,1). Como regla general, los

⁴ Aunque en el caso de muestras de tamaño 100, el que aparentemente obtuvo mejores resultados fue el modelo SVSBE(∞ , ∞).

mejores resultados se obtuvieron con la estimación basada en máxima verosimilitud cuando el modelo estaba especificado correctamente (y siendo superior el método de MVR sobre el de MV en la mayoría de los casos), empeorando también cuando el modelo paramétrico escogido no coincide con el teórico.

Respecto a las coberturas de los intervalos de predicción, se observaron unos valores bastante bajos (al igual que en los resultados de validación cruzada) al utilizar el modelo SVSBE(1,1) con el estimador empírico, mejorando también de forma considerable con el estimador lineal local. Con los demás métodos de estimación se obtuvieron unos resultados bastante similares.

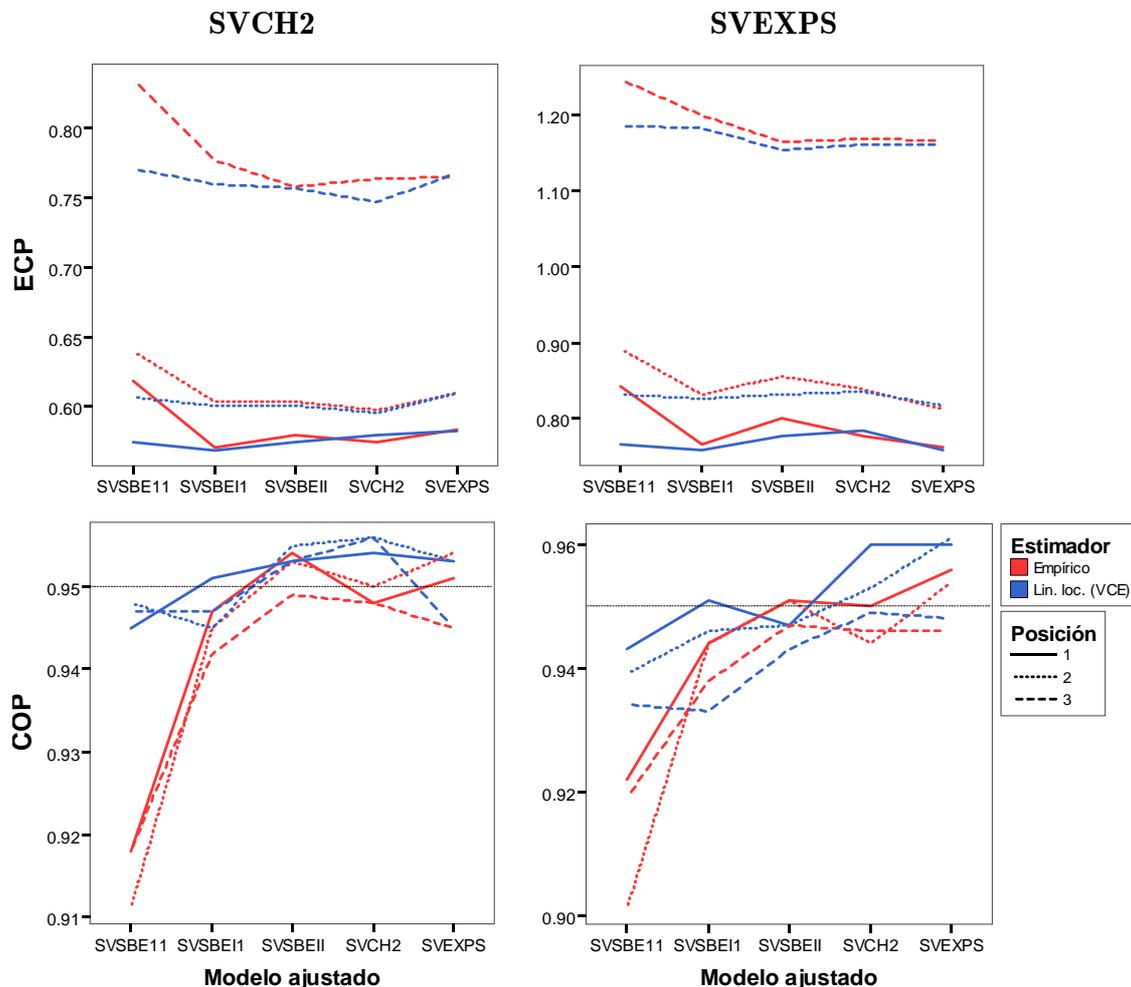


Figura 9.5: Valores medios de ECP y COP en las posiciones de predicción obtenidos con los métodos m.c.p (la línea horizontal representa el valor teórico 0.95).

Finalmente se estudió también la posible existencia de diferentes comportamientos de los modelos ajustados dependiendo de la situación (interior o exterior) de la posición de predicción. Por ejemplo, en la figura 9.5 se muestra un resumen de los valores obtenidos con los diferentes métodos m.c.p. para el caso de $n = 250$ (y utilizando el criterio CVE para la selección de la ventana en la estimación lineal local). Aunque no se apreciaron grandes diferencias en los valores ECP obtenidos con los modelos SVSBE, en general se observó que en los puntos interiores (posiciones 1 y 2) el modelo SVSBE($\infty, 1$) es el que presentaba un mejor comportamiento, mientras que en el exterior de las observaciones (posición 3) el modelo SVSBE(∞, ∞) resultó ser el aparentemente preferible. En la mayoría de los casos se observó también que con el estimador lineal local se obtenían mejores resultados. Respecto a las coberturas de los intervalos de predicción, destacan también los valores bajos al utilizar el estimador empírico con el modelo SVSBE(1,1), especialmente en la posición 2 (mejorando de forma muy significativa al considerar el estimador lineal local).

9.4 CONCLUSIONES

Teniendo en cuenta las observaciones y resultados anteriores podemos realizar las siguientes afirmaciones:

- La combinación de los modelos de Shapiro-Botha extendidos con el estimador lineal local del semivariograma puede ser un buen método para resolver el problema del modelado de la dependencia espacio-temporal de un conjunto de datos.
- Esta aproximación puede llevarse a cabo fácilmente en la práctica, utilizando en la estimación piloto del semivariograma herramientas ya disponibles para la estimación no paramétrica de superficies y realizando el posterior ajuste de los modelos SVSBE mediante programación cuadrática (lo que evita los problemas relacionados con la minimización no lineal multidimensional que pueden aparecer al utilizar otros métodos).
- Las comparaciones realizadas en el estudio de la simulación nos muestran que los resultados obtenidos utilizando la aproximación propuesta son similares (y en ocasiones mejores) a los obtenidos utilizando metodologías

paramétricas tradicionales. Aunque otros métodos, como por ejemplo los basados en máxima verosimilitud, pueden producir mejores resultados cuando la distribución de los datos es normal y el modelo paramétrico está especificado correctamente, al utilizar estos métodos en la práctica es importante asegurarse de que las suposiciones paramétricas son adecuadas. Sin embargo, con la aproximación propuesta se evitan problemas de este tipo.

- En el estudio de simulación se observa también que, en la estimación piloto del semivariograma, el estimador lineal local proporciona resultados claramente mejores a los obtenidos con el estimador empírico. Además, como no se observaron grandes diferencias entre los dos criterios para la selección de la ventana, se podría recomendar en principio el uso del método CVE cuando el número de observaciones es grande (ya que requiere de mucho menos tiempo de computación). La selección de la ventana sin embargo es todavía un tema abierto; por ejemplo, el empleo de ventanas locales puede mejorar de forma significativa la eficiencia de las estimaciones piloto obtenidas. Sería importante realizar un estudio sobre este estimador similar al efectuado por García-Soidán et al. (2003b) para el caso espacial isotrópico; además de ser de utilidad para la determinación de mejores criterios para la selección de la ventana, podría permitir el ajuste de un modelo válido mediante m.c.g. (algo que seguramente también mejoraría los resultados obtenidos con los modelos SVSBE).
- En el ajuste de los modelos SVSBE, la selección de los puntos de discretización a partir de los ceros de funciones de Bessel puede ser una buena opción en la práctica cuando el número de estimaciones piloto es grande; siendo en este caso especialmente recomendable el empleo del estimador lineal local del semivariograma (también puede ser conveniente utilizar modelos no muy flexibles). Cuando se considera un modelo muy flexible (p.e. el modelo SVSBE(1,1) o SVSBE(2,1)), se suelen obtener buenos ajustes en torno al origen empleando el primer criterio mostrado en la sección 6.2.3 (puntos regularmente espaciados), siempre que el tiempo de computación no sea prohibitivo (el número de nodos debería ser pequeño). Todavía es necesaria una mayor investigación sobre este tema que permita

disponer de criterios suficientemente rápidos para la selección “óptima” de los puntos de discretización (incluyendo valores para I y J).

- Finalmente, esta aproximación puede extenderse fácilmente para el caso de tres (o más) componentes anisotrópicos; por ejemplo, para el caso espacio-temporal con anisotropía en las coordenadas espaciales. También se podrían utilizar de forma análoga los modelos completamente anisotrópicos mostrados en la sección 4.5.4; aunque todavía es un problema abierto el encontrar modelos flexibles más simples que permitan modelar asimetrías espacio-temporales (ver sección 6.2).

BIBLIOGRAFÍA

- Abramowitz, M. y I.A. Stegun (1965). *HandBook of Mathematical Functions*. Dover, New York.
- Albert, M., Orts, V. y J. Mateu (2002). Statistical tools for spatial economics. *Spatial Statistics Through Applications*, J. Mateu y F. Montes (Eds.), WITPress, Southampton, UK., 313-345
- Angulo, J.M., W. González-Manteiga, M. Febrero-Bande y F.J. Alonso (1998). Semi-parametric statistical approaches for space-time process prediction. *Environmental and Ecological Statistics*, 5, 297-316.
- Armstrong, M. (1984). Problems with universal kriging. *Mathematical Geology*, 16, 101-108.
- Baczkowski, A.J. y K.V. Mardia (1987). Approximate log-normality of the sample semivariogram under a Gaussian process. *Communications in Statistics. Simulation and Computation*, 16, 447-461.
- Ball, J.S. (2000). Automatic computation of zeros of Bessel functions and other special functions. *Journal of Scientific Computing*, 21, 1458-1464.
- Barnes, R.J. y T.B. Johnson (1984). Positive kriging. *Geostatistics for Natural Resources Characterization*. G. Verly, M. David, A. Journel y A. Marechal (Eds.). Reidel, Dordrecht, 1, 231-244.
- Barry, R.P. y J.M. Ver Hoef (1996). Blackbox kriging: spatial prediction without specifying variogram models, *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297-322.
- Bastin, G. y M. Gevers (1985). Identification and optimal estimation of random fields from scattered point-wise data. *Automatica*, 21, 139-155.
- Beckers, F. y P. Bogaert (1998). Nonstationarity of the mean and unbiased variogram estimation: extension of the weighted least square method. *Mathematical Geology*, 30, 223-240.

- Benes, V. y K. Bodlák (2003). Application of log Gaussian Cox processes in disease mapping. *Proceedings of ISI International Conference on Environmental Statistics and Health*. J. Mateu, D. Holland y W. Gonzalez-Manteiga (Eds.), Universidade de Santiago de Compostela, 95-105.
- Bengtsson, T. y D. Nychka (2001). Adaptive methods in numerical weather prediction. *Spatio-Temporal Modelling of Environmental Processes. Proceedings of the I Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes*. J. Mateu y F. Montes (Eds.). Editorial Universitat Jaume I, Castellón, 15-28.
- Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. University of California Press, Berkeley and Los Angeles, CA.
- Carrol, R.J. y D. Ruppert (1982). A comparison between maximum likelihood and generalized least squares in a heterocedastic linear model. *JASA*, 77, 878-882
- Carrol, R.J., R. Chein, E.I. George, T.H. Li, H.J. Newton, H. Schmiediche y N. Wang (1997). Ozone exposure and population density in Harris County, Texas (con discusión). *JASA*, 92, 392-415.
- Chao, L. (1998). Spectral estimation in space and time domain by nonstationary minimum variance spectral estimator. *Mathematical Geology*, 30, 675-693.
- Cherry, S. (1997). Non-parametric estimation of the sill in geostatistics. *Environmetrics*, 8, 13-27.
- Cherry, S., J. Banfield y W.F. Quimby (1996). An evaluation of a non-parametric method of estimating semivariograms of isotropic spatial processes. *Journal of Applied Statistics*, 23, 435-449.
- Chilès, J.P. y P. Delfiner (1999). *Geostatistics: modeling spatial uncertainty*. Wiley, New York.
- Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research*, 20, 251-265.
- Christakos, G. (1992). *Random Fields Models in Earth Sciences*. Academic Press.
- Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*, Oxford University Press, New York.

- Christensen, R. (1991). *Linear models for multivariate, time series, and spatial data*. Springer-Verlag, New York.
- Clark, I., K.L. Basinger y W.V. Harper (1989). MUCK: A novel approach to co-kriging. *Proceedings of the Conference on Geostatistical, Sensitivity and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling*, B.E. Buxon (Ed.), Battelle Press, Columbus, OH, 473-493.
- Cramer, H. (1940). On the theory of stationary random processes. *Annals of Mathematics*, 41, 215-230.
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17, 563-586.
- Cressie, N. (1986). Kriging nonstationary data. *Journal of the American Statistical Association*, 81, 625-634.
- Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22, 239-252.
- Cressie, N. (1993). *Statistics for Spatial Data*. Ed. rev. John Wiley & Sons.
- Cressie, N. y C.K. Wikle (1998). The variance-based cross-variogram: you can add apples and oranges. *Mathematical Geology*, 30, 789-799.
- Cressie, N. y D.M. Hawkins (1980). Robust estimation of the variogram, I. *Mathematical Geology*, 12, 115-125.
- Cressie, N. y H-L. Huang (1999). Classes of non-separable, spatio-temporal stationary covariance functions, *JASA*, 94, 1330-1340.
- Cressie, N. y M.O. Grondona, (1992). A comparison of variogram estimation with covariogram estimation. *The Art of Statistical Science*, K.V. Mardia (Ed.), Wiley, Chichester, 191-208.
- Cressie, N., H.S. Stern y D.R. Wright (2000). Mapping rates associated with polygons. *Journal of Geographical Systems*, 2, 61-69.
- Davis, B.M. y L.E. Borgman (1979). Some exact sampling distributions for variogram estimators. *Mathematical Geology*, 11, 643-653.
- De Cesare, L., D.E. Myers y D. Posa (2001). Estimating and modeling space-time correlation structures. *Statistics & Probability Letters*, 51, 9-14.
- De Iaco, S., D.E. Myers y D. Posa (2001). Space-time analysis using a general product-sum model. *Statistics & Probability Letters*, 52, 21-28.
- De Iaco, S., D.E. Myers y D. Posa (2002a). Nonseparable space-time covariance models: some parametric families. *Mathematical Geology*, 34, 23-42.

- De Iaco, S., D.E. Myers y D. Posa (2002b). Space-time variograms and a functional form for total air pollution measurements. *Computational Statistics & Data Analysis*, 41, 311-328.
- De Iaco, S., D.E. Myers y D. Posa (2003). The linear coregionalization model and the product-sum space-time variogram. *Mathematical Geology*, 35, 25-38.
- Deutsch, C.V. y A.G. Journel (1992). *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York.
- Dibiasi, A. y A.W. Bowman (2001). On the use of the variogram in checking for independence in spatial data. *Biometrics*, 57, 211-218.
- Dimitrakopoulos, R. y X. Luo (1994). Spatiotemporal modeling: covariances and ordinary kriging systems. *Geostatistics for the next century*, R. Dimitrakopoulos (Ed.), Kluwer Academic Publishers, 88-93.
- Dutter, R. (1996). On robust estimation of variograms in geostatistics, *Lecture Notes in Statistics*, 109, Springer, New York, 153-171.
- Egbert, G.D. y D.P. Lettenmaier (1986). Stochastic modeling of the space-time structure of atmospheric chemical deposition. *Water Resources Research*, 22, 165-179.
- Fan, J. y I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Febrero-Bande, M., P. García-Soidán y W. González-Manteiga (1998). El método núcleo en la estimación del variograma. Comunicación en el *XXIV Congreso Nacional de Estadística e Investigación Operativa.*, Almería.
- Fernandes y Rivoirard (1998). A geostatistical analysis of the spatial distribution and abundance of cod, haddock and whiting in Nort Scotland. *Proceedings of the Second European Conference on Geostatistics for Environmental Applications, GEOENV'98*. Universidad Politécnica de Valencia, 169-180.
- Fernández-Casal, R. (1995). *Análisis estadístico de datos espaciales*. Tesina del departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela.
- Fernández-Casal, R. (2000). Algunas notas de interés sobre la validación cruzada de un modelo de variograma en kriging. Comunicación en el *XXV Congreso Nacional de Estadística e Investigación Operativa.*, Vigo 4-7 de Abril.

- Fernández-Casal, R., W. González-Manteiga y M. Febrero-Bande (2001). General classes of flexible spatio-temporal stationary variogram models. *Spatio-Temporal Modelling of Environmental Processes. Proceedings of the I Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes*. J. Mateu y F. Montes (Eds.). Editorial Universitat Jaume I, Castellón, 51-69.
- Fernández-Casal, R., W. González-Manteiga y M. Febrero-Bande (2003a). Flexible spatio-temporal stationary variogram models. *Statistics and Computing*, 13, 127-136.
- Fernández-Casal, R., W. González-Manteiga y M. Febrero-Bande (2003b). Space-time dependency modeling using general classes of flexible stationary variogram models. *Journal of Geophysical Research-Atmosphere*, 108 (D24), 8779, special issue: Application of Recent Advances in Space-Time Statistics to Atmospheric Data.
- Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, 89, 197-210.
- Fuentes, M. y R.L. Smith (2001). A new class of nonstationary spatial models. *Manuscrito no publicado* (descargado de www4.stat.ncsu.edu/~fuentes/sdf.ps).
- Fuller, W.A. (1996). *Introduction to Statistical Time Series*. 2ª ed. Wiley, New York.
- García-Soidán, P., M. Febrero-Bande y W. González-Manteiga (2003a). Nonparametric kernel estimation of an isotropic semivariogram. *Journal of Statistics and Planning Inference*. *En prensa*.
- García-Soidán, P., M. Febrero-Bande y W. González-Manteiga (2003b). Local linear regression estimation of the semivariogram. *Statistics & Probability Letters*, 64, 169-179.
- Gelfand, A.E., H.J. Kim y C.F. Sirmans (2001). Spatial modeling with spatially varying coefficient processes. *Spatio-Temporal Modelling of Environmental Processes. Proceedings of the I Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes*. J. Mateu y F. Montes (Eds.). Editorial Universitat Jaume I, Castellón, 95-110.

- Genton, M.G. (1998). Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Mathematical Geology*, 30, 323-345.
- Genton, M.G. y D.J. Gorsch (2003). Nonparametric variogram and covariogram estimation with Fourier-Bessel matrices. *Computational Statistics and Data Analysis*, Special issue on Matrix Computations and Statistics, 41, 47-57.
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *JASA*, 97, 590-600.
- Golberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57, 369-375.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Goovaerts, P. (1998). Ordinary cokriging revisited. *Mathematical Geology*, 30, 21-42.
- Gotway, C. A. y N. Cressie (1993). Improved multivariate prediction under a general linear model. *Journal of Multivariate Analysis*, 45, 56-72.
- Goulard, M. (1989). Inference in a Coregionalization Model. *Geostatistics*, M. Armstrong (Ed.), Kluwer Academic Publishers, 1, 397-408.
- Goulard, M. y M. Voltz (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24, 269-286.
- Guttorp, P., y P. Sampson (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. Handbook of statistics, G.P. Patil y C.R. Rao (Eds.), Elsevier Science, North-Holland, Amsterdam, 12, 661-689.
- Guttorp, P., P.D. Sampson y K. Newman (1992). Non-parametric estimation of spatial covariance with application to monitoring network evaluation. *Statistics in the Environmental and Earth Sciences*. Edward Arnold, London, 39-51.
- Haas, T.C. (1990). Lognormal and moving window methods of estimating acid deposition. *JASA*, 85, 950-963.

- Haas, T.C. (1995). Local prediction of a spatio-temporal process with an application to wet sulphate deposition. *JASA*, 90, 1189-1199.
- Hall, P. y P. Patil (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab. Theory Related Fields*, 99, 399-424.
- Hall, P., N.I. Fisher y B. Hoffman (1994). On the nonparametric estimation of covariance functions. *Annals of Statistics*, 22, 2115-2134.
- Handcock, M.S. y J.R. Wallis (1994). An approach to statistical spatial-temporal modeling of meteorological fields (con discusión). *JASA*, 89, 368-390.
- Härdle, W. y P. Vieu (1992). Kernel regression smoothing of time series. *J. Time Ser. Anal.*, 13, 209-232.
- Hawkins, D. M. y N. Cressie (1984). Robust kriging - a proposal. *Mathematical Geology*, 16, 3-18.
- Helterbrand, J.D. y N. Cressie (1994). Universal cokriging under intrinsic coregionalization. *Mathematical Geology*, 26, 205-226.
- Høst, G., H. More y P. Switzer (1995). Spatial interpolation errors for monitoring data. *JASA*, 90, 853-861.
- Irwin, M.E., N. Cressie y G. Johannesson (2002). Spatial-temporal nonlinear filtering based on hierarchical statistical models. *Test*, 11, 249-302.
- Isaaks, E.H. y R.M. Srivastava (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York.
- Jones, R.H. y Y. Zhang (1997). Models for continuous stationary space-time processes. *Modeling Longitudinal and Spatially Correlated Data. Lecture Notes in Statistics*, 122, Springer, New York, 289-298.
- Journel, A.G. (1986). Geostatistics: models and tools for the earth sciences. *Mathematical Geology*, 18, 119-140.
- Journel, A.G. (1999). Markov models for cross-covariances. *Mathematical Geology*, 31, 955-964.
- Journel, A.G. y C.J. Huijbregts (1978). *Mining Geostatistics*. Academic Press, New York.
- Künsch, H.R., A. Papritz y F. Bassi, (1997). Generalized cross-covariances and their estimati3n. *Mathematical Geology*, 29, 779-799.
- Kyriakidis, P.C. y A.G. Journel (1999). Geostatistical space-time models: a review. *Mathematical Geology*, 30, 589-615.

- Lahiri, S.N., Y. Lee y N. Cressie (2003). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. Para aparecer en *Journal of Statistical Planning and Inference*.
- Long, A.E. y D.E. Myers (1997). A new form of the cokriging equations. *Mathematical Geology*, 29, 685-703.
- Ma, C. (2002). Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology*, 34, 965-975.
- Ma, C. (2003). Nonstationary covariance functions that model space-time interactions. *Statistics & Probability Letters*, 61, 411-419.
- Mardia, K.V. y A.J. Watkins (1989) On multimodality of the likelihood in the spatial linear model. *Biometrika*, 76, 289-295.
- Mardia, K.V. y R.J. Marshall (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71, 135-146.
- Mateu, J. y P.J.Jr. Ribeiro (1998). Geostatistical data versus point process data: analysis of second-order characteristics. *Proceedings of the Second European Conference on Geostatistics for Environmental Applications, GEOENV '98*. Universidad Politécnica de Valencia, 181-192.
- Matheron, G. (1962). Traite de geostatistique appliquee, Tome I. *Memories du Bureau de Recherches Geologiques et Minières*, 14. Ediciones Bureau de Recherches Geologiques et Minières, Paris.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58, 1246-1266.
- Matheron, G. (1971). The theory of regionalized variables and its applications. *Cahiers du Centre de Morphologie Mathématique*, 5. Fontainebleau, France.
- Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, 5, 439-468.
- Matheron, G. (1978). *Estimateur et Choisir*. Ecoles des Mines de Paris, Fontainebleau, France. Traducción al inglés: *Estimating and Choosing*. Springer-Verlag, New York, 1989.
- Mugglin, A.S., N. Cressie y I. Gemmell (2002). Hierarchical statistical modeling of influenza-epidemic dynamics in space and time. *Statistics in Medicine*, 21, 2703-2721.

- Müller, W.G. (2003). Spatial design methods in case of correlated observations: a comparison. *Proceedings of ISI International Conference on Environmental Statistics and Health*. J. Mateu, D. Holland y W. Gonzalez-Manteiga (Eds.), Universidade de Santiago de Compostela, 21-32.
- Myers D.E. y A. Journel (1990). Variograms with zonal anisotropies and non-invertible kriging systems. *Mathematical Geology*, 22, 779-785
- Myers, D.E. (1982). Matrix formulation of co-kriging. *Mathematical Geology*, 14, 249-257.
- Myers, D.E. (1991). Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology*, 23, 805-816.
- Neuman, J. e I.J. Schoenberg (1941). Fourier integrals and metric geometry. *Transactions of the American Mathematical Society*, 50, 226-251.
- Neuman, S.P. y E.A. Jacobson (1984). Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. *Mathematical Geology*, 16, 499-521.
- Niu, X.F. (1996). Nonlinear additive models for environmental time series, with applications to ground-level ozone data analysis. *JASA*, 91, 1310-1321.
- Papritz, A., H.R. Künsch y R. Webster (1993). On the pseudo cross-variogram. *Mathematical Geology*, 25, 1015-1026.
- Pardo-Igúzquiza, E. (1997). MLREML: A computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. *Computers & Geosciences*, 23, 153-162.
- Pardo-Igúzquiza, E. (1998). Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, 20, 95-108.
- Rehman, S.U. y A. Shapiro (1996). An integral transform approach to cross-variogram modeling. *Computational Statistics and Data Analysis*, 22, 213-233.
- Ripley, B.D. (1988). *Statistical inference for spatial processes*. Cambridge University Press, Cambridge.
- Rivoirard, J. (1987). Two key parameters when choosing the kriging neighborhood. *Mathematical Geology*, 19, 851-886.
- Rivoirard, J. (1994). *Introduction to Disjunctive Kriging and Non-Linear Geostatistics*. Oxford University Press, New York.

- Rouhani, S. y D.E. Myers (1990). Problems in Space-Time Kriging of Hydrogeological data. *Mathematical Geology*, 22, 611-623
- Rouhani, S. y T.J. Hall (1989), Space-time kriging of groundwater data. *Geostatistics*, M. Armstrong (Ed.), Kluwer Academic Publishers, 2, 639-651.
- Ruppert, D. y M.P. Wand (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22, 1346-1370.
- Samper Calvete, F.J. y J. Carrera Ramirez (1990). *Geoestadística. Aplicaciones a la hidrología subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería, Barcelona.
- Samper, F.J. y S.P. Neuman (1989). Estimation of spatial covariance structures by adjoint state maximum likelihood cross-validation: 1. Theory. *Water Resources Research*, 25, 351-362.
- Sampson, P.D. y P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *JASA*, 87, 108-119
- Schoenberg, I.J. (1938). Metric spaces and completely monotone functions. *Annals of mathematics*, 39, 811-841.
- Searle, S.R. (1971). *Linear models*. Wiley, New York.
- Shapiro, A. y J.D. Botha (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis*, 11, 87-96.
- Shimizu, K. y K. Iwase (1987). Unbiased estimation of the autocovariance function in a stationary generalized lognormal process. *Communications in Statistics. Theory and Methods*, 16, 2145-2154.
- Smith, R.L., S. Kolenikov y L.H. Cox (2003). Spatio-temporal modeling of PM2.5 data with missing values. *Journal of Geophysical Research-Atmosphere*, special issue: Application of Recent Advances in Space-Time Statistics to Atmospheric Data. *En prensa*. (disponible en www.stat.unc.edu/faculty/rs/papers/RLS_Papers.html).
- Solow, A.R. y S.M. Gorelick (1986). Estimating missing streamflow values by cokriging. *Mathematical Geology*, 18, 785- 809.
- Stein, M.L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Annals of Statistics*, 16, 55-63.

- Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- Stein, M.L. (2003). Space-time covariance functions. *Manuscrito en preparación* (descargado de www.stat.uchicago.edu/~cises).
- Stoica, R.S., M.N.M. van Lieshout, X. Descombes y J. Zerubia (2002). An application of marked point processes to the extraction of linear networks from images. *Spatial Statistics Through Applications*, J. Mateu y F. Montes (Eds.), WITPress, Southampton, UK., 287-312.
- Turlach, B.A. y M.P. Wand (1996). Fast computation of auxiliary quantities in local polynomial regression. *Journal of Computational and Graphical Statistics*, 5, 337-350.
- Vargas-Guzmán, J.A., A.W. Warrick y D.E. Myers (2002). Coregionalization by linear combination of nonorthogonal components. *Mathematical Geology*, 34, 405-419.
- Ver Hoef, J.M. y Barry, R.P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69, 275-294.
- Ver Hoef, J.M. y N. Cressie (1993). Multivariable spatial prediction. *Mathematical Geology*, 25, 219-240.
- Ver Hoef, J.M. y R.P. Barry (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69, 275-294.
- Ver Hoef, J.M., N. Cressie y R.P. Barry (2003). Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT). Enviado al *Journal of Computational and Graphical Statistics*.
- Wackernagel, H. (1998). *Multivariate Geostatistics*. 2ª ed. Springer, Berlin.
- Wand, M.P. (1994). Fast Computation of Multivariate Kernel Estimators. *Journal of Computational and Graphical Statistics*, 4, 433-445.
- Wand, M.P. y M.C. Jones (1995). *Kernel Smoothing*. Chapman and Hall, London.
- Wang, Y., H.A. Chipman y W.J. Welch (2003). Modeling the relationship between chemical structure and activity for drug discovery. *Proceedings of ISI International Conference on Environmental Statistics and Health*. J. Mateu, D. Holland y W. Gonzalez-Manteiga (Eds.), Universidade de Santiago de Compostela, 173-181.

- Wikle, C.K. y N. Cressie (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86, 815-829.
- Xu, W., T. Tran, R.M. Srivastava y A.G. Journel (1992). Integrating seismic data in reservoir modeling: The collocated cokriging alternative. SPE paper # 24742.
- Yaglom, A.M. (1986). *Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results*. Springer-Verlag, New York
- Yakowitz, S.J. y F. Szidarovszky (1985). A comparison of kriging with nonparametric regression methods. *Journal of multivariate analysis*, 16, 21-53.
- Yao, T. y A.G. Journel (1998). Automatic modeling of (cross) covariance tables using fast fourier transform. *Mathematical Geology*, 30, 589-615.
- Yu, K. y J. Mateu (2002). Nonparametric nearest-neighbour variogram estimation. *Spatial Statistics Through Applications*, J. Mateu y F. Montes (Eds.), WITPress, Southampton, UK., 103-125.
- Zhu, J., S.N. Lahiri y N. Cressie (2002). Asymptotic inference for spatial CDFs over time. *Statistica Sinica*, 12, 843-861
- Zimmerman, D.L. y M.B. Zimmerman (1991). A Monte Carlo comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, 33, 77-91.