Geoestadística para datos espaciales, espacio-temporales y funcionales

Campos aleatorios espaciales

Martha Bohorquez

Facultad de Ciencias
Sede Bogotá

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Geoestadística para datos espaciales, espacio-temporales y funcionales

Campos aleatorios espaciales

Geoestadística para datos espaciales, espacio-temporales y funcionales

Campos aleatorios espaciales

Martha Bohorquez

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Bogotá, D. C., Colombia

Primera edición, 2025
ISBN 978-628-503-048-2 (digital)

Edición
Alejandro Cano Moreno
Coordinación de Publicaciones, Facultad de Ciencias, Sede Bogotá
coorpub_fcbog@unal.edu.co

Corrección de estilo
Diva Marcela Piamba Tulcán

Diseño de carátula
Damián Crofort

Maqueta LaTeX
Camilo Cubides

Editado en Bogotá, D. C., Colombia

Catalogación en la publicación Universidad Nacional de Colombia

Bohórquez Castañeda, Martha Patricia, 1972-
Geoestadística para datos espaciales, espacio-temporales y funcionales : campos aleatorios espaciales / Martha Bohorquez. -- Primera edición. -- Bogotá : Universidad Nacional de Colombia. Facultad de Ciencias : Coordinación de publicaciones Facultad de Ciencias, 2025.
1 recurso en línea (268 páginas) : ilustraciones (principalmente a color), diagramas, mapas. -- (Colección Textos. Estadística)

Incluye referencias bibliográficas e índice
ISBN 978-628-503-048-2 (digital)

1. Estadística matemática -- Estudio y enseñanza superior -- Problemas, ejercicios, etc. 2. Modelos lineales (Estadística) 3. Teorías no lineales 4. Teoría de la predicción -- Metodología 5. Procesos estocásticos 6. Estimación estadística -- Metodología -- Ejercicios y problemas 7. Análisis espacial (Estadística) -- Aplicaciones 8. Análisis de datos -- Métodos estadísticos 9. Geoestadística 10. Estadística espacial -- Procesamiento de datos 11. Campos aleatorios 12. Correlación (Estadística) 13. Autocorrelación (Estadística) 14. Distribución (Teoría de probabilidades) 15. Análisis multivariante 16. Krigeaje 17. Variograma 18. Análisis de covarianza 19. Análisis de regresión 20. Método de las diferencias de las variables 21. Muestreo (Estadística) 22. Minería de datos espaciales 23. Base de datos espacio-temporal 24. Paquetes de R 25. Análisis espacial (Estadística) -- Proceso de datos -- Obras educativas I. Título II. Serie

CDD-23 519.50711 / 2025

Contenido

Lista de figuras

Lista de tablas

Prefacio

En muchos fenómenos de la naturaleza, las variables de interés son observadas a través del tiempo, del espacio, del espacio-tiempo o de otras dimensiones equivalentes, tales como la frecuencia o la profundidad. Estos fenómenos suelen presentar estructura de autocorrelación en cada una de sus dimensiones, así como interacción entre ellas: cada momento del espacio-tiempo da información de los demás y viceversa. Los objetivos al analizar este tipo de fenómenos son diferentes a aquellos propuestos cuando los datos son independientes. Como consecuencia de ello, y de que en general no se cumplen los supuestos requeridos para los métodos usuales, se requiere una teoría especial para su estudio. Así surge la estadística espacio-temporal, que permite encontrar predicciones de las variables en lugares donde no hay observaciones, determinar cómo se interrelacionan o interactúan las variables de interés en diferentes ubicaciones y en diferentes momentos del tiempo, extender los modelos lineales y no lineales a este contexto, construir diseños de muestreo para poblaciones no finitas, por nombrar algunos casos.

Este texto hace un recorrido por la geoestadística desde sus inicios hasta la actualidad. Introduce al lector, empezando con la geoestadística espacial tanto univariada como multivariada, para luego presentar su generalización a la geoestadística espacio-temporal y a la geoestadística funcional, lo cual hace posible la aplicación de estos métodos sin limitaciones en presencia de grandes volúmenes de datos. Se construyen en detalle todos los predictores y los estimadores de las estructuras de covarianza y semivarianza. Se comentan aspectos prácticos que se deben tener en cuenta en todas las etapas del método, y se utilizan problemáticas reales para ilustrar su aplicación. Todas sus aplicaciones son construidas en el software libre R cran. La aplicación de todos los métodos abordados se puede llevar a cabo usando el paquete de R, SpatFD, (https://cran.r-project.org/web/packages/SpatFD/index.html) y el código y ejemplos disponibles en https://rpubs.com/mpbohorquezc y https://mpbohorquezc.github.io/SpatFD-Functional-Geostatistics/, donde también se encuentran varios conjuntos de datos de prueba.

Este libro se encuentra dirigido a todo aquel que se interese en el análisis de datos espaciales, temporales o datos similares definidos en dominios como la frecuencia; sin duda, es mejor abordarlo contando con conocimientos previos de regresión y estadística multivariada. Este libro es el producto de muchos años de docencia, investigación y aplicación en el área de geoestadística, y existe gracias al trabajo conjunto con los estudiantes de los cursos tanto de pregrado como de posgrado de estadística espacial, y a las discusiones con colegas en diferentes escenarios académicos, especialmente en la Universidad Nacional de Colombia.

Parte I

Geoestadística escalar espacial

Capítulo

uno

Introducción a los campos aleatorios espaciales

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

En este capítulo se presentan los conceptos generales de los campos aleatorios adaptados a la geoestadística, los tipos de estacionariedad y los parámetros de interés.

1.1. Motivación y definiciones básicas

Definición 1 (Proceso geoestadístico espacial (o campo aleatorio espacial)). *Se le llama proceso geoestadístico a un fenómeno que ocurre de manera continua en el espacio. Esto es, no hay lugares donde no exista el fenómeno. Lo que hace que el proceso sea continuo en el espacio son las características del conjunto índice en el que ocurre y no el tipo o los tipos de variables que se observan y modelan. Sea Y la variable de interés y sea $\boldsymbol{s}$ la ubicación espacial donde existe Y, el proceso espacial es el proceso estocástico*

$$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D_{\boldsymbol{s}} \subset \mathbb{R}^d\}$$

donde $d \in \mathbb{N}$, $D_{\boldsymbol{s}}$ está formado por todas las ubicaciones $\boldsymbol{s}$ y es el conjunto índice del proceso. La ubicación espacial $\boldsymbol{s}$ puede estar en una, dos o más dimensiones. Cuando $\boldsymbol{s}$ es un vector, esto es, a partir de $\mathbb{R}^2$, al proceso se le suele llamar campo aleatorio.

Lo más usual es analizar procesos en $\mathbb{R}^2$, ignorando la tercera dimensión, que puede ser la altitud o la profundidad, dado que los modelos se construyen localizados para áreas de interés pequeñas que se pueden considerar aproximadamente planas. No es adecuado tratar de abarcar con un mismo modelo áreas muy grandes, porque las condiciones del fenómeno pueden presentar características distintas. Para áreas grandes, se requiere incluir la tercera dimensión espacial, porque, de otro modo, la idealización del plano estaría muy lejos de la realidad. Además, puede ser necesario tener en cuenta otras covariables y las condiciones locales específicas de cada zona. Así, es posible que el comportamiento general de la variable sea muy diferente entre zonas y no sea correcto utilizar el mismo modelo global para toda la región, lo que implicaría también la necesidad de usar otros tipos de distancias.

Los procesos espaciales cuyo valor corresponde a un área en lugar de un punto, como por ejemplo tasas o indicadores según la división política de un país, tienen asociado un conjunto índice espacial discreto y requieren diferentes tipos de análisis. Usualmente, se ajustan modelos explicativos que

dependen de estructuras de vecindad o criterios de conexión entre las áreas [1].

Ejemplo 1. Reservorio acuífero *En el marco de un proyecto de conservación del medio ambiente, se requiere revisar y verificar en el terreno los informes relacionados con los cálculos de las reservas explotables y el avance del fenómeno de intrusión marina en un reservorio acuífero, con el fin de establecer las recomendaciones necesarias para definir un programa de manejo. Los datos recolectados son ubicados mediante coordenadas bidimensionales (Este, Norte). En una notación más universal, se puede pensar en un plano cartesiano en el cual dichas coordenadas se notan* (x, y) *(ver Figura 1.1).*

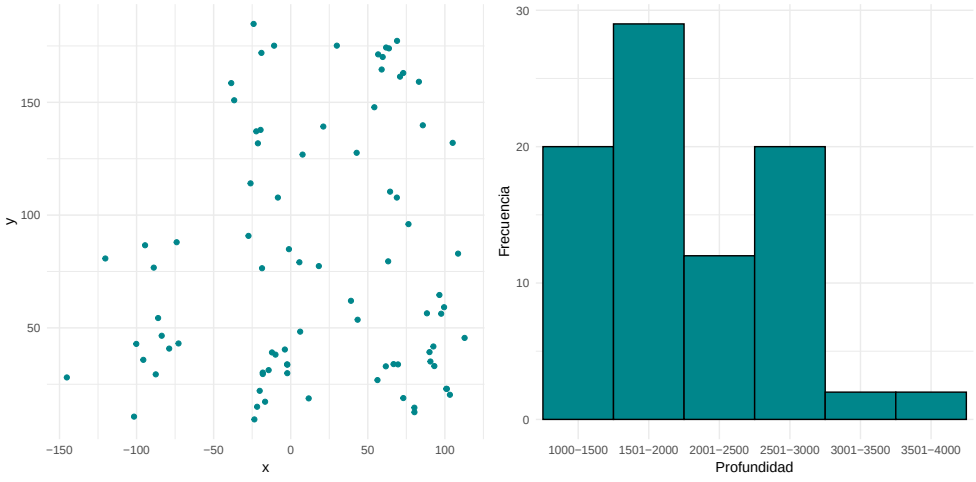


Figura 1.1. Panel izquierdo: ubicaciones $\boldsymbol{s} = (\mathrm{x}, \mathrm{y})$ en las cuales fue medida la altura piezométrica. Fuente [24]. Panel derecho: distribución de frecuencias de las observaciones de la altura piezométrica.

Aquí:

- $Y(\boldsymbol{s})$*: Profundidad piezométrica por debajo del nivel del mar en la ubicación* $\boldsymbol{s}$*.*
- $D_s \subset \mathbb{R}^2$*;* $\boldsymbol{s} \in \mathbb{R}^2$*,* $\boldsymbol{s} = (\mathrm{x}, \mathrm{y})$ *con un punto de origen arbitrario en* $(0, 0)$*, fijado en el momento en que se diseña el muestreo.*

Así, $\boldsymbol{s}$ *es un elemento de un conjunto* D_s*, el cual, en general, es un subconjunto de* $\mathbb{R}^d$*. El área* D_s *no es demasiado extensa para no mezclar procesos espaciales, así que*

no hay cambio de altitud, por lo que es suficiente utilizar una proyección al plano este-norte, $\boldsymbol{s} = (\mathrm{x}, \mathrm{y})$*, entonces* $\boldsymbol{s} \in D_{\boldsymbol{s}} \subset \mathbb{R}^2$*. En la tabla 1.1 se puede observar la profundidad piezométrica medida en algunas coordenadas [24]. El proceso espacial de interés es* $Y(s)$*: Altura piezométrica,* $y(\boldsymbol{s})$ *denota la realización de* $Y(s)$*.*

x	y	$y(\boldsymbol{s}) = y(\mathrm{x}, \mathrm{y})$
42.782	127.622	1464
-27.396	90.787	2553
-1.1628	84.896	2158
-18.618	76.451	2455
96.465	64.580	1756
108.562	82.923	1702
88.363	56.453	1805
90.042	39.258	1797
93.172	33.058	1714
97.610	56.278	1466

Tabla 1.1. Altura piezométrica medida en cada coordenada $\boldsymbol{s} = (\mathrm{x}, \mathrm{y})$, Fuente:[24].

Definición 2 (Proceso temporal). *Sea* Y *la variable de interés, y sea* t *el momento del tiempo en el que ocurre* Y*. Así, se tiene el proceso estocástico*

$$\{Y(t) : t \in D_T\}$$

donde $D_T \subset \mathbb{R}$ *es el conjunto de todos los tiempos y es su conjunto índice.*

El conjunto índice de un proceso temporal tiene una sola dimensión y es un caso particular del proceso espacial. Nótese que D_T puede ser discreto o continuo, esto es, existen procesos temporales en tiempo discreto como el precio de una acción en la bolsa de valores y procesos temporales en tiempo continuo como la temperatura. Uno de los objetivos principales, en el caso de series de tiempo, es encontrar pronósticos, que es una forma de extrapolación. Esto contrasta con los objetivos para datos espaciales en los que se requiere predecir el valor de la o las variables estudiadas, en lugares

no observados de la región de interés usando métodos de interpolación. Esta es la razón por la cual los métodos desarrollados en ambas áreas difieren.

1.1.1. Funciones de distribución finito-dimensionales

El valor observado en cada punto $\boldsymbol{s}$ es una realización $y(\boldsymbol{s})$ de una variable aleatoria espacial $Y(\boldsymbol{s})$. En términos matemáticos, la familia de todas estas variables aleatorias se denomina ***función aleatoria***, ***proceso estocástico*** o ***campo aleatorio*** [67]. Dado que la variable $y(\boldsymbol{s})$ se observa en n lugares en el espacio, los análisis se basan en un vector aleatorio con n componentes. Por lo tanto, aunque solo sea una variable, el proceso espacial es siempre multivariado y todas las ubicaciones de la zona de interés interactúan entre sí. En el caso geoestadístico, el proceso espacial ocurre en un conjunto índice continuo, es decir, el proceso ocurre en todos los infinitos puntos de su dominio, así que no es posible formalizar de manera directa los conceptos de distribuciones de probabilidad multivariadas. Es necesario recurrir al concepto fundamental de distribución de probabilidad finito-dimensional, con el cual una función aleatoria puede ser caracterizada por la distribución de probabilidad de sus subconjuntos finito-dimensionales. Esto es, solo es posible estudiar la distribución de probabilidad conjunta a partir de un vector aleatorio que contiene un subconjunto finito de variables espaciales

$$\mathbb{Y} = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)) .$$

y que corresponde al conjunto de las n ubicaciones donde se observa el fenómeno. Este conjunto de ubicaciones es seleccionado con algún criterio específico entre todos los conjuntos posibles de ubicaciones $\boldsymbol{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n\}$ (ver Sección 11.2 para detalles de cómo diseñar un muestreo espacial óptimo). La función de distribución n-variada del vector aleatorio $\mathbb{Y}$ es

$$F_{\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_n}(y_1, y_2, \ldots, y_n) = P\left[Y(\boldsymbol{s}_1) \leq y_1, Y(\boldsymbol{s}_2) \leq y_2, \ldots, Y(\boldsymbol{s}_n) \leq y_n\right] .$$

1.1.2. Parámetros de un campo aleatorio

Los parámetros del campo aleatorio son fundamentales para entender y caracterizar la forma en que el proceso depende de la ubicación espacial. A continuación se presentan las funciones de media, varianza, autocovarianza y semivarianza, con base en las cuales se desarrollan los conceptos de estacionariedad fuerte, de segundo orden e intrínseca.

1. **Función de media** $E\left[Y(\boldsymbol{s})\right]$ La media o momento de primer orden del proceso espacial es la esperanza matemática definida como

$$E\left[Y\left(\boldsymbol{s}\right)\right] = \mu\left(\boldsymbol{s}\right).$$

Es conocida como la función de *media*, *deriva* o la *tendencia* del proceso, debido a que usualmente el valor medio de la variable depende de manera determinística de la ubicación espacial.

2. **Función de varianza,** $\text{Var}[Y(\boldsymbol{s})]$ La *varianza* o momento centrado de segundo orden del proceso espacial $Y\left(\boldsymbol{s}\right)$ con media $\mu\left(\boldsymbol{s}\right)$ se define como

$$\text{Var}\left[Y\left(\boldsymbol{s}\right)\right] = E\left[Y\left(\boldsymbol{s}\right) - \mu\left(\boldsymbol{s}\right)\right]^2.$$

La varianza puede depender de la ubicación espacial $\boldsymbol{s}$. Sin embargo, considerar que la varianza puede variar continuamente, punto a punto a través de la zona, implicaría un proceso demasiado inestable, imposible de modelar. En general, lo que sí tiene sentido es subdividir la zona en sub-zonas, cada una de ellas con varianza constante, σ^2_{Y}, y encontrar modelos locales. En este caso, es necesario definir cómo se tratan los bordes en el momento de juntar los modelos.

3. **Función de autocovarianza,** $\text{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right]$.

Se le llama autocovarianza porque es la covarianza de una misma variable espacial en dos ubicaciones diferentes, esto es,

$$\text{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right] = E\left[\left(Y(\boldsymbol{s}_i) - \mu(\boldsymbol{s}_i)\right)\left(Y(\boldsymbol{s}_j) - \mu(\boldsymbol{s}_j)\right)\right].$$

Se asume que la varianza del proceso es finita, $\text{Var}[Y(\boldsymbol{s})] \leq \infty$, y la covarianza se suele modelar como una función del vector de separación entre las dos ubicaciones espaciales $\boldsymbol{s}_i$ y $\boldsymbol{s}_j$, esto es, C es una función de $\boldsymbol{s}_i - \boldsymbol{s}_j$,

$$\text{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right] = C(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}) \geq 0.$$

donde $\boldsymbol{\Theta}$ es un vector de parámetros fijos. Una función de autocovarianza $\boldsymbol{C}(.; \boldsymbol{\Theta})$ es una función *definida positiva*. Es decir, para cualquier número finito m de ubicaciones espaciales

$$\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m.$$

y cualquier conjunto de números reales, $\{a_1, a_2, \ldots, a_m\}$ con $m \in \mathbb{Z}^+$, $\boldsymbol{C}$ debe satisfacer la siguiente desigualdad:

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j C(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}) \geq 0\ . \tag{1.1}$$

Si la función $C(.;\boldsymbol{\Theta})$ cumple esta condición, se garantiza una varianza de error de predicción no negativa.

El predictor usado en geoestadística, conocido como kriging, es una suma ponderada de los n valores observados. Se denota como $Y^*(\boldsymbol{s}_0)$ y está dado por

$$Y^*(\boldsymbol{s}_0) = \sum_{i=1}^{n} \lambda_i Y(\boldsymbol{s}_i). \tag{1.2}$$

$\lambda_i;\ i = 1, \ldots, n$ es la ponderación correspondiente del valor observado de la variable en el lugar $\boldsymbol{s}_i$. Por lo tanto, la varianza del predictor kriging depende de las covarianzas entre todos los posibles pares de ubicaciones observadas:

$$\text{Var}\left[Y^*(\boldsymbol{s}_0)\right] = \text{Var}\left[\sum_{i=1}^{n} \lambda_i Y(\boldsymbol{s}_i)\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \text{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right]$$

Así, dado que $C(.;\boldsymbol{\Theta})$ sea definida positiva, (ver (6.3)) y que sea función de la separación entre dos lugares, se tiene que

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \text{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_i)\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}) \geq 0.$$

Como aclaración final, es obvio que la $\text{Var}\left[Y^*(\boldsymbol{s}_0)\right]$ es positiva, pero desconocida. En el momento en el que se quiere estimar es cuando se reconoce la necesidad de modelar la función de covarianza y, a partir de este modelo, encontrar su estimación. Lo que se enfatiza aquí es que en este proceso de estimar un parámetro desconocido se puede llegar a resultados absurdos si no se utilizan correctamente los elementos teóricos. Para un tratamiento más detallado de funciones definidas positivas, se puede consultar [9] y [15].

4. **Función de autocorrelación,** $\text{Cor}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right]$.

   La función de *autocorrelación* de la variable aleatoria Y en dos ubicaciones espaciales $Y(\boldsymbol{s}_i)$ y $Y(\boldsymbol{s}_j)$ se define como:

$$\text{Cor}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right] = \frac{\text{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j\right]}{\sqrt{\text{Var}\left[Y(\boldsymbol{s}_i)\right] \text{Var}\left[Y(\boldsymbol{s}_j)\right]}}.$$

   y dadas las consideraciones tenidas en cuenta en los ítems 2 y 3, se considera varianza constante, por lo que la función de *autocorrelación* $\rho\left(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}\right)$ del proceso es

$$\text{Cor}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right] = \frac{C\left(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}\right)}{\sigma_Y^2} = \rho\left(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}\right).$$

   dado que

$$\text{Var}\left[Y(\boldsymbol{s}_i)\right] = C\left(\boldsymbol{s}_i, \boldsymbol{s}_i; \boldsymbol{\Theta}\right) = C\left(\mathbf{0}; \boldsymbol{\Theta}\right) = \sigma_Y^2.$$

   Así, la función de autocorrelación espacial se encuentra acotada por 1 como es usual. En este contexto el numerador es siempre positivo. Con estos métodos se modelan las funciones de autocovarianza y autocorrelación para zonas en las que se puede suponer varianza constante. Nótese que estas funciones utilizan directamente la variable de interés $Y(\boldsymbol{s})$.

   Las funciones de autocovarianza y de autocorrelación requieren que el proceso tenga una varianza acotada. A continuación, se presenta otra función con la cual es posible modelar procesos cuya varianza no es acotada.

5. **Función de varianza de los incrementos,** $\text{Var}[Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)]$.

   A diferencia de las funciones de autocovarianza y autocorrelación que usan directamente la variable de interés, esta función modela la varianza de los incrementos. La variable incrementos es la variable resultante de la diferencia del proceso en dos lugares, y está dada por

$$Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j).$$

Entonces, el parámetro de interés es

$$\text{Var}[Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)].$$

que se modela de nuevo bajo el supuesto de que esta varianza depende de la separación espacial entre las ubicaciones y se denota como

$$\text{Var}[Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)] = 2\gamma\left(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}\right).$$

lo que equivale a

$$\frac{1}{2}\text{Var}[Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j)] = \gamma\left(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta}\right).$$

De donde proviene su nombre de función de semivarianza. $\boldsymbol{\Theta}$ es un vector de parámetros fijos. El dos es una constante conveniente que se justifica en la siguiente sección de estacionariedad (Sección 1.1.3) a través de la relación existente en algunos casos entre las funciones de autocovarianza y de semivarianza.

Si bien no requiere que la varianza del proceso sea finita, sí tiene algunas restricciones sobre su velocidad de crecimiento, además de requerir que los incrementos sean de media cero (Ver Definición 6).

1.1.3. Estacionariedad

La estacionariedad es un conjunto de propiedades que cumple un proceso estocástico y que permiten extraer información relevante acerca de este, en algunos casos inferencial, y en otros casos descriptiva. Se distinguen tres tipos de estacionariedad según los elementos del proceso estocástico que involucra cada una de ellas.

Definición 3 (Estacionariedad fuerte (o de primer orden)). *Se define en términos de las funciones de distribución. Un campo aleatorio es estacionario si su función de distribución no varía con una traslación $\boldsymbol{h} \in D_s$ del vector aleatorio. Esto es, si para todo $\boldsymbol{h} \in D_s$*

$$\mathbb{Y}(\boldsymbol{s}) = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))' \quad y \quad \mathbb{Y}(\boldsymbol{s}+\boldsymbol{h}) = (Y(\boldsymbol{s}_1+\boldsymbol{h}), \ldots, Y(\boldsymbol{s}_n+\boldsymbol{h}))'.$$

son tales que

$$F_{\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_n}(y_1, y_2, \ldots, y_n) = F_{\boldsymbol{s}_1+\boldsymbol{h},\boldsymbol{s}_2+\boldsymbol{h},\ldots,\boldsymbol{s}_n+\boldsymbol{h}}(y_1, y_2, \ldots, y_n).$$

De tal manera que las propiedades probabilísticas, tanto de la distribución como de sus parámetros, son iguales sin importar dónde se realicen las mediciones. Son invariantes a traslaciones y a rotaciones. Un proceso es estacionario fuerte si las relaciones dentro de cualquier subconjunto de puntos se mantienen idénticas, independientemente del lugar donde residen los puntos en el espacio. Este supuesto se refiere a que la estructura probabilística del campo aleatorio es similar en cualquier parte del dominio espacial $D_{\boldsymbol{s}}$, es decir, que el proceso alcanza un estado de equilibrio. Si se repite en el espacio el mismo comportamiento sistemáticamente, es posible hacer, por ejemplo, inferencia estadística. Sin embargo, es bastante difícil validar este supuesto a partir de una única realización del proceso. Nótese que aquí n observaciones de $Y(\boldsymbol{s})$ no constituyen una muestra aleatoria de tamaño n, sino que representan una muestra de tamaño 1 de un proceso estocástico n-variado. Además, el hecho de que los elementos del vector aleatorio estén correlacionados implica que contiene menos información que si hubiera independencia. Así, dado que la estacionariedad fuerte es un supuesto difícil de tener y de probar, es necesario utilizar otros conceptos que, aunque sean menos informativos, son más flexibles y permiten llevar a cabo los análisis para extraer la información requerida de los fenómenos en estudio. En particular, para algunos objetivos de los métodos geoestadísticos, como la estimación de la covarianza espacial y la predicción en lugares no muestreados, es suficiente con que se pueda asumir algún tipo de estacionariedad que involucre solo los dos primeros momentos.

Definición 4 (Estacionariedad débil (o de segundo orden)). *Se define en términos de la media y la covarianza del proceso. El proceso estocástico* $Y(\boldsymbol{s})$ *es un proceso débilmente estacionario o estacionario de segundo orden si la varianza es finita,* $\mathrm{Var}[Y(\boldsymbol{s})] \leq \infty$ *para* $\boldsymbol{s} \in D_{\boldsymbol{s}}$*, y además cumple las siguientes dos propiedades:*

- *La media es constante a través del dominio* $D_{\boldsymbol{s}}$*. Es decir, existe y no depende de la posición* $\boldsymbol{s}$

$$E[Y(\boldsymbol{s})] = \mu, \quad \forall \boldsymbol{s} \in D_{\boldsymbol{s}}, \mu \in \mathbb{R}.$$

- *Para toda pareja*

$$\big(Y(\boldsymbol{s}),\ Y(\boldsymbol{s}+\boldsymbol{h})\big).$$

  la función de covarianza del proceso espacial denotada como $\mathrm{Cov}[Y(\boldsymbol{s}),\ Y(\boldsymbol{s}+\boldsymbol{h})]$ *existe y es función única del vector de separación*

espacial $\boldsymbol{h}$*, con* $\boldsymbol{h} \in D_s$*. Esto es,*

$$\text{Cov}\,[Y(\boldsymbol{s}),\ Y(\boldsymbol{s}+\boldsymbol{h})] = C(\boldsymbol{h};\boldsymbol{\Theta}).$$

$\boldsymbol{\Theta}$ *es un vector de parámetros fijos.* $\boldsymbol{h}$ *es conocido como el rezago espacial.*

Bajo estacionariedad de segundo orden, la función de autocovarianza en dos ubicaciones espaciales depende de la distancia de separación y no de sus ubicaciones absolutas ni del valor de la variable de interés. Así, dado que las coordenadas absolutas no muestran ninguna influencia sobre la ocurrencia de la variable y no importa dónde están localizados los puntos, sino solo la separación entre ellos, la covarianza se puede estimar usando una función de la distancia absoluta $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}$ con algunos parámetros fijos. En consecuencia, todos los pares de puntos que tengan el mismo rezago espacial $\boldsymbol{h}$ pueden contribuir a la estimación.

Definición 5 (El modelo geoestadístico)**.** *El modelo geoestadístico asume que el proceso estocástico de interés se puede descomponer en la suma de dos partes:*

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + Z(\boldsymbol{s}), \quad \mathbb{Y}(\boldsymbol{s}) \sim (\boldsymbol{\mu}(\boldsymbol{s}), \Sigma), \quad \mathbb{Z}(\boldsymbol{s}) \sim (\mathbf{0}, \Sigma).$$

una determinística, que es la función media $\mu(\boldsymbol{s})$*, y otra estocástica, que es* $Z(\boldsymbol{s})$*.* $Z(\boldsymbol{s})$ *es un proceso centrado que conserva la estructura de covarianza del proceso original de interés gracias a la aditividad. Estos dos componentes son independientes.* Σ *es una matriz de covarianza definida positiva de dimensión* $n \times n$

$$\Sigma = \text{Var}[\mathbb{Y}(\boldsymbol{s})] = \text{Var}[\mathbb{Z}(\boldsymbol{s})]$$

De tal forma que el procedimiento para encontrar las predicciones en los lugares sin datos se puede llevar a cabo utilizando el proceso centrado asociado $Z(\boldsymbol{s})$. Los sitios en los que no se tiene observación y en los cuales se requiere predecir, se denotan $\boldsymbol{s}_0$. Aquí, se encuentra la predicción del proceso centrado en el lugar no observado $\boldsymbol{s}_0$. Es decir, si denotamos la predicción de $Z(\boldsymbol{s}_0)$ como $Z^*(\boldsymbol{s}_0)$, esta última se sustituye en el modelo

$$Y^*(\boldsymbol{s}_0) = \mu(\boldsymbol{s}_0) + Z^*(\boldsymbol{s}_0). \tag{1.3}$$

Las coordenadas x_0, y_0 se sustituyen en el modelo para la media, y así se encuentra la predicción del proceso espacial original $Y(\boldsymbol{s}_0)$.

Nota 1 (Notación). *De aquí en adelante se usará la notación $Z(\boldsymbol{s})$ para referirse al proceso centrado.*

Nota 2 (Relación entre las funciones de covarianza y de semivarianza bajo estacionariedad de segundo orden). *Cuando el campo aleatorio es estacionario de segundo orden, se tiene que*

$$\boldsymbol{\gamma}\left(\boldsymbol{s}_i-\boldsymbol{s}_j;\boldsymbol{\Theta}\right)=C\left(\mathbf{0};\boldsymbol{\Theta}\right)-C\left(\boldsymbol{s}_i-\boldsymbol{s}_j;\boldsymbol{\Theta}\right). \tag{1.4}$$

donde $\boldsymbol{\Theta}$ es el vector de parámetros del modelo y $C\left(\mathbf{0};\boldsymbol{\Theta}\right)=\sigma_Z^2$. Para demostrarlo, se utiliza la definición de semivarianza del proceso centrado $\boldsymbol{Z}(\boldsymbol{s})$

$$\begin{aligned}\operatorname{Var}\left[Z(\boldsymbol{s}_i)-Z(\boldsymbol{s}_j)\right]=\gamma\left(\boldsymbol{s}_i-\boldsymbol{s}_j;\boldsymbol{\Theta}\right)&=\frac{1}{2}E\left[Z\left(\boldsymbol{s}_i\right)-Z\left(\boldsymbol{s}_j\right)\right]^2\\&=\frac{1}{2}E\left[\left(Z\left(\boldsymbol{s}_i\right)-\mu\right)-\left(Z\left(\boldsymbol{s}_j\right)-\mu\right)\right]^2\end{aligned}$$

$$=\frac{1}{2}E\left[\left(Z\left(\boldsymbol{s}_i\right)-\mu\right)^2+\left(Z(\boldsymbol{s}_j)-\mu\right)^2-2\left(Z\left(\boldsymbol{s}_i\right)-\mu\right)\left(Z\left(\boldsymbol{s}_j\right)-\mu\right)\right].$$

Si se desarrolla el cuadrado del binomio y se toman esperanzas, se obtiene

$$\begin{aligned}&\frac{1}{2}\Big(E\left[\left(Z\left(\boldsymbol{s}_i\right)-\mu\right)^2\right]+E\left[\left(Z\left(\boldsymbol{s}_j\right)-\mu\right)^2\right]-2E\left[\left(Z\left(\boldsymbol{s}_i\right)-\mu\right)\left(Z\left(\boldsymbol{s}_j\right)-\mu\right)\right]\Big)\\&\qquad=\frac{1}{2}\Big(\operatorname{Var}\Big[Z\left(\boldsymbol{s}_i\right)\Big]+\operatorname{Var}\Big[Z\left(\boldsymbol{s}_j\right)\Big]-2\operatorname{Cov}\Big[Z(\boldsymbol{s}_i),Z(\boldsymbol{s}_j)\Big]\Big).\end{aligned}$$

Dado que la varianza es constante y que la covarianza es función de $\boldsymbol{s}_i-\boldsymbol{s}_j$ se encuentra que

$$\frac{1}{2}\Big(\boldsymbol{C}\left(\mathbf{0};\boldsymbol{\Theta}\right)+\boldsymbol{C}\left(\mathbf{0};\boldsymbol{\Theta}\right)-2\boldsymbol{C}\left(\boldsymbol{s}_i-\boldsymbol{s}_j;\boldsymbol{\Theta}\right)\Big)=\boldsymbol{C}\left(\mathbf{0};\boldsymbol{\Theta}\right)-\boldsymbol{C}\left(\boldsymbol{s}_i-\boldsymbol{s}_j;\boldsymbol{\Theta}\right).$$

lo cual se puede escribir en términos del rezago espacial $\boldsymbol{h}=\boldsymbol{s}_i-\boldsymbol{s}_j$, así:

$$\boldsymbol{\gamma}\left(\boldsymbol{h};\boldsymbol{\Theta}\right)=\boldsymbol{C}\left(\mathbf{0};\boldsymbol{\Theta}\right)-\boldsymbol{C}\left(\boldsymbol{h};\boldsymbol{\Theta}\right). \tag{1.5}$$

En consecuencia, la función de autocorrelación se puede expresar como

$$\rho\left(\boldsymbol{h};\Theta\right)=\frac{C\left(\boldsymbol{h};\Theta\right)}{C\left(\mathbf{0};\Theta\right)}=1-\frac{\gamma\left(\boldsymbol{h};\Theta\right)}{C\left(\mathbf{0};\Theta\right)}.$$

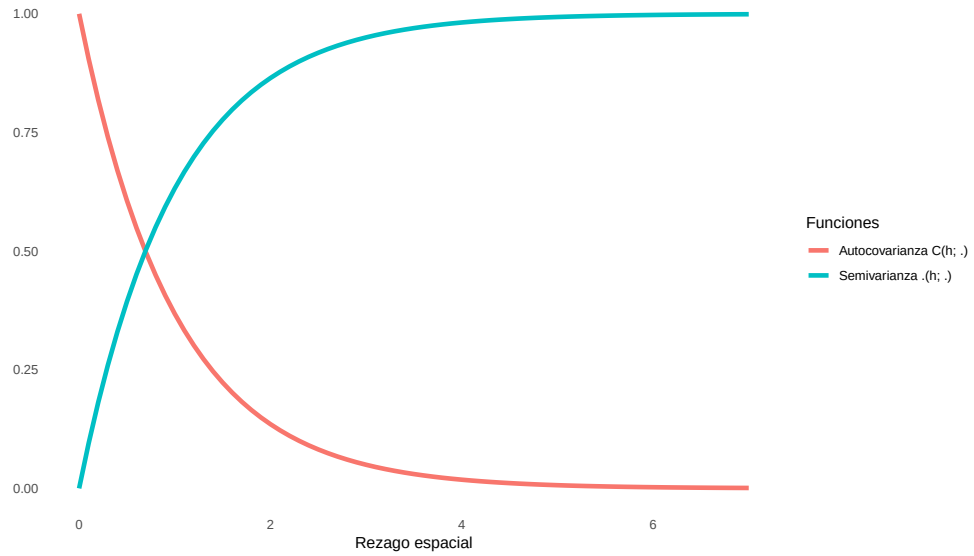


Figura 1.2. Relación entre la función de covarianza (autocovarianza) y la función de semivarianza bajo estacionariedad de segundo orden. Cuando $\|\boldsymbol{h}\| \longrightarrow \infty$ se tiene que $C(\boldsymbol{h};\boldsymbol{\Theta}) \longrightarrow 0$, lo cual implica $\gamma(\boldsymbol{h};\boldsymbol{\Theta}) \longrightarrow C(\mathbf{0})$.

El valor de la función de autocorrelación para cada rezago $\boldsymbol{h}$ mide relación lineal en el dispersograma rezagado $Z(\boldsymbol{s})$ Vs. $Z(\boldsymbol{s}+\boldsymbol{h})$.

En geoestadística es muy común usar la variable llamada incrementos

$$Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}).$$

haciendo una analogía con la diferenciación de una serie de tiempo cuando no se tiene estacionariedad en media para $Z(\boldsymbol{s})$. En este caso, se puede interpretar la variable incrementos como el cambio de la variable de interés después de un desplazamiento $\boldsymbol{h}$. Así, modelar la varianza de la variable incrementos es una forma alternativa de modelar la dependencia espacial. La estacionariedad formulada en términos de los incrementos se conoce como estacionariedad intrínseca. Esta es más flexible que la estacionariedad de segundo orden, dado que en este caso no se requiere varianza acotada.

Definición 6 (Estacionariedad intrínseca (o de los incrementos)). *Se define en términos de la media y la varianza de los incrementos del proceso.*

- *La media de la variable incrementos es cero, es decir:*

$$E\left[Z(\boldsymbol{s}+\boldsymbol{h})-Z(\boldsymbol{s})\right]=0 \;\; \forall \boldsymbol{s},\boldsymbol{h}\in D_s.$$

- *La varianza de $Z(\boldsymbol{s})-Z(\boldsymbol{s}+\boldsymbol{h})$ es una función del vector de separación $\boldsymbol{h}$.*

$$\frac{1}{2}\text{Var}\left[Z(\boldsymbol{s}+\boldsymbol{h})-Z(\boldsymbol{s})\right]=\gamma(\boldsymbol{h};\Theta).$$

Donde $\gamma(\boldsymbol{h};\Theta)$ es la función de semivarianza o semivariograma del proceso. Θ es un vector de parámetros fijo. Si bien bajo estacionariedad intrínseca $\gamma(\boldsymbol{h};\Theta)$ admite varianza infinita, existe un límite para su velocidad de crecimiento. La condición es que el semivariograma no crezca más rápido que una ecuación de segundo grado, esto es,

$$\frac{\gamma(\boldsymbol{h};\Theta)}{\|\boldsymbol{h}\|^2}\to 0 \qquad \textit{cuando} \quad \boldsymbol{h}\to\infty.$$

Un proceso estacionario de segundo orden también es estacionario intrínseco. El recíproco no se tiene en general, dado que para que exista estacionariedad de segundo orden es preciso que la varianza sea finita. Así que, si la función de semivarianza no es acotada, $\boldsymbol{C}(\mathbf{0};\Theta)$ no existe y, en consecuencia, no existe la función de covarianza ni la equivalencia (1.4).

Definición 7 (Isotropía). *Se tiene isotropía cuando las funciones de covarianza y semivarianza dependen únicamente de la magnitud $\|\boldsymbol{h}\|$ y no del ángulo. Esto es, las funciones de autocovarianza y semivarianza tiene como argumento la norma del vector de rezago espacial*

$$\text{Cov}\left[Z(\boldsymbol{s}),Z(\boldsymbol{s}+\boldsymbol{h})\right]=C\left(\|\boldsymbol{h}\|;\Theta\right)$$

y

$$\frac{1}{2}\text{Var}\left[Z(\boldsymbol{s}+\boldsymbol{h})-Z(\boldsymbol{s})\right]=\gamma(\|\boldsymbol{h}\|;\Theta).$$

La estacionariedad permite agrupar los datos en conjuntos de parejas que estén separadas por un mismo vector. Además, si a cualquier dirección el comportamiento de la función es el mismo y el vector diferencia puede ser reemplazado por su magnitud, usando para su cálculo, por ejemplo, la distancia euclidiana, entonces el campo aleatorio es isotrópico. Por lo tanto,

en este caso, las funciones de covarianza, correlación y semivarianza no dependen de la dirección del vector. En contraste, si un campo aleatorio no es isotrópico, el proceso se desarrolla de manera diferente en cada dirección del espacio y los modelos de covarianza son funciones tanto de la distancia como del ángulo entre cualquier par de puntos. En términos geométricos, la estacionariedad y la isotropía son propiedades de invarianza; la estacionariedad es invarianza bajo traslación y la isotropía es invarianza bajo rotaciones y reflexiones.

1.1.4. Media del campo aleatorio

En general, la media del proceso en la zona de interés se asume como una función determinística de la ubicación espacial. Esta media puede ser conocida y constante, o desconocida y constante o desconocida y no constante. A continuación se presenta en detalle cada uno de estos casos.

1.1.4.1. Media constante

Tanto la estacionariedad de segundo orden como la estacionariedad intrínseca requieren media constante. Observe que

- Si la media es conocida, no es necesario que sea constante, porque es posible centrar todas las observaciones y obtener así un proceso de media cero. Si se conoce $\mu(\boldsymbol{s}) \quad \forall \boldsymbol{s} \in D_S$ Entonces se puede construir el proceso centrado $Z(\boldsymbol{s})$, con $E[Z(\boldsymbol{s})] = 0$ dado que

$$Z(\boldsymbol{s}) = Y(\boldsymbol{s}) - \mu(\boldsymbol{s})$$

- Si la media no es conocida, es necesario estimarla sea constante o no. Si es constante, en una primera etapa se estima usando el promedio para obtener el proceso centrado,

$$Z(\boldsymbol{s}) = Y(\boldsymbol{s}) - \overline{Y}(\boldsymbol{s})$$

  y posteriormente cuando se tenga una estimación de la covarianza se re-estima usando el estimador de mínimos cuadrados generalizados para la media (ver expresión (1.6)).

  Es importante notar que en presencia de autocorrelación espacial, $\overline{Y}$ no es el estimador correcto, ya que no involucra dicha estructura. Dado que la media sea constante en D_s, su estimador correcto está

dado por el estimador de mínimos cuadrados generalizados

$$\hat{\mu} = \frac{\mathbf{1}'\Sigma^{-1}\mathbb{y}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \tag{1.6}$$

con varianza dada por

$$\text{Var}(\hat{\mu}) = \left(\mathbf{1}'\Sigma^{-1}\mathbf{1}\right)^{-1} \tag{1.7}$$

donde Σ es una matriz estructurada y definida positiva, que puede ser construida a partir de cualquier modelo válido de los presentados en el capítulo 2. No obstante, usar una media estimada genera restricciones en los procesos de predicción espacial.

1.1.4.2. Modelos de regresión polinómicos para la tendencia

Si la media no es constante y hay tendencia, $\mu(\boldsymbol{s})$ se estima con modelos de regresión sencillos, usualmente polinomios de segundo o tercer grado máximo. Como es usual, se prefieren modelos parsimoniosos. Los residuales del modelo de regresion seleccionado son el proceso centrado de media constante con el que se continúa el proceso geoestadístico. Este procedimiento de regresión, tiene la ventaja de ser paramétrico, y así el modelo estimado en función de las coordenadas x, y, permite reconstruir el valor de la media en cualquier lugar del dominio espacial, sin importar si en ese lugar hay o no observación (ver (1.3)). A continuación, algunos ejemplos de modelos polinómicos usados para la media.

$$\begin{aligned}
\mu(\boldsymbol{s}) &= \beta_0 + \beta_1 \mathrm{x} + \beta_1 \mathrm{y} \\
\mu(\boldsymbol{s}) &= \beta_0 + \beta_1 \mathrm{x} + \beta_2 \mathrm{y} + \beta_3 \mathrm{xy} \\
\mu(\boldsymbol{s}) &= \beta_0 + \beta_1 \mathrm{x} + \beta_2 \mathrm{x}^2 \\
\mu(\boldsymbol{s}) &= \beta_0 + \beta_1 \mathrm{x} + \beta_2 \mathrm{y} + \beta_3 \mathrm{x}^2 + \beta_4 \mathrm{xy} + \beta_5 \mathrm{y}^2 \\
\mu(\boldsymbol{s}) &= \beta_0 + \beta_1 \mathrm{x} + \beta_2 \mathrm{y} + \beta_3 \mathrm{x}^2 + \beta_4 \mathrm{xy} + \beta_5 \mathrm{y}^2 + \beta_6 \mathrm{x}^2\mathrm{y}^2.
\end{aligned}$$

Si la media no es constante pero no se ve una tendencia clara o hay datos atípicos que no permiten identificarla, se puede recurrir al pulimiento de medianas (ver sección 1.1.4.3).

1.1.4.3. Pulimiento de medianas

En ocasiones existen datos atípicos o se observa que la media del proceso no es constante a lo largo de las coordenadas, pero no se encuentra un modelo polinómico que la describa adecuadamente. El pulimiento de medianas asume que $\mu(\boldsymbol{s})$ se puede descomponer aditivamente en componentes direccionales. El pulimiento de medianas fue planteado originalmente para análisis de datos a dos vías, en el que un factor con p niveles se ubica en la fila y otro factor con q niveles se ubica en la columna, y en su intersección se observa una variable numérica que es la variable respuesta [43].

La adaptación al caso espacial supone una grilla regular en la cual cada coordenada horizontal x_p, $p = 1, \ldots, P$ es un nivel de un factor y cada coordenada vertical y_q $q = 1, \ldots Q$ es un nivel del otro factor, siendo las observaciones los $y(\boldsymbol{s}_i)$ $i = 1, \ldots, n$.. El modelo en este caso es el siguiente

$$\mu(\mathrm{x}_p, \mathrm{y}_q) = a + r_p + c_q + e_{pq} \tag{1.8}$$

- a es un efecto global, común a todas las observaciones,
- r_p es el efecto de la $p - \acute{e}sima$ fila $p = 1, \ldots, P$,
- c_q el efecto de la $q - \acute{e}sima$ columna $q = 1, \ldots, Q$ y
- e_{pq} denota el respectivo residuo.

Si existe interacción entre fila y columna, se puede incluir el término $r_p c_q$. Las estimaciones de r_p y c_q se obtienen mediante un proceso iterativo de sustracción de las medianas por fila y luego de las medianas por columna, hasta que las medianas de filas y columnas convergen a cero. Usualmente con dos iteraciones es suficiente para que esto ocurra. El modelo geoestadístico considerado es el mismo (ver Definición 5) solo que la media se estima ahora con este método y una vez se tiene el efecto global y los efectos por filas y columnas, se obtiene un proceso residual de media cero, por lo tanto, constante.

Ejemplo 2. *La Figura 1.3 muestra gráficas de dispersión para los datos del Ejemplo 1, así como para los residuales que quedan de dos modelos estimados.* $Y(\boldsymbol{s})$ *es la profundidad a la que se encuentra el agua y se grafica en términos de las coordenadas* x *y* y. *Se observa que no se puede asumir que la media es constante ni en el eje* X *ni en el eje* Y. *Por el contrario, la variable disminuye a medida que hacia se avanza hacia el oriente y hacia el norte. En consecuencia, en este caso, el modelo de media depende tanto de* X *como de* Y. *El objetivo del modelo de regresión es capturar la tendencia global, de tal manera que los*

residuales $Z(\boldsymbol{s})$ asociados no muestren tendencia. La estimación del modelo de regresión se lleva a cabo utilizando el método de mínimos cuadrados ordinarios. Los residuales resultantes del modelo más parsimonioso, que no muestran tendencia, serán el proceso espacial centrado denotado $Z(\boldsymbol{s})$. En general, $Z(\boldsymbol{s})$ es un proceso con autocorrelación espacial, por lo tanto, es preciso continuar el modelamiento. Es importante notar que este procedimiento es solo la primera etapa del proceso y es específicamente para remover tendencia. El modelo de regresión no cumple el supuesto de independencia, pero este modelo no será usado con fines predictivos, sino únicamente de reconstrucción de la tendencia global. A esta tendencia global se le suma la predicción encontrada con métodos geoestadísticos, de la variable Z en el lugar $\boldsymbol{s}_0$ en el que no se tiene observación. Con respecto a los parámetros del modelo de regresión, efectivamente son estimados. Sin embargo, tener en cuenta la incertidumbre de estas estimaciones complicaría mucho el método y no mejoraría el resultado. Ver [76].

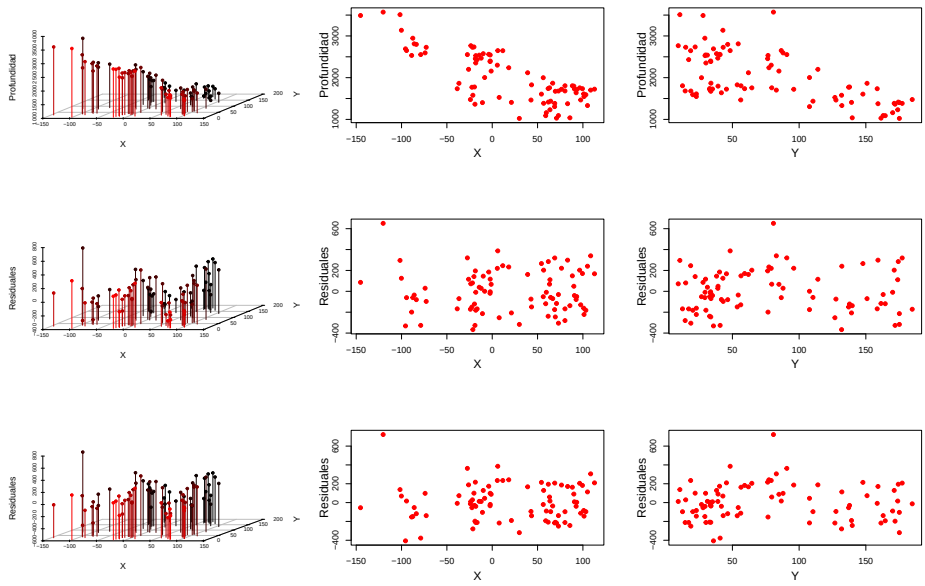


Figura 1.3. En la primera fila se muestran los gráficos de dispersión de la variable profundidad $Y(\boldsymbol{s})$ en términos de las coordenadas espaciales horizontal, x y vertical, y. La tendencia es evidente en ambas direcciones. Adicionalmente, el gráfico en las tres dimensiones sugiere evaluar el término de interacción. La segunda fila muestra los residuales del modelo $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 \mathrm{x} + \beta_2 \mathrm{y}$ y la tercera fila muestra los residuales del modelo $\mu(\boldsymbol{s}) = \beta_0 + \beta_1 \mathrm{x} + \beta_2 \mathrm{x} + \beta_3 \mathrm{xy}$, es decir, las realizaciones de $Z(\boldsymbol{s})$. La tendencia es eliminada por ambos modelos de acuerdo con los gráficos de los residuales; sin embargo, por parsimonia, es mejor elegir el modelo sin interacción.

1.2. Ejercicios

1. Se ha tomado una muestra de la altura en (cm) alcanzada por los árboles en un cultivo de mango (Ver Tabla 1.2).

Y				
999850				270
999840	420	400	470	270
999830	470	325	300	320
999820	400	340	330	270
999810	590	460	560	460
999800	390		270	415
X	**9564**	**9574**	**9584**	**9594**

Tabla 1.2. Diseño espacial de muestreo y resultado de las realizaciones en cada ubicación.

   1.1 ¿Para un $\|h\| = 10$ con $\phi = 0°$, esto es, en dirección horizontal, para $\boldsymbol{s} = (9564, 999800)$; $Z(\boldsymbol{s})$ =? y $Z(\boldsymbol{s} + \boldsymbol{h})$ =?

   1.2 ¿Para $\phi = 45°$, es posible encontrar pares de datos? En caso afirmativo, encuentre $\|h\|$ y $n(\|h\|)$.

   1.3 La media es constante e igual a 460. Si la covarianza espacial es un modelo esférico sin pepita y con silla 25000 y rango 15, ¿Cuál es el valor del coeficiente de autocorrelación para $\|h\| = 30$ con $\phi = 90°$?

2. Demuestre o refute:

   2.1 La estacionariedad intrínseca implica la estacionariedad de segundo orden. Ilustre con ejemplos.

   2.2 La estacionariedad de segundo orden implica la estacionariedad intrínseca. Ilustre con ejemplos.

3. Se tomaron 547 muestras de la elevación (en metros) de un terreno. El objetivo es generar el mapa topográfico de este. Ver Figura 1.4. A continuación, algunos resultados:

3.1 ¿Cómo es el comportamiento de la elevación del terreno en cada una de las coordenadas geográficas? ¿Existe tendencia?

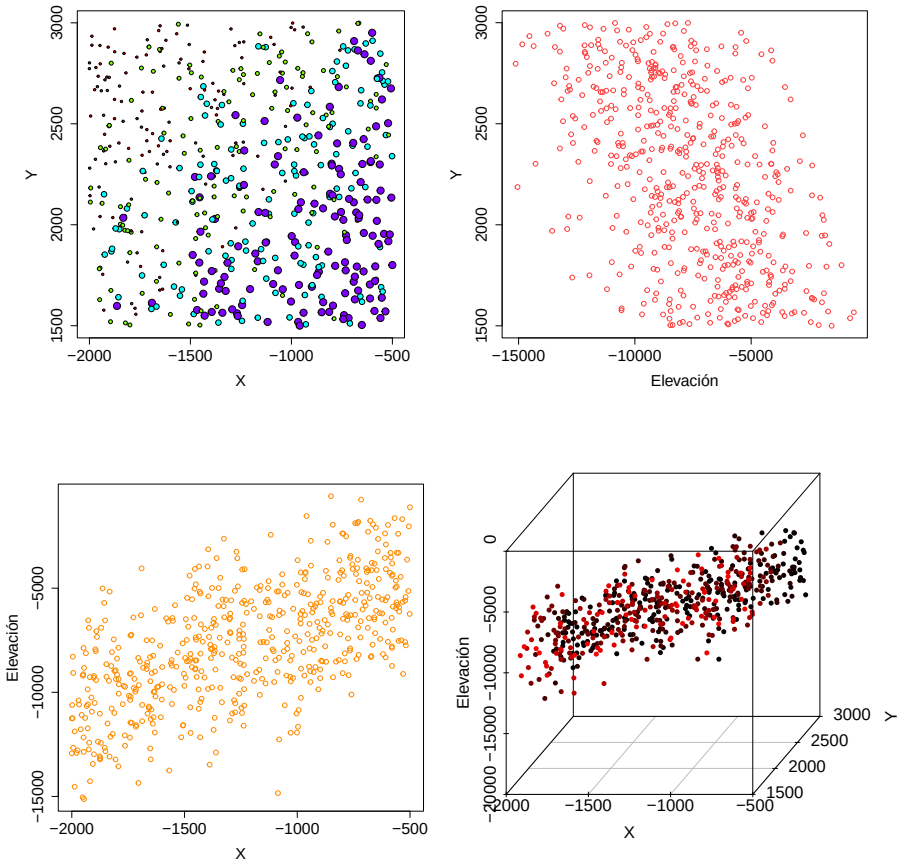
3.2 Encuentre la correlación de Pearson entre la elevación y cada una de las coordenadas espaciales. Interprete.

3.3 ¿Considera que es necesario incluir el análisis de la interacción entre las coordenadas geográficas este y norte en el modelo de regresión? Justifique.

3.4 ¿Considera usted que el proceso espacial posee media estacionaria? Justifique.

3.5 Interprete los colores en el gráfico del panel superior izquierdo. Es posible analizar tendencia en este gráfico?.

3.6 Escriba formalmente el modelo de regresión estimado. Si el valor de un residual en un lugar s_0 es 10, cuál es el valor de la variable de interés en ese lugar?



Coeficiente de Correlación		Coeficiente de determinación	Coeficiente de determinación ajustado		Error estándar
0,899		0,8091	0,808		8,21
ANOVA	**Grado de libertad**	**Suma de cuadrados**	**Cuadrados medios**	**Estadístico F**	**Valor p**
Regresión	3	155111,02	51703,67	766,9	9,8xE-195
Residuos	543	36608,9	67,42		
Total	546	191719,92			
Resultados regresión lineal múltiple	**Parámetros**	**Coeficientes**	**Error estándar**	**Estadístico t**	**Valor p**
	Intercepto	734,7	19,83	37,06	8,5x10-151
	Este	0,25	0,015	16,59	2,63xE-50
	Norte	-0,21	0,01	-20,35	7,61xE-69
	Este*Norte	-0,00015	7,50xE-6	-20,62	3,23xE-70

Figura 1.4. Panel superior izquierdo: ubicaciones $\boldsymbol{s}$ = (x, y), en las cuales fue medida la elevación del terreno, clasificadas por cuartiles. Panel superior derecho: gráfico de dispersión de la elevación Vs. la coordenada Y. Panel medio izquierdo: gráfico de dispersión de la coordenada X Vs. la elevación. Panel medio derecho: gráficos de dispersión en 3D de la elevación Vs. las coordenadas espaciales. Panel inferior: resultados de un modelo de regresión para ajustar la tendencia.

Capítulo

dos

Funciones de semivarianza y autocovarianza espacial

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

Para los propósitos de la geoestadística es suficiente con la información que aportan la media, la varianza del proceso de incrementos y la covarianza del proceso, cuando esta existe. Estos son los elementos fundamentales para encontrar predicciones espaciales óptimas. Por esta razón, los métodos se desarrollan bajo estacionariedad de segundo orden o estacionariedad intrínseca. Este capítulo se dedica al estudio del modelamiento de la estructura de dependencia espacial de los campos aleatorios, a través de la estimación de las funciones de covarianza y de semivarianza, definidas en la Sección 1.1.2. En la función de semivarianza el prefijo semi solo es agregado porque se utiliza por conveniencia la constante $\frac{1}{2}$.

La función de semivarianza $\gamma(\boldsymbol{h};\Theta)$ cuantifica la varianza de una variable espacial en dos ubicaciones separadas un vector $\boldsymbol{h}$.

$$\gamma(\boldsymbol{h};\Theta) = \frac{1}{2}\text{Var}\left[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})\right]. \tag{2.1}$$

2.1. Estimación empírica del semivariograma

La ventaja de usar la variable incrementos es que se pueden encontrar varias realizaciones, que corresponden a las diferencias de todos los pares de datos que se encuentran separados a una misma distancia $\boldsymbol{h}$ y a un mismo ángulo ϕ. Con base en estas realizaciones de los incrementos, se propone un estimador muestral para la semivarianza y a este se le da el nombre de *estimador empírico del semivariograma*, dado que, al no ser posible tener realizaciones directas, se construye con base en las realizaciones de la variable de interés. Un estimador muy natural está basado en el método de momentos y es conocido como el estimador clásico. Consiste en la estimación de la varianza muestral de los incrementos y está dado por:

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2|N(\boldsymbol{h})|}\sum_{N(\boldsymbol{h})}\Big(z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s})\Big)^2 \quad \boldsymbol{h}\in\mathbb{R}^d. \tag{2.2}$$

donde $N(\boldsymbol{h}) \equiv \left\{(\boldsymbol{s}_i,\boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\right\}$ es el conjunto de todos los pares de ubicaciones cuya separación corresponde a un vector $\boldsymbol{h}$, y $|N(\boldsymbol{h})|$ es el cardinal de $N(\boldsymbol{h})$. Los valores del semivariograma empírico se toman como la variable respuesta, y así un modelo de semivariograma teórico se estima con los métodos usuales para regresión. Estos elementos se ilustran en el Ejemplo 2.2 y en la Tabla 2.1. Este estimador presenta los inconvenientes

ya conocidos de la varianza muestral, principalmente su sensibilidad a datos atípicos, [21]. Por eso, se han propuesto estimadores resistentes a datos atípicos tales como

$$\tilde{\gamma}(\boldsymbol{h}) = \frac{1}{2(0.457 + \frac{0.494}{|N(\boldsymbol{h})|} + \frac{0.045}{|N^2(\boldsymbol{h})|}} \left(\frac{\sum_{N(\boldsymbol{h})} |z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s})|^{1/2}}{|N(\boldsymbol{h})|} \right)^4 . \quad (2.3)$$

Finalmente, según la teoría de datos resistentes, es posible cambiar la media por la mediana con lo que se obtiene un estimador alternativo dado por

$$\check{\gamma}(\boldsymbol{h}) = \frac{\left(me|z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s})|^{1/2} \right)^4}{2(0.457)}. \quad (2.4)$$

En los denominadores se puede apreciar en cada caso la corrección por sesgo [24]. Adicionalmente, los sumandos

$$|z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s})|^{1/2}$$

son mas similares entre sí que los

$$(z(\boldsymbol{s}+\boldsymbol{h}) - z(\boldsymbol{s}))^2.$$

La escala requerida se alcanza cuando toda la sumatoria se eleva a la cuatro. En la Figura 2.1 se aprecia la diferencia entre los sumandos de los estimadores del semivariograma. En el panel superior se observa que al elevar al cuadrado la diferencia aparecen muchos incrementos atípicos y hay mayor variabilidad. Estos efectos son mejor controlados por la operación de elevar a la 1/2 como se observa en el panel inferior. Si se opta por la estimación de la función de covarianza, se puede recurrir a la definición clásica de covarianza muestral:

$$\widehat{C}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} (z(\boldsymbol{s}+\boldsymbol{h}) - \overline{z})(z(\boldsymbol{s}) - \overline{z}). \quad (2.5)$$

Nótese que para el semivariograma, basta con que la media sea constante, aunque no se conozca, pero para la covarianza se requiere la estimación de la media muestral. A partir de los semivariogramas o covariogramas empíricos, se ajustan los modelos paramétricos. En la siguiente sección se hace un recorrido por los métodos de estimación más usados.

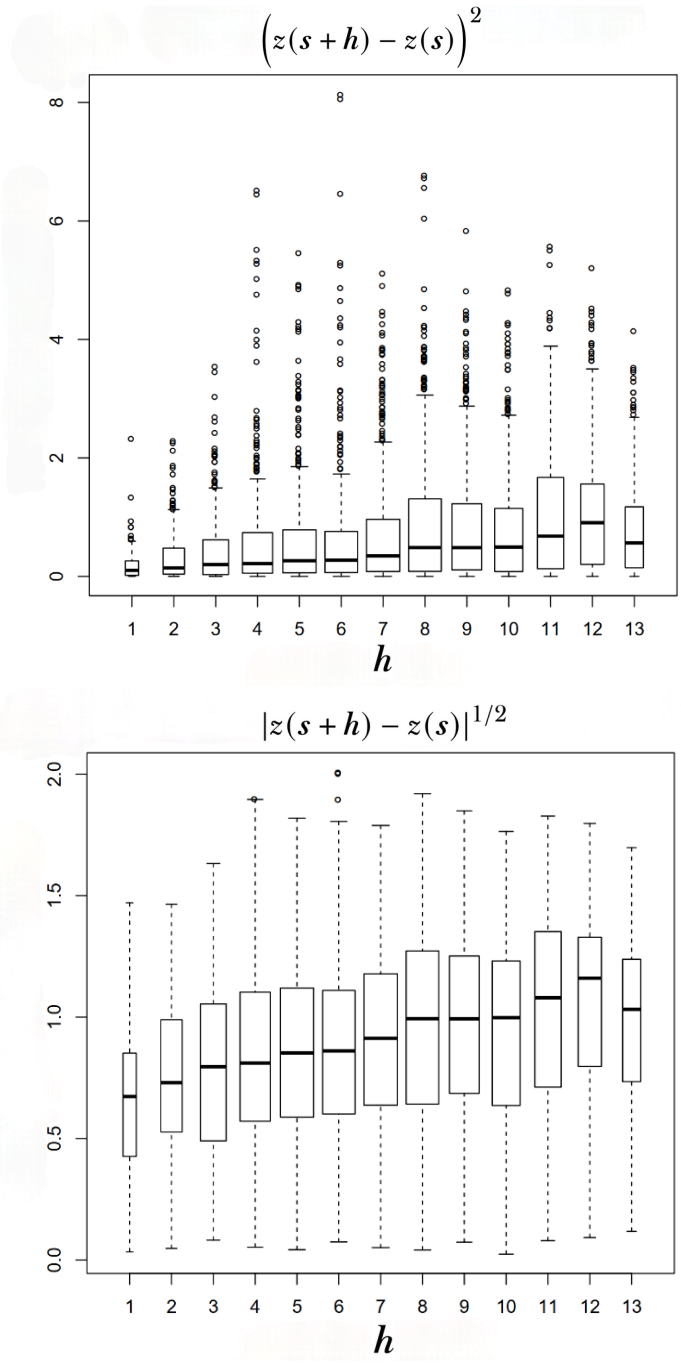


Figura 2.1. Diagramas de caja de las nubes de puntos usadas para la estimación clásica $(z(s+h)-z(s))^2$ (panel superior) y $|z(s+h)-z(s)|^{1/2}$ resistente (panel inferior) del semivariograma.

Si existe tendencia en los datos, y esta no es eliminada antes de encontrar la estimación empírica del semivariograma o del covariograma, estos estimadores quedarán contaminados y no se podrá identificar su verdadero comportamiento. Es muy importante verificar el cumplimiento del supuesto relativo a la media constante del proceso para proponer un buen modelo. Por lo tanto, el estimador empírico del semivariograma se calcula con $Z(\boldsymbol{s})$ y no con $Y(\boldsymbol{s})$.

> El estimador empírico del semivariograma se calcula con las realizaciones del proceso centrado $Z(\boldsymbol{s})$ y no con las de $Y(\boldsymbol{s})$, a menos que el proceso $Y(\boldsymbol{s})$ tenga media constante en D_s, caso en el cual ambos cálculos son equivalentes.

El comportamiento natural de un semivariograma empírico es que sea una función creciente de la distancia. A mayor distancia, mayor variabilidad, pues las variables lejanas se parecen menos que las cercanas. Pueden existir fluctuaciones o altibajos, como es usual en las realizaciones de pares de datos con base en los cuales se ajusta un modelo. Un comportamiento diferente puede verse al final para distancias grandes, debido a la poca cantidad de pares de datos. En este caso, sería necesario modelar el semivariograma hasta una distancia menor al máximo rezago para el que se tienen datos. Posteriormente, en la etapa de predicción solo se tienen en cuenta datos en esta vecindad espacial.

2.2. Aspectos prácticos de la estimación del semivariograma empírico

El rezago espacial $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$ es la distancia entre las ubicaciones $\boldsymbol{s}_i$ y $\boldsymbol{s}_j$. La Figura 2.2 muestra a la izquierda una idealización de una grilla regular en la que es posible tomar mediciones a distancias iguales, mientras que a la derecha un diseño irregular en el que todos los pares de puntos están separados por una distancia diferente.

El diseño de la izquierda puede ser un cultivo organizado en el que se siembra en surcos en un área homogénea. Se tienen distancias fijas y ángulos fijos entre puntos. El rezago espacial más pequeño al que se encuentran pares de observaciones consecutivas es $\|\boldsymbol{h}_1\| = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, seguido de $\|\boldsymbol{h}_2\| = 2\|\boldsymbol{h}_1\|$ y $\|\boldsymbol{h}_3\| = 3\|\boldsymbol{h}_1\|$. Hay $K = 3$ rezagos posibles. Con este procedimiento, se construyen conjuntos de pares de datos que llevan a

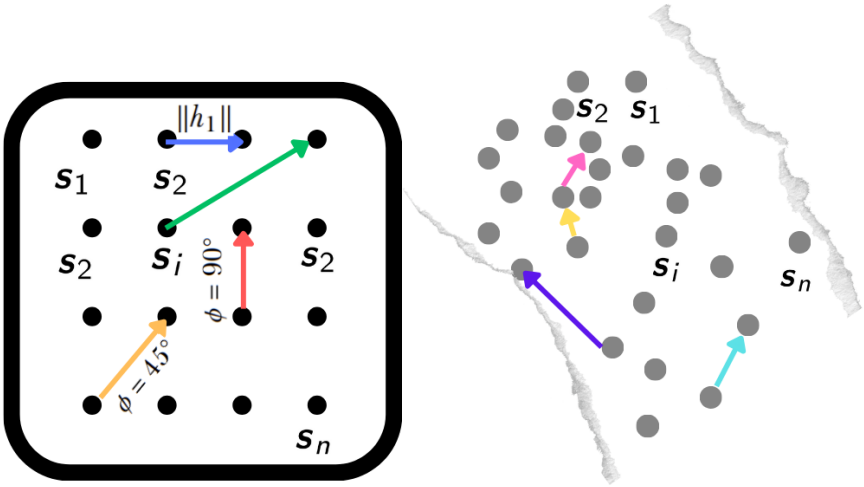


Figura 2.2. Dos diseños de muestreo espaciales. A la izquierda, un diseño regular, todos los puntos son equidistantes. A la derecha, un muestreo irregular donde, en general, todos los pares de datos tienen diferente separación y diferente ángulo.

realizaciones de la variable incrementos, de tal forma que la lista de rezagos y el número de pares encontrados, tanto en dirección horizontal, $\phi = 0°$, como en dirección vertical, $\phi = 90°$, son los siguientes:

$$\begin{aligned}
&\{z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j); \|\boldsymbol{h}_1\| = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|\}, \quad |N(\|\boldsymbol{h}_1\|)| = 12\\
&\{z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j); \|\boldsymbol{h}_2\| = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|\}, \quad |N(\|\boldsymbol{h}_2\|)| = 8\\
&\{z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j); \|\boldsymbol{h}_3\| = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|\}, \quad |N(\|\boldsymbol{h}_3\|)| = 4.
\end{aligned}$$

En este caso ilustrativo coinciden por la distribución de la grilla, pero en general no tiene porque ser así. Nótese que, a mayor distancia hay menor cantidad de pares de datos. Así, disminuye notoriamente la cantidad de puntos incluidos en la estimación del semivariograma. Por esta razón, en general, se usan los rezagos espaciales hasta la mitad de la máxima distancia. Además, hay que tener todas las direcciones en cuenta. Para cada dirección los rezagos pueden ser distintos (ver por ejemplo, las flechas verde y amarilla en la Figura 2.2).

El diseño de la derecha puede ser un conjunto de estaciones climáticas o ambientales en el que se requiere mayor densidad de mediciones en algunas zonas, debido a diferencias en la correlación espacial. Las zonas son

heterogéneas y hay lugares de difícil acceso, como zonas muy montañosas, selváticas, rocas o cuerpos de agua donde no se pueden tomar observaciones o instalar dispositivos. En estos casos, todos los pares de puntos se encuentran a diferentes distancias. El muestreo en grilla regular no es ni posible ni adecuado.

Por lo tanto, encontrar pares de puntos con separación exacta $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}_k$ es muy difícil; se acostumbra a definir una tolerancia tanto de longitud como de ángulo, de tal forma que $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| \in (\boldsymbol{h}_k - \xi, \boldsymbol{h}_k + \xi)$ y el ángulo entre $\boldsymbol{s}_i$ y $\boldsymbol{s}_j$ está entre $\phi - \omega$ y $\phi + \omega$. Los pares de datos están en cada conjunto si su distancia y ángulo se encuentran en los intervalos $\|\boldsymbol{h}\| \pm \xi$ y $\phi \pm \omega$, respectivamente.

Rezago espacial	$\boldsymbol{h}_1$	$\boldsymbol{h}_2$	…	$\boldsymbol{h}_K$
Total de pares a cada rezago espacial	$\lvert N(\boldsymbol{h}_1)\rvert$	$\lvert N(\boldsymbol{h}_2)\rvert$	…	$\lvert N(\boldsymbol{h}_K)\rvert$
Semivariograma empírico	$\hat{\gamma}(\boldsymbol{h}_1)$	$\hat{\gamma}(\boldsymbol{h}_2)$	…	$\hat{\gamma}(\boldsymbol{h}_K)$
Modelo de semivariograma teórico	$\gamma(\boldsymbol{h}_1;\Theta)$	$\gamma(\boldsymbol{h}_2;\Theta)$	…	$\gamma(\boldsymbol{h}_K;\Theta)$
Modelo de semivariograma estimado	$\gamma(\boldsymbol{h}_1,\hat{\Theta})$	$\gamma(\boldsymbol{h}_2,\hat{\Theta})$	…	$\gamma(\boldsymbol{h}_K,\hat{\Theta})$

Tabla 2.1. Ilustración de los pasos en la estimación de la función de varianza de los incrementos. Los incrementos se construyen de acuerdo con los rezagos espaciales. El semivariograma empírico se calcula con base en los incrementos y luego se aplican métodos de regresión para estimar los parámetros de la función de semivarianza. Con base en estas estimaciones se puede estimar la varianza de los incrementos para cualquier distancia.

Nota 3 (Justificación intuitiva del semivariograma empírico resistente a datos atípicos.[24]). *Bajo el supuesto de normalidad marginal del campo aleatorio* $Z(\boldsymbol{s}) \sim N(\mu;\sigma^2), \forall \boldsymbol{s} \in D_s$ *se tiene que* $(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})) \sim N(0; 2\gamma(\boldsymbol{h}))$ *y entonces* $(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))/\sqrt{2\gamma(\boldsymbol{h})} \sim \chi^2_1$. *Por lo tanto el cuadrado de los incrementos* $(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2 \sim 2\gamma(\boldsymbol{h})\chi^2_1$, *y* $E\left[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})\right]^2 = 2\gamma(\boldsymbol{h})$. *Ahora se transforma el cuadrado de los incrementos a una variable de distribución simétrica. El valor determinado a partir la transformación Box-Cox es* $\lambda = 1/4$, *con lo cual se obtiene una nueva variable con coeficiente de asimetría 0,08 y coeficiente de curtosis 2,48 [23]. Nótese que ni el supuesto de normalidad ni otro supuesto distribucional son necesarios para la geoestadística. Sin embargo, partir de estos supuestos permiten generar algunas ideas exploratorias muy útiles.*

2.3. Modelos válidos para la función de semivarianza espacial de la variable incrementos.

En esta sección se presentan los aspectos teóricos de los modelos usados como funciones de semivarianza o covarianza espacial. Si bien es un problema de regresión, no cualquier función se puede usar como modelo teórico de semivarianza o covarianza espacial, porque de estas funciones dependen las predicciones espaciales y sus respectivas varianzas. Estas últimas deben garantizarse positivas, lo que lleva a restringir el conjunto de modelos teóricos al conjunto de funciones definidas positivas para la covarianza y condicionalmente definidas negativas para la semivarianza, como se ve a continuación. Observe las siguientes propiedades obtenidas de las dobles sumatorias y la propiedad distributiva desarrollando el cuadrado del binomio.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \left(Z(s_i) - Z(s_j)\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \left(Z^2(s_i) - 2Z(s_i)Z(s_j) + Z^2(s_j)\right)$$

$$= \sum_{i=1}^{n} a_i Z^2(s_i) \sum_{j=1}^{n} a_j + \sum_{j=1}^{n} a_j Z^2(s_j) \sum_{i=1}^{n} a_i - 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Z(s_i)Z(s_j) \quad (2.6)$$

Ahora, siempre es cierto que el cuadrado de una sumatoria se puede expresar como una doble sumatoria.

$$\left(\sum_{i=1}^{n} a_i Z(s_i)\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Z(s_i)Z(s_j). \quad (2.7)$$

Si además se cumple la condición

$$\sum_{i=1}^{n} a_i = 0$$

y se toman esperanzas en la expresión (2.6), se obtiene

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (Z(s_i) - Z(s_j))^2\right] = -2E\left[\sum_{i=1}^{n} a_i Z(s_i)\right]^2.$$

Finalmente, sustituyendo la definición de la función de varianza de los incrementos, el resultado indica que el lado izquierdo de la expresión 2.3 es de signo negativo para que la varianza de la combinación lineal al lado derecho sea positiva.

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j 2\gamma(s_i - s_j)\right) = -2\text{Var}\left[\sum_{i=1}^{n} a_i Z(s_i)\right].$$

Se concluye que bajo la condición

$$\sum_{i=1}^{n} a_i = 0$$

la función $\gamma(.)$ debe ser negativa para garantizar que las varianzas de las predicciones de las variables espaciales sean positivas (Ver ecuación 1.2).

$$\text{Var}\left[\sum_{i=1}^{n} a_i Z(s_i)\right] = -\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(s_i - s_j)\right).$$

A continuación, se define formalmente una función de valor real condicionalmente definida negativa.

Definición 8 (Función condicionalmente definida negativa). *Una función $f : \mathbb{R}^d \to \mathbb{R}$ es condicionalmente definida negativa si $f(\boldsymbol{w}) = f(-\boldsymbol{w})$ para todo $\boldsymbol{w} \in \mathbb{R}^d$, y si para cualquier conjunto de m números reales $a_1, a_2, \ldots, a_m$ tales que $\sum_{i=1}^{m} a_i = 0$, con $m \in \mathbb{Z}^+$, y para m elementos del dominio $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m$, la función f satisface*

$$\sum_{i=1}^{m}\sum_{i=1}^{m} a_i a_j f(\boldsymbol{w}_i - \boldsymbol{w}_j) \leq 0\,. \tag{2.8}$$

La función $\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\Theta})$ debe ser condicionalmente definida negativa para ser un modelo válido de semivariograma y garantizar así varianzas de predicción positivas. En la siguiente sección se presenta una lista no exhaustiva de modelos válidos. Aunque esto puede parecer restrictivo, los modelos tienen gran flexibilidad y muchas posibilidades de comportamientos diferentes. Además, también son válidas las combinaciones lineales con coeficientes positivos de estos modelos, dando una inmensa cantidad de opciones adicionales, (ver Sección 3.7).

2.4. Elementos del semivariograma bajo estacionariedad de segundo orden

El semivariograma de un proceso estacionario de segundo orden tiene una asíntota en $C(\mathbf{0})$. Esto significa que su varianza es finita. Esta es una herramienta para verificar el supuesto de estacionariedad de segundo orden. Si el semivariograma empírico tiende a una línea horizontal cuando se incrementa la separación entre los puntos, es porque está acotado superiormente, y esta cota es la varianza del proceso. Un semivariograma de un proceso estacionario de segundo orden posee al menos los siguientes dos parámetros:

- **Silla:** es la cota superior del semivariograma. Algunos modelos la alcanzan de manera exacta, mientras que otros la alcanzan solo de manera asintótica. La silla se denota como $C(\mathbf{0})$,

$$C(\mathbf{0}) = \sigma_Y^2 = \sigma_Z^2$$

  y también es conocida como meseta. En la Figura 1.2 la silla es 1.

- **Rango o alcance:** es la distancia a partir de la cual el proceso espacial ya no se encuentra autocorrelacionado. Variables espaciales separadas una distancia inferior al rango muestran autocorrelación espacial. Si el proceso tiene rango finito, esto es, soporte compacto, la correlación es cero para distancias mayores al rango. En otros casos, la cota es una asíntota, así que la correlación disminuye a medida que la semivarianza se acerca a ella. En la Figura 1.2 el rango es 6.

Algunos modelos solo tienen estos dos parámetros. Otros modelos tienen parámetros adicionales, como parámetros de suavizamiento que los hacen más flexibles. Pero bajo estacionariedad de segundo orden todos los modelos tienen silla y rango.

- **Efecto pepita:** Hay un parámetro adicional que surge como consecuencia de un mal diseño de muestreo o de errores en la toma de datos. Es el llamado efecto pepita. De la definición del semivariograma se puede ver que para $\boldsymbol{h} = \mathbf{0}$, debería ocurrir que $\gamma(\boldsymbol{h}) = 0$. Sin embargo, en general existe discontinuidad en el origen, $\gamma(\boldsymbol{h}) \to c_0$ cuando $\boldsymbol{h} \to \mathbf{0}$. Ocurre muy frecuentemente que a partir del semivariograma empírico se encuentra evidencia de que la función no pasa por $(\mathbf{0}, 0)$.

La primera causa posible es la presencia de un error de medida (EM); cuando no es posible repetir una medida en la ubicación $\boldsymbol{s}$ sin error, entonces allí se evidencia su variabilidad. Esto se puede atribuir a dispositivos mal calibrados o a errores de registro de observaciones. La segunda causa posible es el llamado efecto de microescala o microestructura (MS), el cual se genera porque el proceso espacial varía a distancias más cortas de las consideradas en el diseño de muestreo. Las observaciones en este caso no aportan información sobre el proceso espacial, porque fueron hechas a distancias iguales o superiores al rango. Por lo tanto, si cualquiera de las dos componentes (EM o MS) es diferente de cero, el semivariograma presentará una discontinuidad puntual en el origen. La magnitud de esta discontinuidad es llamada el *efecto pepita* (c_0) y se define como:

$$c_0 = \sigma^2_{EM} + \sigma^2_{MS}. \tag{2.9}$$

Esto genera un inconveniente adicional y es que la silla $C(\mathbf{0})$ ahora se encuentra dividida en una parte que corresponde al error, que es el efecto pepita, y otra que corresponde al proceso, que es la silla parcial.

- **S**illa parcial: Si un semivariograma tiene efecto pepita c_0 y silla $C(\mathbf{0})$, la diferencia

$$\sigma^2_p = C(\mathbf{0}) - c_0$$

es llamada la silla parcial del semivariograma. Esta representa la varianza que se puede atribuir a la variable $Z(\boldsymbol{s})$ $(Y(\boldsymbol{s}))$, quitándole el efecto pepita.

En presencia del efecto pepita, el proceso está contaminado por un proceso estacionario $\mathrm{V}(s)$, [29]. Así, sean ϵ_i, $i = 1, \ldots, n$, mutuamente independientes e idénticamente distribuidos, esto es, $\forall i\ \epsilon_i \sim N(0; c_0)$. Se tiene que

$$Z(s_i) = \mathrm{V}(s_i) + \epsilon_i \quad i = 1, \ldots, n.$$

Entonces, la función de covarianza de $Z(\boldsymbol{s})$ $(Y(\boldsymbol{s}))$ es $C(\boldsymbol{h}) = \sigma^2_p \rho(\boldsymbol{h})$, tal que $\rho(\mathbf{0}) = 1$, pero dado que $\mathrm{Var}(Z(\boldsymbol{s}_i)) = \sigma^2_p + c_0$ no es posible modelar toda la estructura de correlación del proceso de interés

$$\frac{\sigma^2_p \rho(\boldsymbol{h})}{\sigma^2_p + c_0} < 1 \qquad \text{cuando} \qquad h \longrightarrow 0.$$

Este comportamiento en el origen que genera el efecto pepita en el semivariograma, tiene unas implicaciones directas en las propiedades en L_2 del proceso [23].

- $2\gamma(\boldsymbol{h})$ es continuo en el origen, entonces $Z(\boldsymbol{s})$ es continuo en L_2, continuo en media cuadrática:

$$E[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})]^2 \longrightarrow \mathbf{0}$$

  si y solo si $2\gamma(\boldsymbol{h}) \longrightarrow \mathbf{0}$ cuando $\boldsymbol{h} \longrightarrow \mathbf{0}$

- Si $2\gamma(\boldsymbol{h})$ no tiende a 0 cuando $\boldsymbol{h} \longrightarrow \mathbf{0}$, entonces $Z()$ no es continuo en L_2. Esta discontinuidad es el efecto pepita.

- $2\gamma(\boldsymbol{h})$ es positivo y constante (excepto en el origen donde es cero). Entonces $Z(\boldsymbol{s}_i)$ y $Z(\boldsymbol{s}_j)$ no están espacialmente correlacionados para ningún par $\boldsymbol{s}_i \neq \boldsymbol{s}_j$, sin importar qué tan cerca estén. Este fenómeno es conocido como ruido blanco o efecto pepita puro.

Con la existencia del efecto pepita, se altera la relación entre el covariograma y el semivariograma. Se habla de covarianza a partir del punto en el que termina el error, es decir, en el que termina el efecto pepita. Así, tanto la covarianza como la semivarianza quedan desplazadas hacia arriba. (Ver Figura 2.3).

$$\gamma(\boldsymbol{h};\Theta) = c_0 + \sigma_p^2\,\rho(\boldsymbol{h};\Theta)$$

y

$$C(\boldsymbol{h};\Theta) = \begin{cases} \sigma_p^2(1-\rho(\boldsymbol{h};\Theta)) & \text{si} \quad \boldsymbol{h} > \mathbf{0} \\ c_0 + \sigma_p^2 & \text{si} \quad \boldsymbol{h} = \mathbf{0}. \end{cases}$$

Donde $\rho(\boldsymbol{h};\Theta)$ es la función de autocorrelación y representa el modelo espacial, σ_p^2 es la silla parcial y c_0 es el efecto pepita. Es importante tener esto en cuenta para no mezclar el error de medida o de microestructura con los parámetros del proceso.

> Si los errores de medición o de microestructura son muy altos, la estimación de los modelos no es buena, afectando la calidad de las predicciones. Lo ideal en este caso es tomar la muestra existente como una muestra piloto inicial, con base en la cual se diseña un muestreo espacial óptimo, para recolectar nuevos datos y así reunir la información suficiente para cumplir con los objetivos del estudio, (ver Sección 11.2).

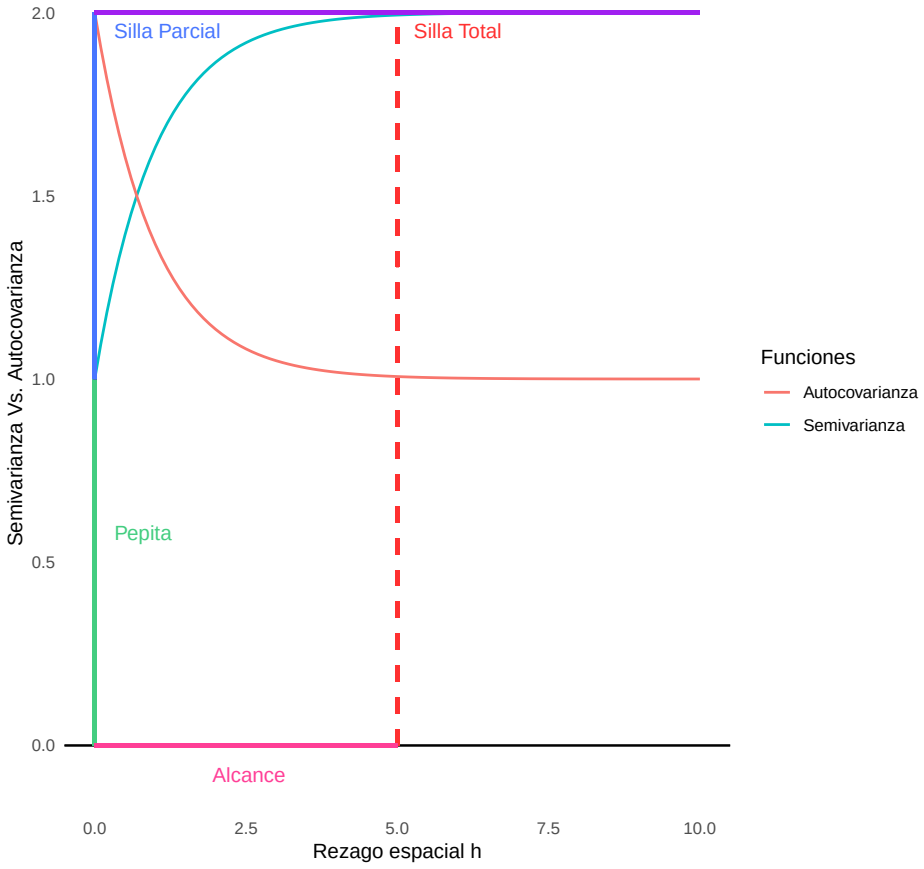


Figura 2.3. Pepita, silla total, silla parcial y rango bajo estacionariedad de segundo orden. El modelo es acotado, pero presenta discontinuidad en el origen. La silla se alcanza de manera asintótica.

En la siguiente sección, se presentan algunos de los modelos teóricos válidos más usados para ajustar los semivariogramas empíricos. Los semivariogramas son funciones de la distancia que miden la varianza de los incrementos, por lo tanto, son crecientes. La interpretación es que a mayor distancia hay mayor diferencia entre los datos y esto genera mayor variabilidad. Hay algunos casos especiales tales como el modelo efecto pepita, que es una función constante de la distancia, indicando que no existe autocorrelación espacial, o los modelos que presentan ondas y que alcanzan la silla oscilando alrededor de ella, indicando que permiten autocorrelaciones negativas.

Todos los modelos se pueden mezclar para construir nuevos modelos, usando productos y combinaciones lineales con coeficientes positivos. Esto se conoce como el modelo lineal de regionalización y amplía mucho las opciones, ([19, 24, 47]). Este procedimiento se detalla en la sección (3.4). Todos los modelos acotados de semivarianza, tienen su modelo de autocovarianza equivalente a partir de la relación (1.5). Con base en un análisis riguroso de los semivariogramas empíricos y el conocimiento detallado de las características de los modelos teóricos, se selecciona el modelo correcto y se estiman sus parámetros usando los métodos usuales tales como mínimos cuadrados o máxima verosimilitud (ver Sección 3).

Las Figuras 2.4, 2.5 y 2.6 y la Tabla 2.2 muestran varios ejemplos de funciones teóricas de semivarianza y covarianza.

2.5. Modelos de semivariograma acotados

Modelo efecto pepita. Válido en $\mathbb{R}^d$. Si el semivariograma ajustado es un efecto pepita, entonces no existe autocorrelación espacial.

$$\gamma(h;\Theta) = \begin{cases} 0 & \text{si } \|h\| = 0; \\ \sigma^2 & \text{si } \|h\| \neq 0. \end{cases}$$

Hay que evaluar esta situación, ya que pudo haber ocurrido un problema grave de error de medición o de microestructura. Si no hay razones para pensar en independencia o error de medición, la microestructura indica que la variación espacial ocurre a distancias menores a las observadas. Es necesario tomar nuevas observaciones a distancias cortas, que es donde ocurre la variación espacial del proceso (ver Sección 11.2 y Figura 2.4).

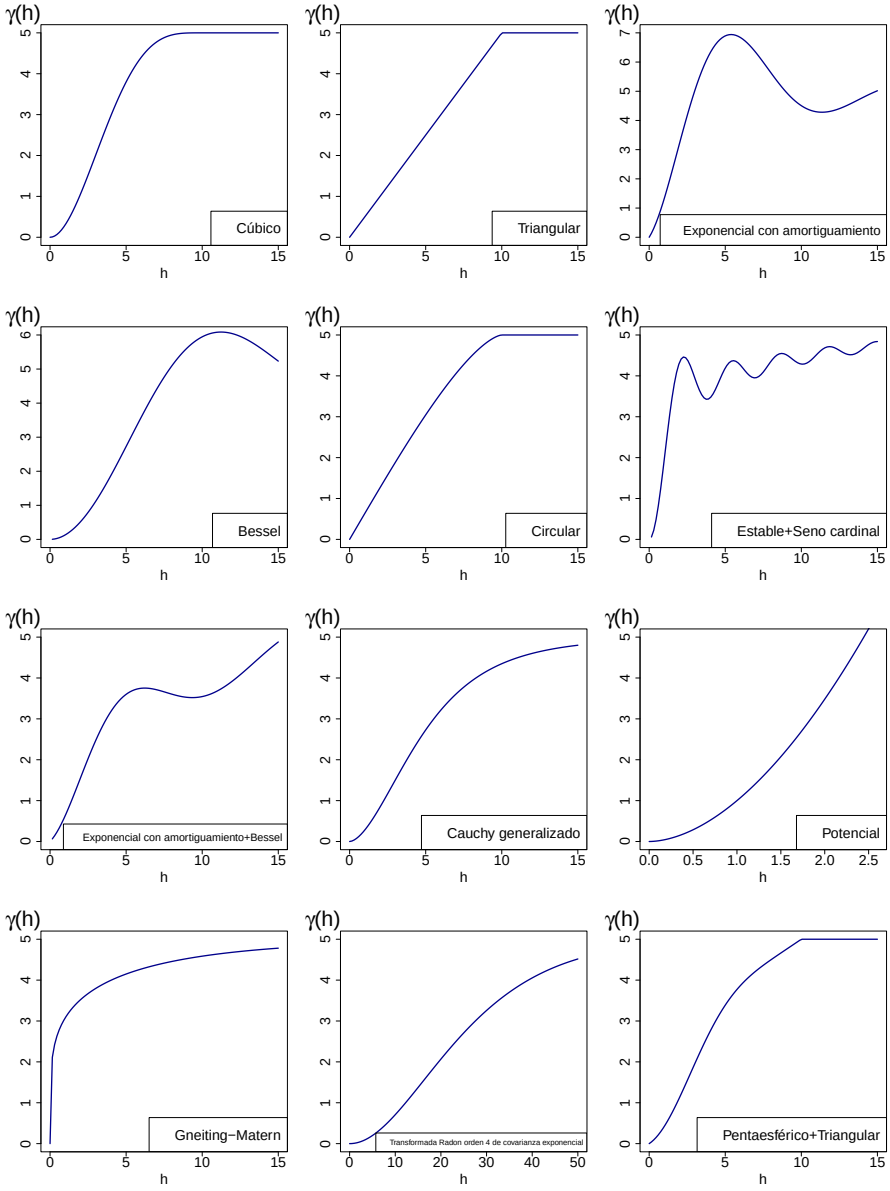


Figura 2.4. Algunos modelos válidos de semivariogramas y combinaciones lineales de ellos, usando el modelo lineal de regionalización (Ver Sección 3.7).

En todos los modelos presentados a continuación, a representa el alcance espacial y por lo tanto $a > 0$. Los modelos esférico, circular, pentaesférico, triangular y cúbico muestran un alcance finito; la función se estabiliza perfectamente.

Modelo esférico. Válido en $\mathbb{R}^d,\ d = 1, 2, 3.$

$$\gamma(h;\Theta) = \begin{cases} \sigma^2 \left(\frac{3}{2}\frac{\|h\|}{a} - \frac{1}{2}\left(\frac{\|h\|}{a}\right)^3 \right) & \text{si} \quad 0 \le \|h\| \le a \\ \sigma^2 & \text{si} \quad \|h\| \ge a. \end{cases}$$

Modelo circular. Válido en $\mathbb{R}^d,\ d = 1, 2.$

$$\gamma(h;\Theta) = \begin{cases} \sigma^2 \left(1 + \frac{2||h||}{a\pi}\sqrt{1 - \frac{||h||^2}{a^2}} - \frac{2}{\pi}\cos^{-1}\left(\frac{||h||}{a}\right) \right) & \text{si} \quad 0 \le ||h|| \le a \\ \sigma^2 & \text{si} \quad \|h\| \ge a. \end{cases}$$

Este comportamiento puede ser un poco irreal, ya que la dependencia espacial va decreciendo, pero es difícil imaginarse una situación práctica en la que esta se vaya a cero exactamente de un punto a otro, en un dominio espacial continuo. Sin embargo, estos modelos tienen unas inmensas ventajas computacionales, dado que generan matrices de covarianza con gran cantidad de posiciones con valor cero, 0, y esto agiliza enormemente la inversión de las matrices que suele requerirse en los procesos de estimación, usualmente iterativos por la ausencia de linealidad en los parámetros de los modelos.

Modelo pentaesférico. Válido en $\mathbb{R}^d$, $d = 1, 2, 3$

$$\gamma(h;\Theta) = \begin{cases} \sigma^2 \left(\frac{15}{8}\frac{||h||}{a} - \left(\frac{5}{4}\frac{||h||}{a}\right)^3 + \left(\frac{3}{8}\frac{||h||}{a}\right)^5 \right) \\ \text{si} \quad 0 < ||h|| \le a \\ \\ \sigma^2 \quad \text{si} \quad \|h\| \ge a. \end{cases}$$

Modelo triangular. Válido en $\mathbb{R}^1$

$$\gamma(h;\Theta) = \begin{cases} \sigma^2 \left(\frac{||h||}{a} \right) & \text{si} \quad 0 < ||h|| \le a \\ \sigma^2 & \text{si} \quad ||h|| \ge a \end{cases}$$

Modelo Cúbico. Válido en $\mathbb{R}^d$, $d = 1, 2, 3$

$$\gamma(h;\Theta) = \begin{cases} \sigma^2 \left(\frac{7||h||^2}{a^2} - \frac{35}{4}\frac{||h||^3}{a^3} + \frac{7}{2}\frac{||h||^5}{a^5} - \frac{3}{4}\frac{||h||^7}{a^7} \right) & \text{si} \quad 0 < ||h|| \leq a \\ \sigma^2 & \text{si} \quad ||h|| \geq a \end{cases}$$

Los modelos presentados a continuación son estacionarios de segundo orden, pero, a diferencia de los anteriores, alcanzan la silla de manera asintótica. Uno de los modelos más utilizados debido a la gran flexibilidad dada por su parámetro de suavizamiento κ es el modelo de Matern, (ver Figura 2.5). Si $\kappa = 0,5$ coincide con el modelo exponencial y si $\kappa \to \infty$ coincide con el modelo gaussiano. Así, el rango no es el valor exacto de a, pues después de a aún puede existir correlación espacial importante. Es usual seguir llamándolo rango, pero se debe identificar para qué valor de a la distancia entre la curva y la línea a la altura de la silla es muy pequeña. Esta distancia se denomina rango o alcance práctico. La parametrización de estos modelos, por lo tanto, no es única. Por ejemplo, algunos incluyen el tres en el denominador y a $3a$ lo llaman el alcance práctico.

El modelo gaussiano es uno de los más continuos cerca del origen. De hecho, es infinitamente diferenciable cerca de 0. Cuando los modelos son más suaves cerca del origen que lejos del origen, generan predicciones erráticas [9].

Modelo Matérn. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \left\{ \sigma^2 \left(1 - \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \right) \left(\frac{\|h\|}{a} \right)^{\kappa} K_{\kappa}\left(\frac{\|h\|}{a} \right) \right.$$

donde $K_{\kappa}(\frac{\|h\|}{a})$ es la función de Bessel modificada de tercera clase de orden κ, $\kappa > 0$. Ver Figura 2.5.

Modelo exponencial. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2 \left(1 - \exp\left(-\frac{||h||}{a} \right) \right)$$

Modelo Gaussiano. Válido en $\mathbb{R}^d$

$$\gamma(h;\Theta) = \sigma^2 \left(1 - \exp\left(-\frac{||h||^2}{a^2} \right) \right)$$

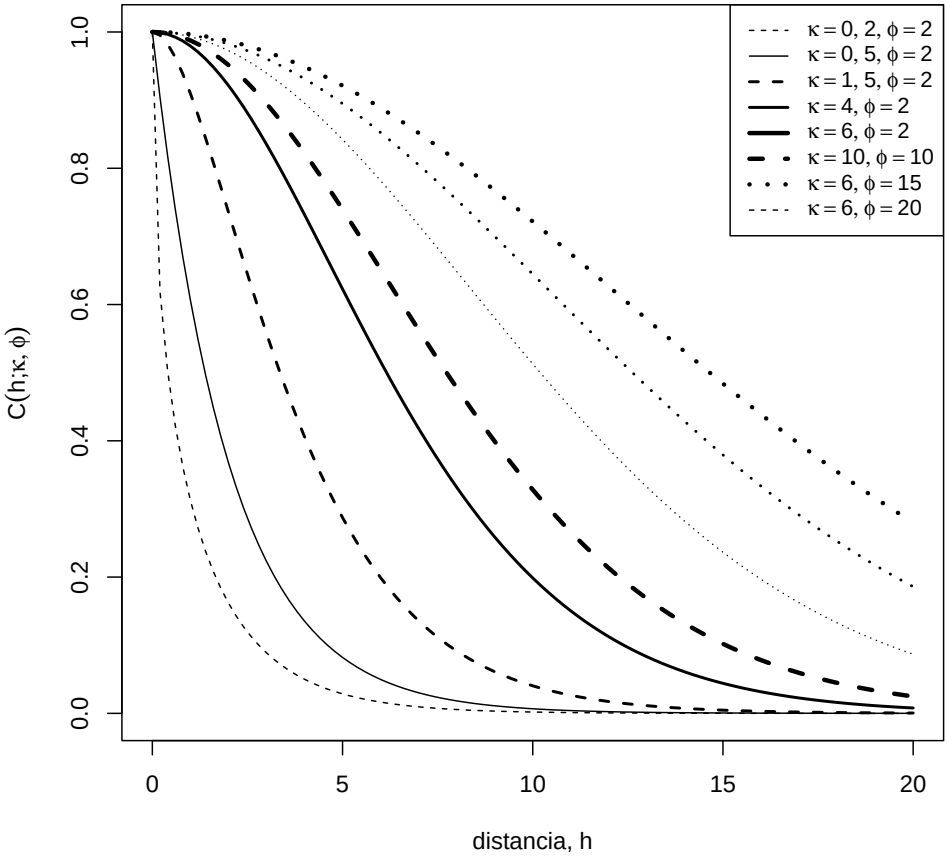


Figura 2.5. Modelo de autocovarianza Matérn para diferentes valores de los parámetros de suavizamiento y de rango. Es un modelo muy flexible.

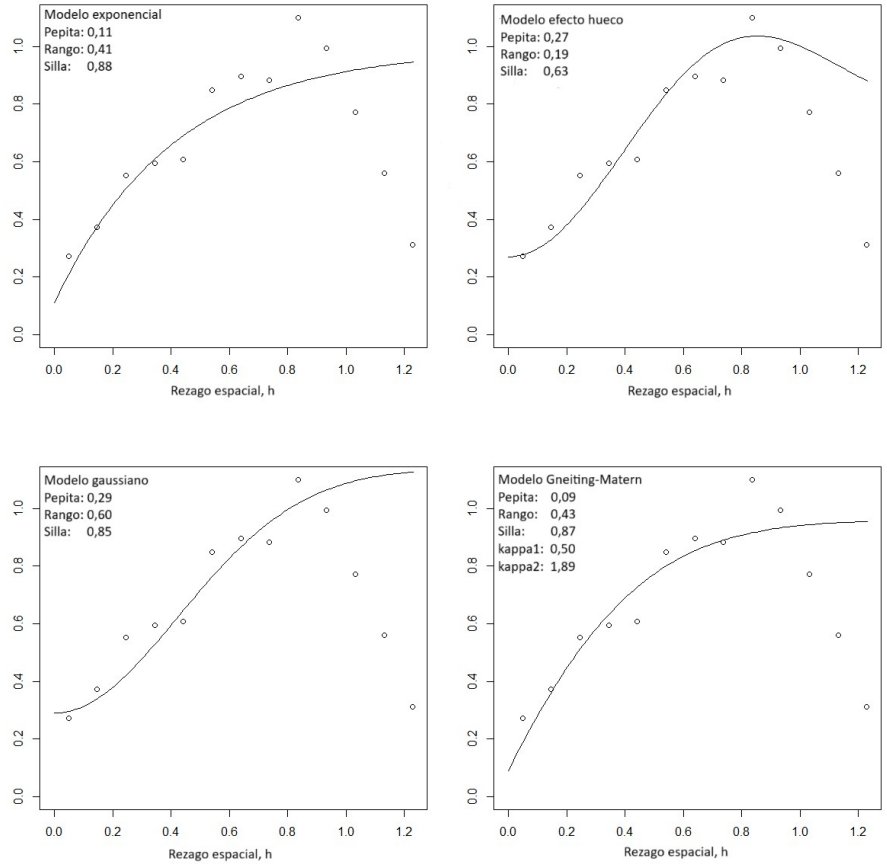


Tabla 2.2. Ilustración de un semivariograma empírico y algunos posibles modelos teóricos ajustados. Note al final el comportamiento decreciente del semivariograma empírico. Es común que haya menor cantidad de pares de datos para distancias grandes, lo que genera un comportamiento errático de la estimación empírica. En estos casos, lo más adecuado es modelar solo hasta la distancia a la cual existe suficiente información y tener en cuenta este radio para la predicción.

Modelo cuadrático racional. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2\left(\frac{||h||^2}{1+\frac{||h||^2}{a}}\right)$$

Modelo estable. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2\left(1-\exp\left(-\left(\frac{||h||}{a}\right)^{\delta}\right)\right) \quad \forall h, \quad 0 \leq \delta < 2.$$

Transformada de Radon de orden 2 del modelo exponencial. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2\left(1-\left(1+\frac{||h||}{a}\right)\exp\left(-\frac{||h||}{a}\right)\right)$$

Transformada de Radon de orden 4 del modelo exponencial. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2\left(1-\left(1+\frac{||h||}{a}+\frac{1}{3}\frac{||h||^2}{a^2}\right)\exp\left(-\frac{||h||}{a}\right)\right)$$

Modelo de Cauchy generalizado. Válido en $\mathbb{R}^d$. $\kappa_1 > 0$, $0 < \kappa_2 \leq 2$

$$\gamma(h;\Theta) = \sigma^2\left(1-\left(1+\left(\frac{||h||}{a}\right)^{\kappa_2}\right)^{-\kappa_1/\kappa_2}\right)$$

Modelo de Lantuéjoul. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2\left(1-\sin\left(\frac{\pi}{2}\exp\left(-\frac{|h|}{a}\right)\right)\right)$$

Los modelos presentados a continuación admiten ondas, esto es, admiten correlación espacial negativa. Estos modelos siguen siendo estacionarios de segundo orden, pero las ondas van disminuyendo su amplitud cuando aumenta el rezago espacial pero continúan oscilando alrededor de la silla.

Modelo del seno cardinal (Efecto hueco). Válido en $\mathbb{R}^d$, $d = 1, 2, 3$.

$$\gamma(h;\Theta) = \begin{cases} 0 & \text{si } ||h|| = 0; \\ \sigma^2\left(1-\frac{a}{||h||}\sin\left(\frac{||h||}{a}\right)\right) & \text{si } ||h|| > 0. \end{cases}$$

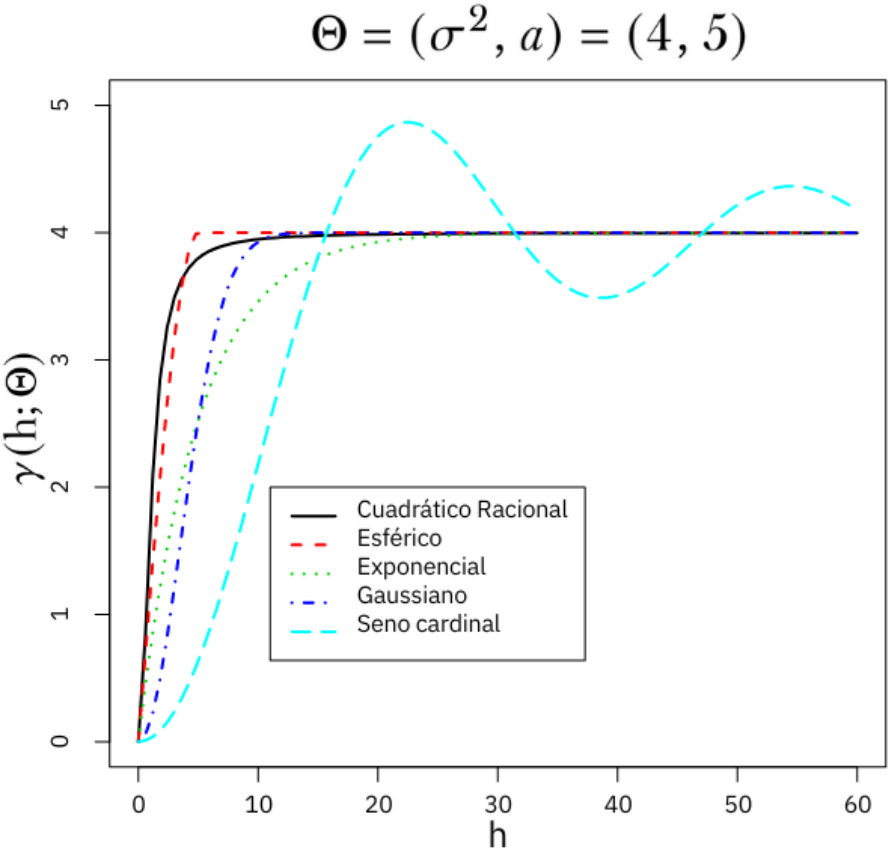


Figura 2.6. Semivariograma teórico para cinco modelos acotados distintos, pero con los mismos parámetros de silla y rango. Nótese las diferencias de comportamiento entre ellos.

Modelo de Bessel. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \sigma^2 - \sigma^2\left(2^{d/2-1}\Gamma\left(\frac{d}{2}\right)\left(\frac{||h||}{a}\right)^{1-d/2} J_{d/2-1}\left(\frac{||h||}{a}\right)\right)$$

$J_\cdot()$ es la función de Bessel de la primera clase.

Modelo exponencial con amortiguamiento. Válido en $\mathbb{R}^1$. Válido en $\mathbb{R}^2$ si y solamente si $a_2 \geq a_1$. Válido en $\mathbb{R}^3$ si y solamente si $a_2 \geq a_1\sqrt{3}$, $a_1, a_2 > 0$.

$$\gamma(h;\Theta) = \sigma^2 - \sigma^2\left(\exp\left(-\frac{||h||}{a_1}\right)\cos\left(\frac{||h||}{a_2}\right)\right)$$

Modelo coseno. El máximo efecto hueco es obtenido con el modelo coseno. Sin embargo, este modelo es solo definido no negativo en $\mathbb{R}^1$, por esta razón se usa más el modelo exponencial con amortiguamiento.

$$\gamma(h;\Theta) = \sigma^2 - \sigma^2\left(\cos\left(\frac{||h||}{a}\right)\right).$$

2.6. Modelos de semivariograma no acotados

Los modelos presentados a continuación no tienen varianza finita, esto es, no alcanzan silla y, por lo tanto, no tienen rango. Son válidos, ya que cumplen los requisitos de la estacionariedad intrínseca, esto es, son funciones condicionalmente definidas negativas (ver Sección 2.3). Estos modelos se reconocen a partir de los semivariogramas empíricos porque el valor de la semivarianza espacial nunca se estabiliza, esto es, nunca alcanza silla (ver Figura 2.4).

Modelo potencial. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = b||h||^\delta, \qquad b > 0.$$

Este modelo es válido para $0 < \delta < 2$, ya que en otro caso crece demasiado rápido y con esto violaría la hipótesis de estacionariedad intrínseca (ver Definición 6). Un caso particular es el modelo lineal.

Modelo lineal Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = b\|h\|, \qquad b > 0.$$

Modelo logarítmico modificado o modelo de Wijsian-c^2. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \frac{3}{2} b \log\left(||h||^2 + c^2\right), \qquad c \neq 0 \qquad 0 \leq b < 1 \qquad (2.10)$$

El modelo de Wijsian tiene varias versiones y propiedades muy útiles e interesantes. Ver detalles en [45] y [19]. Es un modelo no acotado ya que aplica logaritmo a la distancia, $\log(||h||)$, pero esto tiene el inconveniente de que no pasa por el origen, de ahí la modificación presentada en 2.10.

Finalmente, a continuación, se presenta un modelo muy flexible y que tiene como casos particulares tanto modelos acotados como modelos no acotados, como por ejemplo el Cauchy generalizado y el Wijsian-c^2, ver [68] para un tratamiento detallado del mismo.

Modelo puente entre modelos acotados y no acotados. Válido en $\mathbb{R}^d$.

$$\gamma(h;\Theta) = \frac{b\left(1 + ||h||^\alpha\right)^{\beta/\alpha} - 1}{2^{\beta/\alpha} - 1}, \qquad 0 < \alpha \leq 2, \quad -\infty < \beta \leq 2$$

α controla la suavidad tanto del semivariograma como del campo aleatorio asociado y β el comportamiento de largo-alcance. A mayor alcance mayor continuidad espacial y viceversa. Es acotado si y solamente si $\beta < 0$.

La suavidad de las funciones de dependencia espacial muestra la velocidad a la cual la variable va cambiando a través del espacio. Entre más suaves son, más continuo es el proceso. A distancias cortas, la función de autocorrelación toma valores más altos, es decir, su semivariograma toma valores más bajos.

La predicción en un lugar no muestreado es más fácil y más precisa si el proceso no presenta cambios bruscos. Si el proceso es más irregular y su variabilidad es mayor, no es tan clara la estructura y existe menos información sobre $Z(\boldsymbol{s})$ en los lugares vecinos a $\boldsymbol{s}$. Cuando los modelos vuelven a presentar una suavidad similar lejos del origen que cerca del origen, las técnicas de predicción espacial son imprecisas y confunden la autocorrelación cerca y lejos del origen [9], dado que pueden generar ponderaciones altas para variables alejadas de la ubicación objetivo.

Nota 4. *En lo sucesivo se utilizan h y ||h|| indistintamente.*

2.7. Ejercicios

1. Grafique los modelos de covarianza y semivarianza:

   1.1 Fije la silla y varíe el rango.

   1.2 Fije el rango y varíe la silla.

   1.3 Fije silla y rango y varíe parámetro de suavidad cuando aplique.

   1.4 Compare los comportamientos de diferentes modelos de covarianza y semivarianza cuando para todos se usan los mismos parámetros.

2. Demuestre o refute: la covarianza entre un proceso espacial en la ubicación $\boldsymbol{s}_i$ y el proceso valor promedio del fenómeno sobre el área B es igual al promedio de las covarianzas entre el proceso espacial en la ubicación $\boldsymbol{s}_i$ y los procesos espaciales existentes en todos los puntos dentro de B.

3. Compare para varios conjuntos de datos los 3 estimadores empíricos del semivariograma. ¿Existe alguna relación de orden entre los 3? ¿Existe alguna relación de orden entre 2 de ellos? o, definitivamente, no existe ningún orden entre estos estimadores?

4. Deduzca la correlación entre dos estimaciones del semivariograma empírico $\gamma(\boldsymbol{h}_k)$ y $\gamma(\boldsymbol{h}_l)$ para el modelo exponencial. Use la información dada en la nota 5.

5. A continuación, se muestran las funciones de covarianza y densidad espectral para un modelo de covarianza gaussiano. Utilice el Teorema 1 para demostrar que una puede obtenerse a partir de la otra,

   Función de covarianza:

$$C(h) = \sigma^2 \exp(-\alpha^2)$$

   Función de densidad espectral:

$$f(\omega) = \frac{1}{2}\sigma^2(\pi\alpha)^{-\frac{1}{2}} \exp\left(-\frac{\omega^2}{4\alpha}\right)$$

Teorema 1 (Teorema Bochner). *Una función continua es definida no-negativa si y solamente si puede ser representada como sigue:*

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^2} \exp(ih'\omega) dF(\omega)$$

donde F es una función real, acotada y no decreciente. Si F es absolutamente continua (con respecto a la medida de Lebesgue):

$$C(\boldsymbol{h}) = \int_{\mathbb{R}^2} \exp(ih'\omega) f(\omega) d\omega$$

6. Seleccione otra función de covarianza espacial y aplique el Teorema 1 para encontrar la función de densidad espectral asociada. ¿La densidad espectral siempre existe? Justifique su respuesta.

7. Haga una tabla con la siguiente estructura para todos los modelos en los que aplique: Columna 1, función de autocovarianza; Columna 2, función de semivarianza; Columna 3, función de autocorrelación. ¿Existen expresiones para las 3 funciones en todos los modelos? Si no, ¿en qué casos no existen expresiones para las 3 columnas y porqué?

8. Realice un análisis geoestadístico para los datos presentados en la Tabla 2.3. A continuación se sugieren los pasos para llevar a cabo un análisis de estacionariedad e isotropía así como de predicción espacial.

    8.1 Encuentre la matriz de correlación de la variable y sus coordenadas geográficas. Analice.

    8.2 Realice gráficos de dispersión de la variable contra cada una de las coordenadas geográficas. Analice.

    8.3 Realice gráficos de dispersión de la media (mediana) de cada columna con respecto a su coordenada horizontal.

    8.4 Realice gráficos de dispersión de la media (mediana) de cada fila con respecto a su coordenada vertical.

    8.5 Realice gráficos de dispersión de la varianza de cada columna con respecto a su coordenada horizontal.

    8.6 Realice gráficos de algunos dispersogramas. Encuentre el coeficiente de autocorrelación para cada caso. Analice.

    8.7 Realice gráficos de dispersión de la varianza de cada fila con respecto a su coordenada vertical.

8.8 Realice al menos un gráfico de medias móviles que no sea ni por fila ni por columna.

8.9 Existe tendencia en alguna dirección? En caso afirmativo, explore varios modelos de regresión e intente ajustarla y posteriormente verifique que la tendencia fue eliminada realizando gráficos de dispersión de los residuales del modelo de regresión seleccionado.

8.10 Grafique el semivariograma empírico horizontal de la variable centrada.

8.11 Grafique el semivariograma empírico vertical de la variable centrada.

8.12 Grafique el semivariograma empírico en dirección 45 grados de la variable centrada.

8.13 Grafique el semivariograma empírico omnidireccional, esto es, sin tener en cuenta la dirección, sino únicamente la distancia de la variable centrada.

8.14 Grafique, si es posible, un variograma con dirección diferente a las anteriores de la variable centrada.

8.15 ¿Tienen un comportamiento similar los semivariogramas en todas las direcciones? ¿En todas las direcciones el semivariograma empírico sugiere ajustar un modelo acotado? Esto es, se observa que existen silla y rango? o por el contrario la varianza no parece estabilizarse?

8.16 Ajuste un modelo de semivariograma teórico omnidireccional. Según lo encontrado en el análisis a diferentes ángulos, determine si es necesario aplicar corrección por anisotropía.

8.17 Con los datos originales, aplique dos iteraciones del pulimiento de medianas y obtenga los residuales. Encuentre al menos 10 predicciones utilizando kriging con pulimiento de medianas. Sumele a las predicciones los efectos fila, columna y global para devolver los datos a su escala original.

8.18 Realice la validación cruzada (al menos para 10 lugares observados) y calcule:

8.18.1 Media (mediana) de los errores. Media (mediana) estandarizada.

8.18.2 Grafique observados Vs. predichos y calcule su coeficiente de correlación lineal de Pearson.

Norte										
	29		19,6		20,8		19,6	19,8	21,5	
	27		18,8	19,9	20,3	20,6	19,9	20,1	19,8	19,9
	25		17,9	20,2	20,5	19,8	21.0	20,1	20,2	20,9
	23	14,5	18,5	19,9	20,6	20,2	20,4	20,9	21,9	22,1
	21	16,5	18,6	19,9	20,5	20,3	21,6	20,9	19,8	19,4
	19	19,5	19,0	20,8	20,8	21,5	19,8	20,8	19,5	21,2
	17	18,2	19,2	20,6	20,9	19,8	19,9	20,8		19,9
	15	17,4	18,5	20,5	20,9	19,9	22,1	20,3		21,1
	13	16,5	19,0	19,9	20,8	21,9	20,9	21,5		20,2
	11	19,5	19,2	20,8		20,9	21,6	19,9		
	9	18,2	18,5	20,6		21,5	21,7	22,1		
	7	17,4	18,5	20,5		20,7	20,9	20,9		
	5	14,5		20,2			21,6			
	3	16,5		19,9						
	1									
		1	2	3	4	5	6	7	8	9

Este

Tabla 2.3. Datos de sensación térmica

Capítulo

tres

Estimación teórica de las funciones de semivarianza y de autocovarianza

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$
$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$

En este capítulo se muestran algunos métodos de estimación de parámetros de modelos de regresión, adaptados al caso de de la semivarianza o de la covarianza espacial. En términos generales, es importante recordar que estos métodos estiman los parámetros, asumiendo que el modelo propuesto es el verdadero. Esto es, si se propone un modelo exponencial, por ejemplo, el método encuentra los parámetros que optimizan un criterio bajo el supuesto de que el modelo exponencial es el modelo correcto de semivarianza. De allí la importancia del semivariograma empírico, también conocido como experimental. Para su estimación es importante tener en cuenta:

- Debe ser construido con los datos libres de tendencia, es decir, con media constante. En caso de encontrarse tendencia en los datos, esta tendencia se modela con modelos de regresión polinómicos o con análisis a dos vías y el semivariograma se construye con los residuales obtenidos en cualquiera de estos casos. Estos residuales ya son de media cero.

- Se debe usar un estimador resistente a datos atípicos en caso de detectar su presencia.

- Es importante elegir una distancia máxima, para así incluir solo pares de datos con distancias menores a un h específico, en caso de que se observe un comportamiento errático a distancias mayores.

- Es necesario familiarizarse con las características de los modelos válidos y compararlas con el comportamiento del semivariograma empírico para elegir el modelo más adecuado.

- Usualmente, los modelos de semivarianza o autocovarianza no son lineales en los parámetros. Por lo tanto, se usan métodos iterativos y se requieren valores iniciales. Estos se eligen a partir de la observación-exploración del semivariograma empírico. Esto se conoce como estimación a ojo o a sentimiento.

3.1. Mínimos cuadrados ordinarios

Los valores del semivariograma empírico son las realizaciones de la variable respuesta, en función de la separación $\boldsymbol{h}$ (ver Tabla 2.1). El modelo $\gamma(\boldsymbol{h};\Theta)$ es un modelo válido elegido de acuerdo con el comportamiento visto en el semivariograma empírico (ver sección 2.3).

La estimación por el método de los mínimos cuadrados ordinarios consiste en encontrar el valor del parámetro Θ que minimice $\mathbb{Q}$ dado por

$$\mathbb{Q} = \sum_{k=1}^{K} \left(\hat{\gamma}(\boldsymbol{h}_k) - \gamma(\boldsymbol{h}_k; \Theta)\right)^2 .$$

Si la cantidad de pares de datos para calcular cada $\hat{\gamma}(\boldsymbol{h}_k)$ y el cambio entre varianzas de cada grupo no es tan importante, el método de mínimos cuadrados ordinarios puede funcionar bien. En otro caso, es mejor usar mínimos cuadrados ponderados.

3.2. Mínimos cuadrados ponderados.

Dado que cada punto del semivariograma es estimado con diferente cantidad de datos y que las varianzas entre semivariogramas empíricos son notoriamente distintas de un rezago espacial a otro, es adecuado usar mínimos cuadrados ponderados (ver Tabla 2.1).

La adaptación de este método de estimación al caso de la estimación del semivariograma consiste en construir la matriz de ponderación $\mathrm{W}(\Theta)$, usando la varianza del estimador empírico del semivariograma. [23] encuentra una aproximación a esta varianza bajo el supuesto de normalidad al menos marginal de $Z(s)$. Así, se tiene que para cualquier $\boldsymbol{h}$,

$$(Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s}))^2 \sim 2\gamma(\boldsymbol{h};\Theta)\,\chi^2_1.$$

y en consecuencia para cada rezago $\boldsymbol{h}_k$

$$\mathrm{Var}[Z(\boldsymbol{s}+\boldsymbol{h}_k) - Z(\boldsymbol{s})]^2 = 2(2\gamma(\boldsymbol{h}_k;\Theta))^2.$$

Incluyendo el denominador del semivariograma empírico para ponderar también por la cantidad de pares encontrados para cada rezago $\boldsymbol{h}_k$, la matriz diagonal de ponderación se puede construir como sigue

$$\mathrm{W}(\Theta) \simeq \mathrm{Diag}\left(\frac{2(2\gamma(\boldsymbol{h}_k;\Theta))^2}{|N(\boldsymbol{h}_k)|}\right).$$

con

$$N(\boldsymbol{h}_k) = \{(i,j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}_k\}.$$

Para las ubicaciones $i, j = 1, \ldots, n$, que generan los primeros k rezagos espaciales $k = 1, \ldots, K$, [22]. $|N(\boldsymbol{h}_k)|$ es el cardinal del conjunto $N(\boldsymbol{h}_k)$.

La estimación de los parámetros del semivariograma por MCP, consiste en construir la matriz de ponderación $\mathrm{W}(\Theta)$ usando la varianza del estimador empírico del semivariograma y $|N(\boldsymbol{h}_k)|$ que es el cardinal del conjunto $N(\boldsymbol{h}_k)$. Entonces

$$\mathrm{W}(\Theta) \simeq \mathrm{Diag}\left(\frac{2(2\gamma(\boldsymbol{h}_k;\Theta))^2}{|N(\boldsymbol{h}_k)|}\right).$$

y la expresión a minimizar es

$$(2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\Theta}))' \, \mathrm{W}^{-1}(\Theta) \, (2\hat{\boldsymbol{\gamma}} - 2\boldsymbol{\gamma}(\boldsymbol{\Theta})) \, .$$

$\hat{\boldsymbol{\gamma}} = (\hat{\gamma}(\boldsymbol{h}_1), \ldots, \hat{\gamma}(\boldsymbol{h}_K))$ y $\boldsymbol{\gamma}(\Theta) = (\gamma(\boldsymbol{h}_1;\Theta), \ldots, \gamma(\boldsymbol{h}_K;\Theta))$.

Esta opción es usada muy frecuentemente y en general arroja buenos resultados en la calidad de la estimación. Es importante resaltar las siguientes características de este método de estimación:

- La matriz de ponderación depende del parámetro a estimar Θ. Por consiguiente, cada etapa de la iteración tiene una matriz de ponderación distinta obtenida al reemplazar el valor de $\hat{\Theta}$ resultante de la iteración inmediatamente anterior.

- Los modelos válidos de semivarianza en su mayoría son no lineales en sus parámetros, y los sistemas de ecuaciones de estimación derivados, no permiten despejar el valor de cada parámetro en términos de los demás. En el contexto de estimación de modelos de regresión no lineales, se requiere de procedimientos de estimación iterativos, partiendo de unos valores iniciales dados a los parámetros.

- Existen otras alternativas de ponderación como por ejemplo

$$\mathrm{W}(\Theta) = \mathrm{Diag}\left(\frac{1}{|N(h_k)|}\right) \quad \text{o} \quad \mathrm{W}(\Theta) = \mathrm{Diag}\left(\frac{h_k^2}{|N(h_k)|}\right)$$

que también funcionan bien en la práctica y ayudan a equilibrar el hecho de tener diferentes $|N(\boldsymbol{h}_k)|$ para cada $\boldsymbol{h}_k$. Nótese que en estas alternativas de ponderación, la matriz $\mathrm{W}(\Theta)$ no cambia de una iteración a otra.

Los métodos de estimación presentados hasta ahora se basan en la estimación del semivariograma empírico. A partir de este gráfico se eligen el modelo a estimar y sus parámetros iniciales. Este procedimiento de estimación tiene las desventajas de que el número de rezagos y su longitud son elegidos subjetivamente. Si se tienen pocas observaciones, la cantidad de datos en cada una de estas clases es muy pequeña, de tal forma que para los rezagos menores se pueden tener suficientes datos para la estimación de cada $\hat{\gamma}(\boldsymbol{h}_k)$, pero no así para los rezagos mayores. Estos son problemas compartidos por el histograma con el número de clases o por el estimador de densidad kernel con el ancho de banda.

Los métodos presentados en la siguiente sección se basan en el conocimiento de la función de verosimilitud de $\mathbb{Z}$ o $\mathbb{Y}$, excepto por sus parámetros, y *en teoría* no requieren de la estimación del semivariograma empírico. Sin embargo, no existe otra forma de aproximarnos al comportamiento de la variabilidad espacial del proceso y a sus características principales. Así que, los métodos de máxima verosimilitud y máxima verosimilitud restringida pueden arrancar con cualquier matriz de covarianza no estructurada, que no tengan ninguna relación con el fenómeno, como valor inicial. Sin embargo, esto puede llevar a estimaciones pobres, a demoras en la convergencia o a que la optimización se quede en óptimos locales. Además, se requeriría imponer en cada iteración la restricción de que la matriz obtenida sea definida positiva y aún quedaría pendiente la forma de estimar el vector de covarianza entre los lugares observados y el lugar a predecir.

En términos prácticos, lo mas adecuado es usar la estimación del semivariograma empírico, el modelo propuesto y el ajuste a sentimiento respectivo para construir la matriz de covarianza inicial y así estimar Θ.

3.3. Máxima verosimilitud.

La estimación de los parámetros de la función de covarianza usando los métodos de máxima verosimilitud (ML) y de máxima verosimilitud restringida (REML), requiere de la especificación de la distribución de probabilidad conjunta del vector

$$\mathbb{Y} = \big(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\big).$$

En general, la estimación se lleva a cabo con base en el supuesto de normalidad multivariada, debido a la posibilidad que da esta distribución de tener expresiones matemáticas cerradas y además porque los parámetros de

esta distribución, coinciden con los parámetros relacionados en el supuesto de estacionariedad de segundo orden. También es importante notar que estos métodos tienen unas propiedades estadísticas excepcionales, ver por ejemplo [17].

En cuanto a los métodos conocidos como de pseudoverosimilitud, son aquellos métodos que relajan el supuesto de la distribución conjunta y construyen las ecuaciones de estimación con base en distribuciones marginales univariadas o bivariadas. Esto es útil, siempre y cuando sea posible encontrar relación entre los parámetros de la distribución resultante y los parámetros de interés.

Para el caso ML se asume que

$$\mathbb{Y} \sim N_n(\mathrm{X}\boldsymbol{\beta}, \Sigma(\Theta)).$$

donde $\Sigma(\Theta) = \mathrm{Cov}[\mathbb{Y}]$ es una matriz de dimensión $n \times n$ y (X es una matriz de dimensión $n \times q$, $q < n$, de variables explicativas, dentro de las cuales comúnmente se encuentran las coordenadas geográficas o funciones polinómicas. El negativo de la función de log-verosimilitud es

$$L(\boldsymbol{\beta}, \Theta) = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln|\boldsymbol{\Sigma}(\Theta)| + \frac{1}{2}(\mathbb{Y} - \mathrm{X}\boldsymbol{\beta})'\Sigma^{-1}(\Theta)(\mathbb{Y} - \mathrm{X}\boldsymbol{\beta}).$$

El elemento ij de la matriz $\Sigma(\Theta)$ corresponde a la covarianza espacial entre las variables $Y(\boldsymbol{s}_i)$ y $Y(\boldsymbol{s}_j)$, esto es,

$$\mathrm{Cov}\left[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)\right] = C(\boldsymbol{s}_i - \boldsymbol{s}_j; \Theta) = C(\boldsymbol{h}; \Theta).$$

$C(\boldsymbol{h}; \Theta)$ es una función definida positiva con la que se construye la matriz $\Sigma(\Theta)$. Por lo tanto, $\Sigma(\Theta)$ es una matriz definida positiva y en consecuencia $\Sigma^{-1}(\Theta)$ existe y es única.

El estimador $\hat{\Theta}$ es sesgado, pero asintóticamente eficiente; sin embargo, paradójicamente, tener una muestra grande genera el inconveniente de tener que realizar gran cantidad de operaciones, debido al cálculo tanto del determinante como de la inversa de la matriz de covarianza en forma iterativa.

Ejemplo 3 (Elementos de la función ML). *Si Y está georreferenciado en $\mathbb{R}^2$, $\boldsymbol{s}_i = (\mathrm{x}_i, \mathrm{y}_i)$, y su media es una función lineal de ambas coordenadas, entonces, los elementos de la función de verosimilitud son los siguientes bajo el modelo $Y(\boldsymbol{s}) =$*

$X(s)\beta + Z(s)$, $E[\mathbb{Y}] = \mathrm{X}\boldsymbol{\beta}$, $E[Z(\boldsymbol{s})] = 0$, $\mathrm{Var}[\mathbb{Y}] = \mathrm{Var}[\mathbb{Z}] = \Sigma(\Theta)$.

$$\mathbb{Y} = \begin{pmatrix} Y(\boldsymbol{s}_1) \\ \vdots \\ Y(\boldsymbol{s}_i) \\ \vdots \\ Y(\boldsymbol{s}_n) \end{pmatrix} \qquad \mathrm{X} = \begin{pmatrix} 1 & \mathrm{x}_1 & \mathrm{y}_1 \\ \vdots & \vdots & \vdots \\ 1 & \mathrm{x}_i & \mathrm{y}_i \\ \vdots & \vdots & \vdots \\ 1 & \mathrm{x}_n & \mathrm{y}_n \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \qquad \mathbb{Z} = \begin{pmatrix} Z(\boldsymbol{s}_1) \\ \vdots \\ Z(\boldsymbol{s}_i) \\ \vdots \\ Z(\boldsymbol{s}_n) \end{pmatrix}.$$

$$\Sigma(\Theta) = \begin{pmatrix} C(\mathbf{0};\Theta) & C(\boldsymbol{s}_1 - \boldsymbol{s}_2;\Theta) & \dots & C(\boldsymbol{s}_1 - \boldsymbol{s}_n;\Theta) \\ C(\boldsymbol{s}_2 - \boldsymbol{s}_1;\Theta) & C(\mathbf{0};\Theta) & \dots & C(\boldsymbol{s}_2 - \boldsymbol{s}_n;\Theta) \\ \vdots & \vdots & \ddots & \vdots \\ C(\boldsymbol{s}_n - \boldsymbol{s}_1;\Theta) & C(\boldsymbol{s}_n - \boldsymbol{s}_2;\Theta) & \dots & C(\mathbf{0};\Theta) \end{pmatrix}.$$

C es la función de covarianza del proceso espacial (ver Sección 2.5). En general, cada fila de la matriz X *es una coordenada o una función usualmente polinómica de ellas.*

Nótese que aquí se plantea la forma general de estimar $(\boldsymbol{\beta}, \Theta)$ simultáneamente. Si se tiene el proceso centrado y la media se conoce o ya se ha estimado, se puede usar directamente $\mathbb{Z}$ y solo se estima Θ, $\mathbb{Z} \sim N_n(\mathbf{0}, \Sigma(\Theta))$.

Ejemplo 4 (Ilustración de los cálculos para la construcción de la función de log-verosimilitud con media constante y semivarianza espacial, exponencial.). *Supongamos con fines ilustrativos que $n = 2$ y que las realizaciones del vector $\mathbb{Y}$ son $\mathbb{y} = (y(\boldsymbol{s}_1), y(\boldsymbol{s}_2)) = (3, 4)$ y $|\boldsymbol{s}_1 - \boldsymbol{s}_2| = 5$. La función de semivarianza exponencial es*

$$\gamma(h, \Theta) = \sigma^2 \left(1 - \exp\left(-\frac{h}{a}\right)\right) \quad , \quad h \geq 0$$

Dado que hay estacionariedad de segundo orden,se puede usar la equivalencia

$$\gamma(h;\Theta) = \sigma^2 - C(h;\Theta)$$

y la función de covarianza correspondiente es

$$C(h, \Theta) = \sigma^2 \exp\left(-\frac{h}{a}\right) \quad \textit{para} \quad h \geq 0$$

Nótese que $\sigma^2 = C(0, \Theta)$. *, así que la matriz de covarianza, su determinante y su inversa, están dadas por:*

- *Matriz de covarianza*

$$\Sigma(\Theta) = \begin{pmatrix} C(\mathbf{0}, \Theta) & C(\boldsymbol{s}_1 - \boldsymbol{s}_2; \Theta) \\ C(\boldsymbol{s}_2 - \boldsymbol{s}_1; \Theta) & C(\mathbf{0}, \Theta) \end{pmatrix}$$

- *Determinante*

$$|\Sigma(\Theta)| = C^2(\mathbf{0}, \Theta) - C(\boldsymbol{s}_2 - \boldsymbol{s}_1; \Theta)C(\boldsymbol{s}_1 - \boldsymbol{s}_2; \Theta) = C^2(\mathbf{0}, \Theta) - C^2(\boldsymbol{s}_1 - \boldsymbol{s}_2; \Theta) \geq 0$$

- *Inversa de la matriz de covarianza*

$$\Sigma^{-1}(\Theta) = \frac{1}{C^2(\mathbf{0}, \Theta) - C^2(\boldsymbol{s}_1 - \boldsymbol{s}_2; \Theta)} \begin{pmatrix} C(\mathbf{0}, \Theta) & -C(\boldsymbol{s}_2 - \boldsymbol{s}_1; \Theta) \\ -C(\boldsymbol{s}_1 - \boldsymbol{s}_2; \Theta) & C(\mathbf{0}, \Theta) \end{pmatrix}.$$

- *Asumiendo media constante el proceso centrado es*

$$\mathrm{y} - \boldsymbol{\mu} = \begin{pmatrix} y(\boldsymbol{s}_1) - \mu \\ y(\boldsymbol{s}_2) - \mu \end{pmatrix}.$$

La función de log-verosimilitud es

$$l(\Theta, \mu; \mathrm{y}) = \frac{2}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma(\Theta)| + \frac{1}{2}(\mathrm{y} - \boldsymbol{\mu})'\Sigma^{-1}(\Theta)(\mathrm{y} - \boldsymbol{\mu}). \tag{3.1}$$

y el término cuadrático de l *en la ecuación* (3.1) *queda*

$$(\mathrm{y} - \boldsymbol{\mu})'\,\Sigma^{-1}(\Theta)\,(\mathrm{y} - \boldsymbol{\mu}) =$$

$$\frac{C(0, \Theta)\left((y(\boldsymbol{s}_1) - \mu)^2 + (y(\boldsymbol{s}_2) - \mu)^2\right) - 2C(\boldsymbol{s}_1 - \boldsymbol{s}_2; \Theta)\,(y(\boldsymbol{s}_1) - \mu)\,(y(\boldsymbol{s}_2) - \mu)}{|\Sigma(\Theta)|}.$$

Reemplazando todo en la función negativo de la log-verosimilitud $l(\Theta, \mu; \mathrm{y})$*, se obtiene*

$$\ln(2\pi) + \frac{\ln\left(\sigma^4 - \left(\sigma^2 \exp(-5/a)\right)^2\right)}{2} + \frac{\sigma^2\left((3-\mu)^2\,(4-\mu)^2\right) - 2\exp(-5/a)\,((3-\mu)\,(4-\mu))}{2}. \tag{3.2}$$

Es importante tener claridad en los elementos y cálculos de todos los métodos tanto con fines de interpretación como de programación. El vector a estimar es $\Theta = (\sigma^2, a)$. La función a optimizar es 3.2. Usualmente, se elige tanto el modelo como los valores iniciales de los parámetros con base en un ajuste a sentimiento del semivariograma empírico.

3.4. Máxima verosimilitud restringida espacial.

Una variación del estimador ML que reduce el sesgo de las estimaciones de los parámetros de la covarianza es el estimador REML, el cual sustituye la maximización de la verosimilitud del vector $\mathbb{Y}$ por la del vector $\mathbf{A}'\mathbb{Y}$ que satisface $E[\mathbf{A}'\mathbb{Y}] = 0$. Se usa el modelo de función aleatoria intrínseca de orden $k = 1$, así que se asume la media constante, [52]. Por eso aquí se presenta en términos de $\mathbb{Y}$. La matriz $\mathbf{A}$ es una matriz de dimensión $n \times (n - p)$, de rango columna completo. Con esta modificación, y dado que $\text{Var}[\mathbf{A}'\mathbb{Y}] = \mathbf{A}'\Sigma(\Theta)\mathbf{A}$, el negativo de la función de log-verosimilitud queda:

$$l(\Theta; \mathrm{y}) = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln|\mathbf{A}'\Sigma(\Theta)\mathbf{A}| + \frac{1}{2}\mathbb{Y}'\mathbf{A}(\mathbf{A}'\Sigma(\Theta)\mathbf{A})^{-1}\mathbf{A}'\mathbb{Y}. \quad (3.3)$$

Este método no usa el modelamiento de la superficie de tendencia, sino que se basa directamente en un vector de incrementos de media $\mathbf{0}$. Sin embargo, a pesar de reducir el sesgo en las estimaciones de Θ, aún requiere una cantidad muy grande de operaciones. $\mathbf{A} = (a_{ij})$ es una matriz $(n-1)\times n$ con los siguientes elementos

$$a_{ij} = \begin{cases} 1 & si \quad i = j \quad \text{y} \quad j = 1, \ldots, n-1 \\ -1 & si \quad i = j+1 \quad \text{y} \quad j = 1, \ldots, n-1 \\ 0 & \text{en otro caso} \end{cases}.$$

Ejemplo 5 (Ilustración método REML para el caso $n = 3$). . *$Y(\boldsymbol{s})$ es un proceso de media constante.*

Nótese que se encuentra de manera alternativa el proceso de incrementos, y que las características son las mismas que las del método de ML aplicado al proceso $\boldsymbol{A}'\mathbb{Y}$.

$$\mathbb{Y} = \begin{pmatrix} Y(\boldsymbol{s}_1) \\ Y(\boldsymbol{s}_2) \\ Y(\boldsymbol{s}_3) \end{pmatrix} \quad y \quad \boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

$$\boldsymbol{A}'\mathbb{Y} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} Y(\boldsymbol{s}_1) \\ Y(\boldsymbol{s}_2) \\ Y(\boldsymbol{s}_3) \end{pmatrix} = \begin{pmatrix} Y(\boldsymbol{s}_1) - Y(\boldsymbol{s}_2) \\ Y(\boldsymbol{s}_2) - Y(\boldsymbol{s}_3) \end{pmatrix}.$$

El método que se presenta a continuación es una combinación de los métodos basados en mínimos cuadrados y los métodos basados en la verosimilitud, lo que lo hace muy atractivo y eficiente.

3.5. Pseudoverosimilitud. Verosimilitud compuesta (CL)

El método *CL* que se usa en este caso [5] para estimar Θ consiste en sumar componentes individuales de log-verosimilitud, y hacer que correspondan a las marginales de las variables de interés. Por lo tanto, este método no requiere el conocimiento de la distribución multivariada de $\mathbb{Y}$, sino que se basa en las distribuciones marginales

$$f(Y(\boldsymbol{s}_i), \Theta)$$

Se asume que existen tanto el gradiente como la matriz Hessiana de f.

Si se suponen conocidas $f(Y(\boldsymbol{s}_i), \Theta)$, excepto por el parámetro Θ; entonces $l(Y(\boldsymbol{s}_i), \Theta) = \ln(f(Y(\boldsymbol{s}_i), \Theta))$ es una función de log-verosimilitud y la función de verosimilitud compuesta es

$$CL(\Theta) = \sum_{i=1}^{n} l(Y(\boldsymbol{s}_i), \Theta)$$

A su gradiente

$$\nabla CL(\Theta) = CS(\Theta)$$

se le llama la función score compuesta.

Así, para encontrar el estimador $\hat{\Theta}$, se resuelve el sistema de ecuaciones

$$CS(\Theta) = \sum_{i=1}^{n} \nabla l(Y(\boldsymbol{s}_i), \Theta) = 0.$$

Asumiendo que el proceso es de media constante, posiblemente después de quitar la tendencia con un modelo de regresión polinómica en las

coordenadas, una implementación de este método para la estimación del semivariograma espacial es la siguiente,

1. El objetivo es estimar los parámetros del semivariograma, y por lo tanto es muy natural la construcción de la variable incrementos, la cual se va a denotar U:

$$U_{ij} = Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j).$$

   La variable U se logra efectuando todas las combinaciones posibles, resultando en $\frac{n(n-1)}{2}$ realizaciones, la cual es una inmensa cantidad de datos. Por lo tanto, es lógico ingresar a la estimación solo aquellos datos que involucran información sobre la dependencia espacial. Se pueden omitir aquellos pares de observaciones que se encuentren muy alejadas, según un criterio definido, tal como el alcance espacial observado en el semivariograma experimental.

2. Se requieren las funciones de verosimilitud marginales de las variables de interés.

   Asumiendo normalidad para las distribuciones marginales de $\mathbb{Y}$, esto es,

$$Y(\boldsymbol{s}) \sim N(\mu; \sigma^2), \forall(\boldsymbol{s}) \in D.$$

   se tiene que

$$U_{ij} \sim N(0, 2\gamma(s_i - s_j, \Theta))$$

   Así que el negativo de la función de log-verosimilitud es

$$l(U_{ij}, \Theta) = \frac{1}{2} \ln\left(2\pi\left(2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j, \Theta)\right)\right) + \frac{U_{ij}^2}{2\left(2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j, \Theta)\right)}.$$

3. Determinar la función de verosimilitud compuesta, sumando todas las funciones de log-verosimilitud marginales de la variable U;

$$\sum_{i=1}^{n-1} \sum_{j>i}^{n} l(U_{ij}, \Theta).$$

4. Para un modelo válido de semivarianza $2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j, \Theta)$, se determina la función score compuesta

$$\sum_{i=1}^{n-1}\sum_{j>i}^{n}\frac{\partial l(U_{ij}, \Theta)}{\partial(\Theta)}.$$

la cual bajo el supuesto de normalidad queda

$$\sum_{i=1}^{n-1}\sum_{j>i}^{n}\frac{\partial \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j, \Theta)}{\partial(\Theta)}\frac{1}{4\gamma^2(\boldsymbol{s}_i - s_j, \Theta)}(U_{ij}^2 - 2\gamma(\boldsymbol{s}_i - s_j, \Theta)).$$

La cual realmente es una estimación por mínimos cuadrados ponderados del modelo

$$U_{ij}^2 = 2\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j, \Theta) + e_{ij}.$$

Nótese que, aunque en un principio, al usar las distribuciones marginales no se involucra la estructura de dependencia espacial existente en los datos, esta es involucrada más adelante a través de la varianza de la variable incrementos. Una de las razones para usar las verosimilitudes marginales es que, aunque inicialmente no se cumpla el supuesto requerido, es posible aproximarse a este a través de alguna transformación.

3.6. Aspectos prácticos en la estimación del semivariograma

- **ML solo si existe la covarianza**. Es importante notar que si la varianza del proceso no es finita, no existe la covarianza y no es posible tener el supuesto de estacionariedad de segundo orden. En consecuencia, no es posible usar el supuesto de normalidad. En estos casos, donde solo se tiene estacionariedad intrínseca, los parámetros se estiman usando MCO, MCP o Pseudoverosimilitud.

- **Errores en el diseño de muestreo.** Dado que tanto el modelo como sus parámetros iniciales se eligen a partir del semivariograma empírico, hay que tener en cuenta todas las posibilidades en el proceso de ensayo y error. Puede ocurrir que en la ventana donde hay observaciones no se estabilice el semivariograma, pero si los modelos lineal o potencial no se ajustan bien, es posible que el proceso alcance la silla más lejos de lo que se ve con los datos en el semivariograma

empírico. Así, se puede probar con un modelo acotado, y si los resultados son buenos, seguramente faltan mediciones más alejadas para identificar correctamente el rango a partir del semivariograma empírico. Si, por el contrario, faltó llevar a cabo observaciones a distancias más cortas, será evidente con la existencia del efecto pepita.

- **Tamaños de las matrices**. Mientras que el método de ML requiere invertir iterativamente una matriz de dimensión $n \times n$, los MCO, MCP o CL solo requieren la inversión iterativa de una matriz cuadrada, cuya dimensión la da el número de parámetros del modelo de covarianza o semivarianza.

- **Matrices de covarianza estructuradas.** Las matrices de covarianza que se usan en este texto se conocen como estructuradas, porque son construidas usando una función definida positiva, usualmente con pocos parámetros. Hay quienes no gustan del uso del semivariograma empírico y en este caso estiman todos los elementos de la matriz sin estructura. Esto puede llevar a problemas de convergencia en un proceso de estimación iterativo, debido a que su tamaño es demasiado grande con respecto al número de observaciones. La cantidad de parámetros a estimar puede resultar excesiva: $\frac{n(n-1)}{2}$. Adicionalmente, no se tendría cómo calcular la covarianza entre lugares diferentes a los de la matriz de observaciones y también se perdería la interpretación de la variabilidad espacial que presenta el semivariograma.

3.7. El modelo lineal de regionalización

En la mayoría de las situaciones, dos o más modelos pueden ser combinados para ajustar la forma del semivariograma empírico o de la función de covarianza.

> Combinaciones lineales con coeficientes positivos, de modelos válidos de semivarianza o covarianza son a su vez modelos válidos.

El modelo lineal de regionalización construye el campo aleatorio $Y(s)$ como una combinación lineal de L campos aleatorios centrados, autocorrelacionados y mutuamente independientes, cada uno con función de autocorrelación $\rho_l(h;\boldsymbol{\theta}_l)$. A continuación se presenta el detalle de esta

descomposición. El modelo lineal de regionalización asume que

$$Y(\boldsymbol{s}) = \sum_{l=0}^{L} b_l Z_l(\boldsymbol{s}) + \mu \qquad \forall l,\ l = 0, \ldots, L \tag{3.4}$$

1. $E[Y(\boldsymbol{s})] = \mu$.

2. $E[Z_l(\boldsymbol{s})] = 0$.

3. $\text{Cor}[Z_l(\boldsymbol{s}), Z_{l'}(\boldsymbol{s}+\boldsymbol{h})]$ está dada por

$$\text{Cor}[Z_l(\boldsymbol{s}), Z_{l'}(\boldsymbol{s}+\boldsymbol{h})] = \begin{cases} \rho_l(\boldsymbol{h};\boldsymbol{\theta}_l) & \text{si } l = l', \\ 0 & \text{en otro caso.} \end{cases} \tag{3.5}$$

La descomposición en 3.4 implica que la función de autocovarianza de $Z(\boldsymbol{s})$ es una combinación lineal de L funciones de autocovarianza $C_l(h;\boldsymbol{\theta}_l)$, y aplicando la independencia mutua entre los Y y (3.5) se obtiene

$$\text{Cov}\,[Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})] = \sum_{l=0}^{L}\sum_{l'=0}^{L} \text{Cov}\left[b_l Z_l(\boldsymbol{s}), b_{l'} Z_{l'}(\boldsymbol{s}+\boldsymbol{h})\right] = \sum_{l=0}^{L} b_l{}^2 \rho_l(\boldsymbol{h};\boldsymbol{\theta}_l). \tag{3.6}$$

Las funciones Z_l presentan autocorrelación espacial pero no correlación cruzada. Finalmente, el modelo de covarianza completo $C(\boldsymbol{h})$ es

$$C(\boldsymbol{h};\Theta) = \sum_{l=0}^{L} b_l^2 \rho_l(\boldsymbol{h};\boldsymbol{\theta}_l)$$

$\Theta = (\boldsymbol{\theta}_1, .., \boldsymbol{\theta}_L)$. $b_l{}^2$ es la silla del modelo básico de covarianza $C_l(\boldsymbol{h};\boldsymbol{\theta}_l)$.

En términos de semivariogramas el modelo lineal de regionalización se puede escribir de manera similar

$$E\left[(Z_l(\boldsymbol{s}) - Z_l(\boldsymbol{s}+\boldsymbol{h}))\,(Z_{l'}(\boldsymbol{s}) - Z_{l'}(s+h))\right] = \begin{cases} 2\gamma_l(h;\boldsymbol{\theta}_l) & \text{si } l = l'; \\ 0 & \text{En otro caso.} \end{cases}$$

Entonces

$$\gamma(\boldsymbol{h};\Theta) = \frac{1}{2} E\,[Y(\boldsymbol{s}+\boldsymbol{h}) - Y(\boldsymbol{s})]^2 = \sum_{l=0}^{L} b_l^2 \gamma_l(\boldsymbol{h};\boldsymbol{\theta}_l).$$

por construcción, es un modelo válido, donde b_l^2 es la contribución a la varianza total del correspondiente proceso Z_l (ver [37]), y

$$C(\mathbf{0};\Theta) = \sum_{l=0}^{L} b_l^2$$

Los $\gamma_l(\boldsymbol{h};\boldsymbol{\theta}_l)$ son definidos con silla 1, así la silla de cada proceso Z_l se ajusta en el coeficiente de la combinación lineal, b_l^2. Es muy importante usar el principio de parsimonia y usar la menor cantidad posible de modelos.

Ejemplo 6 (Modelo de regionalización). *El proceso de interés $Y(\boldsymbol{s})$ se puede descomponer en términos de los siguientes 3 procesos autocorrelacionados espacialmente pero independientes entre sí. Esto es, $L = 3$ y*

$$Y(\boldsymbol{s}) = 0.5Z_0(\boldsymbol{s}) - 4Z_1(\boldsymbol{s}) + 2Z_2(\boldsymbol{s})$$

Los tres modelos de semivarianza correspondientes son:

- *Para $Z_0(s)$ es el efecto pepita, $\gamma_0(h;\boldsymbol{\theta}_0 : \sigma^2 = 1)$*
- *Para $Z_1(s)$ es el modelo esférico de rango 10: $\gamma_1(h;\boldsymbol{\theta}_1 = (1, 10))$.*
- *Para $Z_2(s)$ es el modelo exponencial de rango 8: $\gamma_2(h;\boldsymbol{\theta}_2 = (1, 8))$.*

Nótese que en los tres casos la silla se deja en 1, porque la silla final depende de los coeficientes de la descomposición. La Expresión (3.7) y la Figura 3.1 muestran el modelo de regionalización obtenido.

$$\gamma(\boldsymbol{h};\Theta) = \begin{cases} 0.25 + 16\left(\frac{3}{2}\frac{h}{10} - \frac{1}{2}\left(\frac{h}{10}\right)^3\right) + 4\left(1 - \exp\left(-\frac{h}{8}\right)\right) & Si\ h \le 10; \\ 0.25 + 16 + 4\left(1 - \exp\left(-\frac{h}{8}\right)\right) & Si\ \ h > 10 \end{cases} \tag{3.7}$$

El modelamiento del semivariograma se lleva a cabo con el semivariograma omnidireccional. Esto es, para la estimación de los parámetros del modelo con los métodos presentados en este capítulo se usan los pares $(h, \hat{\gamma}(h))$ y solo se tiene en cuenta la magnitud de h. A continuación, en la Sección 3.8 se describe el procedimiento necesario en caso de que no exista isotropía, es decir, en caso de que el modelo cambie su rango o pendiente con la dirección del semivariograma.

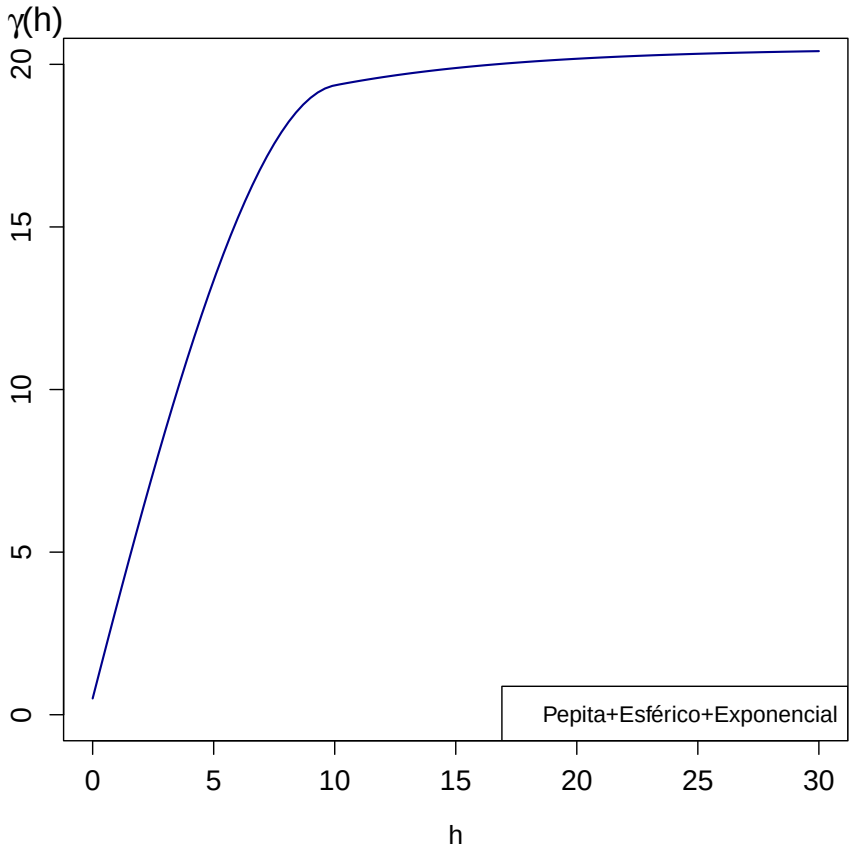


Figura 3.1. Modelo lineal de regionalización obtenido mediante la combinación de 3 campos aleatorios. Ver expresión (3.7)

3.8. Anisotropía

Es posible que el comportamiento del semivariograma presente diferencias si se tiene en cuenta además de la magnitud de h, su dirección. De la Definición 7 se tiene que el proceso $Z(.)$ es anisotrópico, si la dependencia espacial entre $Z(\boldsymbol{s})$ y $Z(\boldsymbol{s}+\boldsymbol{h})$ es una función tanto de la magnitud como de la dirección del vector $\boldsymbol{h}$. En consecuencia, para evaluar si se puede asumir isotropía o no, es necesario estimar ya sea la autocorrelación, la autocovarianza o la semivarianza en diferentes direcciones y comparar sus comportamientos para determinar si el rango o la silla varían con la dirección. Por ejemplo, con base en estos resultados, se puede construir el diagrama de rosa que ayuda a la detección de la llamada anisotropía geométrica, a través de la identificación de los ángulos en los cuales se encuentra el rango mayor y rango menor o bien mayor pendiente y menor pendiente, asociados a la correspondiente estimación del semivariograma (ver Figura 3.2). Una anisotropía es llamada geométrica si los semivariogramas tienen la misma forma y silla pero diferentes valores de rango, y en consecuencia el diagrama de rosa de los rangos, es una elipse. Ver Figura 3.3.

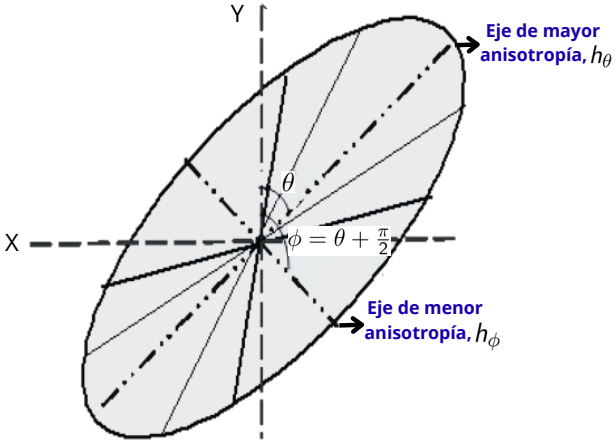


Figura 3.2. Diagrama de rosa para detección de anisotropía geométrica. Se construye a partir de la estimación a sentimiento del rango de cada uno de los semivariogramas empíricos a diferentes direcciones. El eje mayor (menor) de la elipse corresponde a la dirección en la cual el rango o pendiente son los mayores (menores) encontrados.

Entonces,

1. θ : este ángulo es medido en grados desde el eje vertical Y en el sentido de las manecillas del reloj y es el ángulo de mayor anisotropía, esto es mayor rango o pendiente.

2. $\phi = \theta + \frac{\pi}{2}$ es el ángulo de menor anisotropía, esto es menor rango o pendiente.

3. $\lambda = \frac{a_\phi}{a_\theta} < 1$ es la razón de anisotropía

En caso de encontrar anisotropía geométrica el proceso a seguir es el siguiente: En el caso de $\mathbb{R}^2$ se deben transformar las coordenadas originales $\boldsymbol{h} = (h_x, h_y)$ en las nuevas coordenadas $\boldsymbol{h}' = (h'_\phi, h'_\theta)'$, en las cuales el semivariograma es isotrópico y el diagrama de rosa es un círculo.

De esta forma, un modelo de semivariograma anisotrópico, se identifica con un modelo isotrópico en el nuevo sistema de coordenadas, esto es,

$$\gamma(\boldsymbol{h}) \rightarrow \gamma'(|\boldsymbol{h}'|)$$

con

$$|\boldsymbol{h}'| = \sqrt{\left(h'_\phi\right)^2 + \left(h'_\theta\right)^2}$$

donde $\gamma'(|\mathbf{h}'|)$ es un modelo isotrópico con un rango igual al de menor anisotropía a_ϕ, [37].

La isotropía es un caso particular de la anisotropía geométrica cuando la razón de anisotropía es 1, $\lambda = 1$.

Como se puede ver los dos elementos clave de la transformación de coordenadas son el ángulo de máxima anisotropía y la razón de anisotropía. Los pasos a seguir si se encuentra este tipo de anisotropía son los siguientes:

1. Rotación de los ejes coordenados

$$\boldsymbol{h}_\theta = \begin{pmatrix} h_\phi \\ h_\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \operatorname{sen}(\theta) \\ -\operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} h_x \\ h_y \end{pmatrix}$$

2. La elipse es re-escalada a un círculo de radio igual al menor rango de anisotropía

$$\boldsymbol{h}' = \begin{pmatrix} h'_\phi \\ h'_\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \cdot \begin{pmatrix} h_\phi \\ h_\theta \end{pmatrix}$$

El valor del modelo anisotrópico en diferentes direcciones es el valor del modelo isotrópico de rango a_ϕ

Ejemplo 7 (Corrección anisotropía geométrica). *Sea $\gamma(\boldsymbol{h})$ un modelo de semivariograma esférico que presenta anisotropía, entonces el valor del modelo anisotrópico en varias direcciones es igual al valor del modelo isotrópico de rango a_ϕ. El modelo esférico ominidireccional está dado por*

$$\gamma(\boldsymbol{h};\theta) = \begin{cases} \sigma^2 \left[\frac{3}{2}\frac{\|\boldsymbol{h}\|}{a} - \frac{1}{2}\left(\frac{\|\boldsymbol{h}\|}{a}\right)^3 \right] & si\ 0 < \|\boldsymbol{h}\| \leq a \\ \sigma^2 & si\ \|\boldsymbol{h}\| > a \end{cases}$$

y el modelo anisotrópico está dado por

$$\gamma(\boldsymbol{h}') = \begin{cases} \sigma^2 \left[\frac{3}{2}\frac{\|\boldsymbol{h}'\|}{a_\phi} - \frac{1}{2}\left(\frac{\|\boldsymbol{h}'\|}{a_\phi}\right)^3 \right] & si\ 0 < \|\boldsymbol{h}'\| \leq a_\phi \\ \sigma^2 & si\ \|\boldsymbol{h}'\| > a_\phi \end{cases}$$

Además por construcción note que

$$\|\boldsymbol{h}'\| = \sqrt{\left(h'_\phi\right)^2 + \left(h'_\theta\right)^2} = \sqrt{\left(h_\phi\right)^2 + \left(\lambda h_\theta\right)^2}$$

Otra forma de anisotropía es la zonal. Una anisotropía es llamada zonal si los valores de las sillas varían con la dirección. En general, en la dirección específica de mayor rango también hay mayor silla. El algoritmo para incorporar esta característica en el modelo de semivariograma es similar al utilizado para la anisotropía geométrica. Se rotan los ejes coordenados en la dirección perpendicular a la de mayor silla. El primer paso es usar la matriz de rotación:

$$\begin{pmatrix} h_\phi \\ h_\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \operatorname{sen}(\theta) \\ -\operatorname{sen}(\theta) & \cos(\theta) \end{pmatrix} \cdot \begin{pmatrix} h_x \\ h_y \end{pmatrix}$$

Los nuevos ejes son re-escalados al modelo zonal de tal forma que no contribuyan a la direccion de máxima continuidad o máximo rango o pendiente, θ. Así el semivariograma del ángulo de menor varianza solo contribuye en cada ángulo específico.

$$\begin{pmatrix} h'_\phi \\ h'_\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} h_\phi \\ h_\theta \end{pmatrix}$$

Finalmente, es importante notar que entre mas información se tenga

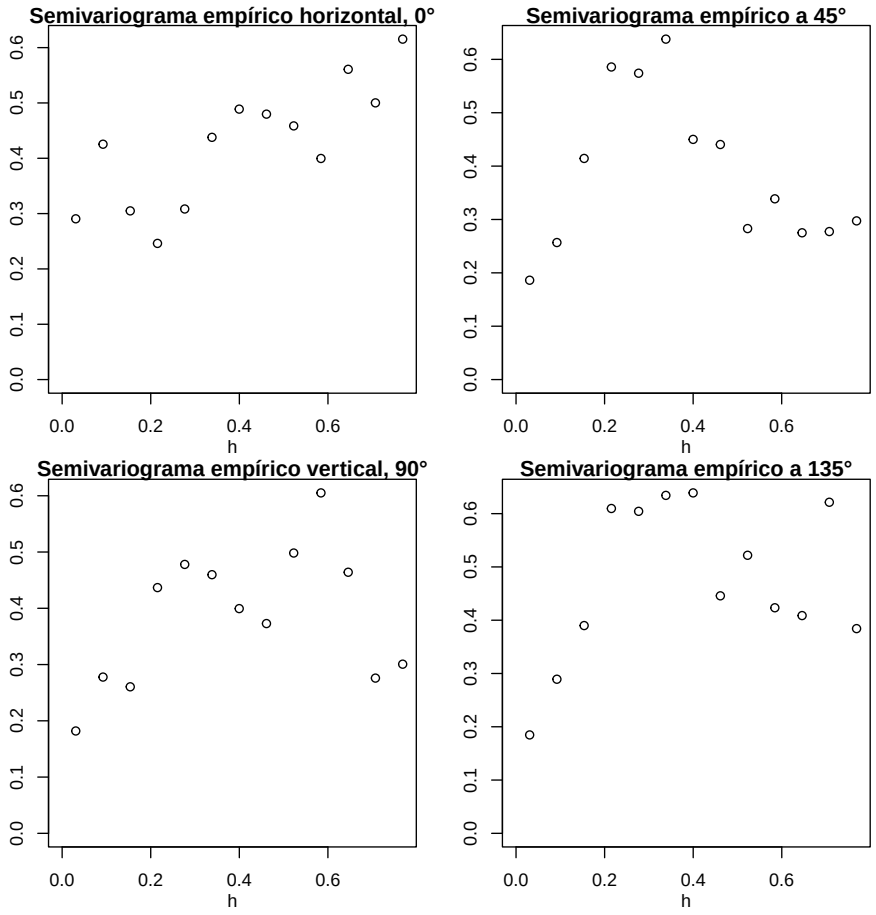


Figura 3.3. Semivariogramas direccionales. Cada semivariograma es construido con una tolerancia angular de $\pi/8$

especialmente en las direcciones de mayor rango o silla, serán mucho más sencillas la detección de la anisotropía, su clase y así su tratamiento. El experto en el área de estudio puede poseer mas información acerca de la dirección y características del ángulo de mayor anisotropía y esto puede usarse como insumo para que el diseño de muestreo sea mas denso en esta dirección.

Nota 5 (Estimación del semivariograma con mínimos cuadrados generalizados)**.** *Si se tiene en cuenta que los incrementos pueden estar correlacionados, se podría considerar el uso de mínimos cuadrados en su forma*

más amplia, esto es, usar mínimos cuadrados generalizados. Así, el objetivo sería minimizar:

$$(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\Theta))' V^{-1} (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}(\Theta)) .$$

Dónde

$$\hat{\boldsymbol{\gamma}} = (\hat{\gamma}(\boldsymbol{h}_1), \ldots, \hat{\gamma}(\boldsymbol{h}_K)).$$

y

$$\boldsymbol{\gamma}(\Theta) = (\gamma(\boldsymbol{h}_1; \Theta), \ldots, \gamma(\boldsymbol{h}_K; \Theta)).$$

Como es usual, la dificultad de este método es el requerimiento de conocer o tener alguna estimación de la matriz de covarianza respectiva, en este caso

$$V = \left(\mathrm{Cov}\left[\hat{\gamma}(\boldsymbol{h}_k), \hat{\gamma}(\boldsymbol{h}_l) \right] \right)_{k,l=1,\ldots,K}$$

Así, se complicaría bastante el procedimiento. La estimación de la matriz de interés $\Sigma = \mathrm{Cov}\left[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j) \right]$ *implicaría la estimación previa de* V. *[24] explora esta opción de igual manera que en la sección 2.3. La correlación entre pares de cuadrados de incrementos se puede calcular usando la siguiente propiedad:*

$$\rho^2(X_1, X_2) = \rho(X_1^2, X_2^2).$$

la cual se cumple bajo el supuesto de normalidad bivariada. Así, utilizando esta equivalencia a partir de la variable incrementos

$$\begin{aligned} &\mathrm{Cor}\left[\left(Z(\boldsymbol{s}_i + \boldsymbol{h}_k) - Z(\boldsymbol{s}_i) \right)^2; \left(Z(\boldsymbol{s}_j + h_l) - Z(\boldsymbol{s}_j) \right)^2 \right] \\ &\qquad = \mathrm{Cor}^2 \left[Z(\boldsymbol{s}_i + \boldsymbol{h}_k) - Z(\boldsymbol{s}_i); Z(\boldsymbol{s}_j + \boldsymbol{h}_l) - Z(\boldsymbol{s}_j) \right] \end{aligned} \quad (3.8)$$

y expresando la función de autocorrelación en términos de la función de semivarianza, se obtiene:

$$\frac{\gamma(\boldsymbol{s}_i - \boldsymbol{s}_j + \boldsymbol{h}_k) + \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j - \boldsymbol{h}_l) - \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j + \boldsymbol{h}_k - h_l) - \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j)}{2\gamma(\boldsymbol{h}_k) 2\gamma(\boldsymbol{h}_l)}.$$

Esto requiere la estimación adicional de esta correlación en el proceso de estimar lo que es realmente de interés: Σ. *Sin embargo, la correlación entre incrementos, (ver expresión 3.8) no suele ser significativa. De modo que usar mínimos cuadrados generalizados para encontrar la estimación de* Σ *no es algo práctico y por lo tanto no se usa.*

3.9. Ejercicios

1. Realice una tabla comparativa entre el método de ecuaciones de estimación generalizada y el método de CL. Encuentre similitudes y diferencias.

2. Suponga que se tiene un modelo de covarianza y se requiere elegir entre dos vectores de parámetros, cuál de ellos es más verosímil con base en el vector de observaciones. Bajo el supuesto de normalidad, describa en general el procedimiento e ilustre para al menos un caso particular.

3. Plantee para cada método y para cada modelo de covarianza o semivarianza la expresión a minimizar:

   3.1 Mínimos cuadrados ordinarios (MCO),

   3.2 Mínimos cuadrados ponderados (MCP) considerando ambas ponderaciones:

      3.2.1 La varianza del estimador empírico del semivariograma

      3.2.2 $N(\boldsymbol{h})$, número de pares de datos para cada rezago espacial considerado $\boldsymbol{h}$

   3.3 ML, con el supuesto de normalidad multivariada conjunta.

   3.4 REML, con el supuesto de normalidad multivariada conjunta.

   3.5 Pseudoverosimilitud (CL), con el supuesto de normalidad multivariada marginal.

   3.6 Diseñe el código para que estos procesos de optimización, funcionen desde la tabla de los datos originales. Puede usar, por ejemplo, la función optim de R cran project. Compare los resultados obtenidos con su código con los obtenidos por las funciones de geoR, variofit, likfit o de gstat, fit.variogram.

   3.7 Determine las dimensiones y la complejidad de cada uno de los métodos. ¿Cuál es la dimensión de las matrices invertidas en cada método? ¿En todos los casos el proceso es iterativo?

4. Para el problema 4 del capítulo 1, encuentre las matrices de covarianza, correlación y semivarianza, si existen. Si alguna no existe, justifique. Verifique las propiedades de las matrices obtenidas para que garanticen varianzas de predicción positivas.

4.1 Si el modelo es un exponencial con rango 50 y silla 1600, sin pepita.

4.2 Si el modelo es un seno cardinal con rango 50 y silla 1600, sin pepita.

4.3 Si el modelo es un cúbico con pendiente 50 y potencia 1600, sin pepita.

4.4 Si el modelo es un potencial con pendiente 50 y potencia 1.6, sin pepita.

4.5 Calcule el valor de la función de verosimilitud según cada modelo y los datos dados. De acuerdo con estos resultados, cuál de estos modelos es mas adecuado para los datos?

4.6 En un modelo de semivariograma anisotrópico, cuál es la magnitud del rezago espacial en la dirección de máxima anisotropía? Cómo queda el modelo de semivariograma en este caso?

4.7 En un modelo de semivariograma anisotrópico, cuál es la magnitud del rezago espacial en la dirección de mínima anisotropía? Cómo queda el modelo de semivariograma en este caso?

4.8 Calcule el valor de la función de pseudoverosimilitud según cada modelo y los datos dados. De acuerdo con estos resultados, cuál de estos modelos es mas adecuado para los datos?

4.9 Escriba cada modelo con la corrección necesaria en presencia de anisotropía.

5. El levantamiento topográfico para la construcción de los Modelos de Elevación Digital (MED) o mapas de curvas de nivel, se realiza mediante el diseño de trayectorias distanciadas 60m, regularmente distribuidas por toda la zona de interés. Para redes hidrológicas, se eligen trayectorias perpendiculares a la dirección de drenaje. Se requiere llevar a cabo un análisis de anisotropía. Se detectó la existencia de una anisotropía de tipo geométrico cuya dirección principal es 25° con un rango máximo de 105m y un rango mínimo de 35m.

5.1 Escriba la matriz de rotación que usted usaría para incluir la anisotropía en el modelo de semivariograma.

5.2 Describa cómo aplicaría la razón de anisotropía.

5.3 Ilustre a través de un caso particular.

Capítulo

cuatro

Predicción espacial escalar univariada. Kriging.

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

Uno de los objetivos principales del análisis geoestadístico de datos espaciales en dominio continuo es la *predicción espacial* en lugares no observados. Se muestrea el campo aleatorio $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ en las ubicaciones $\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_n}$ y con base en los datos se predice en sitios en los que no se tiene dato. El interés se centra en encontrar el mejor predictor de Y en el lugar $\boldsymbol{s}_0$, con base en los valores observados $y(\boldsymbol{s_1}), y(\boldsymbol{s_2}), \ldots, y(\boldsymbol{s_n})$.

Existen muchos métodos determinísticos para obtener valores en lugares no muestreados. Sin embargo, la calidad de las predicciones halladas usando estos métodos determinísticos, solo se pueden evaluar en los lugares con datos, haciendo validación cruzada. Se deja una observación por fuera y se aplica el método de predicción usando las $n - 1$ observaciones restantes para comparar. No existe una medida de incertidumbre asociada, que permita evaluar los resultados en los lugares en los que no se cuenta con observaciones.

Usar los métodos estadísticos de predicción espacial presenta una gran ventaja, ya que junto con cada predicción se obtiene la estimación de su varianza y de la varianza del error de predicción. De hecho, la predicción se encuentra de tal manera que minimice la varianza del error de predicción. Adicionalmente, las medidas usuales para la calidad de la predicción (estadísticos de los residuales o el coeficiente de correlación lineal entre valores observados y predichos) se pueden usar con fines descriptivos y de resumen.

El predictor kriging se construye insesgado y de mínima varianza. Puede ser requerida la predicción solo en algunos sitios específicos o puede que se requiera completar todo el mapa. En este caso, los mapas de predicción generados con kriging se acompañan de los respectivos mapas de varianzas del error de predicción, para poder determinar cuáles zonas tienen predicciones más precisas.

Ejemplo 8 (Calidad del aire). *Se cuenta con mediciones del material particulado por debajo de 10 micras,* $Y(\boldsymbol{s}) = \mathrm{PM10}(\boldsymbol{s})$*, en 10 lugares en la ciudad de Bogotá a una hora particular. Estas variables espaciales conforman el siguiente vector espacial aleatorio*

$$(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_{10})).$$

El PM10 *ha sido observado en los puntos cuyas estaciones son llamadas* $\boldsymbol{s}_1$*=Tunal,* $\boldsymbol{s}_2$*=Sony,...,*$\boldsymbol{s}_{10}$*=EscIng, respectivamente (ver Figura 4.1). Con base en estas observaciones, se requiere predecir el valor del* PM10 *en un lugar en el que no se tiene observación:* $\mathrm{PM10}(\boldsymbol{s}_0) = Y(\boldsymbol{s}_0)$*. El lugar donde se requiere la predicción se ubica más cerca de la estación Fontbn que de cualquier otra ubicación donde se tenga medición; por lo tanto, es lógico pensar que* $Y(\boldsymbol{s}_0)$ *es más parecido al valor*

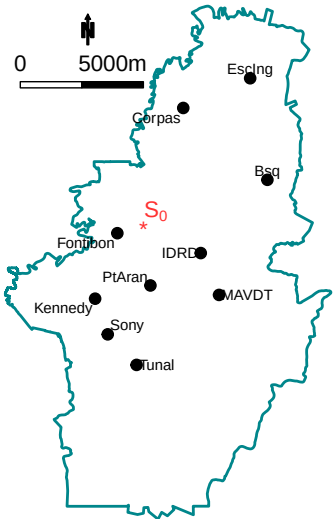


Figura 4.1. Red de calidad del aire de Bogotá. El punto en rojo $\boldsymbol{S}_0$ es un lugar donde no se tiene observación y se requiere predicción.

en esta estación que a cualquiera de las demás. De acuerdo con lo anterior, se puede optar para la predicción, por una media ponderada de las 10 *mediciones, en la cual* $Y\,(Fontbn)$ *tiene mayor peso que cualquier otra, seguida en su orden por* $Y\,(IDRD)$ *y así sucesivamente.*

Denotando $Y^*\,(\boldsymbol{s}_0)$ *el predictor se define como*

$$Y^*\,(\boldsymbol{s}_0) = \lambda_1 Y\,(\boldsymbol{s}_1) + \lambda_2 Y\,(\boldsymbol{s}_2) + \ldots + \lambda_{10} Y\,(\boldsymbol{s}_{10})\,.$$

Donde los λ_i, $i = 1, \ldots, 10$ *son las ponderaciones.*

El predictor kriging es el método que permite encontrar los valores de estas ponderaciones, de tal manera que el predictor sea insesgado y de mínima varianza. Este predictor funciona punto a punto, es decir, se predice un punto a la vez.

4.1. Características del predictor kriging

El predictor kriging de $Y\,(\boldsymbol{s_0})$ se construye con base en el vector observado

$$\mathrm{y} = (y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), \ldots, y(\boldsymbol{s}_n))'$$

el cual es la realización del vector

$$\mathbb{Y} = \Big(Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n)\Big)'$$

y

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$$

es el vector de coeficientes de ponderación. Se construye con las siguientes propiedades:

1. **Linealidad:** El predictor es una combinación lineal de los valores observados.

$$Y^*(\boldsymbol{s}_0) = \sum_{i=1}^{n} \lambda_i y(\boldsymbol{s}_i). \tag{4.1}$$

   o matricialmente

$$Y^*(\boldsymbol{s}_0) = \boldsymbol{\lambda}^T \mathbb{y}. \tag{4.2}$$

2. **Insesgamiento:** El valor esperado del predictor es igual al valor esperado de la variable en el punto de predicción.

$$E\left[Y^*(\boldsymbol{s}_0)\right] = E\left[Y(\boldsymbol{s}_0)\right]. \tag{4.3}$$

3. **Optimalidad. Mínima varianza:** El predictor minimiza la varianza del error de predicción $\mathrm{Var}\left[Y^*(\boldsymbol{s}_0) - Y(\boldsymbol{s}_0)\right]$. Esta es la mínima varianza posible que se puede obtener con un predictor lineal.

De acuerdo con la información que se tenga sobre la media del proceso espacial, se hace una clasificación del kriging. Se le conoce como kriging simple, al caso en el que la media es conocida; como kriging ordinario, al caso en el que la media es constante pero desconocida y por lo tanto es necesario estimarla, y como kriging universal, al caso en el que hay tendencia y es necesario estimar un modelo de regresión en las coordenadas observadas. Es usual en cualquier de estos casos, estimar la media del campo aleatorio, centrar la variable y llevar a cabo el resto del método con los residuales obtenidos, que son de media cero, y al final sumar la media para regresar a la escala de la variable original. A este procedimiento se le conoce como kriging residual.

4.2. Kriging simple

Si la media, $\mu(s) \in \mathbb{R}\ \forall\ \boldsymbol{s} \in D_s$, es conocida, y el proceso espacial se puede expresar de la siguiente manera

$$Y(s) = \mu(s) + Z(\boldsymbol{s}) \qquad s \in D, \quad \mu(s) \in \mathbb{R}. \tag{4.4}$$

se llama kriging simple al procedimiento de centrar la variable, dado que se conoce la media y aplicar el método al proceso residual resultante $Z(\boldsymbol{s})$. Nótese que

$$E[Y(\boldsymbol{s})] = \mu(\boldsymbol{s}), \quad E[Z(\boldsymbol{s})] = 0$$

y

$$\text{Var}[Y(\boldsymbol{s})] = \text{Var}[Z(\boldsymbol{s})].$$

Dado que la media es conocida, para volver a la variable original, solo se evalúa la función $\mu(\boldsymbol{s})$ en $\boldsymbol{s}_0$ y se le suma a la predicción de $Z(\boldsymbol{s}_0)$. Así, la predicción de la variable original en un punto particular $\boldsymbol{s}_0$ se puede reconstruir como

$$Y^*(\boldsymbol{s}_0) = \mu(\boldsymbol{s}_0) + Z^*(\boldsymbol{s}_0). \tag{4.5}$$

Bajo este escenario $E[Z(s_0)] = 0$ y

$$E[Z^*(s_0)] = E\left[\sum_{i=1}^{n} \lambda_i Z(\boldsymbol{s}_i)\right] = \sum_{i=1}^{n} \lambda_i E[Z(\boldsymbol{s}_i)] = 0$$

Entonces siempre es cierto que

$$E[Z(s_0)] = E[Z^*(s_0)]$$

lo que indica que no es necesario imponer ninguna restricción sobre el vector de ponderaciones; el kriging simple siempre es insesgado. Finalmente, se garantiza mínima varianza del error de predicción, esto es, se encuentran las ponderaciones que minimizan la siguiente expresión

$$\text{Var}[Z^*(s_0) - Z(s_0)].$$

En términos de sumatorias y covarianzas, se pueden ver en mayor detalle los pasos y elementos de este procedimiento. Entonces, la varianza del error de predicción usando la variable centrada es,

$$E[Z^*(s_0) - Z(s_0)]^2$$

Desarrollando el cuadrado del binomio y tomando esperanzas se obtiene

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j E\left[Z(s_i)Z(s_j)\right] - 2\sum_{i=1}^{n}\lambda_i E\left[Z(s_i)Z(s_0)\right] + E\left[Z^2(s_0)\right]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j \text{Cov}\left[Z(s_i), Z(s_j)\right] - 2\sum_{i=1}^{n}\lambda_i \text{Cov}\left[Z(s_i), Z(s_0)\right] + \text{Var}\left[Z(s_0)\right]$$

Aplicando las propiedades de estacionariedad de segundo orden las covarianzas están dadas por la función definida positiva $C(h;\Theta)$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j C\left(s_i - s_j;\Theta\right) - 2\sum_{i=1}^{n}\lambda_i C\left(s_i - s_0;\Theta\right) + C\left(0;\Theta\right)). \tag{4.6}$$

Dado que $\mathbb{Z}$ es el vector centrado, las covarianzas se pueden expresar como esperanzas de los productos cruzados respectivos. Observe que

$$E\left[\mathbb{Z}\mathbb{Z}^T\right] = \Sigma$$

$$E\left[Z\left(s_0\right)\mathbb{Z}\right] = \boldsymbol{\sigma}_0$$
$$E\left[Z^2(s_0)\right] = C\left(0;\Theta\right).$$

Σ es la matriz de covarianza del vector que contiene la variable en los sitios con observaciones, $\mathbb{Z}$, ($\mathbb{Y}$), y $\boldsymbol{\sigma}_0$ es el vector que contiene las covarianzas entre $\mathbb{Z}$ y la variable a predecir $Z\left(s_0\right)$. Σ y $\boldsymbol{\sigma}_0$ son calculados con el mismo modelo válido de covarianza encontrado con los métodos vistos en el capítulo 2.

Ahora se aplica el proceso usual de optimización: derivar e igualar a cero. El vector de ponderaciones es la solución de este proceso de optimización. Denótese con $\mathbb{Q}$ la expresión a minimizar

$$\mathbb{Q} = \text{Var}\left[\boldsymbol{\lambda}^T\mathbb{Z} - Z\left(s_0\right)\right]. \tag{4.7}$$

Al derivar 4.7 con respecto a $\boldsymbol{\lambda}$ e igualar a cero, se obtiene

$$\frac{\partial\mathbb{Q}}{\partial\boldsymbol{\lambda}} = 2\Sigma\lambda - 2\boldsymbol{\sigma}_0 \quad\Rightarrow\quad \Sigma\boldsymbol{\lambda} = \boldsymbol{\sigma}_0 \quad\Rightarrow\quad \boldsymbol{\lambda} = \Sigma^{-1}\boldsymbol{\sigma}_0. \tag{4.8}$$

que es el sistema de ecuaciones normales para el kriging simple. El sistema obtenido en (4.8) es un sistema de ecuaciones lineales con solución única, dado que Σ es una matriz definida positiva, por lo tanto, es invertible.

Sustituyendo el predictor encontrado en la expresión (4.7), la varianza del kriging simple está dada por

$$\text{Var}\left[\boldsymbol{\lambda}^T\mathbb{Z} - Z\left(s_0\right)\right] = \boldsymbol{\lambda}^T\Sigma\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^T\boldsymbol{\sigma}_0 + C\left(0;\Theta\right) = C\left(0;\Theta\right) - \boldsymbol{\sigma}_0\Sigma^{-1}\boldsymbol{\sigma}_0. \tag{4.9}$$

El kriging simple está construido aquí utilizando la función de autocovarianza espacial y el supuesto de estacionariedad de segundo orden, pero puede construirse usando la función de semivarianza y el supuesto de estacionariedad intrínseca. El predictor kriging ordinario en la siguiente sección es construido usando la semivarianza con propósitos ilustrativos. De hecho, cualquier tipo de kriging puede ser formulado usando la autocovarianza o la semivarianza, esto es, tanto con modelos acotados como no acotados. Bajo estacionariedad de segundo orden, solo es cuestión de utilizar la relación (1.5), pero es importante recordar que, cuando el modelo de semivarianza no es acotado pero aún es intrínseco, no existe la función de autocovarianza pero si la de semivarianza. Así que los predictores deben ser formulados en términos únicamente de la función $\gamma()$.

El predictor kriging simple de $Y\left(\boldsymbol{s}_0\right)$ con base en el vector observado

$$\mathbb{y} = \left(y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), \ldots, y(\boldsymbol{s}_n)\right)'$$

y su versión centrada

$$\mathbb{z} = \left(z(\boldsymbol{s}_1), z(\boldsymbol{s}_2), \ldots, z(\boldsymbol{s}_n)\right)'$$

es

$$Y^*\left(\boldsymbol{s}_0\right) = \lambda^T\mathbb{z} + \mu\left(\boldsymbol{s}_0\right)$$

donde

$$\boldsymbol{\lambda} = \Sigma^{-1}\boldsymbol{\sigma}_0.$$

La varianza de error del error de predicción del kriging simple es

$$\text{Var}\left[Y^*\left(\boldsymbol{s}_0\right) - Y\left(s_0\right)\right] = C\left(0;\Theta\right) - \boldsymbol{\sigma}_0\Sigma^{-1}\boldsymbol{\sigma}_0.$$

donde Σ es la matriz de covarianza entre las variables en los lugares observados y $\boldsymbol{\sigma}_0$ es el vector de covarianza entre las variables en los lugares observados y la variable en el lugar a predecir. Esto es, $\text{Cov}\left[\mathbb{Y}\right] = \Sigma$, $\text{Cov}\left[Y\left(s_0\right), \mathbb{Y}\right] = \boldsymbol{\sigma}_0$ y $\text{Var}\left[Y\left(s_0\right)\right] = C\left(0;\Theta\right)$.

4.3. Kriging ordinario

El kriging ordinario se usa cuando la media de la variable es desconocida, pero se asume constante. Esto es,

$$Y(\boldsymbol{s}) = \mu + Z(\boldsymbol{s}) \quad E[Y(\boldsymbol{s})] = \mu \quad \forall \boldsymbol{s} \in D, \quad \mu \in \mathbb{R}. \tag{4.10}$$

Es necesario estimarla y, por lo tanto, no se puede trabajar directamente con la variable centrada. Construyendo de nuevo el predictor insesgado, se debe cumplir,

$$E\left[\sum_{i=1}^{n} \lambda_i Y(s_i)\right] = E[Y(s_0)]$$

Tomando esperanzas se obtiene

$$\sum_{i=1}^{n} \lambda_i \mu = \mu.$$

De donde se concluye que, para que se cumpla la propiedad de insesgamiento, se requiere que

$$\sum_{i=1}^{n} \lambda_i = 1$$

Nótese que no hay ningún requerimiento acerca del valor de los λ_i. Estos valores pueden ser de cualquier signo y tomar cualquier valor real; solo tienen que sumar uno 1. A continuación se desarrolla la expresión de la varianza del error de predicción en términos de la función de semivarianza.

$$\begin{aligned} E\left[\sum_{i=1}^{n} \lambda_i Y(s_i) - Y(s_0)\right]^2 &= E\left[\sum_{i=1}^{n} \lambda_i Y(s_i) - \sum_{i=1}^{n} \lambda_i Y(s_0)\right]^2 \\ &= E\left[\sum_{i=1}^{n} \lambda_i \Big(Y(s_i) - Y(s_0)\Big)\right]^2 \\ &= E\left[\sum_{i}^{n}\sum_{j}^{n} \lambda_i \lambda_j \left(Y(s_i) - Y(s_0)\right)\left(Y(s_j) - Y(s_0)\right)\right] \\ &= \sum_{i}^{n}\sum_{j}^{n} \lambda_i \lambda_j E\left[\left(Y(s_i) - Y(s_0)\right)\left(Y(s_j) - Y(s_0)\right)\right]. \end{aligned} \tag{4.11}$$

Por otro lado, note que aunque la media no se conozca, como es constante se tiene que la media de los incrementos es 0. En consecuencia, sumando y restando $Y(s_0)$ en la definición del variograma, se puede expresar como

$$2\gamma(s_i - s_j; \Theta) = E\left[Y(s_i) - Y(s_j)\right]^2 = E\left[(Y(s_i) - Y(s_0)) - (Y(s_j) - Y(s_0))\right]^2$$

Desarrollando el cuadrado del binomio y tomando esperanzas, se encuentra la siguiente equivalencia

$$2\gamma(s_i - s_j; \Theta) = 2\gamma(s_i - s_0; \Theta) + 2\gamma(s_j - s_0; \Theta) - 2E\left[(Y(s_i) - Y(s_0))(Y(s_j) - Y(s_0))\right].$$

Remplazando en (4.11) se obtiene

$$\begin{aligned}\mathbb{Q} &= E\left[\sum_{i=1}^{n} \lambda_i Y(s_i) - Y(s_0)\right]^2 \\ &= \sum\sum \lambda_i \lambda_j \Big(-\gamma(s_i - s_j; \Theta) + \gamma(s_i - s_0; \Theta) + \gamma(s_j - s_0; \Theta) \Big) \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(s_i - s_j; \Theta) + 2\sum \lambda_i \gamma(s_i - s_0; \Theta).\end{aligned}$$

Así, utilizando multiplicadores de Lagrange [51], hay que minimizar Q sujeto a la restricción $\sum \lambda_i = 1$. Esto es,

$$\mathbb{Q} = \frac{1}{2}\mathrm{Q} - \delta\left(\sum \lambda_i - 1\right). \tag{4.12}$$

$$\frac{\partial \mathbb{Q}}{\partial \lambda_i} = \frac{1}{2}\left(-2\sum_{j=1}^{n} \lambda_j \gamma(s_i - s_j; \Theta) + 2\gamma(s_i - s_0; \Theta)\right) - \delta = 0 \qquad i = 1, \ldots, n.$$

Entonces,

$$\sum_{j=1}^{n} \lambda_j \gamma(s_i - s_j; \Theta) + \delta = \gamma(s_i - s_0; \Theta) \qquad i = 1, \ldots, n.$$

Sustituyendo en $\mathbb{Q}$, se obtiene la varianza del kriging ordinario que es la misma para $Z(s_0)$ que para $Y(s_0)$ y que involucra al multiplicador de Lagrange δ.

$$E\left[Y^*(s_0) - Y(s_0)\right]^2 = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0; \Theta) + \delta. \tag{4.13}$$

Se estima la media del proceso al principio con $\overline{Y}$, se centra y se modela la semivarianza o covarianza con $Z(\boldsymbol{s})$. Se encuentra su predicción $Z^*(\boldsymbol{s})$, y al final para la reconstrucción de la variable original, ya se tiene una estimación de Σ que permite estimar correctamente la media usando su estimador de mínimos

cuadrados generalizados, [67]. Así que la predicción final de la variable $Y(\boldsymbol{s}_0)$ queda

$$Y^*(\boldsymbol{s}_0) = \hat{\mu} + Z^*(\boldsymbol{s}_0). \tag{4.14}$$

4.4. Kriging indicador

En algunas ocasiones no se necesita predecir el valor específico de una variable, sino la probabilidad de que exceda o no un umbral determinado. Algunos ejemplos de este caso son el cumplimiento de una norma ambiental o valores de una variable a partir de los cuales se evidencia algún riesgo.

El kriging indicador usa el campo aleatorio $Y(s)$ del cual se tienen las observaciones

$$(y(s_1), \ldots, y(s_n)).$$

para generar otro campo aleatorio definido por un valor de interés y, [24]. El campo aleatorio de interés en este caso es

$$I(s,y) = \begin{cases} 1 & \text{si} \quad Y(s) \leq y, \quad \boldsymbol{s} \in D_s \\ 0 & \text{en otro caso.} \end{cases}$$

Asumiendo que el proceso indicador es estacionario de segundo orden, se encuentran a continuación su media y su varianza

Esperanza:

$$E[I(\boldsymbol{s},y)] = 1\text{Prob}(Y(\boldsymbol{s}) \leq y) + 0\text{Prob}(Y(\boldsymbol{s}) > y) = F(y).$$

Varianza:

$$\begin{aligned} \text{Var}[I(s,y)] &= E\left[I^2(s,y)\right] - E^2[I(s,y)] \\ &= E[I(s,y)] - F^2(y) = F(y) - F^2(y) = F(y)(1 - F(y)). \end{aligned} \tag{4.15}$$

Nótese que la varianza está acotada por 0,25. El predictor propuesto es el usual,

$$I^*(s_0,y) = \sum_{i=1}^{n} \lambda_i I(s_i,y).$$

solo que la interpretación de las variables y los predictores cambia y en lugar de referirse a un valor específico del proceso, se refieren a probabilidades de exceder o no el umbral determinado. Por lo tanto, si se conoce la distribución, se puede usar como predictor espacial el kriging simple. Si no se conoce la distribución, se puede usar como predictor espacial el kriging ordinario, [23]. Para garantizar el

insesgamiento se siguen los mismos pasos usados en la sección 4.3. Es decir,

$$E\left[I^*\left(s_0, y\right) - I\left(s_0, y\right)\right] = 0.$$

Al reemplazar por el predictor propuesto, se obtiene

$$E\left[\sum_{i=1}^{n} \lambda_i I\left(s_i, y\right) - I\left(s_0, y\right)\right] = 0. \tag{4.16}$$

Tomando esperanza, se observa que el predictor es una combinación lineal de términos de la distribución de probabilidad de la variable original. Esto es,

$$I^*\left(s_0, y\right) = \lambda_1\left(P(\boldsymbol{s}_1) < y\right) + \ldots + \lambda_n\left(P(\boldsymbol{s}_n) < y\right) = \lambda_1 F\left(y\right) + \ldots + \lambda_n F\left(y\right).$$

volviendo a (4.16) si no se conoce $F(.)$, para garantizar el insesgamiento es necesaria la restricción.

$$\sum_{i=1}^{n} \lambda_i F\left(y\right) = F\left(y\right) \quad \sum_{i=1}^{n} \lambda_i = 1.$$

Si se conoce $F(.)$ no se requiere restricción, siempre hay insesgamiento (ver Sección 4.2). Ahora, encontremos el semivariograma indicador, $\boldsymbol{h}, \boldsymbol{s} \in \mathbb{R}^d$.

$$\begin{aligned} 2\gamma\left(h; \Theta\right) &= \operatorname{Var}\left[I\left(s+h, y\right) - I\left(s, y\right)\right] \\ &= \operatorname{Var}\left[I\left(s+h, y\right)\right] + \operatorname{Var}\left[I\left(s, y\right)\right] - 2\operatorname{Cov}\left[I\left(s+h, y\right), I\left(s, y\right)\right] \\ &= 2F(y) - 2P\left(Y(s+h) \leq y, Y(s) \leq y\right) \end{aligned}$$

Este resultado es muy importante, pues muestra que el variograma indicador puede ser obtenido directamente de la distribución bivariada de la variable continua observada antes de categorizarla. Nótese que $I^*\left(s_0, y\right)$ es un predictor de la siguiente probabilidad condicional

$$P\left(Y\left(s_0\right) \leq y \mid I\left(s_1, y\right), \ldots, I\left(s_n, y\right)\right).$$

Esta técnica no tiene ningún supuesto, por lo tanto, se considera un método no paramétrico. Si existen varios umbrales de interés, estos se pueden modelar por separado y se obtienen las predicciones de manera independiente para cada una de las variables indicadoras. Sin embargo, esto implica que no hay garantía de monotonicidad ni de que las predicciones resulten entre 0 y 1. Por ejemplo, se pueden construir los mapas de los cuartiles y puede haber traslapes. Otra opción comúnmente usada es modelar el semivariograma de la mediana y usarlo como modelo general para el resto de umbrales. En general, es un método muy sencillo y muy útil.

4.5. Kriging universal

Se usa el predictor kriging universal cuando la media del proceso espacial, $\mu(s)$, es desconocida, pero no es constante a través del dominio espacial, sino que, por el contrario, existe tendencia a lo largo de las coordenadas espaciales, y el modelo para la media consiste de un polinomio en estas coordenadas. El kriging universal encuentra simultáneamente en un mismo procedimiento, la predicción espacial y las estimaciones de los parámetros del modelo para la media.

Esto contrasta con los predictores kriging simple y kriging ordinario, los cuales utilizan los datos centrados para llevar a cabo el modelamiento del semivariograma y la predicción para la variable centrada y al final, a esta predicción se le suma la media. En términos prácticos, siempre es necesario tener la variable centrada para la estimación del semivariograma, a menos que ya se conozca el modelo y no se requiera su estimación.

Se asume el modelo geoestadístico

$$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + Z(\boldsymbol{s})$$

Dónde localmente, la media puede expresarse como una combinación lineal de funciones $f_k(\boldsymbol{s})$, que en general, son términos polinómicos. Es decir, la media es una combinación lineal de términos de la forma

$$\mathrm{x}^p \mathrm{y}^q \quad p, q \in \mathbb{N}$$

Esto es, $E[Z(\boldsymbol{s})] = 0$ y

$$\mu(s) = \sum_{k=1}^{K} \beta_k f_k(s), \qquad s = (\mathrm{x}, \mathrm{y}).$$

Se le llama kriging universal al proceso que encuentra de manera simultánea la estimación de los parámetros de la media,

$$\beta_1, \ldots, \beta_K$$

y la predicción óptima en un lugar $\boldsymbol{s}_0$.

El predictor kriging es el usual y se quiere garantizar su insesgamiento. Es decir, se quiere garantizar que se cumple que

$$E[Y^*(s_0) - Y(s_0)] = 0.$$

El predictor se puede expresar en términos de la media y tomando esperanzas, se tiene que,

$$\begin{aligned} E\left[Y^*\left(s_0\right)\right] &= E\left[\sum_{i=1}^{n} \lambda_i Y\left(s_i\right)\right] \\ &= E\left[\sum_{i=1}^{n} \lambda_i \left(\mu\left(s_i\right) + Z\left(s_i\right)\right)\right]. \\ &= E\left[\sum_{i=1}^{n} \lambda_i \left(\sum_{k=1}^{K} \beta_k f_k\left(s_i\right) + Z\left(s_i\right)\right)\right]. \\ &= \sum_{i=1}^{n} \lambda_i \sum_{k=1}^{K} \beta_k f_k\left(s_i\right). \end{aligned}$$

Igualando la esperanza del predictor con la esperanza de la variable a predecir $Y\left(s_0\right)$ se tiene

$$\sum_{i=1}^{n} \lambda_i \sum_{k=1}^{K} \beta_k f_k\left(s_i\right) = \sum_{k=1}^{K} \beta_k f_k\left(s_0\right).$$

y agrupando de acuerdo a los índices de las sumatorias

$$\sum_{k=1}^{K} \beta_k \sum_{i=1}^{n} \lambda_i f_k\left(s_i\right) = \sum_{k=1}^{K} \beta_k f_k\left(s_0\right).$$

En consecuencia, se generan las siguientes K restricciones:

$$\sum_{i=1}^{n} \lambda_i f_k\left(s_i\right) = f_k\left(s_0\right), \qquad k = 1, \ldots, K. \tag{4.17}$$

Generalizando lo desarrollado para el kriging ordinario (ver Sección 4.3 y expresión 4.12), la expresión a minimizar y su primera derivada son

$$\mathbb{Q} = E\left[Y^*\left(s_0\right) - Y\left(s_0\right)\right]^2 - 2\sum_{k=1}^{K} \delta_k \left(\sum_{i=1}^{n} \lambda_i f_k\left(s_i\right) - f_k\left(s_0\right)\right).$$

$$\frac{\partial \mathbb{Q}}{\partial \lambda_i} = 2\gamma\left(s_i - s_0\right) - 2\sum_{j=1}^{n} \lambda_j \gamma\left(s_i - s_j\right) - 2\sum_{k=1}^{K} \delta_k f_k\left(s_i\right) \qquad i = 1, \ldots, n.$$

Las ecuaciones del kriging universal son

$$\sum_{j=1}^{n} \lambda_j \gamma\left(s_i - s_j\right) + \sum_{k=1}^{K} \delta_k f_k\left(s_i\right) = \gamma\left(s_i - s_0\right) \qquad i = 1, \ldots, n.$$

sujetas a las restricciones dadas en la expresión 4.17 Las ecuaciones generales del kriging universal en forma matricial están dadas por:

$$\begin{pmatrix} \Gamma & F \\ F' & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ \delta \end{pmatrix} = \begin{pmatrix} \gamma_0 \\ f_0 \end{pmatrix}.$$

$\boldsymbol{\lambda}$ es el vector de ponderaciones de las observaciones. $\boldsymbol{\delta}$ denota el vector de multiplicador de Lagrange, γ_0 el vector de semivarianza entre los lugares observados y el lugar a predecir. F es la matriz que contiene las variables polinómicas del modelo de la media evaluada en los lugares observados, y f_0 es el vector que contiene las funciones de la media evaluadas en la coordenada $\boldsymbol{s}_0$ (ver ecuación (4.18)). El detalle de cada submatrix se puede ver a continuación

$$\begin{pmatrix} \gamma(s_1 - s_1) & \gamma(s_1 - s_2) & \cdots & \gamma(s_1 - s_n) & f_1(s_1) & f_2(s_1) & \cdots & f_K(s_1) \\ \gamma(s_2 - s_1) & \gamma(s_2 - s_2) & \cdots & \gamma(s_2 - s_n) & f_1(s_2) & f_2(s_2) & \cdots & f_K(s_2) \\ . & . & & . & . & . & & . \\ . & . & & . & . & . & & . \\ . & . & & . & . & . & & . \\ \gamma(s_n - s_1) & \gamma(s_n - s_2) & \cdots & \gamma(s_n - s_n) & f_1(s_n) & f_2(s_n) & \cdots & f_K(s_n) \\ f_1(s_1) & f_1(s_2) & \cdots & f_1(s_n) & 0 & 0 & \cdots & 0 \\ f_2(s_1) & f_2(s_2) & \cdots & f_2(s_n) & 0 & 0 & \cdots & 0 \\ . & . & & . & . & . & & . \\ . & . & & . & . & . & & . \\ . & . & & . & . & . & & . \\ f_K(s_1) & f_K(s_2) & \cdots & f_K(s_n) & 0 & 0 & \cdots & 0. \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ . \\ . \\ . \\ \lambda_n \\ \delta_1 \\ . \\ . \\ . \\ \delta_K \end{pmatrix} = \begin{pmatrix} \gamma(s_1 - s_0) \\ \gamma(s_2 - s_0) \\ . \\ . \\ . \\ \gamma(s_n - s_0) \\ f_1(s_0) \\ f_2(s_0) \\ . \\ . \\ . \\ f_K(s_0) \end{pmatrix}. \tag{4.18}$$

Para encontrar la varianza del kriging universal se sustituyen las ecuaciones resultantes del proceso de optimización en $\mathbb{Q}$ y se obtiene

$$E\left[Y^*(s_0) - Y(s_0)\right]^2 = \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0) + \sum_{k=1}^{k} \delta_k f_k(s_0)$$

Nótese que el kriging ordinario es un caso particular cuando $k = 1$, $f_1(s) = 1$. La varianza del kriging universal es una extensión natural de la varianza del kriging ordinario (ver expresión 4.13).

Ejemplo 9 (Ilustración Kriging universal.). *Si hay tendencia lineal en ambas coordenadas, el modelo de la media es*

$$\mu(\boldsymbol{s}) = \beta_1 + \beta_2 \mathrm{x} + \beta_3 \mathrm{y}.$$

Entonces, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$. *Hay 3 restricciones,* $f_k(\boldsymbol{s});\ k = 1, 2, 3$ *dadas por*

$$f_1(\boldsymbol{s}) = 1, \quad f_2(s) = \mathrm{x}, \qquad f_3(\boldsymbol{s}) = \mathrm{y}.$$

En la práctica este modelo no se conoce y es necesaria la exploración de los gráficos de dispersión de cada coordenada vs. los valores observados. Para la estimación del semivariograma se requiere la variable centrada. Así que se ajusta un modelo en esta etapa, se modela el semivariograma usando los residuales y posteriormente se estiman de nuevo de manera simultánea los parámetros Θ y $\boldsymbol{\beta}$. El software lleva a cabo las predicciones de los residuales y los valores que devuelve ya corresponden a la suma estas predicciones con su respectiva media.

4.6. Kriging con pulimiento de medianas

En caso de que la falta de estacionariedad en media haya sido modelada utilizando el método de pulimiento de medianas (ver 1.1.4.3) entonces como es usual se modela el semivariograma usando los residuales obtenidos del modelo (7.8), es decir, la variable centrada $Z(\boldsymbol{s})$. Debido a que el pulimiento de medianas se hace sobre una grilla, para sumarle los valores de efecto global, efecto fila y efecto columna a las predicciones $Z^*(\boldsymbol{s}_0)$, se lleva a cabo el siguiente procedimiento:

$$\mu(\boldsymbol{s}_0) = \mu(\mathrm{x}_0, \mathrm{y}_0) = \hat{a} + \hat{r}_p + \frac{\mathrm{y}_0 - \mathrm{y}_q}{\mathrm{y}_{q+1} - \mathrm{y}_q}(\hat{r}_{p+1} - \hat{r}_p) + \hat{c}_q + \frac{\mathrm{x}_0 - \mathrm{x}_p}{\mathrm{x}_{p+1} - \mathrm{x}_p}(\hat{c}_{q+1} - \hat{c}_q). \quad (4.19)$$

los puntos

$$(\mathrm{x}_p, \mathrm{y}_q), (\mathrm{x}_{p+1}, \mathrm{y}_q),$$

$$\mathrm{y}$$

$$(\mathrm{x}_p, \mathrm{y}_{q+1}), (\mathrm{x}_{p+1}, \mathrm{y}_{q+1})$$

son los nodos más cercanos al punto $\boldsymbol{s}_0$.

Este procedimiento es llamado kriging con pulimiento de medianas, y la varianza de las predicciones corresponde a la varianza del kriging simple u ordinario, según el tipo de kriging que se aplique para encontrar $Z(\boldsymbol{s})$. Esto es, si los efectos global, fila y columna se asumen conocidos o no. Observe que en total se encuentran $P + Q + 1$ estimaciones:

- P efectos fila,
- Q efectos columna
- un efecto global.

Nota 6 (Algunos aspectos prácticos del kriging)**.**

La calidad de las predicciones *Para determinar zonas con mejores (peores) resultados, se pueden visualizar simultáneamente los mapas de predicción y de varianza de error de predicción. Además, se puede hacer uso de las estadísticas descriptivas de las varianzas, para detectar lugares cuya predicción es demasiado imprecisa.*

Validación cruzada. *El método más usado para verificar la calidad de los resultados de la predicción es el de validación cruzada. Consiste en extraer de la muestra los n datos,*

pero uno a la vez, para con los $n-1$ *datos restantes predecir el valor de la variable en la ubicación extraída. Se elimina la fila correspondiente a* $Z(\boldsymbol{s}_i)$*, se encuentra* $Z^*(\boldsymbol{s}_i)$ *con los* $n-1$ *datos restantes, y se calcula el residual de la predicción*

$$Z(\boldsymbol{s}_i) - Z^*(\boldsymbol{s}_i)$$

para cada una de las ubicaciones. Este procedimiento se puede llevar a cabo para subconjuntos con mas de una ubicación, dependiendo de la cantidad de datos existente. También se puede llevar a cabo un diagrama de dispersión de los datos observados contra cada una de sus respectivas predicciones. Esto es, graficar todas las parejas $(Z^*(s_i), Z(s_i))$*. Si las predicciones son buenas, esta nube de puntos debería tender a una línea recta de pendiente 45°. En general, funciona cualquier otro método exploratorio similar a los usados para los modelos de regresión.*

Semivarianza vs. kriging. *La estimación de la semivarianza (covarianza) espacial es la parte más importante y compleja, y la que puede llevar a requerimientos computacionales altos por implicar el ajuste de modelos no lineales de manera iterativa. En contraste, con base en estos resultados, se construye el predictor kriging cuya solución es sencilla y directa ya que consiste en un sistema de ecuaciones lineales con única solución.*

Es usual no modelar la semivarianza o covarianza usando hasta el máximo rezago. *Solo se modela el semivariograma hasta distancias menores, debido a que, a mayores distancias, hay menos pares de datos y mayor variabilidad. Este procedimiento es válido, ya que, en el proceso de kriging, los datos con distancias superiores a las elegidas no son incluidos. Esto significa que a partir de esta distancia ya se asume independencia.*

Tipos de kriging según la media. *Suponer la media conocida en ocasiones es factible, por ejemplo, la temperatura en ciertas zonas. En otras ocasiones, esta media debe ser estimada, sea o no constante. Aunque se use kriging universal, es necesario previamente encontrar el proceso centrado asociado para una correcta estimación del semivariograma. Por lo tanto, el primer paso sigue siendo estimar la media. Una vez se tenga la estimación del semivariograma, se puede llevar a cabo el proceso de kriging universal, el cual, de manera simultánea, encuentra la predicción óptima bajo las restricciones que implica la estimación de los parámetros de la media. El resultado es la predicción en su escala original, esto es, devuelve*

$$Y^*(\boldsymbol{s}_0) = Z^*(\boldsymbol{s}_0) + \sum_{k=1}^{K} \hat{\beta}_k f_k(s_0)\,.$$

Note que el kriging universal estima por mínimos cuadrados ordinarios los β_k*;* $k = 1, \ldots, K$*, pero es indispensable darle el modelo de regresión para la media, que es el modelo elegido en la primera etapa del análisis.*

Las ubicaciones de las muestras en la zona. *Si las muestras son tomadas de acuerdo a un diseño de muestreo óptimo, los resultados son mucho mejores, las predicciones son*

más precisas, ya que la varianza es menor, se evita la redundancia espacial y además se optimizan los recursos.

Clases de variables en geoestadística. *Los métodos geoestadísticos se pueden aplicar a cualquier tipo de variable. Por ejemplo, sobre el PM10 existen reglamentaciones legales en cada país que establecen los límites máximos admitidos para este contaminante. Si el interés es evaluar el cumplimiento de la norma legal, se usa kriging indicador (ver Sección 4.16) para Z, dado que esta variable se puede redefinir como una variable indicadora que tome el valor de* 1 *cuando se superen los límites legales establecidos y* 0 *en caso contrario. Esto no afecta la definición del dominio espacial, el proceso ocurre en un dominio espacial continuo, aunque la variable sea dicotómica.*

Ejemplo 10 (Calidad del aire en México. Predicción espacial del ozono). *En primer lugar, se evalúa si se puede asumir que la media del ozono es constante en Ciudad de México. En la Figura 4.2 panel superior, se observa tendencia en la dirección sur-norte, así que se estima el modelo de regresión presentado en la Tabla 4.1. Con los residuales de este modelo se construyen de nuevo los diagramas de dispersión para ver si la tendencia desaparece (ver Figura 4.2 panel inferior). Este modelo es suficiente para recoger la tendencia observada y encontrar un proceso residual con media constante. Dado que los diagramas de dispersión de los residuales ya no muestran tendencia, se elige este como el modelo de* $\mu(\boldsymbol{s})$ *y con sus residuales se construyen los semivariogramas empíricos (ver Figura 4.3).*

	$\hat{\boldsymbol{\beta}}$	s.e $\hat{\boldsymbol{\beta}}$	Valor p
Intercepto	-1564	600	0,0161
y	0,0007581	0,000279	0,01259

Tabla 4.1. Resultados de la estimación del modelo de regresión $Y(s) = \hat{\beta}_0 + \hat{\beta}_1 y$ para la media.

4.7. Ejercicios

1. Suponga que $Z(\boldsymbol{s})$ es una función aleatoria estacionaria de segundo orden que representa la precipitación en los cuatro bordes de un trapecio recto. Las realizaciones registradas del fenómeno en estas cuatro ubicaciones son $z_1(0,0) = 15$; $z_2(100,0) = 16$; $z_3(0,50) = 16$ y $z_4(70,50) = 15$. El modelo de autocorrelación del fenómeno es exponencial con silla 4 y rango 100 m. Encuentre

    1.1 La predicción con su respectiva varianza de error de predicción, esto es, la varianza del kriging simple para la ubicación $Z(75,25)$ si la media es conocida e igual a 15,7. Use para sus cálculos la función de covarianza.

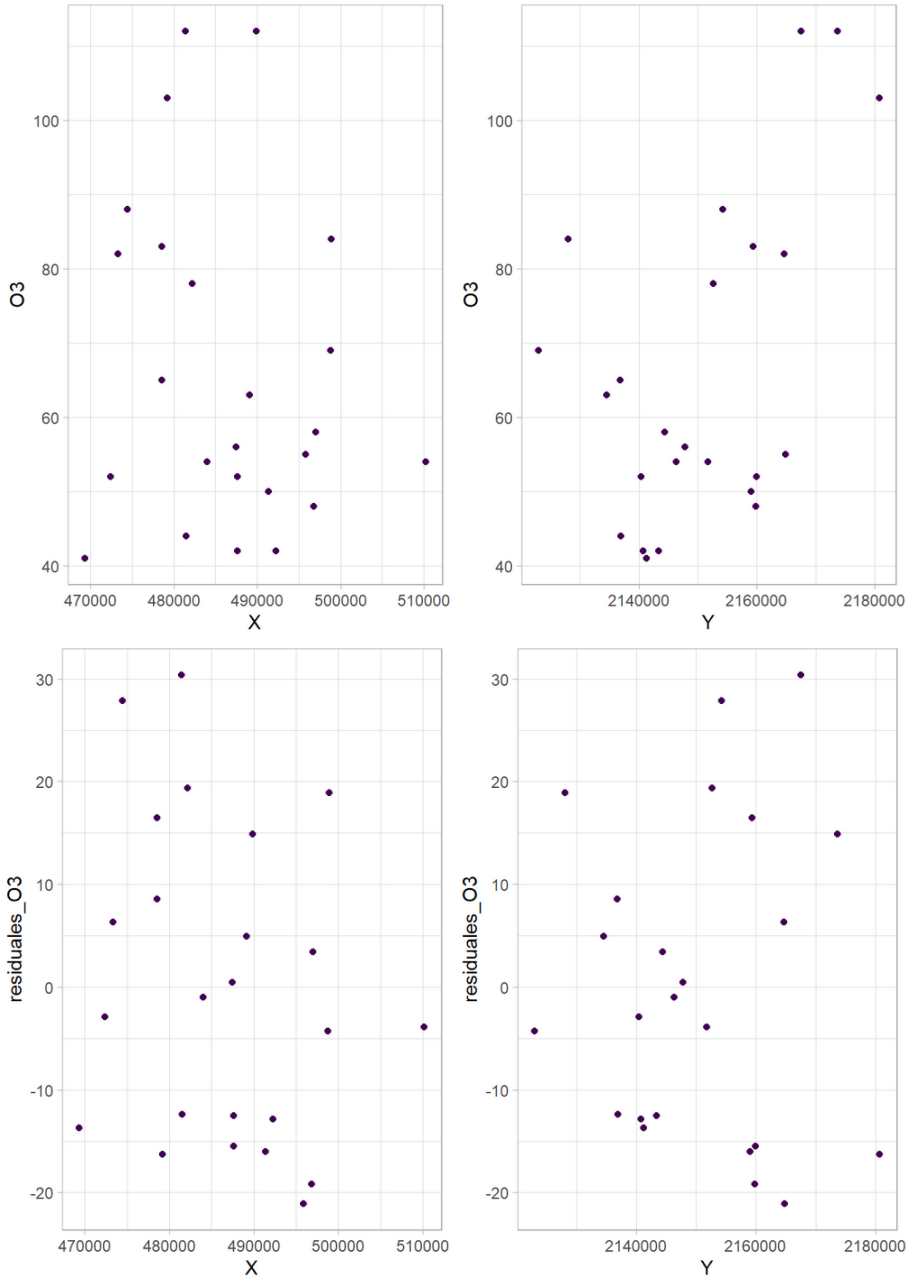


Figura 4.2. Diagramas de dispersión de la concentración de ozono en muestras de aire en Ciudad de México a lo largo de cada coordenada espacial. Panel superior: datos observados. Panel inferior: residuales del modelo mostrado en la Tabla 4.1.

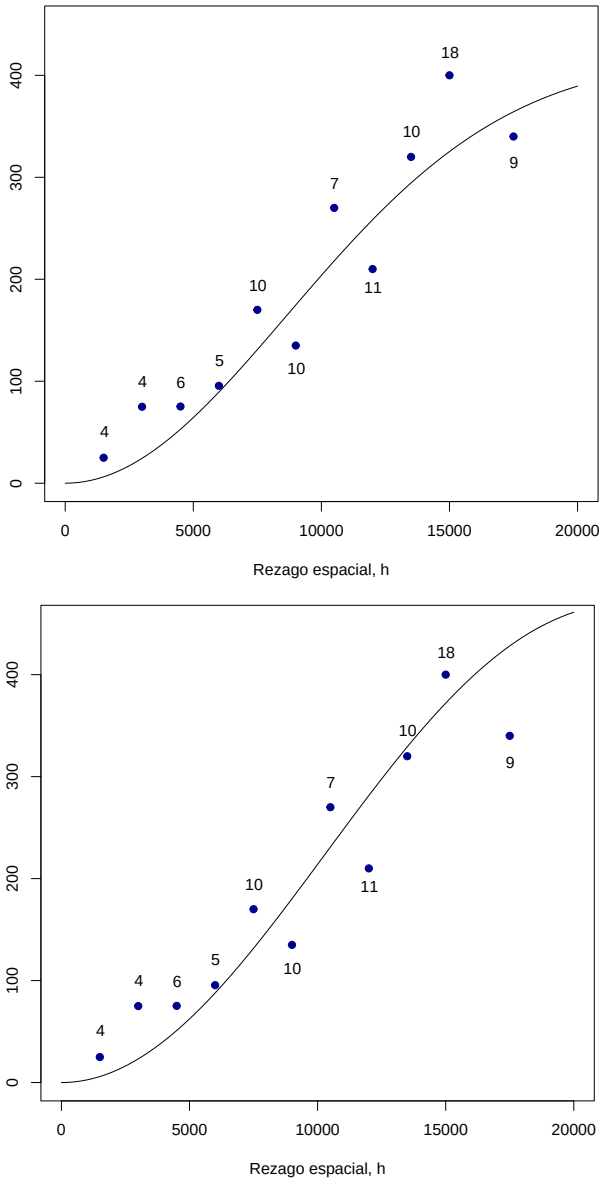


Figura 4.3. Semivariogramas empírico y teórico para el ozono en la Ciudad de México. **Panel superior:**, Modelo exponencial $\hat{\Theta} = (418, 4; 12239, 25)$. **Panel inferior:** Modelo seno cardinal $\hat{\Theta} = (422, 9; 6579, 15)$. Los números al lado de cada punto indican la cantidad de pares existentes para cada rezago, $|N(h)|$. Los ajustes de estos dos modelos al semivariograma empírico son muy similares. Se elige el modelo exponencial porque da unas varianzas de error de predicción ligeramente menores.

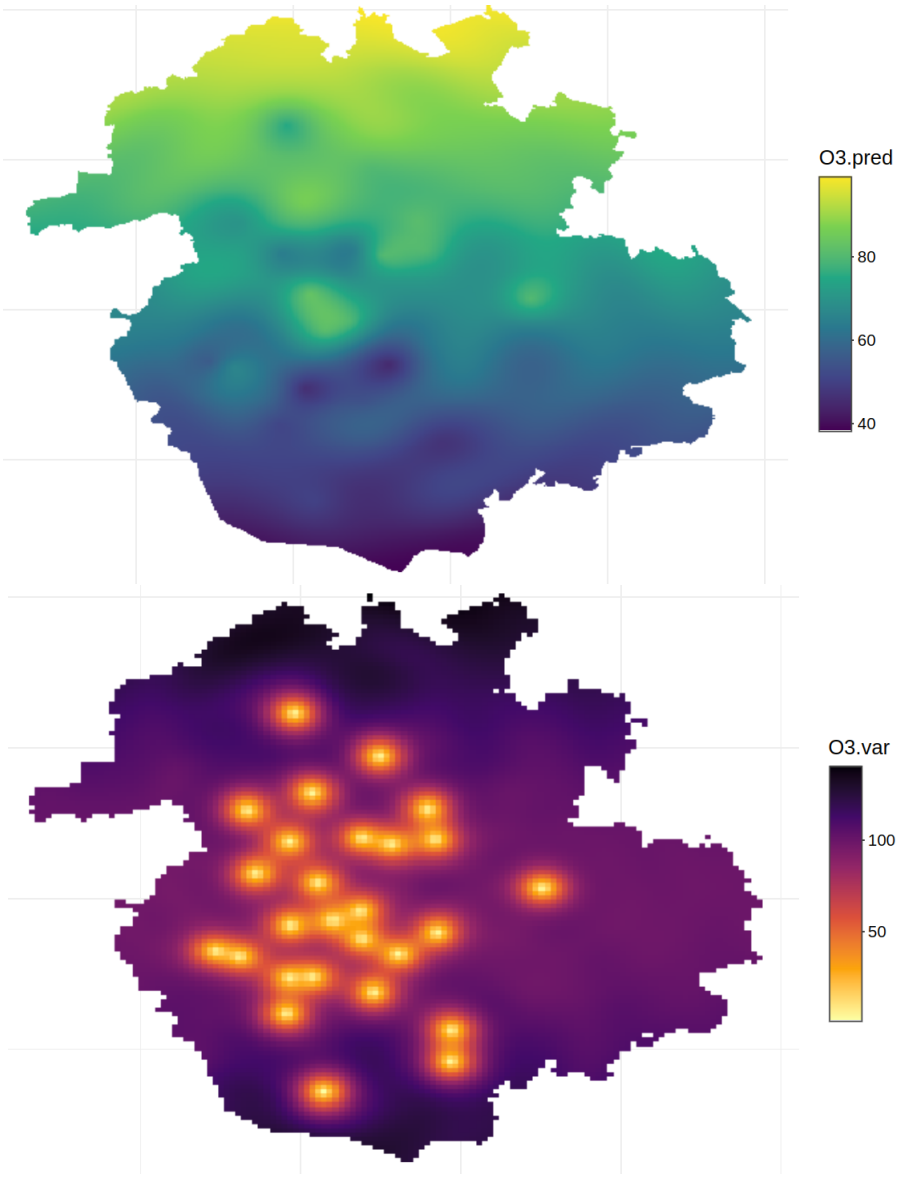


Figura 4.4. Resultados del proceso de predicción del ozono en la Ciudad de México. **Panel superior:** mapa de Ciudad de México con la predicción del ozono usando kriging simple sobre los residuales del modelo para la media, (ver Tabla 4.1) y un semivariograma exponencial $\hat{\Theta} = (418, 4; 12239, 25)$. A cada predicción $Z^*(s_0)$ se le suma su media $\mu(s_0)$. **Panel inferior:** mapa de la varianza del error de predicción (varianza del kriging simple). Nótese que el kriging es un predictor exacto, por lo que, la varianza es cero en los puntos con observaciones. Una alternativa es usar la varianza de la predicción obtenida usando validación cruzada en cada uno de estos puntos.

1.2 Si la media es desconocida, pero constante. Use para sus cálculos la función de covarianza.

1.3 Si la media es desconocida, pero constante. Use para sus cálculos la función de semivarianza.

1.4 Suponga que el proceso es normal multivariado. Calcule el intervalo de confianza del 95 % para la predicción encontrada en cada caso.

1.5 Interprete y compare tanto los resultados obtenidos como los procedimientos realizados.

2. Suponga que

$$\mathscr{W}(s) = \Psi(Y(s)) \qquad s \in D$$

Donde $Y(s)$ es una variable aleatoria con distribución de probabilidad normal y media constante y conocida μ_y. Demuestre que

$$\mathscr{W}^*(s_0) = \Psi(Y^*(s_0)) + \Psi''(\mu_y)\left\{\frac{\text{Var}[Y(s) - Y^*(s_0)]}{2}\right\}$$

Esto es, el predictor kriging aplicado a una transformación de la variable espacial es sesgado. Sugerencia: Utilice el desarrollo en series de Taylor.

3. Encuentre el predictor kriging si la media es constante y conocida y el modelo de covarianza es un modelo efecto pepita.

4. Encuentre el predictor kriging si la media es un polinomio de primer grado e incluye ambas coordenadas, (x, y), y el modelo de covarianza es un modelo efecto pepita.

5. Suponga un campo aleatorio espacial normal multivariado, con observaciones en n ubicaciones y media y covarianza conocidos. Encuentre la esperanza condicional y la respectiva varianza $Z(\boldsymbol{s}_0)$ dado el vector $\big(Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n)\big)$. Compare con las expresiones de predicción y varianza del error de predicción del kriging simple.

6. Para los datos de aquifer que puede encontrar en https://mpbohorquezc.github.io/SpatFD-Functional-Geostatistics/GeoestadisticaUnivariadaConGeoR.html, genere dos mapas de predicciones. Para el primer mapa utilice el predictor kriging y para el segundo mapa utilice predicción basada en modelos de regresión en términos de las coordenadas (x, y) = $(Este, Norte)$, e incluyendo la dependencia espacial a través de la estimación de los parámetros del modelo, vía máxima verosimilitud. Compare el desempeño predictivo del modelo de regresión cuando se usan mínimos cuadrados ordinarios ignorando la dependencia espacial y cuando se usa máxima verosimilitud. Posteriormente, compare el desempeño de la regresión con el kriging.

7. Simule campos aleatorios espaciales bajo el supuesto de normalidad multivariada, para varios tipos de autocovarianza espacial. Grafique los

dispersogramas rezagados variando modelos y parámetros. Esto es, en cada caso, grafique $Y(\boldsymbol{s})$ Vs $Y(\boldsymbol{s}+\boldsymbol{h})$, para varios valores de h. y en cada caso calcule el coeficiente de correlación de Pearson. Analice.

8. Se requiere llevar a cabo predicción espacial para un proceso espacial cuya distribución marginal es lognormal con media conocida. ¿Es posible encontrar un predictor insesgado? Demuestre o refute.

9. Generalice el resultado encontrado en el problema anterior, al caso en el que se transforma la variable usando una transformación Box-Cox.

Capítulo

cinco

Predicción espacial escalar con covariables: Cokriging.

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

Hasta ahora se ha llevado a cabo la predicción espacial de un proceso $Y(\boldsymbol{s})$ utilizando únicamente su propia información. Sin embargo, los fenómenos son en general multivariados. Este capítulo describe cómo llevar a cabo la predicción espacial de un proceso

$$Y_r(\boldsymbol{s}_0)$$

utilizando su propia información y la de covariables que se encuentren espacialmente correlacionadas con este, es decir, utilizando

$$Y_1(\boldsymbol{s}), \cdots, Y_r(\boldsymbol{s}) \cdots, Y_P(\boldsymbol{s}).$$

Este método es conocido como cokriging. Para la aplicación de este método no es necesario que todas las variables estén medidas en las mismas ubicaciones espaciales. En general, cada $Y_r(s)$ puede estar medida en un número diferente de sitios n_r. De hecho es una de las ventajas del método, debido a que en ocasiones existen variables mas costosas o difíciles de medir, así que se toman mas datos de aquellas variables cuyas observaciones son económicas o sencillas.

5.1. Cokriging para el caso de P variables

Si todos los datos se encuentran medidos en la misma grilla de n ubicaciones espaciales, los datos forman una matriz $n \times P$ con (i, j)-ésimo elemento $Y_p(\boldsymbol{s}_i)$, $i = 1, .., n$, $p = 1, ..., P$. La i-ésima fila de la matriz de datos corresponde a las mediciones de todas las variables en la ubicación $\boldsymbol{s}_i$.

$$\mathbb{y}(\boldsymbol{s}_i) = (y_1(\boldsymbol{s}_i), y_2(\boldsymbol{s}_i), ..., y_p(\boldsymbol{s}_i))'. \tag{5.1}$$

El interés es encontrar un predictor insesgado y de mínima varianza para el vector

$$\mathbb{Y}(\boldsymbol{s}_0) \equiv (Y_1(\boldsymbol{s}_0), ..., Y_P(\boldsymbol{s}_0))'. \tag{5.2}$$

No es una predicción multivariada dado que no se realiza simultáneamente para todas las variables. Por el contrario, la predicción es realizada para una variable a la vez (ver [24]), utilizando todas las variables.

El predictor para

$$Y_r(\boldsymbol{s}_0), \quad 1 \leq r \leq P$$

con base en todas las P variables, es la siguiente combinación lineal de los datos observados:

$$Y_r^*(\boldsymbol{s}_0) = \sum_{i=1}^{n_p} \sum_{p=1}^{P} \lambda_i^{rp} y_p(\boldsymbol{s}_i). \tag{5.3}$$

El cokriging se puede construir bajo estacionariedad de segundo orden o bajo estacionariedad intrínseca como el kriging. Bajo estacionariedad de segundo orden se asume que,

Media constante $E[Y_r(\boldsymbol{s})] = \mu_r$, con $\boldsymbol{\mu} = (\mu_1, ..., \mu_P)'$, para todo $\boldsymbol{s} \in D_s$.

Covarianza La estructura de covarianza para cada par de ubicaciones es una matriz $\boldsymbol{\Sigma}_{(\boldsymbol{s}_i,\boldsymbol{s}_{i'})}$ de dimensión $P \times P$, que solo depende de las ubicaciones espaciales (5.5). Esto es, la covarianza de dos procesos espaciales Y_r y $Y_{r'}$, $r, r' = 1, \ldots, P$ en ubicaciones distintas $\boldsymbol{s}_i, \boldsymbol{s}_{i'} \in D_{\boldsymbol{s}}$, se conoce como covarianza cruzada y es función de $\boldsymbol{s}_i$ y $\boldsymbol{s}_{i'}$, esto es,

$$\text{Cov}(Y_r(\boldsymbol{s}_i), Y_{r'}(\boldsymbol{s}_{i'})) = C_{rr'}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) \tag{5.4}$$

La matriz de covarianza del proceso general entre dos ubicaciones espaciales está dada por

$$\boldsymbol{\Sigma}_{(\boldsymbol{s}_i,\boldsymbol{s}_{i'})} = \begin{pmatrix} C_{11}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) & C_{12}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) & \ldots & C_{1p}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) \\ C_{21}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) & C_{22}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) & \ldots & C_{2p}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) \\ \vdots & \vdots & \ddots & \vdots \\ C_{P1}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) & C_{P2}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta) & \ldots & C_{PP}(\boldsymbol{s}_i, \boldsymbol{s}_{i'}; \Theta). \end{pmatrix} \tag{5.5}$$

la cual no es necesariamente simétrica porque, en general, no existe ninguna razón para asumir que

$$\text{Cov}(Y_r(\boldsymbol{s}_i), Y_{r'}(\boldsymbol{s}_{i'})) = \text{Cov}(Y_{r'}(\boldsymbol{s}_i), Y_r(\boldsymbol{s}_{i'})). \tag{5.6}$$

Así, en general,

$$\boldsymbol{\Sigma}_{(\boldsymbol{s}_i,\boldsymbol{s}_{i'})} \neq \boldsymbol{\Sigma}_{(\boldsymbol{s}_{i'},\boldsymbol{s}_i)} \qquad \text{y} \qquad \boldsymbol{\Sigma}_{(\boldsymbol{s}_i,\boldsymbol{s}_{i'})} \neq \boldsymbol{\Sigma}'_{(\boldsymbol{s}_i,\boldsymbol{s}_{i'})}$$

aunque se satisface que

$$\boldsymbol{\Sigma}'_{(\boldsymbol{s}_i,\boldsymbol{s}_{i'})} = \boldsymbol{\Sigma}_{(\boldsymbol{s}_{i'},\boldsymbol{s}_i)}$$

Por lo tanto, si todas las variables son medidas en las mismas n ubicaciones, la matriz de covarianza completa

$$\text{Cov}[\mathbb{Y}] = \sum$$

con todas las variables y ubicaciones observadas es

$$\sum = \begin{pmatrix} \boldsymbol{\Sigma}_{(\boldsymbol{s}_1,\boldsymbol{s}_1)} & \boldsymbol{\Sigma}_{(\boldsymbol{s}_1,\boldsymbol{s}_2)} & \cdots & \boldsymbol{\Sigma}_{(\boldsymbol{s}_1,\boldsymbol{s}_n)} \\ \boldsymbol{\Sigma}_{(\boldsymbol{s}_2,\boldsymbol{s}_1)} & \boldsymbol{\Sigma}_{(\boldsymbol{s}_2,\boldsymbol{s}_2)} & \cdots & \boldsymbol{\Sigma}_{(\boldsymbol{s}_2,\boldsymbol{s}_n)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{(\boldsymbol{s}_n,\boldsymbol{s}_1)} & \boldsymbol{\Sigma}_{(\boldsymbol{s}_n,\boldsymbol{s}_2)} & \cdots & \boldsymbol{\Sigma}_{(\boldsymbol{s}_n,\boldsymbol{s}_n)} \end{pmatrix}. \tag{5.7}$$

La cual es una matriz definida positiva, que se puede construir con una combinación lineal de los modelos de covarianza de cada una de las variables involucradas (ver [37] y [73]). La expresión 5.7 tiene las submatrices organizadas

por pares de ubicaciones espaciales con fines ilustrativos. Sin embargo, es mas usual y práctico organizarla usando las submatrices de cada variable así como de los respectivos cruces de variables como se puede ver a continuación en el Ejemplo 11.

Ejemplo 11 (Ilustración 1, $p = 2$, $n = 3$, Las dos variables son observadas en las mismas ubicaciones). *Cada una de las submatrices se define como:* $\mathbf{\Sigma}_{11} = \text{Cov}(\mathbb{Y}_1, \mathbb{Y}_1)$, $\mathbf{\Sigma}_{22} = \text{Cov}(\mathbb{Y}_2, \mathbb{Y}_2)$ *y* $\mathbf{\Sigma}_{12} = \text{Cov}(\mathbb{Y}_1, \mathbb{Y}_2) = \Sigma'_{21}$. *Es necesario estimar de manera conjunta la matriz* $\sum$ *del proceso bivariado con todas sus submatrices para que sea definida positiva.*

$$\sum = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$= \left(\begin{array}{ccc|ccc} C_{11}(s_1,s_1) & C_{11}(s_1,s_2) & C_{11}(s_1,s_3) & C_{12}(s_1,s_1) & C_{12}(s_1,s_2) & C_{12}(s_1,s_3) \\ C_{11}(s_2,s_1) & C_{11}(s_2,s_2) & C_{11}(s_2,s_3) & C_{12}(s_2,s_1) & C_{12}(s_2,s_2) & C_{12}(s_2,s_3) \\ C_{11}(s_3,s_1) & C_{11}(s_3,s_2) & C_{11}(s_3,s_3) & C_{12}(s_3,s_1) & C_{12}(s_3,s_2) & C_{12}(s_3,s_3) \\ \hline C_{21}(s_1,s_1) & C_{21}(s_1,s_2) & C_{21}(s_1,s_3) & C_{22}(s_1,s_1) & C_{22}(s_1,s_2) & C_{22}(s_1,s_3) \\ C_{21}(s_2,s_1) & C_{21}(s_2,s_2) & C_{21}(s_2,s_3) & C_{22}(s_2,s_1) & C_{22}(s_2,s_2) & C_{22}(s_2,s_3) \\ C_{21}(s_3,s_1) & C_{21}(s_3,s_2) & C_{21}(s_3,s_3) & C_{22}(s_3,s_1) & C_{22}(s_3,s_2) & C_{22}(s_3,s_3) \end{array}\right).$$

Ejemplo 12 (Ilustración 2, $p = 2$, $n_1 = 3$, $n_2 = 2$). *Se supone que el proceso* $Y_1(\boldsymbol{s})$ *fue medido en los lugares* $\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3$ *y que el proceso* $Y_2(\boldsymbol{s})$ *fue medido en los lugares* $\boldsymbol{s}'_1, \boldsymbol{s}'_2$, *posiblemente diferentes.*

$$\sum = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$= \left(\begin{array}{ccc|cc} C_{11}(s_1,s_1) & C_{11}(s_1,s_2) & C_{11}(s_1,s_3) & C_{12}(s_1,s'_1) & C_{12}(s_1,s'_2) \\ C_{11}(s_2,s_1) & C_{11}(s_2,s_2) & C_{11}(s_2,s_3) & C_{12}(s_2,s'_1) & C_{12}(s_2,s'_2) \\ C_{11}(s_3,s_1) & C_{11}(s_3,s_2) & C_{11}(s_3,s_3) & C_{12}(s_3,s'_1) & C_{12}(s_3,s'_2) \\ \hline C_{21}(s'_1,s_1) & C_{21}(s'_1,s_2) & C_{21}(s'_1,s_3) & C_{22}(s'_1,s'_1) & C_{22}(s'_1,s'_2) \\ C_{21}(s'_2,s_1) & C_{21}(s'_2,s_2) & C_{21}(s'_2,s_3) & C_{22}(s'_2,s'_1) & C_{22}(s'_2,s'_2) \end{array}\right).$$

El predictor (5.3) se construye insesgado,

$$E\left[Y_r^*(\boldsymbol{s}_0) - Y_r(\boldsymbol{s}_0)\right] = 0$$

Esto es,

$$E\left[Y_r^*(\boldsymbol{s}_0) - Y_r(\boldsymbol{s}_0)\right] = E\left[\sum_{i=1}^{n_p}\sum_{p=1}^{P}\lambda_i^{rp}Y_p(\boldsymbol{s}_i) - Y_r(\boldsymbol{s}_0)\right] = 0. \tag{5.8}$$

Si las medias se consideran conocidas, de manera análoga al caso del kriging simple, el predictor siempre es insesgado y por lo tanto, no es necesario imponer ninguna restricción. Si la media se considera constante y desconocida, tomando esperanza y aplicando el supuesto de media constante se obtiene

$$\sum_{i=1}^{n_p}\sum_{p=1}^{P}\lambda_i^{rp}\mu_p = \mu_r. \tag{5.9}$$

$$\sum_{i=1}^{n_1}\lambda_i^{r1}\mu_1 + \sum_{i=1}^{n_2}\lambda_i^{r2}\mu_2 + \ldots + \sum_{i=1}^{n_p}\lambda_i^{rp}\mu_p = \mu_r. \tag{5.10}$$

Así, las condiciones necesarias y suficientes para que el predictor cokriging sea insesgado son:

1. $\sum_{i=1}^{n_r}\lambda_i^{rr} = 1$.

2. $\sum_{i=1}^{n_p}\lambda_i^{rp} = 0$ para $p = 1, \ldots, r-1, r+1, \ldots, P$.

El mejor predictor lineal insesgado se calcula usando los λ_i^{rp} que minimizan

$$E\left[Y_r(\boldsymbol{s}_0) - \sum_{i=1}^{n_p}\sum_{p=1}^{P}\lambda_i^{rp}Y_p(\boldsymbol{s}_i)\right]^2. \tag{5.11}$$

sujeto a las restricciones para garantizar insesgamiento. Todos los conceptos aquí son similares a los desarrollados para el kriging en el Capítulo 4.

5.2. Cokriging para el caso de dos variables

A continuación, se desarrolla el cokriging para el caso de dos variables para detallar todos los elementos y procedimientos involucrados.

Sean $Y_1(\boldsymbol{s})$ y $Y_2(\boldsymbol{s})$ dos procesos estocásticos espaciales medidos en n_1 y n_2 puntos, respectivamente, nótese que todos los puntos podrían eventualmente ser distintos. Los vectores observados son

$$\mathbb{y}_1 = \left(y_1(\boldsymbol{s}_1), \ldots, y_1(\boldsymbol{s}_{n_1})\right).$$

$$\mathbb{y}_2 = \left(y_2\left(\boldsymbol{s}_1'\right), \ldots, y_2\left(\boldsymbol{s}_{n_2}'\right)\right).$$

Se requiere predecir $Y_1(\boldsymbol{s}_0)$ usando $\mathbb{y}_1(\boldsymbol{s})$ y $\mathbb{y}_2(\boldsymbol{s})$. El predictor cokriging es

$$Y_1^*(\boldsymbol{s}_0) = \sum_{i=1}^{n_1} \lambda_i y_1(\boldsymbol{s}_i) + \sum_{j=1}^{n_2} \beta_j y_2(\boldsymbol{s}_j) = \lambda \mathbb{y}_1 + \beta \mathbb{y}_2.$$

Si las medias del campo aleatorio son constantes y desconocidas, y se requiere garantizar que el predictor sea insesgado, entonces hay que garantizar que

$$E\left[Y_1(\boldsymbol{s}_0) - Y_1^*(\boldsymbol{s}_0)\right] = 0.$$

Tomando esperanza

$$E[Y_1(\boldsymbol{s}_0)] = \sum_{i=1}^{n_1} \lambda_i E(Y_1(\boldsymbol{s}_i)) + \sum_{j=1}^{n_2} \beta_j E(Y_2(\boldsymbol{s}_j)).$$

$$\mu_1 = \sum_{i=1}^{n_1} \lambda_i \mu_1 + \sum_{j=1}^{n_2} \beta_j \mu_2.$$

Por lo tanto, se deben imponer las siguientes restricciones

$$\sum_{i=1}^{n_1} \lambda_i = 1 \qquad \text{y} \qquad \sum_{j=1}^{n_2} \beta_j = 0.$$

Si se conocen la medias, se construyen los campos aleatorios centrados,

$$Y_1(\boldsymbol{s}) - \mu_1(\boldsymbol{s}) = Z_1(\boldsymbol{s}), \quad Y_2(\boldsymbol{s}) - \mu_2(\boldsymbol{s}) = Z_2(\boldsymbol{s}).$$

entonces $E[Z_1(\boldsymbol{s})] = 0$ y $E[Z_2(\boldsymbol{s})] = 0$ y, por lo tanto, el predictor cokriging simple siempre es insesgado y no requiere restricciones sobre las ponderaciones. Se lleva a cabo el procedimiento del cokriging con las realizaciones de las variables centradas $\mathbb{Z}_1$ y $\mathbb{Z}_2$, estas son,

$$\mathbb{z}_1 = \left(z_1(\boldsymbol{s}_1), \ldots, z_1(\boldsymbol{s}_{n_1})\right).$$

$$\mathbb{z}_2 = \left(z_2\left(\boldsymbol{s}_1'\right), \ldots, z_2\left(\boldsymbol{s}_{n_2}'\right)\right).$$

y posteriormente se le suma la media respectiva. El predictor cokriging $Z_1^*(\boldsymbol{s}_0)$ de la variable $Z_1(\boldsymbol{s}_0)$, se encuentra minimizando

$$\text{Var}[Z_1^*(\boldsymbol{s}_0) - Z_1(\boldsymbol{s}_0)]^2.$$

que según el tipo de estacionariedad se puede formular en términos de semivarianzas o de covarianzas.

A continuación se presentan los sistemas de ecuaciones para cada tipo de cokriging en términos de covarianzas para predecir $Z_1(\boldsymbol{s}_0)$. Nótese que estos casos son generalizaciones de los tipos de kriging. $\boldsymbol{\lambda}$ y $\boldsymbol{\beta}$ son los vectores de ponderaciones de las observaciones de Z_1 y Z_2, respectivamente. δ denota un

multiplicador de Lagrange, los vectores de covarianza entre cada una de las variables en los lugares observados y el lugar a predecir son

$$\boldsymbol{\sigma}_{10} = \text{Cov}\left(\mathbb{Z}_1, Z_1(\boldsymbol{s}_0)\right)$$

y

$$\boldsymbol{\sigma}_{20} = \text{Cov}\left(\mathbb{Z}_2, Z_1(\boldsymbol{s}_0)\right).$$

F es en cada caso la matriz que contiene las variables polinómicas de los modelos de medias evaluadas en los lugares observados, y $\mathbb{f}_0$ es el vector que contiene las funciones de la media de Z_1 evaluadas en la coordenada $\boldsymbol{s}_0$ (ver ecuación (4.18)).

Cokriging simple

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{10} \\ \boldsymbol{\sigma}_{20} \end{pmatrix}. \tag{5.12}$$

Cokriging ordinario

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \mathbf{1}_{n_1} & 0 \\ \Sigma_{21} & \Sigma_{22} & 0 & \mathbf{1}_{n_2} \\ \mathbf{1}'_{n_1} & 0 & 0 & 0 \\ 0 & \mathbf{1}'_{n_2} & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \delta_1 \\ \delta_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{10} \\ \boldsymbol{\sigma}_{20} \\ 1 \\ 0 \end{pmatrix}. \tag{5.13}$$

Cokriging universal

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} & | & F_1 & 0 \\ \Sigma_{21} & \Sigma_{22} & | & 0 & F_2 \\ - & - & - & - & - \\ F_1' & 0 & | & 0 & 0 \\ 0 & F_2' & | & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ - \\ \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{10} \\ \boldsymbol{\sigma}_{20} \\ - \\ \mathbb{f}_{10} \\ 0 \end{pmatrix}. \tag{5.14}$$

Nótese que es muy difícil aplicar en la práctica el cokriging universal. En la matriz de coeficientes del cokriging universal se evidencia la necesidad del supuesto de que las variables usadas en los modelos de medias no tengan ninguna relación, para que la matriz sea de rango completo y de que todas las variables sean medidas en los mismos lugares, lo cual no es realista.Este método comparte todas las características con el kriging universal (ver Sección 4.5), el cual resulta ser un caso particular del cokriging universal. Lo más práctico es usar cokriging simple con los residuos de las variables centradas y al final sumarle a las predicciones la media extraída, o usar cokriging ordinario.

Además,como en cualquier método multivariado, las dimensiones de las matrices de covarianza crecen muy rápido a medida que aumenta la cantidad variables. Por lo tanto, modelar muchos campos aleatorios simultáneamente no es conveniente, ya que se puede complicar la estimación de la matriz $\sum$ o se pueden encontrar modelos válidos, pero no de buena calidad. Hay que ajustar modelos individuales y bivariados de tal manera que la matriz completa resulte definida positiva. Por eso el principio de parsimonia aquí es fundamental (el modelamiento de $\sum$ es tratado en la Sección 5.4). Finalmente, si no existe correlación espacial cruzada, el predictor cokriging se reduce al predictor kriging.

5.3. Medidas de dependencia espacial cruzada

La covarianza y la correlación cruzada modelan y cuantifican la dependencia que tiene una variable espacial en un lugar, $Y_r(s)$, con otra variable espacial $Y_{r'}(s+h)$ en otro lugar. Son una generalización natural de los conceptos de autocovarianza y autocorrelación al caso de dos procesos espaciales. La función de covarianza cruzada puede tomar valores negativos y, además, puede mostrar covarianza máxima para $h \neq 0$.

El modelo lineal de corregionalización del campo aleatorio espacial multivariado $\mathbb{Y} = (Y_1, \ldots, Y_P)$, se construye con base en la representación de cada campo aleatorio $Y_r(\boldsymbol{s})$ como combinación lineal de l campos aleatorios centrados, autocorrelacionados y mutuamente independientes Z_r^l, así

$$Y_r(\boldsymbol{s}) = \sum_{l=0}^{L} \sum_{k=1}^{m_l} b_{rk}^l Z_k^l(\boldsymbol{s}) + \mu_r \qquad \forall r = 1, \ldots, P.$$

- $E[Y_r(\boldsymbol{s})] = \mu_r$
- $E[Z_k^l(\boldsymbol{s})] = 0 \qquad \forall l = 0, \ldots, L$

$$\text{Cov}[Z_k^l(\boldsymbol{s}), Z_{k'}^{l'}(\boldsymbol{s}+\boldsymbol{h})] = \begin{cases} C_l(\boldsymbol{h}; \boldsymbol{\theta}_l) & \text{Si } k = k' \quad y \quad l = l'; \\ 0 & \text{En otro caso.} \end{cases}$$

Con base en la independencia mutua entre Zs se obtiene la covarianza entre dos elementos de $\mathbb{Y}$,

$$\begin{aligned} \text{Cov}[Y_r(\boldsymbol{s}), Y_{r'}(\boldsymbol{s}+\boldsymbol{h})] &= \sum_{l=0}^{L} \sum_{l'=0}^{L} \sum_{k=1}^{m_l} \sum_{k'=1}^{m_l} b_{rk}^l b_{r'k'}^{l'} \text{Cov}[Z_k^l(\boldsymbol{s}), Z_{k'}^{l'}(\boldsymbol{s}+\boldsymbol{h})] \\ &= \sum_{l=0}^{L} \sum_{k=1}^{m_l} b_{rk}^l b_{r'k}^l C_l(\boldsymbol{h}; \boldsymbol{\theta}_l) = \sum_{l=0}^{L} d_{rr'}^l C_l(\boldsymbol{h}; \boldsymbol{\theta}_l). \end{aligned} \tag{5.15}$$

donde

$$\sum_{k=1}^{m_l} b_{rk}^l b_{r'k}^l = d_{rr'}^l.$$

$d_{rr'}^l$ es la silla del modelo $C_l(\boldsymbol{h})$ y su contribución al modelo completo de covarianza cruzada entre Y_r y $Y_{r'}$; $r, r' = 1, \ldots, P$. Los coeficientes de (5.15) forman una matriz semidefinida positiva de dimensión $P \times P$ que es llamada la matriz de corregionalización.

$$D_l = \left[d_{rr'}^l \right].$$

Como en el caso univariado, en el proceso de estimación de la función de dependencia espacial bivariada, no es muy generalizado el uso del covariograma, ya que requiere para su estimación las medias y varianzas del proceso, y, además, que las variables estén medidas en los mismos lugares. Una alternativa presentada por [23] es el pseudo variograma cruzado, pero su inconveniente es la falta de sentido de las unidades del resultado, ya que se define como el cuadrado de la resta entre diferentes variables. Aunque una posibilidad es estandarizar las variables, esta práctica puede enmascarar características de las variables originales. Puede ser muy útil para comparar la misma variable cuando ha sido medida en diferentes condiciones. Por esta razón, es mejor utilizar medidas que se basen en la variable incrementos. La medida más utilizada es el semivariograma cruzado que no tiene ninguna de estas restricciones.

El semivariograma cruzado aunque implica simetría, es una función par y es invariante bajo traslación como su análogo univariado;

$$\gamma_{rr'}(0;\Theta) = 0, \quad \gamma_{rr'}(\boldsymbol{h};\Theta) = \gamma_{rr'}(-\boldsymbol{h};\Theta).$$

A continuación, se presentan una serie de medidas que se pueden usar para describir el comportamiento general de diferentes variables en diferentes ubicaciones.

Semivariograma cruzado o semicovarianza de los incrementos

$$\begin{aligned}\mathrm{Var}[Y_r(\boldsymbol{s}), Y_{r'}(\boldsymbol{s}+\boldsymbol{h})] &= \frac{1}{2} E\left[Y_r(\boldsymbol{s}+\boldsymbol{h}) - Y_r(\boldsymbol{s})\right]\left[Y_{r'}(\boldsymbol{s}+\boldsymbol{h}) - Y_{r'}(\boldsymbol{s})\right] \\ &= \gamma_{rr'}(\boldsymbol{h};\Theta).\end{aligned}$$

$\gamma_{rr'}(\boldsymbol{h})$ es la covarianza de los incrementos de dos variables espaciales y se encuentra acotada por el producto de las varianzas de los incrementos correspondientes [73]. Es decir, se cumple que

$$\gamma_{rr}(h;\Theta)\gamma_{r'r'}(h;\Theta) \geq |\, \gamma_{rr'}(\boldsymbol{h};\Theta)\,|^2 \tag{5.16}$$

Covarianza cruzada

$$\begin{aligned}\text{Cov}[Y_r(\boldsymbol{s}), Y_{r'}(\boldsymbol{s}+\boldsymbol{h})] &= E\left[\left(Y_r(\boldsymbol{s}) - E\left[Y_r(\boldsymbol{s})\right]\right)\left(Y_{r'}(\boldsymbol{s}+\boldsymbol{h}) - E\left[Y_{r'}(\boldsymbol{s}+\boldsymbol{h})\right]\right)\right] \\ &= C_{rr'}(\boldsymbol{h};\Theta).\end{aligned}$$

Correlación cruzada

$$\text{Cor}[Y_r(\boldsymbol{s}), Y_{r'}(\boldsymbol{s}+\boldsymbol{h})] = \frac{C_{rr'}(\boldsymbol{h};\Theta)}{\sigma_r\sigma_{r'}} = \rho_{rr'}(\boldsymbol{h};\Theta).$$

El pseudovariograma cruzado

$$\phi_{rr'}(\boldsymbol{h};\Theta) = \frac{1}{2}E\left[Y_r(\boldsymbol{s}+\boldsymbol{h}) - Y_{r'}(\boldsymbol{s})\right]^2.$$

Codispersión

$$\upsilon_{rr'}(\boldsymbol{h};\Theta) = \frac{\gamma_{rr'}(\boldsymbol{h};\Theta)}{\sqrt{\gamma_{rr}(\boldsymbol{h};\Theta)}\sqrt{\gamma_{r'r'}(\boldsymbol{h};\Theta)}} \qquad \in [-1, 1].$$

El intervalo es válido mientras se usen los mismos datos para todos. El coeficiente de codispersión es interpretado como el coeficiente de correlación entre los incrementos y representa la pendiente de la regresión lineal de $(Y_{r'}(\boldsymbol{s}+\boldsymbol{h}) - Y_{r'}(\boldsymbol{s}))$ en términos de $(Y_r(\boldsymbol{s}+\boldsymbol{h}) - Y_r(\boldsymbol{s}))$.

$$\frac{\gamma_{rr'}(\boldsymbol{h};\Theta)}{\gamma_r(\boldsymbol{h};\theta)}.$$

5.4. Modelo lineal de corregionalización, MLC

La matriz de covarianza en el caso del cokriging es una matriz de definida positiva que contiene las matrices de covarianzas individuales y cruzadas de todos los pares de variables incluidas, en todos los sitios donde estas fueron observadas. Para $r, r' = 1, \ldots, P$; $i = 1, \ldots, n_r$ y $j = 1, \ldots, n_{r'}$

$$\sum = (\Sigma_{rr'}) = \left(\text{Cov}[Y_r(\boldsymbol{s}_i), Y_{r'}(\boldsymbol{s}_j)]\right) \tag{5.17}$$

Es una generalización del modelo lineal de regionalización presentado en la Sección 3.7 [37], involucrando los modelos de covarianza directos y cruzados. Los modelos no pueden ser construidos independientemente uno de otro, pues se tiene que garantizar que la matriz $\sum$ presentada en (5.17) sea definida positiva.

En términos prácticos el modelo lineal de corregionalización es una combinación lineal de L modelos válidos de covarianza (semivarianza) cuyos coeficientes son matrices definidas positivas de dimensión $P \times P$.

Los subíndices l y k se usan respectivamente, para el número de modelos incluidos y el número de campos aleatorios independientes en los que se hace la descomposición. Esta indexación es necesaria dado que la covarianza de un campo aleatorio puede requerir mas de un modelo o varios campos aleatorios pueden compartir el mismo modelo de covarianza. De hecho, cada campo aleatorio tiene su propio modelo lineal de regionalización, y este se usa para construir el modelo lineal de corregionalización. Acerca de los $Z_k^l(\boldsymbol{s})$, solo es necesario conocer su estructura de covarianza (semivarianza), esto evidencia su existencia y explícitamente no se necesita mas información de estos campos aleatorios.

Ejemplo 13 (Modelo lineal de corregionalización). *Sean $Y_1(\boldsymbol{s})$ y $Y_2(\boldsymbol{s})$ los campos aleatorios originales que se van a modelar. Ambos se construyen como combinaciones lineales del mismo conjunto de campos aleatorios independientes $k = 1, 2$ y con tres estructuras de semivarianza $l = 0, 1, 2$ dadas por $\gamma_0(\boldsymbol{h}; \boldsymbol{\theta}_0)$, un modelo efecto pepita, $\gamma_1(\boldsymbol{h}; \boldsymbol{\theta}_1 = (1, 7))$ un modelo esférico de rango 7 y $\gamma_2(\boldsymbol{h}; \boldsymbol{\theta}_2 = (1, 3))$ un modelo exponencial de rango 3.*

$$\left\{Z_k^l(\boldsymbol{s}), k = 1, 2, \quad l = 0, 1, 2\right\}$$

con medias constantes μ_1 y μ_2. Los procesos originales expresados en los campos aleatorios independientes y centrados son

$$Y_1(\boldsymbol{s}) = b_{11}^0 Z_1^0(\boldsymbol{s}) + b_{12}^0 Z_2^0(\boldsymbol{s}) + b_{11}^1 Z_1^1(\boldsymbol{s}) + b_{12}^1 Z_2^1(\boldsymbol{s}) + b_{11}^2 Z_1^2(\boldsymbol{s}) + b_{12}^2 Z_2^2(\boldsymbol{s}) + \mu_1.$$
$$Y_2(\boldsymbol{s}) = b_{21}^0 Z_1^0(\boldsymbol{s}) + b_{22}^0 Z_2^0(\boldsymbol{s}) + b_{21}^1 Z_1^1(\boldsymbol{s}) + b_{22}^1 Z_2^1(\boldsymbol{s}) + b_{21}^2 Z_1^2(\boldsymbol{s}) + b_{22}^2 Z_2^2(\boldsymbol{s}) + \mu_2.$$

En general, estas constantes se deducen de la silla o pendiente de los modelos de covarianza o semivarianza. Asumiendo unas constantes arbitrarias b_{rk}^l se ilustra la descomposición a continuación.

$$Y_1(s) = 5Z_1^0(\boldsymbol{s}) + 0Z_2^0(\boldsymbol{s}) + 2Z_1^1(\boldsymbol{s}) + 4Z_2^1(\boldsymbol{s}) + 1.6Z_1^2(\boldsymbol{s}) + 0Z_2^2(\boldsymbol{s}) + \mu_1.$$
$$Y_2(\boldsymbol{s}) = 0Z_1^0(\boldsymbol{s}) + 3Z_2^0(\boldsymbol{s}) + 5Z_1^1(\boldsymbol{s}) + 2Z_2^1(\boldsymbol{s}) + 0.7Z_1^2(\boldsymbol{s}) + 1Z_2^2(\boldsymbol{s}) + \mu_2.$$

Por construcción $b_{ij}^l = b_{ji}^l$. Para este caso el modelo lineal de corregionalización en notación matricial es:

$$\begin{pmatrix} Y_1(\boldsymbol{s}) & Y_2(\boldsymbol{s}) \end{pmatrix} \begin{pmatrix} Y_1(\boldsymbol{s}) \\ Y_2(\boldsymbol{s}) \end{pmatrix} = \begin{pmatrix} b_{11}^0 & b_{12}^0 \\ b_{21}^0 & b_{22}^0 \end{pmatrix} \begin{pmatrix} Z_1^0(\boldsymbol{s}) \\ Z_2^0(\boldsymbol{s}) \end{pmatrix} + \begin{pmatrix} b_{11}^1 & b_{12}^1 \\ b_{21}^1 & b_{22}^1 \end{pmatrix} \begin{pmatrix} Z_1^1(\boldsymbol{s}) \\ Z_2^1(\boldsymbol{s}) \end{pmatrix}$$
$$+ \begin{pmatrix} b_{11}^2 & b_{12}^2 \\ b_{21}^2 & b_{22}^2 \end{pmatrix} \begin{pmatrix} Z_1^2(\boldsymbol{s}) \\ Z_2^2(\boldsymbol{s}) \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad (5.18)$$

$$\mathbb{Y}(\boldsymbol{s}) = B_0\mathbb{Z}^0(\boldsymbol{s}) + B_1\mathbb{Z}^1(\boldsymbol{s}) + B_2\mathbb{Z}^2(\boldsymbol{s}) + \boldsymbol{\mu}.$$

B_0, B_1 y B_2 son semidefinidas positivas. A continuación, los semivariogramas resultantes $\gamma_{11}(.)$ y $\gamma_{22}(.)$, y el semivariograma cruzado $\gamma_{12}(.)$ son definidos como sumas ponderadas de los 3 modelos básicos $\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0)$, $\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1)$ y $\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2)$, según el MLC dado en (13).

$$\begin{aligned}
\gamma_{11}(h;\Theta_1) &= d_{11}^0\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + d_{11}^1\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + d_{11}^2 1\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2) \\
\gamma_{22}(h;\Theta_2) &= d_{22}^0\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + d_{22}^1\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + d_{22}^2\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2) \\
\gamma_{12}(h;\Theta_{12}) &= d_{12}^0\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + d_{12}^1\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + d_{12}^2\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2).
\end{aligned} \tag{5.19}$$

con coeficientes dados por

$$\begin{aligned}
d_{11}^0 &= b_{11}^0 b_{11}^0 + b_{12}^0 b_{12}^0 = 25 & d_{11}^1 &= b_{11}^1 b_{11}^1 + b_{12}^1 b_{12}^1 = 20 \\
d_{11}^2 &= b_{11}^2 b_{11}^2 + b_{12}^2 b_{12}^2 = 2.56 & d_{22}^0 &= b_{21}^0 b_{21}^0 + b_{22}^0 b_{22}^0 = 9 \\
d_{22}^1 &= b_{21}^1 b_{21}^1 + b_{22}^1 b_{22}^1 = 29 & d_{22}^2 &= b_{21}^2 b_{21}^2 + b_{22}^2 b_{22}^2 = 1.49 \\
d_{12}^0 &= b_{11}^0 b_{21} + b_{12}^0 b_{22}^0 = 0 & d_{12}^1 &= b_{11}^1 b_{21}^1 + b_{12}^1 b_{22}^1 = 18 \\
d_{12}^2 &= b_{11}^2 b_{21}^2 + b_{12}^2 b_{22}^2 = 1.12
\end{aligned}$$

El modelo definitivo es

$$\begin{aligned}
\gamma_{11}(\boldsymbol{h};\Theta_1) &= 25\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + 20\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + 2.56\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2). \\
\gamma_{22}(\boldsymbol{h};\Theta_2) &= 9\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + 29\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + 1.49\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2). \\
\gamma_{12}(\boldsymbol{h};\Theta_{12}) &= 18\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + 1.12\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2).
\end{aligned} \tag{5.20}$$

Dado que $Z_1^0(\boldsymbol{s})$ y $Z_2^0(\boldsymbol{s})$ son independientes, el semivariograma cruzado no tiene pepita. Estos coeficientes son los elementos de las matrices de corregionalización D_0, D_1, D_2, las cuales son semidefinidas positivas y además son las sillas de los respectivos modelos. El modelo final para este caso particular es

$$\begin{pmatrix} \gamma_{11}(\boldsymbol{h};\Theta_1) & \gamma_{12}(\boldsymbol{h};\Theta_{12}) \\ \gamma_{21}(\boldsymbol{h};\Theta_{21}) & \gamma_{22}(\boldsymbol{h};\Theta_2) \end{pmatrix} = \begin{pmatrix} d_{11}^0 & d_{12}^0 \\ d_{12}^0 & d_{22}^0 \end{pmatrix}\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + \begin{pmatrix} d_{11}^1 & d_{12}^1 \\ d_{12}^1 & d_{22}^1 \end{pmatrix}\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + \begin{pmatrix} d_{11}^2 & d_{12}^2 \\ d_{12}^2 & d_{22}^2 \end{pmatrix}\gamma_2(h;\boldsymbol{\theta}_2).$$

$$\Gamma(\boldsymbol{h}) = D_0\gamma_0(\boldsymbol{h};\boldsymbol{\theta}_0) + D_1\gamma_1(\boldsymbol{h};\boldsymbol{\theta}_1) + D_2\gamma_2(\boldsymbol{h};\boldsymbol{\theta}_2),$$

Las condiciones suficientes para que (5.15) constituya un modelo de corregionalización válido son:

1. Las funciones $C_l(\boldsymbol{h};\boldsymbol{\theta}_l)$ son modelos válidos.
2. Las matrices B_l son todas semidefinidas positivas. Cada modelo puede tener su propio rango y su propio patrón de anisotropía.

3. En términos de semivariogramas, el MLC $\gamma_{rr'}(\boldsymbol{h};\Theta)$ está definido como:

$$E\left[Z_k^l(\boldsymbol{s}+\boldsymbol{h})-Z_k^l(\boldsymbol{s})\right]\left[Z_{k'}^{l'}(\boldsymbol{s}+\boldsymbol{h})-Z_{k'}^{l'}(\boldsymbol{s})\right]=\begin{cases}2\gamma_l(\boldsymbol{h};\boldsymbol{\theta}_l) & \text{Si } k=k',\ l=l';\\ 0 & \text{En otro caso.}\end{cases}$$

$$\gamma_{rr'}(\boldsymbol{h};\Theta)=\sum_{l=0}^{L} d_{rr'}^l\gamma_l(\boldsymbol{h};\boldsymbol{\theta}_l)\qquad \forall r,r'=1,\ldots,P.$$

Los $d_{rr'}^l$ son las sillas para modelos acotados en caso de estacionariedad de segundo orden o también pueden ser las pendientes de los $\gamma_l(\boldsymbol{h};\boldsymbol{\theta}_l)$ en los casos de modelos no acotados bajo estacionariedad intrínseca.

Si fuera un vector aleatorio espacial $\mathbb{Y}=(Y_1(\boldsymbol{s}),Y_2(\boldsymbol{s}),Y_3(\boldsymbol{s}))$ con L modelos de semivarianza el modelo quedaría

$$\begin{bmatrix}\gamma_1(\boldsymbol{h};\Theta_1) & \gamma_{12}(\boldsymbol{h};\Theta_{12}) & \gamma_{13}(\boldsymbol{h};\Theta_{13})\\ \gamma_{21}(\boldsymbol{h};\Theta_{21}) & \gamma_2(\boldsymbol{h};\Theta_2) & \gamma_{23}(\boldsymbol{h};\Theta_{23})\\ \gamma_{31}(\boldsymbol{h};\Theta_{31}) & \gamma_{32}(\boldsymbol{h};\Theta_{32}) & \gamma_3(\boldsymbol{h};\Theta_3)\end{bmatrix}=\sum_{l=0}^{L}\begin{bmatrix}d_{11}^l & d_{12}^l & d_{13}^l\\ d_{21}^l & d_{22}^l & d_{23}^l\\ d_{31}^l & d_{32}^l & d_{33}^l\end{bmatrix}\gamma_l(\boldsymbol{h};\boldsymbol{\theta}_l).$$

Las matrices D_l deben ser semidefinidas positivas. Los elementos de la diagonal son las sillas de los semivariogramas individuales o directos de cada modelo y por lo tanto son no negativos. Ver [50] y [37] para un tratamiento detallado definición y propiedades de las matrices semidefinidas positivas. La suma total de todas las ubicaciones rr en las L matrices es La silla total de la variable espacial $Y_r(\boldsymbol{s})$.

5.5. Aspectos prácticos del cokriging

- Algunos de los campos Z_k^l pueden tener la misma estructura de covarianza y aun así ser independientes.
- No se requiere que los semivariogramas directos o cruzados involucren todas las estructuras básicas. Algunos $b_{kk'}^l$ pueden ser 0.
- Por construcción, $d_{rr'}^l$ y $d_{r'r}^l$ son idénticos, por lo tanto $C_{rr'}(\boldsymbol{h})=C_{r'r}(\boldsymbol{h})$
- El cokriging usa todo el vector $\mathbb{Y}$ para llevar a cabo la predicción de cada una de las variables $Y_r(\boldsymbol{s}_0)$, una a la vez, y luego con estas predicciones se forma el vector de predicción en el lugar $\boldsymbol{s}_0$. Por lo tanto, no es estrictamente una predicción multivariada.

  Se han buscado alternativas para la predicción simultánea pero esto ha resultado muy complejo y poco eficiente. [72] propone una metodología para la predicción multivariada simultánea de todas las variables espaciales, pero requiere una condición especial de simetría, $C_{rr'}(\boldsymbol{s}_i-\boldsymbol{s}_{i'})=C_{r'r}(\boldsymbol{s}_i-\boldsymbol{s}_{i'})$, que es demasiado restrictiva. [72] muestra que la calidad de la predicción es mejor sin la restricción de simetría y usando el variograma cruzado.

Así, el cokriging basado en el MLC sigue siendo el método más utilizado para la predicción espacial cuando existe la posibilidad de usar covariables espaciales.

- [55] presenta matricialmente el desarrollo detallado del cokriging para varios casos. Esta notación es compacta y muy práctica. Para P variables observadas en n_r sites, $r = 1, \ldots, P$, la matriz de covarianza

$$\sum = \text{Var}(Z_1(\boldsymbol{s}_1), \ldots, Z_1(\boldsymbol{s}_{n_1}), \ldots, , Z_1(\boldsymbol{s}_{nP}), \ldots, Z_P(\boldsymbol{s}_{nP})).$$

de dimensión $\sum_{r=1}^{P} n_r \times \sum_{r=1}^{P} n_r$. Λ_i para cada $\boldsymbol{s}_0$ es una matriz $P \times P$, formada por las ponderaciones $\Lambda_i^{rr'}$, que representan la contribución de la $r-$ésima variable en la ubicación $\boldsymbol{s}_i$ a la predicción de la $r'-$ésima variable.

$$\Lambda_i = \begin{pmatrix} \lambda_{11}^i & \lambda_{12}^i & \cdots & \lambda_{1P}^i \\ \lambda_{21}^i & \lambda_{22}^i & \cdots & \lambda_{2P}^i \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{P1}^i & \lambda_{P2}^i & \cdots & \lambda_{PP}^i \end{pmatrix}. \tag{5.21}$$

Finalmente, el mejor predictor lineal insesgado se encuentra minimizando cualquiera de las expresiones en (5.22). La suma es mas tratable computacionalmente.

$$\text{máx}\, \text{Var}\left[Z_r(\boldsymbol{s}_0) - Z_r^*(\boldsymbol{s}_0)\right] \tag{5.22}$$

o

$$\sum \text{Var}\left[Z_r(\boldsymbol{s}_0) - Z_r^*(\boldsymbol{s}_0)\right].$$

Ejemplo 14 (Modelo lineal de corregionalización para los contaminantes PM10 y NOX Bogotá). *Se analizan los datos de la red para la calidad del aire en la Ciudad de México durante la temporada seca, dado que en la temporada de lluvias todos los contaminantes del aire disminuyen. La red de calidad del aire, monitorea cada hora material particulado de hasta 10 micrómetros de tamaño (PM10) y óxido de nitrógeno (NOX), entre otros. El material particulado (PM) es un componente importante de la contaminación del aire. El NOX es un contaminante gaseoso del aire, producido por el tráfico y otros procesos de combustión de combustibles fósiles y contribuye a la formación y modificación de otros contaminantes del aire como el material particulado. Además, tanto el PM10 como el NOX dañan los materiales y son las principales causas de la reducción de la visibilidad (SEDEMA).*

Se ha aplicado el método de cokriging simple para estos dos contaminantes usando las observaciones del 15 de abril 2024, a las 8am.

Ninguno de los dos procesos muestra tendencia, así que en ambos casos se centraron con su media aritmética. A partir de los procesos centrados se lleva a cabo la estimación de los semivariogramas empíricos y teóricos. El MLC para los contaminantes NOX y PM10 está dado por:

$$\begin{bmatrix} 17{,}9993193 & 10{,}962510 \\ 10{,}962510 & 6{,}855200 \end{bmatrix} \gamma_1(\boldsymbol{h}, \Theta_1) + \begin{bmatrix} 22{,}3786663 & 27{,}179734 \\ 27{,}179734 & 55{,}560817 \end{bmatrix} \gamma_2(\boldsymbol{h}, \Theta_2).$$

Dónde $\gamma_1(\boldsymbol{h}, \Theta_1)$ es un modelo efecto hueco, con rango 2322,42 y $\gamma_2(\boldsymbol{h}, \Theta_2)$ es un modelo gaussiano con rango 20180,25. Las matrices contienen los valores de las sillas para cada modelo. Los tres modelos son:

$$\begin{aligned} \gamma_{NOX}(h) &= 17{,}9993193\gamma_1(\boldsymbol{h}, \Theta_1) + 22{,}3786663\gamma_2(\boldsymbol{h}, \Theta_2) \\ \gamma_{PM10}(h) &= 6{,}855200\gamma_1(\boldsymbol{h}, \Theta_1) + 55{,}560817\gamma_2(\boldsymbol{h}, \Theta_2) \\ \gamma_{NOX,PM10}(h) &= 10{,}962510\gamma_1(\boldsymbol{h}, \Theta_1) + 27{,}179734\gamma_2(\boldsymbol{h}, \Theta_2) \end{aligned}$$

La Figura 5.1 muestra el MLC ajustado, así como los dos semivariogramas empíricos para cada una de las variables NOX y PM10 y el semivariograma empírico cruzado de NOX y PM10. Posteriormente, usando la matriz de covarianza, se vuelve a estimar la media de cada proceso, por MCG, para sumarla a los predictores cokriging de los respectivos procesos residuales.

Los resultados de la predicción para ambas variables, usando cokriging simple sobre los residuales y sumándole las medias correspondientes en cada caso, se muestran en la Figura 5.2. Cada mapa de predicción va acompañado del mapa de la varianza del error de predicción respectivo.

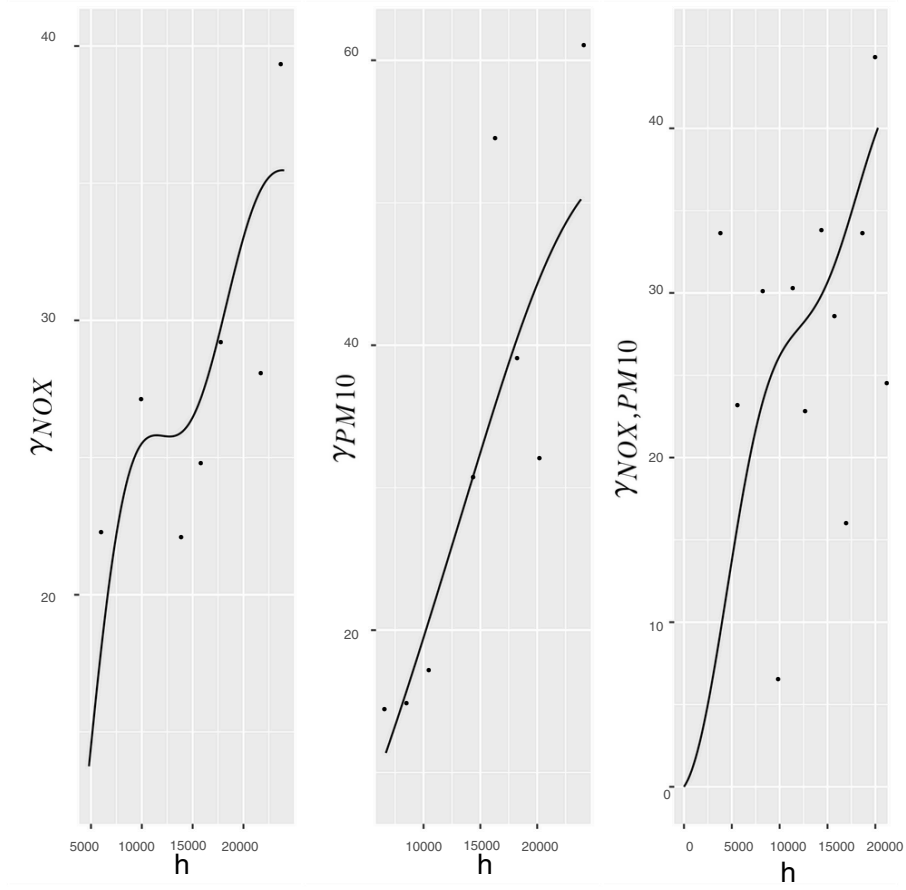


Figura 5.1. Modelo lineal de corregionalización para los contaminantes NOX y PM10. 15 de abril 2024, 8am.

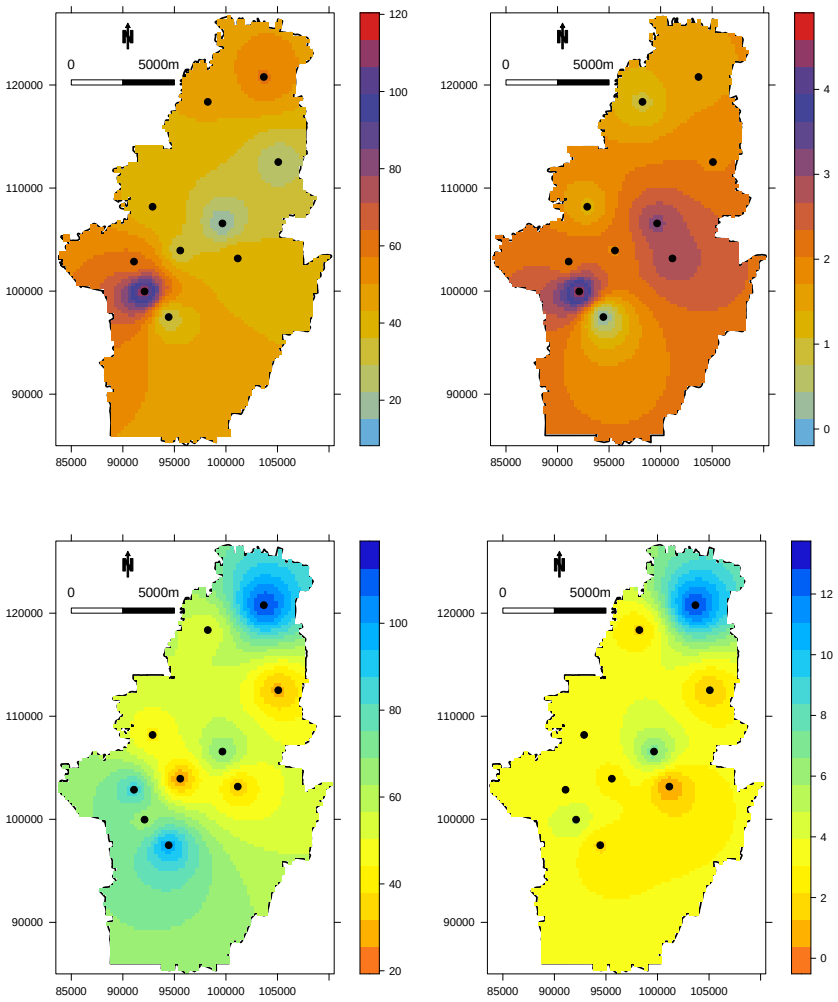


Figura 5.2. Mapas de las predicciones obtenidas utilizando cokriging. Panel superior izquierdo: Predicción del PM10 utilizando NOX como covariable. Panel superior derecho: Mapa de la varianza del error de predicción del cokriging para PM10. Panel inferior izquierdo: Predicción del NOX utilizando PM10 como covariable. Panel superior derecho: Mapa de la varianza del error de predicción del cokriging para NOX. Las varianzas en los puntos observados son las obtenidas por validación cruzada en ambos casos.

5.6. Enfoque jerárquico bajo normalidad, caso bivariado

La especificación de un modelo paramétrico válido para la covarianza de $\mathbb{Y}$, denotada $\sum$, se resuelve en la mayoría de los casos con el modelo lineal de corregionalización [37], el cual supone simetría. Sin embargo, bajo el supuesto de normalidad, [66] propone usar la distribución condicional, basada en los dos primeros momentos de la distribución conjunta. Cuando existe un conocimiento profundo del fenómeno, este modelo permite encontrar la estructura de dependencia involucrada en cada etapa, mediante el enfoque jerárquico.

Sea $(Y_1(\boldsymbol{s}), Y_2(\boldsymbol{s}))$; $i, i' = 1, \ldots, n$ un campo aleatorio bivariado . No necesariamente tienen que ser medidas ambas variables en los mismos sitios. El enfoque jerárquico se basa en la especificación de los primeros dos momentos de las distribuciones.

$$Y_1|Y_2 \quad \text{y} \quad Y_2.$$

Si la esperanza condicional es lineal , entonces la primera etapa del modelo es

$$\boldsymbol{Y}_1|\boldsymbol{Y}_2 \sim \left(\boldsymbol{\mu}_1 + \boldsymbol{B}(\boldsymbol{Y}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{1|2}\right).$$

$\boldsymbol{B}$ es una $m \times m$-matriz de coeficientes de regresión, y parametrizar la dependencia entre $\boldsymbol{Z}_1$ y $\boldsymbol{Z}_2$, usando la matriz $\boldsymbol{B}$, permite un modelamiento flexible. $\boldsymbol{\Sigma}_{1|2}$ es una matriz condicionalmente definida positiva de la covarianza de $\boldsymbol{Y}_1|\boldsymbol{Y}_2$, y

$$\boldsymbol{Y}_2 \sim (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}).$$

donde $\boldsymbol{\Sigma}_{22}$ es una matriz de covarianza definida positiva. Las medias y las covarianzas conjuntas entre $\boldsymbol{Y}_1$ y $\boldsymbol{Y}_2$ bajo esta especificación son

$$\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{pmatrix} = \left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}'_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right). \tag{5.23}$$

$$\begin{pmatrix}\boldsymbol{Y}_1\\ \boldsymbol{Y}_2\end{pmatrix} = \left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{1|2} + \boldsymbol{B}\boldsymbol{\Sigma}_{22}\boldsymbol{B}' & \boldsymbol{B}\boldsymbol{\Sigma}_{22}\\ (\boldsymbol{B}\boldsymbol{\Sigma}_{22})' & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right). \tag{5.24}$$

La formulación jerárquica se basa en la especificación y estimación de $\boldsymbol{\Sigma}_{1|2}$ y $\boldsymbol{B}$, que son matrices de dimensión $n \times n$. Las dependencias cruzadas son capturadas en la distribución $[\boldsymbol{Y}_1|\boldsymbol{Y}_2]$. La especificación de un modelo marginal y condicional implica la estructura de covarianza cruzada. El paso final es especificar $\boldsymbol{\Sigma}_{22}$, que es también una matriz de dimensión $n \times n$. El modelo conjunto obtenido es válido por construcción. En este caso, la matriz $2n \times 2n$ es definida positiva y, en general, la matriz $Pn \times Pn$ también lo es.

5.7. Ejercicios

1. Suponga dos campos aleatorios espaciales correlacionados. Encuentre, si existe, alguna relación entre la varianza de predicción del kriging simple y la del cokriging simple. Si no existe ninguna relación, justifique su respuesta.
2. Demuestre o refute para cada caso. Todas las clases de kriging y cokriging son predictores exactos.
3. Se tienen 2 procesos espaciales $Z_1(\boldsymbol{s})$ y $Z_2(\boldsymbol{s})$ autocorrelacionados, pero no correlacionados entre ellos. Escriba el predictor cokriging con su respectiva varianza para predecir Z_1 en un lugar no observado. Interprete los resultados.
4. Determine si los siguientes son modelos válidos de corregionalización.

   4.1

$$\begin{aligned}\gamma_1(\boldsymbol{h}) &= 16\text{Exponencial}(a = 20) + 4\text{Esférico}(a = 100)\\ \gamma_2(\boldsymbol{h}) &= 25\text{Gaussiano}(a = 20)\\ \gamma_{12}(\boldsymbol{h}) &= 12\text{Matern}(a = 55)\end{aligned}$$

   4.2

$$\begin{aligned}\gamma_1(\boldsymbol{h}) &= 4\text{Esférico}(a = 10)\\ \gamma_2(\boldsymbol{h}) &= 25\text{Esférico}(a = 4)\\ \gamma_{12}(\boldsymbol{h}) &= 12\text{Esférico}(a = 7)\end{aligned}$$

   4.3

$$\begin{aligned}\gamma_1(\boldsymbol{h}) &= 36 + 64\text{Cauchy}(a = 520)\\ \gamma_2(\boldsymbol{h}) &= 36 + 64\text{ExpAmort}(a = 220)\\ \gamma_{12}(\boldsymbol{h}) &= 38.4\text{Cos}(a = 100)\end{aligned}$$

5. Falso o verdadero, demuestre o refute. Se asume estacionariedad de segundo orden.

   5.1 $\gamma_{rr'}(\boldsymbol{h}) = \gamma_{rr'}(-\boldsymbol{h})$

   5.2 $\gamma_{rr'}(\boldsymbol{h}) = C_{rr'}(\mathbf{0}) - \frac{1}{2}\left(C_{rr'}(-\boldsymbol{h})C_{rr'}(+\boldsymbol{h})\right)$

   5.3 $\gamma_{rr'}(\boldsymbol{h}) + \gamma_{rr'}(-h) = \gamma_{rr'}\mathbf{0}) + 2\gamma_{rr'}(\boldsymbol{h})$

   5.4 $2\gamma_{rr'}(\boldsymbol{h}) = C_r(0) + C_{r'}(0) - 2C_{rr'}(\boldsymbol{h})$

   5.5 $\gamma_{rr}(\boldsymbol{h})\gamma_{r'r'}(\boldsymbol{h}) \geq |\ \gamma_{rr'}(\boldsymbol{h})\ |^2$

6. Considere los siguientes campos aleatorios: $Z_1(\boldsymbol{s})$ y $Z_2(\boldsymbol{s}) = Z_1(\boldsymbol{s}+\boldsymbol{h})$. ¿Existe alguna distancia para la cual sea máxima la covarianza cruzada? Justifique en cualquier caso.

Parte II

Geoestadística escalar espacio-temporal

Capítulo

seis

Predicción escalar espacio-temporal: Kriging espacio-temporal

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

En este capítulo se aplican los métodos vistos en el caso espacial, al caso espacio-tiempo. Las ideas son muy similares, pero ganan complejidad debido a que se adicionan dos elementos fundamentales: la dimensión temporal y la interacción espacio-temporal. Sin embargo, en el espacio-tiempo sigue existiendo el interés en modelar la covarianza o la semivarianza y hacer predicciones. Las predicciones en el espacio permiten la generación de mapas, así que con los métodos presentados en este capítulo se pueden ver series de mapas a lo largo del tiempo, lo cual es muy atractivo. De todos modos, es importante notar que estos métodos no son de extrapolación; los métodos geoestadísticos son de interpolación y son descriptivos, así que no son diseñados particularmente con fines de pronóstico. Los métodos presentados hasta este capítulo funcionan bien para cantidades moderadas de datos. Cuando la cantidad de datos aumenta notoriamente y desborda estas técnicas, es necesario usar geoestadística para datos funcionales. Estos métodos se presentan en el Capítulo 9.

6.1. Conceptos preliminares

Definición 9 (Proceso geoestadístico espacio-temporal). *Un proceso geoestadístico espacio-temporal es un proceso estocástico*

$$\{Y(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D_{\boldsymbol{s}} \times D_T\}$$

donde el conjunto $D_{\boldsymbol{s}} \subset \mathbb{R}^d$ es el conjunto índice de la ubicación espacial $\boldsymbol{s}$ y el conjunto $D_T \subset \mathbb{R}$ es el conjunto índice del tiempo t; es decir,

$$(\boldsymbol{s}, t) \in D_{\boldsymbol{s}} \times D_T \in \mathbb{R}^d \times \mathbb{R}$$

$D_{\boldsymbol{s}} \times D_T \subset \mathbb{R}^d \times \mathbb{R}$ es el conjunto índice del proceso espacio-tiempo. Los conjuntos $D_{\boldsymbol{s}}$ y D_T son continuos en el contexto geoestadístico.

La Tabla 6.1 ilustra cómo se ven en general los registros de un proceso espacio-tiempo observado en n ubicaciones espaciales. En cada ubicación espacial, la longitud de la serie de tiempo puede ser diferente, y en un momento determinado, los lugares con registro no necesariamente son los mismos, aunque pueden existir coincidencias. Esto es consecuencia de que se reúne información de diferentes fuentes, o existen lugares donde no se encuentra el registro del dato, o se eliminó porque el dato era de mala calidad. Por ejemplo, si los datos provienen de estaciones que miden variables relacionadas con el clima o con el medio ambiente, alguna estación puede presentar daños y se deben sacar los datos en una fase de validación, porque la estación queda fuera de control. Incluso mientras se lleva a cabo la reparación o reemplazo de la estación, esta puede durar un largo tiempo sin registrar datos.

Además, nótese que no en todos los lugares donde se miden los datos se tiene la misma frecuencia. Si se unen datos con diferente fuente, es posible que la frecuencia temporal difiera. De igual manera ocurre si se tienen distintas variables.

Por ejemplo, el brillo solar no suele medirse tan densamente en el espacio ni en el tiempo como la temperatura. En estos contextos, la cantidad de los datos es un factor de suma importancia y que debe ser considerado en el momento de diseñar los análisis. Si una variable se mide cada hora durante muchos años, la longitud de las series de tiempo no permite el uso de varios modelos debido a la dificultad computacional que genera este volumen de información en los procesos de optimización. En el caso de series de tiempo de imágenes, el problema es el mismo, pero los datos en este caso están densamente medidos en el espacio y no en el tiempo, debido a la alta resolución de muchas imágenes.

t; $\boldsymbol{s}$	$\boldsymbol{s}_1 = (\mathrm{x}_1, \mathrm{y}_1)$	...	$\boldsymbol{s}_j = (\mathrm{x}_j, \mathrm{y}_j)$	...	$\boldsymbol{s}_n = (\mathrm{x}_n, \mathrm{y}_n)$
1	$y(\boldsymbol{s}_1, 1)$	...	$y(\boldsymbol{s}_j, 1)$	...	$y(\boldsymbol{s}_n, 1)$
2	$y(\boldsymbol{s}_1, 2)$	...		...	$y(\boldsymbol{s}_n, 2)$
⋮	⋮	⋮	⋮	⋮	⋮
t_n		...	$y(\boldsymbol{s}_j, t_n)$	...	$y(\boldsymbol{s}_n, t_n)$
⋮	⋮	⋮	⋮	⋮	
t_1	$y(\boldsymbol{s}_1, t_1)$	...	$y(\boldsymbol{s}_j, t_1)$	⋮	
⋮		⋮	⋮	⋮	
t_j		...	$y(\boldsymbol{s}_j, t_j)$	...	

Tabla 6.1. Ilustración de notación para las coordenadas de un proceso espacio-tiempo. $s \in \mathbb{R}^2$, $t \in \mathbb{R}$. $Y(s_i; t_{ij})$; $j = 1, ..., T_i$, $i = 1, ..., n$. Debido a la gran cantidad de ubicaciones espacio-tiempo, es usual que existan datos faltantes o datos erróneos, y que las series de tiempo en cada sitio puedan tener diferente cantidad de datos.

Uno de los objetivos de la geoestadística espacio-temporal es la predicción óptima del valor de la variable aleatoria $Y(\boldsymbol{s_0}, t_0)$, en la coordenada espacio-tiempo $(\boldsymbol{s_0}, t_0)$ con base en el vector

$$\mathbb{Y} = \big(Y(\boldsymbol{s_1}, t_1), \ldots, Y(\boldsymbol{s_N}, t_N)\big)$$

observado en un número finito N de coordenadas espacio-tiempo $(\boldsymbol{s_1}, t_1), \ldots, (\boldsymbol{s_N}, t_N)$.

Las N observaciones espacio-tiempo corresponden a $N = nT$, donde n es la cantidad de ubicaciones espaciales y T es la longitud de la serie de tiempo. Usualmente, no se cuenta con la misma cantidad de observaciones en todas las ubicaciones.

Una notación mas general es

$$\mathbb{Y} = \big(Y(s_i, t_{ij})\big)\,;\ j = 1, \ldots, T_i,\ \ i = 1, \ldots, n.$$

El total de observaciones en este caso es $N = \sum_{i=1}^{n} T_i$.

6.2. Kriging espacio-tiempo

Dado que el modelo geoestadístico es

$$Y(\boldsymbol{s_0}; t_0) = \mu(\boldsymbol{s_0}; t_0) + Z(\boldsymbol{s_0}; t_0)(\mathbf{Y} - \boldsymbol{\mu}) \tag{6.1}$$

De igual forma que en el caso espacial, el predictor lineal óptimo para $Y(\boldsymbol{s_0}; t_0)$, conocido como kriging simple espacio tiempo, está dado por el predictor kriging de la variable asociada centrada mas su media, esto es,

$$Y^*(\boldsymbol{s_0}; t_0) = \mu(\boldsymbol{s_0}; t_0) + \boldsymbol{\sigma}_0'\Sigma^{-1}(\mathbb{Y} - \boldsymbol{\mu}) \tag{6.2}$$

dónde $\Sigma = \text{Cov}(\mathbb{Y})$, $\boldsymbol{\sigma}_0 = \text{Cov}(\mathbb{Y}, Y(\boldsymbol{s_0}; t_0))$, $\boldsymbol{\mu} \equiv E[\mathbb{Y}]$ y $\mathbb{Z} = \mathbb{Y} - \boldsymbol{\mu}$. $\boldsymbol{\lambda}' = \boldsymbol{\sigma}_0'\Sigma^{-1}\mathbb{Z}$. Por lo tanto, la varianza del predictor $Y^*(\boldsymbol{s_0}; t_0)$ es

$$\boldsymbol{\sigma}_0'\Sigma^{-1}\boldsymbol{\sigma}_0$$

y la varianza del kriging simple espacio tiempo o varianza del error de predicción está dada por

$$\text{Var}[Y(\boldsymbol{s_0}; t_0) - Y^*(\boldsymbol{s_0}; t_0)] = \sigma_Y^2 - \boldsymbol{\sigma}_0'\Sigma^{-1}\boldsymbol{\sigma}_0$$

dado que $Y(\boldsymbol{s_0}; t_0)$ sea estacionario de segundo orden [25]. Nótese que las expresiones son idénticas a las obtenidas para el caso espacial, solo que ahora se cuenta con la coordenada adicional del tiempo. Así se observan el proceso espacial y el temporal como casos particulares del espacio-tiempo. A continuación, se extienden las definiciones de estacionariedad y de funciones definidas positivas al caso espacio-tiempo.

Estacionariedad de segundo orden El proceso $\{Y(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D \times T\}$ es, al menos, estacionario de segundo orden; esto es,

$$E[Y(\boldsymbol{s}, t)] = \boldsymbol{\mu} \quad \text{y} \quad \text{Cov}[Y(\boldsymbol{s_i}, t), Y(\boldsymbol{s_j}, t')] = C(\boldsymbol{s_i} - \boldsymbol{s_j}; t - t'; \Theta)$$

dado que $\text{Var}[Y(\boldsymbol{s}; t)] < \infty$. Es decir que el proceso tiene media constante, varianza finita y la función de covarianza depende únicamente de las separaciones o rezagos espacial $\boldsymbol{s_i} - \boldsymbol{s_j}$ y temporal $t - t'$.

6.3. Funciones válidas de covarianza espacio-temporal

Sean $\boldsymbol{h} = \boldsymbol{s_i} - \boldsymbol{s_j}$ y $u = t - t'$. Si el proceso es estacionario de segundo orden, se puede escribir la función de covarianza del proceso espacio-temporal como sigue:

$$\text{Cov}(Y(\boldsymbol{s}, t), Y(\boldsymbol{s}+\boldsymbol{h}, t+u)) = \text{Cov}(Z(\boldsymbol{s}, t), Z(\boldsymbol{s}+\boldsymbol{h}, t+u)) = \boldsymbol{C}(\boldsymbol{h}; u; \Theta)$$

$\boldsymbol{C}$ es la función covarianza del proceso espacio-temporal; $\boldsymbol{h}$ es el rezago espacial y u es el rezago temporal. Θ es el vector de parámetros de la función $\boldsymbol{C}$.

$C(\boldsymbol{h}; u)$ **es una función definida positiva.** La función $\boldsymbol{C}$ debe ser definida positiva para garantizar una varianza de error de predicción no negativa. Esto es, para cualquier número finito m de ubicaciones espacio-temporales

$$(\boldsymbol{s_1}, t_1), (\boldsymbol{s_2}, t_2), \ldots, (\boldsymbol{s_m}, t_m)$$

y cualquier conjunto de números reales, $\{a_1, a_2, \ldots, a_m\}$ con $m \in \mathbb{Z}^+$, $\boldsymbol{C}$ debe satisfacer

$$\sum_{i=1}^{m}\sum_{i=1}^{m} a_i a_j \boldsymbol{C}(\boldsymbol{s_i} - \boldsymbol{s_j}; t - t') \geq 0 \tag{6.3}$$

Para asegurar el cumplimiento de (6.3), con frecuencia se especifica una función de covarianza $\boldsymbol{C}$ que pertenezca a una familia paramétrica $\boldsymbol{C}^0$, en la cual todas las funciones cumplan dicha propiedad; se asume que para todo vector de parámetros Θ

$$\text{Cov}(Z(\boldsymbol{s_i}, t), Z(\boldsymbol{s_j}, t')) = \boldsymbol{C}^0(\boldsymbol{s_i} - \boldsymbol{s_j}; t - t' | \Theta)$$

$\boldsymbol{C}$ es una función definida positiva.

Como en el caso espacial, se debe seleccionar la mejor función de covarianza entre algunas familias, pues la estructura de dicha matriz difícilmente puede resultar conocida a priori.

Para la construcción de más funciones de covarianza se usan con frecuencia las siguientes dos propiedades conocidas como aditiva y multiplicativa, respectivamente [67]:

- Si $\boldsymbol{C}_l(\cdot)$ son modelos de covarianza válidos, con $l = 1, \ldots, k$, entonces $\sum_{l=1}^{k} b_l \boldsymbol{C}_l(\cdot)$ es un modelo de covarianza válido con $b_l \geq 0, \forall l$.
- Si $\boldsymbol{C}_l(\cdot)$ son modelos de covarianza válidos, con $l = 1, \ldots, k$, entonces $\prod_{l=1}^{k} \boldsymbol{C}_j(\cdot)$ es un modelo de covarianza válido.

A continuación, se presenta un supuesto que con frecuencia se hace sobre funciones de covarianza espacio-tiempo para simplificar los procesos de estimación [46].

6.4. Construcción de funciones de covarianza espacio tiempo separables

Con base en las propiedades aditiva y multiplicativa, se definen nuevas familias de modelos de covarianza espacio-temporales, conocidas como separables. Sean $\boldsymbol{C}_s(\boldsymbol{s}_i, \boldsymbol{s}_j|\Theta_1)$ una función definida positiva en $\mathbb{R}^d$ y $\boldsymbol{C}_t(t, t'|\Theta_2)$ una función definida positiva en $\mathbb{R}$, un proceso espacio-temporal tiene función de covarianza separable en cualquiera de los siguientes dos casos, con $\Theta = (\Theta_1, \Theta_2)$

- su función de covarianza se puede descomponer en la suma de un componente puramente espacial y uno puramente temporal:

$$\boldsymbol{C}^0(\boldsymbol{s}_i, \boldsymbol{s}_j; t, t'|\Theta) = \boldsymbol{C}_s(\boldsymbol{s}_i, \boldsymbol{s}_j|\Theta_1)) + \boldsymbol{C}_t(t, t'|\Theta_2)$$

- su función de covarianza se puede descomponer en el producto de un componente puramente espacial y uno puramente temporal:

$$\boldsymbol{C}^0(\boldsymbol{s}_i, \boldsymbol{s}_j; t, t'|\Theta) = \boldsymbol{C}_s(\boldsymbol{s}_i, \boldsymbol{s}_j|\Theta_1)\boldsymbol{C}_t(t, t'|\Theta_2)$$

En la construcción de estos modelos es indispensable tener en cuenta que si $\boldsymbol{C}_l(\cdot)$ es una función de covarianza válida en $\mathbb{R}^d$, es válida en $\mathbb{R}^p$ para $p < d$.

La complejidad del modelamiento de la variación simultánea espacial y temporal de un proceso hace atractiva la idea de separabilidad. Sin embargo, este supuesto implica la ausencia de interacción entre espacio y tiempo, hecho que en general, no coincide con la realidad. En muchos casos, el efecto de un desplazamiento en la ubicación espacial para un fenómeno de interés, no es el mismo para diferentes momentos en el tiempo, lo que provoca que el modelo de dependencia espacial varíe y muestre que existe una interacción que debe ser tenida en cuenta. En el caso de la contaminación ambiental, por ejemplo, debido a factores dinámicos como los automóviles, su variación en dos lugares distintos $\boldsymbol{s}_i$ y $\boldsymbol{s}_j$ depende de la hora a la cual se realice la comparación; esto es, cambios simultáneos en ubicación y tiempo generan diferentes relaciones entre los valores de contaminación. Sin embargo, al usar un modelo separable, no queda incluido este efecto de interacción espacio-tiempo.

Nota 7. *Las propiedades aditiva y multiplicativa no se cumplen cuando alguna de las funciones depende solo de un subconjunto de componentes del vector. [65] presenta un ejemplo en el que luego de modelar de forma separada los componentes espacial y temporal, el modelo resultante de la suma de éstos, no es válido.*

6.5. Construcción de funciones de covarianza espacio tiempo no separables

A continuación se presentan varios métodos para construir familias de covarianza espacio temporales válidas no separables, así como algunos casos particulares de cada uno de ellos.

6.5.1. Método basado en representaciones espectrales de las funciones de covarianza

Con base en este método se construyen familias paramétricas de covarianzas estacionarias, encontrando transformadas inversas de Fourier de densidades espectrales (ver [25]). Este procedimiento no impone restricciones de separabilidad, aunque podrían surgir modelos separables como casos particulares. El procedimiento es el siguiente:

Sea la función $g(\boldsymbol{w}, \tau)$ *que se puede escribir como una transformada escalar de Fourier en* u*, así*

$$g(\boldsymbol{w}, \tau) = (2\pi)^{-1} \int e^{-iu\tau} h(\boldsymbol{w}, u) du \tag{6.4}$$

donde

$$h(\boldsymbol{w}, u) = (2\pi)^{-d} \int e^{-i\boldsymbol{h}'\boldsymbol{w}} C(\boldsymbol{h}, u) d\boldsymbol{h} \tag{6.5}$$

la cual es la transformada inversa escalar de Fourier de $g(\boldsymbol{w}, \tau)$

$$h(\boldsymbol{w}, u) = \int e^{i\tau u} g(\boldsymbol{w}, \tau) d\tau \tag{6.6}$$

Entonces la construcción de la función de covarianza espacio temporal $C(\boldsymbol{h}, u)$ *depende de la selección de la función* $h(\boldsymbol{w}, u)$ *y*

$$C(\boldsymbol{h}, u) = \int e^{i\boldsymbol{h}'\boldsymbol{w}} h(\boldsymbol{w}, u) d\boldsymbol{w}$$

Asumiendo que $h(\boldsymbol{w}, u)$ *se puede escribir como el producto de dos funciones,*

$$h(\boldsymbol{w}, u) = k(\boldsymbol{w}) \rho(\boldsymbol{w}, u)$$

donde $k(\boldsymbol{w})$ *es la densidad espectral de un proceso espacial puro y* $\rho(\boldsymbol{w}, u)$ *una función de autocorrelación temporal para cada* $\boldsymbol{w} \in \mathbb{R}^d$*; que satisfacen:*

1. $\rho(\boldsymbol{w}, .), \ \forall \boldsymbol{w} \in \mathbb{R}^d$ *es una función de autocorrelación temporal continua, con* $\int \rho(\boldsymbol{w}, u) du < \infty$ *y* $k(\boldsymbol{w}) > 0$
2. $\int k(\boldsymbol{w}) d\boldsymbol{w} < \infty$

Se encuentra una fórmula general para la construcción de las funciones de covarianza espacio tiempo así:

$$C(\boldsymbol{h}, u) = \int e^{ih'\boldsymbol{w}} k(\boldsymbol{w}) \rho(\boldsymbol{w}, u) d\boldsymbol{w}$$

Por lo tanto, esta metodología requiere pares de transformadas de Fourier que posean una forma cerrada.

Ejemplo 15 (Función espacio-temporal de Cressie 1, [26]). *Sean las funciones* $\rho(\boldsymbol{w}, u)$ *y* $k(\boldsymbol{w})$

$$\rho(w; u) = e^{-\|w\|^2 u^2/4} e^{-\delta u^2} \quad y \quad k(w) = e^{-c_0\|w\|^2/4}$$

con $c_0 > 0$. *La función de covarianza espacio-temporal resultante es*

$$C(h; u) \alpha \frac{1}{(c_0 + u^2)^{d/2}} e^{-\frac{h^2}{c_0+u^2} - \delta u^2}$$

para $\delta > 0$, *con base en la cual resulta el siguiente modelo de tres parámetros:*

$$C(h; u|\Theta) = \frac{\sigma^2}{(a^2u^2 + 1)^{d/2}} e^{-\frac{b^2h^2}{a^2u^2+1}}$$

para $\delta \to 0$ *y* $\Theta = (a, b, \sigma^2)$, *y* $a \geq 0$, $b \geq 0$ *parámetros de escala temporal y espacial respectivamente y* $\sigma^2 = C(\mathbf{0}, 0|\Theta)$. *Se ilustran algunos casos de esta función de covarianza espacio-temporal en la Figura 6.1.*

6.5.2. Método basado en la composición de funciones con propiedades de monotonicidad

[35] propone una clase de funciones de covarianza para procesos espacio-tiempo que, a diferencia de la propuesta por [25], no depende de la existencia de la transformada de Fourier. La construcción de esta clase de funciones es resumida en el Teorema 2.

Teorema 2 (Teorema de Gneiting). *Para* $r \geq 0$, *sean* $\varphi(r)$ *una función completamente monótona y* $\phi(r)$ *una función positiva con una derivada completamente monótona. Entonces, para* $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{\phi(u^2)^{d/2}} \varphi\left(\frac{\|\boldsymbol{h}\|^2}{\phi(u^2)}\right)$$

es una función de covarianza espacio-tiempo.

Estas familias de funciones de covarianza presentan toda clase de posibilidades: pueden ser cóncavas o convexas, tienen diferentes velocidades de cambio; unas decrecen de manera uniforme con respecto a los dos ejes, mientras

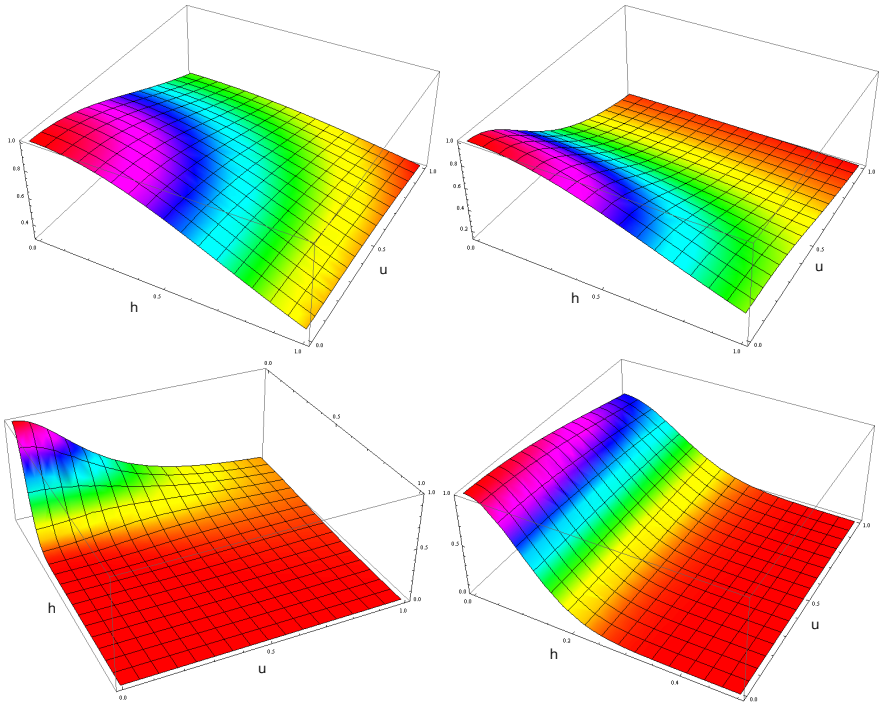


Figura 6.1. Cressie-Huang, ejemplo 15, $\Theta = (\sigma^2, a, b)$, $0 \leq \|h\| \leq 1$ y $0 \leq u \leq 1$. a. $\Theta = (1; 1; 1)$, b. $\Theta = (1; 5; 1)$, c. $\Theta = (1; 10; 50)$, d. $\Theta = (1; 0,5; 50)$

que otras decrecen más rápido con respecto a uno de ellos, lo que permite variados comportamientos para todas las posibles combinaciones de rezagos espacio-tiempo. Esto se puede observar en las superficies de la Figura 6.1 y en las curvas de nivel presentadas en la Figura 6.2, donde se aprecia que algunos gráficos tienen sus isolíneas muy separadas, como en el panel superior derecho, indicando un cambio lento en el eje vertical, mientras que en otras ocasiones las isolíneas se encuentran muy juntas, indicando que la covarianza cambia muy rápidamente, como en el panel inferior izquierdo. La Figura del panel superior izquierdo muestra que la variación es más fuerte en el eje vertical, mientras que en el panel inferior izquierdo se observa que la covarianza cambia en forma similar con respecto a ambos ejes. La Figura del panel inferior derecho ilustra un modelo separable; la variación solo se presenta en el eje horizontal. Las familias de modelos son muy flexibles; permiten prácticamente todos los comportamientos que se puedan presentar en la práctica.

Ejemplo 16 (Modelo de covarianza espacio temporal de Gneiting 1, [35]). *Sean las siguientes dos funciones que cumplen los requisitos del Teorema 2:*

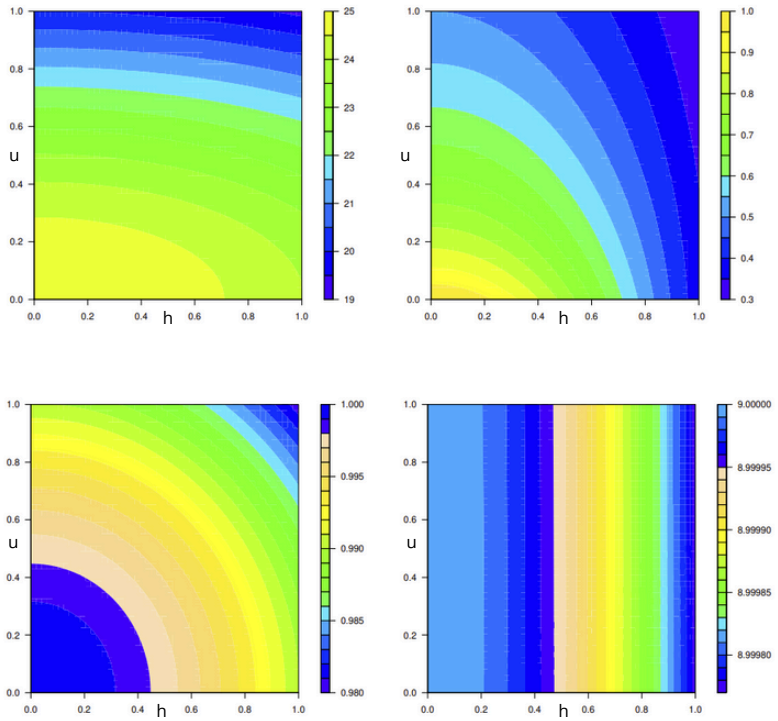


Figura 6.2. Modelo Cressie-Huang (ver Ejemplo 15). $\Theta = (\sigma^2, a, b)$, $0 \leq \|h\| \leq 1$ y $0 \leq u \leq 1$. Panel superior izquierdo. $\Theta = (1; 1; 1)$, Panel superior derecho. $\Theta = (1; 5; 1)$, Panel inferior izquierdo. $\Theta = (1; 10; 50)$, Panel inferior derecho. $\Theta = (1; 0, 5; 50)$.

$$\varphi(r) = e^{(-cr^{\gamma})} \quad y \quad \phi(r) = (1 + ar^{\alpha})^{\beta}$$

Para $c > 0$, $a > 0$, $0 < \gamma \leq 1$, $0 < \alpha \leq 1$ *y* $0 \leq \beta \leq 1$. *La familia paramétrica de funciones de covarianza generada es*

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{(1 + au^{2\alpha})^{\frac{\beta d}{2}}} e^{\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(1+au^{2\alpha})^{\beta\gamma}}\right)} \tag{6.7}$$

donde a y c (no negativos) son parámetros de escala de tiempo y espacio respectivamente. Los parámetros α *y* γ *controlan la suavidad de la función, y el parámetro* β *corresponde a la interacción espacio-tiempo;* σ^2 *es la varianza del proceso. El Ejemplo 17 muestra un caso particular de esta familia.*

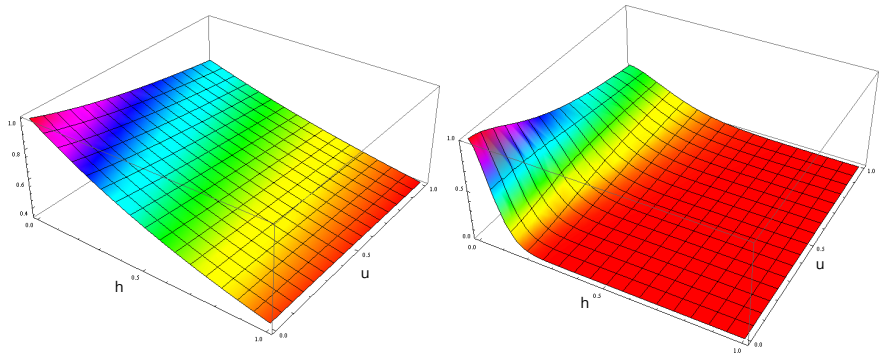


Figura 6.3. Dos casos de la familia de funciones de covarianza Gneiting (ver Teorema 2 y Ejemplo 16): $\Theta = (\sigma^2, a, \alpha, \beta, c, \gamma)$, Panel izquierdo. $\Theta = (1; 1; 0,5; 0,5; 1; 0,5)$, Panel derecho. $\Theta = (1; 0,5; 0,8; 0,8; 1,0; 8)$.

Ejemplo 17 (Modelo de covarianza espacio temporal de Gneiting-Matern, [35]). *Modelo Gneiting-Matern En este ejemplo se usa la función de covarianza de Matern como función monótona $\varphi(r)$ (ver Sección 2.5).*

$$\varphi(r) = \frac{1}{2^{v-1}\Gamma(v)}(cr^{\frac{1}{2}})^{v}\mathcal{K}_{v}(cr^{\frac{1}{2}}) \quad y \quad \phi(r) = (ar^{\alpha}+1)^{\beta}$$

La familia paramétrica de funciones de covarianza generada es

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{2^{v-1}\Gamma(v)(au^{2\alpha}+1)^{\delta+\beta d/2}}\left(\frac{c|\boldsymbol{h}|}{(au^{2\alpha}+1)^{\beta/2}}\right)^{v} \times \mathcal{K}_{v}\left(\frac{c|\boldsymbol{h}|}{(au^{2\alpha}+1)^{\beta/2}}\right) \tag{6.8}$$

con $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$.

Donde a y c (no negativos) son parámetros de escala de tiempo y espacio respectivamente, $0 < \alpha \leq 1$ es el parámetro de suavidad del tiempo, $v > 0$ es el parámetro de suavidad del espacio; $0 \leq \beta \leq 1$ es el parámetro de interacción, $\delta \geq 0$, $\sigma^2 > 0$ y $\mathcal{K}_v$ es la función de Bessel modificada de la segunda clase de orden v.

6.5.3. Otros métodos para encontrar familias de funciones de covarianza espacio tiempo válidas.

Existen varias otras metodologías para la construcción de familias de covarianzas espacio-tiempo no separables, en general partiendo de funciones válidas tanto en el espacio como en el tiempo. A continuación, se citan algunas:

La familia de funciones de covarianza suma-producto, [28].

$$C_{s,t}(\boldsymbol{h}, u) = aC_s(\boldsymbol{h}, \Theta_s)C_t(u, \Theta_t) + bC_s(\boldsymbol{h}, \Theta_s) + cC_t(u, \Theta_t)$$

donde a, b, c son coeficientes no negativos y $C_s(\boldsymbol{h}, \Theta_s)$ y $C_t(u, \Theta_t)$ son funciones de covarianza válidas en el espacio y en el tiempo, respectivamente.

Integral del producto [28]. (Teorema 3.)

Teorema 3. *Sea μ una medida finita no negativa sobre un conjunto no vacío Θ. Sean $C_s(\boldsymbol{h}, \Theta_s)$ y $C_t(u, \Theta_t)$ funciones de covarianza válidas en $\mathbb{R}^d$ y $\mathbb{R}$ respectivamente, y dado que $C_s(\boldsymbol{h}, \Theta_s)C_t(u, \Theta_t)$ es integrable con respecto a la medida μ sobre $\boldsymbol{\Theta} = (\Theta_s, \Theta_t)$, entonces*

$$C(\boldsymbol{h}, u) = \int C_s(\boldsymbol{h}, \Theta_s)C_t(u, \Theta)d\mu(\Theta_t), \quad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

es una función de covarianza en $\mathbb{R}^d \times \mathbb{R}$

Integral del producto con medida gamma. En el Teorema 3, μ es una medida que sigue una distribución gamma y tanto la covarianza espacial como la temporal corresponden a un modelo estable. Esto es,

$$C_s(\boldsymbol{h}, \Theta_1) = \sigma_1^2 e^{-(h/a)^\alpha} \quad C_s(\boldsymbol{h}, \Theta_2) = \sigma_2^2 e^{-(u/b)^\beta} \tag{6.9}$$

donde $a > 0$, $b > 0$ y $\gamma > 0$, $\alpha \in (0, 2)$, $\beta \in (0, 2)$, se llega como caso particular a la familia

$$C(\boldsymbol{h}; u) = \sigma^2 \left(1 + \left\|\frac{\boldsymbol{h}}{a}\right\|^\alpha + \left|\frac{u}{b}\right|^\beta\right)^{-\gamma} \tag{6.10}$$

En [49] se muestran otras familias que pueden ser construidas a partir del Teorema 3.

Función de covarianza de Dagum. Se encuentra una función de covarianza espacio-tiempo mediante una adaptación de la función de supervivencia de Dagum, ecuación (6.11) [2].

$$\varphi(r) = \begin{cases} 1, & \text{si} \quad \text{r=0}, \\ 1 - \frac{1}{(1+\varrho r^{-\zeta})^\varepsilon}, & \text{si} \quad \text{r>0} \end{cases}. \tag{6.11}$$

donde los parámetros $\varrho, \zeta, \varepsilon$ son estrictamente positivos. Θ, ε son parámetros de suavizamiento y ϱ es un parámetro de escala.

En [61], se estudia como una función de base radial y se construye una función definida positiva en $\mathbb{R}^d$, para $d = 1, 2, \ldots$.

Se consideran las siguientes dos alternativas:

- Una función de covarianza separable, obtenida como el siguiente producto tensorial

$$C_{1(s,t)}(\boldsymbol{h}, u) = \varphi(\|h\|)\varphi(|u|)$$

- Una función de covarianza no separable, obtenida como la siguiente suma convexa

$$C_{2(s,t)}(\boldsymbol{h}, u) = \vartheta\varphi(\|h\|) + (1-\vartheta)\varphi(|u|)$$

donde $\vartheta \in [0, 1]$. Esta función de covarianza espacio-tiempo es válida siempre que $\Theta < \frac{7-\varepsilon}{1+5\varepsilon}$ y $\varepsilon < 7$.

Además, [70] presenta un método de construcción de funciones de covarianza espacio-temporal no estacionarias a partir de la función de covarianza de Matern. [2] trata funciones de covarianza espacio-tiempo estacionarias y anisotrópicas haciendo particiones en el dominio espacial y muestra algunas formas generalizadas de las funciones de Gneiting al caso anisotrópico.

La definición extendida al caso espacio tiempo del semivariograma $2\gamma(h, u|\Theta)$ y su relación con la covarianza bajo estacionariedad de segundo orden, se presentan a continuación

$$2\gamma(\mathbf{h}, u|\Theta) = \text{Var}\left[\mathbf{Z}(\mathbf{s}+\mathbf{h}; t+u) - \mathbf{Z}(\mathbf{s}; t)\right] = 2\left(C(\mathbf{0}; 0|\Theta) - C(\mathbf{h}; u|\Theta)\right)$$

para $\mathbf{h} \in \mathbb{R}^d$, $u \in \mathbb{R}$; $\sigma^2 = C^0(\mathbf{0}; 0|\Theta)$.

6.6. Ejercicios

1. Se tiene un proceso estacionario espacio-tiempo de segundo orden, con media constante y con matriz de autocovarianza conocida Σ. Si se tienen mediciones en n ubicaciones espaciales y T momentos en el tiempo, encuentre el estimador de la media y su varianza, bajo el supuesto de normalidad multivariada.
2. Suponga estacionariedad de segundo orden. Seleccione una función de covarianza espacio-temporal separable y una función de covarianza espacio-temporal no separable. Para cada una, verifique usando un conjunto cualquiera de ubicaciones espacio-temporales y un conjunto de números reales, que es definida positiva y que su semivariograma asociado es condicionalmente definido negativo.
3. Construya una función de covarianza espacio-temporal no separable, aplicando el Teorema 2. Encuentre su semivariograma asociado.
4. Construya el predictor kriging en el caso en el que no se encuentra autocovarianza espacio-temporal.

Capítulo

siete

Estimación teórica de la función de autocovarianza espacio-temporal

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

En este capítulo se describen los métodos de estimación que fueron presentados en 4 para el caso espacial, extendidos al caso espacio-tiempo. Por lo tanto, se presentan de forma más concreta. Para más detalles, se puede consultar la Sección 3 en la que se ilustran todos los elementos para el caso espacial. El método fue propuesto originalmente por [48] y posteriormente ha sido adaptado al caso espacial por [27] y al caso espacio-tiempo por [3]. Los conceptos son los mismos, pero los procesos de optimización en este caso son más complejos computacionalmente por la cantidad de datos.

7.1. Mínimos cuadrados ponderados espacio-tiempo

Este método se puede generalizar al caso espacio-temporal de la siguiente forma: se estima Θ para un variograma espacio-temporal, minimizando

$$(2\hat{\gamma} - 2\gamma(\Theta))'\mathrm{W}^{-1}(\Theta)(2\hat{\gamma} - 2\gamma(\Theta))$$

La matriz de ponderación está dada por

$$\mathrm{W}(\Theta) = \mathrm{Diag}(\mathrm{Var}[2\hat{\gamma}(h_k, u))] \simeq \mathrm{Diag}\left(\frac{2(2\gamma(h_k, u|\Theta))^2}{N(h_k; u)}\right)$$

con

$$N(h_k; u) = \{(i, j, t, t') : s_i - s_j = h_k; \mid t - t' \mid = u\}$$

para los primeros u rezagos temporales, $u = 0, \ldots, U$, y para las ubicaciones $i, j = 1, \ldots, n$, que generan los primeros k rezagos espaciales, $k = 1, \ldots, K$; en general, se usan los rezagos espaciales hasta la mitad de la máxima distancia entre cualquier par de ubicaciones, del mismo modo que en el caso espacial, debido a que, para ubicaciones muy separadas, ya sea en tiempo o en espacio, disminuye notoriamente la cantidad de puntos incluidos en la estimación del variograma. En el tiempo es común que los datos se encuentren igualmente espaciados, pero en el espacio, en general, las muestras son tomadas en una grilla irregular, por lo que es difícil encontrar suficientes pares de puntos con separación exacta $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}_k$. Se acostumbra definir tolerancia de rezagos en el tiempo, y en el espacio tolerancias de longitud y de ángulo. Así, $\|s_i - s_j\| \in (h_k - \xi, h_k + \xi)$, $\|t - t'\| \in (t - v, t' + v)$ y el ángulo entre $\boldsymbol{s}_i$ y $\boldsymbol{s}_j$ está entre $\phi - \omega$ y $\phi + \omega$.

Los estimadores empíricos del semivariograma son una generalización directa del caso espacial. Por ejemplo, el estimador clásico está definido como

$$\hat{\gamma}(\boldsymbol{h}, u) = \frac{1}{2|N(\boldsymbol{h}, u)|} \sum_{N(\boldsymbol{h},u)} \Big(z(\boldsymbol{s} + \boldsymbol{h}, t + u) - z(\boldsymbol{s}, t)\Big)^2, \quad \boldsymbol{h} \in \mathbb{R}^d, u \in \mathbb{R} \qquad (7.1)$$

Desafortunadamente, en este caso la visualización es más difícil por estar en tercera dimensión y el semivariograma empírico ya no es tan contundente para

ayudar en la selección de modelos, si bien sigue siendo una excelente herramienta exploratoria. En este contexto, es necesario ensayar muchos modelos y muchos valores iniciales hasta encontrar la mejor opción, según los resultados tanto de la estimación de la estructura de dependencia espacio-temporal como de las predicciones.

La aproximación de $\text{Var}[2\hat{\gamma}(h_k, u|\Theta)]$ se obtiene bajo el supuesto de que $Z(\mathbf{s}, t) \sim N(\mu; \sigma^2)\ \forall \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}$, y que, por lo tanto,

$$\Big(Z(\mathbf{s}+\mathbf{h}; t+u) - Z(\mathbf{s}; t)\Big)^2 \sim 2\gamma(\mathbf{h}, u) \cdot \chi_1^2$$

Ver Capítulo 3 donde se construye para el caso espacial que es similar. La desventaja que presenta el uso de los métodos de mínimos cuadrados es la necesidad de definir clases de rezagos. En este caso, al ser espacio-temporales $(\boldsymbol{h}, u)$, se acentúa la dificultad de encontrar suficiente cantidad de observaciones para cada sector. Nótese que es un sector circular en 3D. Si la cantidad de datos en cada una de estas clases es muy pequeña, solo los menores rezagos pueden tener suficientes datos para la estimación de cada $\hat{\gamma}(h_k, u)$.

7.2. Métodos basados en la función de verosimilitud espacio tiempo

La estimación de parámetros de la función de covarianza usando los métodos de máxima verosimilitud (ML) y máxima verosimilitud restringida (REML) requiere la especificación de la distribución del vector

$$\mathbb{Y} = \big(Y(s_i, t_{ij})\big)\ ;\ j = 1, \ldots, T_i,\ \ i = 1, \ldots, n.$$

$N = \sum_{i=1}^{n} T_i$. En general, se asume normalidad multivariada. Para el caso ML se tiene que

$$\mathbb{Y} \sim \boldsymbol{N}_N(X\beta, \boldsymbol{\Sigma}(\Theta))$$

donde $\boldsymbol{\Sigma}(\Theta) = \text{Cov}(\mathbb{Y})$ es una matriz de dimensión $N \times N$ y X es una matriz de dimensión $N \times q$ con $q < N$, de variables explicativas, dentro de las cuales comúnmente se encuentran las coordenadas geográficas y el tiempo, o funciones de ellas. El negativo de la función log-verosímil es

$$L(\beta, \Theta) = \frac{N}{2}\log(2\pi) + \frac{1}{2}\log|\boldsymbol{\Sigma}(\Theta)| + \frac{1}{2}(\mathbb{Y} - \boldsymbol{X}\boldsymbol{\beta})'\Sigma^{-1}(\Theta)(\mathbb{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

donde $\boldsymbol{\Sigma}(\Theta) = \text{Cov}[Y(s_i, t), Y(s_j, t')] = C(s_i - s_j, t - t'; \Theta)$. Cada elemento ij corresponde a la covarianza espacio-temporal entre las variables $Y(s_i, t)$ y $Y(s_j, t')$. El estimador $\hat{\Theta}$ es sesgado, pero asintóticamente eficiente. Sin embargo, la cantidad de datos en el caso espacio-tiempo crece rápidamente y esto puede generar inconvenientes computacionales. El estimador REML, sustituye la maximización

de la verosimilitud del vector $\mathbb{Y}$ por la del vector $\mathbf{A}'\mathbb{Y}$ de la misma forma que en el caso espacial (ver Sección 3.5).

7.3. Pseudoverosimilitud CL espacio tiempo

Este método no tiene los problemas de dimensionalidad de los métodos anteriores. Se describe en mayor detalle por ser el que funciona mejor para el caso espacio tiempo. El paquete [4] presenta una implementación de este método.

El método de *verosimilitud compuesta* (CL), [5], consiste en sumar las funciones log-verosimilitud individuales correspondientes a las marginales de las variables de interés. Por lo tanto, este método no requiere el conocimiento de la distribución multivariada de $\mathbb{Y}$; solo se requieren las marginales $f(Y(s_i, t_i), \Theta)$, pues se asume que existen tanto el gradiente como la matriz Hessiana de f.

Supongamos conocidas $f(Y(s_i, t_i), \Theta)$, excepto por el parámetro Θ; entonces $l(Y(s_i, t_i), \Theta) = ln(f(Y(s_i, t_i), \Theta))$ es una función log-verosímil y la función de verosimilitud compuesta es

$$CL(\Theta) = \sum_{i=1}^{n} l(Y(s_i, t_i), \Theta)$$

A su gradiente $\nabla CL(\Theta) = CS(\Theta)$ se le llama la función score compuesta. Así, para encontrar el estimador $\hat{\Theta}$ se resuelve el sistema de ecuaciones

$$CS(\Theta) = \sum_{i=1}^{n} \nabla l(Y(s_i, t_i), \Theta) = 0$$

Una implementación de este método para la estimación del semivariograma espacio-temporal se presenta a continuación.

7.4. Construcción de la función de estimación

1. El objetivo es estimar los parámetros del semivariograma y, por lo tanto, es muy natural la construcción de la variable incrementos espacio-tiempo, la cual se va a notar V. Se usa la variable centrada $Z(s_i, t)$.

$$V(i, j, t, t') = Z(s_i, t) - Z(s_j, t')$$

   La variable V se logra efectuando todas las combinaciones posibles: con ubicación fija, variando en el tiempo; con tiempo fijo, variando la ubicación y variando tanto tiempo como ubicación. Esto genera una inmensa cantidad de datos, lo que constituye una de las fortalezas del método, ya que el tamaño de la matriz involucrada en los procesos de optimización en CL,

depende solo de la dimensión del parámetro Θ de la función de semivarianza o autocovarianza, como ocurre en los métodos de mínimos cuadrados y a diferencia de ML y REML cuya dimensión es $N \times N$.

Además, es suficiente ingresar a la estimación solo aquellos datos que involucran información sobre la dependencia espacio-tiempo, y pueden omitirse aquellos pares de observaciones que se encuentren muy alejados, de acuerdo con algún criterio definido, tal como el alcance espacial o el alcance temporal observados en los semivariogramas o covariogramas.

2. Dado que se requieren las funciones de verosimilitud marginales de las variables de interés, asumiendo normalidad para las distribuciones marginales de $\mathbb{Y}$,

$$Y(\boldsymbol{s}, t) \sim N(\mu; \sigma^2), \forall (\boldsymbol{s}, t) \in D \times T$$

se tiene que

$$V(i, j, t, t') \sim N(0, 2\gamma(s_i - s_j, t - t', \Theta))$$

Así que el negativo de la función log-verosímil es

$$l(V(i, j, t, t'), \Theta) = \frac{\log 2\pi}{2} + \frac{\log 2\gamma(s_i - s_j, t - t', \Theta)}{2} + \frac{V^2(i, j, t, t')}{4\gamma(s_i - s_j, t - t', \Theta)}$$

3. Se determina la función de verosimilitud compuesta $CL(\Theta)$, sumando todas las funciones log-verosímil marginales de la variable V;

$$CL(\Theta) = \sum_{t=1}^{T} \sum_{t'>t}^{T} \sum_{i=1}^{n} \sum_{j>i}^{n} l(v(i, j, t, t'), \Theta) + \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j>i}^{n} l(v(i, j, t, t), \Theta) + \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{t'>t}^{T} l(v(i, i, t, t'), \Theta)$$

4. Se propone un modelo válido para el variograma espacio-tiempo $2\gamma(s_i - s_j, t - t', \Theta)$

5. Se debe establecer un umbral que permita seleccionar los pares de datos cuya separación espacio-temporal no supere los rangos respectivos.

6. Se resuelve el sistema de ecuaciones

$$CS(\Theta) = \nabla CL(\Theta)$$

donde un término genérico para las ecuaciones del sistema resultante se puede escribir como:

$$CS(\Theta; i, j, t, t') = \frac{\partial \gamma(s_i - s_j, t - t', \Theta)/\partial \Theta}{4\gamma^2(s_i - s_j, t - t', \Theta)} \left(v^2(i, j, t, t') - 2\gamma(s_i - s_j, t - t', \Theta)\right). \tag{7.2}$$

La expresión (7.2) es una función de estimación ponderada por el cociente,

$$\frac{\partial \gamma(s_i - s_j, t - t', \Theta)/\partial \Theta}{4\gamma^2(s_i - s_j, t - t', \Theta)}.$$

el cual disminuye a medida que los rezagos espacio-tiempo aumentan. Esto justifica la selección de los datos según umbrales espacio-tiempo determinados, más allá de los cuales no exista dependencia, y se incluyen solamente las observaciones $v(i, j, t, t')$, cuyos rezagos no superen dichos umbrales, para optimizar el procedimiento.

Características

- $CL(\Theta)$ es una función de estimación insesgada, ya que cada uno de sus componentes es una verosimilitud:

  $$E(V^2(i, j, t, t')) = E((Z(s_i, t) - Z(s_j, t'))^2) = 2\gamma(s_i - s_j, t - t', \Theta).$$

  independientemente del supuesto que se haga sobre la distribución de $Z(s, t)$;

- Bajo condiciones de regularidad, [27] muestran que la función objetivo cumple la desigualdad de información de Kullback-Leibler, esto es,

  $$E(-l(V(i, j, t, t'; \Theta)) > E(-l(V(i, j, t, t'; \Theta_0))). \tag{7.3}$$

  y que existe una solución consistente, siempre y cuando los dos primeros momentos del proceso hayan sido bien especificados.

- La varianza de $CS(\Theta)$ no coincide con la matriz de información de Fisher, sino que está dada por la matriz de información de Godambe $G(\Theta)$ expresión (7.4) [71]:

  $$G(\Theta) = H(\Theta)J^{-1}(\Theta)H(\Theta) \tag{7.4}$$

  donde

  $$H(\Theta) = E(-\nabla CS(\Theta))$$

  y

  $$J(\Theta) = Var(CS(\Theta)).$$

  [32] encuentra que $I(\Theta) - G(\Theta)$, donde $I(\Theta)$ es la matriz de información de Fisher, es una matriz semidefinida positiva; por lo tanto, la estimación CL es menos eficiente que la estimación ML. Sin embargo, [48] muestra que esta eficiencia aumenta cuando se utilizan ponderaciones; por lo que, de acuerdo con lo comentado de la expresión 7.2, en este contexto se usan ponderaciones dicotómicas para seleccionar aquellas observaciones que no

superan simultáneamente el alcance espacial y el alcance temporal. Así se logra, entonces, mejorar la eficiencia de CL y al mismo tiempo optimizar los datos ingresados al proceso de estimación. Las ponderaciones se definen como:

$$w_{i,j,t,t'} = \begin{cases} 0 & \text{si } s_i - s_j > \hbar \quad \text{y} \quad t - t' > u \\ 1 & \text{en caso contrario.} \end{cases}$$

Por lo tanto, la función a minimizar finalmente será:

$$\begin{aligned} CL(\Theta) = &\sum_{t=1}^{T}\sum_{t'>t}^{T}\sum_{i=1}^{n}\sum_{j>i}^{n} w_{i,j,t,t'} l(v(i,j,t,t'),\Theta) \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j>i}^{n} w_{i,j,t,t'} l(v(i,j,t,t),\Theta) \\ &+ \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t'>t}^{T} w_{i,j,t,t'} l(v(i,i,t,t'),\Theta) \end{aligned}$$

7.5. Errores estándar de la estimación vía CL

El proceso de maximización involucra únicamente una matriz $p \times p$, donde p es el número de parámetros del modelo de covarianza propuesto. Se requieren operaciones eficientes, vectoriales y matriciales que permitan generar todos los rezagos espacio-tiempo necesarios; pero una vez encontrados, la estimación de unos pocos parámetros se basa en una gran cantidad de datos. Sin embargo, no existe aún una manera específica para hallar las medidas de calidad de estos estimadores. Una propuesta para este fin se presenta a continuación.

La estimación de los errores estándar para los parámetros estimados por el método CL está aún sin resolver, incluso en el caso puramente espacial. Las especificidades de cada uno de los contextos de aplicación son diferentes y así también las formas de adaptar el procedimiento. Por lo tanto, a continuación se presenta una propuesta derivada del submuestreo de ventana para encontrar estos estimadores y que hasta ahora se ha aplicado a datos de área [42]. En este trabajo, se realiza una adaptación al caso de un dominio continuo espacio-tiempo, usando las vecindades generadas para alcanzar una independencia aproximada.

Análogo al caso de la matriz de información de Fisher, la covarianza asintótica de $\widehat{\Theta}$ está dada por

$$G^{-1}(\Theta) = H^{-1}(\Theta) J(\Theta) H^{-1}(\Theta)$$

por lo cual es indispensable obtener estimadores consistentes tanto de la matriz $J(\Theta)$ como de la matriz $H(\Theta)$. Un estimador empírico para $H(\Theta)$ es la matriz hessiana evaluada en la estimación, [71]:

$$\widehat{H}(V) = \frac{1}{M}\sum w_{i,j,t,t'}\nabla CS(\widehat{\Theta}; i, j, t, t')$$

con $M = \sum w_{i,j,t,t'}$; en ambos casos, los subíndices varían como en la función de verosimilitud compuesta.

Para obtener un estimador consistente de J, se requiere dividir la región de estudio en subregiones que generen grupos independientes. Así, el estimador para la matriz $\widehat{J}(V)$ es

$$\widehat{J}(V) = \frac{1}{kl}\sum_{p=1}^{k}\sum_{q=1}^{l}|\mathbb{A}_{pq}|CS(\widehat{\Theta}, v_{i,j,t,t'} \in \mathbb{A}_{pq})CS(\widehat{\Theta}, v_{i,j,t,t'} \in \mathbb{A}_{pq})'$$

donde $\mathbb{A}_{pq}$ son las vecindades espacio-tiempo generadas buscando garantizar independencia entre cada uno de estos subconjuntos con el fin de obtener una estimación consistente de la varianza de $\widehat{\Theta}$. En total, se cuenta con kl vecindades que corresponden a k vecindades espaciales y a l saltos en el tiempo (ver detalles en [6]). $|\mathbb{A}_{pq}|$ es la cantidad de datos en la vecindad $\mathbb{A}_{pq}$.

7.6. Selección del modelo

Con base en la estimación propuesta para las matrices $H(V)$ y $J(V)$, se cuenta con un criterio de información análogo al criterio de Akaike para el método de verosimilitud compuesta. Este criterio se conoce como $CLIC$ y fue propuesto por [71]. Su expresión se deduce de la sustitución de la desigualdad de Kullback Leibler para la función CL (ecuación (7.3)) en la deducción del AIC. El $CLIC$ selecciona el modelo que maximiza (7.5)

$$CLIC = CL(\widehat{\Theta}; V) - tr((\widehat{J}(V))\widehat{H}^{-1}(V)) \tag{7.5}$$

donde la penalización por la dimensión del vector de parámetros está dada por $tr((\widehat{J}(V))\widehat{H}^{-1}(V))$, [6].

7.7. Modelos jerárquicos

Si existe suficiente conocimiento científico acerca del fenómeno de interés, se pueden usar los modelos jerárquicos. El modelo se puede formular como un modelo dinámico basado en distribuciones condicionales. Por ejemplo, si el proceso de interés $Y(s_i; t_{ij})$ no puede ser observado directamente, el modelo es

$$Y(s_i; t_{ij}) = \mathfrak{L}(s_i; t_{ij}) + \varepsilon(s_i; t_{ij}) \qquad j = 1, \ldots, T_i \quad i = 1, \ldots, n \tag{7.6}$$

donde$\{\varepsilon(s_i; t_{ij})\} \sim iid(0, \sigma_\varepsilon^2)$ es independiente de $Y(\cdot;\cdot)$ y representa el error de medida.

Así, este puede ser formulado como un modelo dinámico (ver modelo (7.6)). El primer y segundo nivel respectivamente son

$$[\mathbb{Y}|\mathfrak{Z}(\cdot;\cdot),\Theta_D]$$

y

$$[\mathbb{Y}(\cdot;\cdot)|\Theta_P]$$

donde Θ_D son parámetros en el modelo de los datos y Θ_P son parámetros del modelo del proceso. Ahora, con el Teorema de Bayes se puede hacer inferencia acerca de $\mathbb{Y}(\cdot;\cdot)$; existen dos opciones:

1. Modelo jerárquico empírico (EHM). Estimar $\Theta = \{\Theta_D, \Theta_P\}$ de los datos $\mathfrak{Z}$ y reemplazar este valor en la distribución a posteriori (plug-in).

$$[\mathbb{Y}(\cdot;\cdot)|\mathfrak{Z},\Theta]\alpha[\mathfrak{Z}|\mathbb{Y}(\cdot;\cdot),\Theta_D][\mathbb{Y}(\cdot;\cdot)|\Theta_P]$$

2. Modelo jerárquico Bayesiano. Agregar otro nivel para la estructura del parámetro Θ

$$[\mathbb{Y}(\cdot;\cdot)|\mathfrak{Z},\Theta]\alpha[\mathfrak{Z}|\mathbb{Y}(\cdot;\cdot),\Theta_D][\mathbb{Y}(\cdot;\cdot)|\Theta_P][\Theta]$$

Por ejemplo, si el índice espacial es continuo, este puede ser una representación de un kriging bayesiano jerárquico donde la distribución de los parámetros de covarianza está en el último nivel.

Si el espacio es considerado discreto, el modelo puede ser espacio-temporal autorregresivo de media móvil, (STARMA). Por ejemplo, un modelo VAR(1),

$$\mathbb{Y}_t = M\mathbb{Y}_{t-1} + W_t \tag{7.7}$$

donde M es la matriz de propagación sobre Y_t de la variable en cada lugar observado en el tiempo anterior, $\{W_t\}$ es un proceso de ruido blanco $n - variado$, $\{W_t\} \sim iid(0, Q)$. En este caso, la clave es la estimación de M. Es necesario reducir la dimensión, y se hace necesario encontrar formas de reducir la cantidad de parámetros. Una ventaja de este modelo particular es que facilita los análisis debido a que algunas ecuaciones diferenciales estocásticas pueden ser reemplazadas por modelos VAR(p).

Nota 8. *Una alternativa para el proceso espacio-temporal es expresarlo con una ubicación que no distinga las coordenadas espacio y tiempo, tal como*

$$\{Y(\boldsymbol{v}) : \boldsymbol{v} \in D \subset \mathbb{R}^{d+1}\}$$

La ventaja de la representación dada en la Definición 9 es la posibilidad de identificar los parámetros asociados al proceso espacial, al proceso temporal y a su interacción, además de que no se mezclan escalas en el cálculo de las distancias.

Ejemplo 18. *Predicción espacial-temporal del PM10 en Bogotá Se muestra una aplicación de la metodología para el modelamiento geoestadístico (espacio-tiempo), con estructuras de covarianza no separables.*

El problema a abordar es la distribución del material particulado (PM) en la ciudad de Bogotá, el cual es un componente importante de la contaminación del aire; sus características físico-químicas y sus efectos sobre la salud humana hacen que su monitoreo y control sean prioritarios. Este material está compuesto por partículas líquidas o sólidas que provienen de procesos como la erosión, las erupciones volcánicas y los incendios, así como del uso de combustibles fósiles en la industria y el transporte, entre otros. Una de las características físicas más importantes de este material es el diámetro por partícula, puesto que parte de él puede ingresar al tracto respiratorio y producir daños en los tejidos y órganos que lo conforman o servir como vehículo para bacterias y virus. Las partículas PM10 son aquellas cuyo tamaño es menor o igual a 10 micras; varios estudios han mostrado evidencia de que estas partículas están asociadas con la mortalidad y morbilidad de la población [39, 58, 60].

La Secretaría Distrital de Ambiente actualmente opera la red de monitoreo de calidad del aire de Bogotá (RMCAB), la cual cuenta con 14 estaciones ubicadas en puntos estratégicos de la ciudad, que monitorean, cada hora las concentraciones de material particulado (PM10,PM2.5, PST), de gases contaminantes (SO2, NO2, CO, O3) y los parámetros meteorológicos de precipitaciones, vientos, temperatura, radiación solar y humedad relativa. Su objetivo es obtener, procesar y divulgar información de la calidad del aire, de forma confiable y clara, para evaluar el cumplimiento de estándares y para la definición de políticas de control de contaminación [69].

La Tabla 7.1 resume las características de los sectores de la ciudad en donde se encuentran ubicadas las estaciones de la RMCAB. En este caso, no se tiene en cuenta la altitud, dado que es aproximadamente constante en la ciudad. Nótese que el PM10 existe en todos los infinitos puntos del dominio espacial y en todos los infinitos puntos del dominio temporal. Es decir, este proceso varía continuamente en el espacio-tiempo $D_s \times D_t \subset \mathbb{R}^2 \times \mathbb{R}$.

Análisis exploratorio

A continuación, se analizan los datos de la red de calidad del aire en la ciudad de Bogotá, modelando la dependencia espacio-temporal con fines predictivos. Los datos usados para la ilustración de la metodología corresponden a 73 horas consecutivas, que van desde el 16 de agosto del 2024 a las 6 de la tarde hasta el 18 de agosto del 2024 a las 11 de la noche. En Bogotá existían en ese momento 14 estaciones ambientales, cuya ubicación se muestra en la Figura 7.1. Se toma esta fracción de datos por ser la única franja para la cual se cuenta con información en todas las estaciones. Existe gran dificultad para tener series largas debido a las continuas fallas en las estaciones, pues a veces solo hay 5 o 6 estaciones registrando datos simultáneamente.

En la resolución 0601 de 2006 del Ministerio de Ambiente, Vivienda y Desarrollo Territorial se establecen, entre otros, los niveles máximos permisibles para el PM10: $70\mu g/m^3$ *promedio anual, y* $150\mu g/m^3$ *en 24 horas. Además, en el primer parágrafo se hace la aclaración de que estos niveles máximos disminuirían a* $60\mu g/m^3$ *en el año 2009 y a* $50\mu g/m^3$ *en el año 2011.*

Sector	Estación	Características
Norte	Universidad Bosque, Escuela Ingeniería	Zona residencial de baja densidad poblacional y alto tráfico vehicular
Nor-occidente	Carrefour calle 80, Universidad Corpas, Fontibón	Alto tráfico vehicular y uso residencial y comercial
Sur	Hospital del Olaya, Central de Mezclas	Alto tráfico vehicular, uso residencial, comercial
Sur-occidente	Sony Music, Cazucá	Zona industrial con alto tráfico vehicular y uso residencial
Central	MMA, Universidad Nacional, Universidad Santo Tomás	Alto tráfico vehicular y uso residencial, comercial e institucional
Centro-occidente	Cade-Energía, Merck	Zona industrial con alto tráfico vehicular y uso residencial

Tabla 7.1. Estaciones de la red de calidad del aire de Bogotá, Colombia.

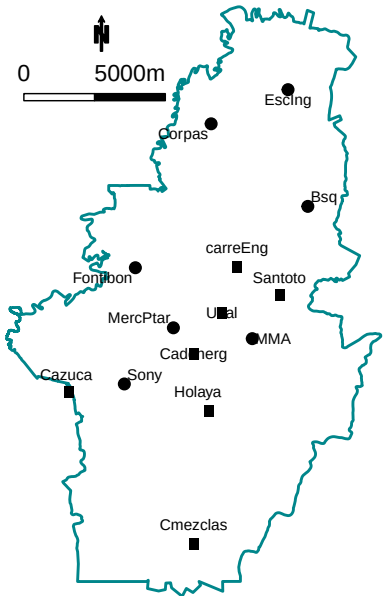


Figura 7.1. Red de calidad del aire en la ciudad de Bogotá.

El análisis de este tipo de datos, en general, se hace solamente temporal, como una serie de tiempo en la ubicación espacial $\boldsymbol{s}$*, o solamente espacial, como una distribución espacial en un tiempo fijo* t_i *y se compara a lo largo del tiempo. Se evalúa la necesidad de usar modelos de covarianza no separables que incluyan la interacción entre estas dos estructuras. En lo que sigue, se lleva a cabo un análisis exploratorio a través del cual se evalúan la homocedasticidad y la estacionariedad en media del proceso PM10 en la ciudad de Bogotá.*

En la Figura 7.2 se evidencia la diferencia entre los comportamientos del contaminante en los diferentes sectores de la ciudad. Se encuentran valores atípicos en Suba, pero, en general, los valores son muy altos con respecto a lo que indica la norma; se notan mayores varianzas a medida que las medias aumentan, lo que evidencia la presencia de heterocedasticidad. Esta es confirmada por la Figura 7.3 a., donde se muestra una fuerte relación entre la media y la varianza, razón por la cual se lleva a cabo una transformación Box-Cox con $\lambda = -0,1$*; los datos transformados ya no muestran dicha relación (Figura 7.3, b.).*

En la Figura 7.3 c., se muestra un gráfico de dispersión para las $10a.m.$ *con los datos resultantes de la transformación Box-Cox. Las características son similares para diferentes horas. Se puede ver que la media no es constante; para su estimación y debido a que la cantidad de ubicaciones espaciales es pequeña (*14*), se usa el pulimiento de medianas, que es un método resistente (ver Sección 7.8). Se consideran los datos como una tabla a dos vías, en la cual el factor fila corresponde a la hora de observación* (t) *y el factor columna a la estación*

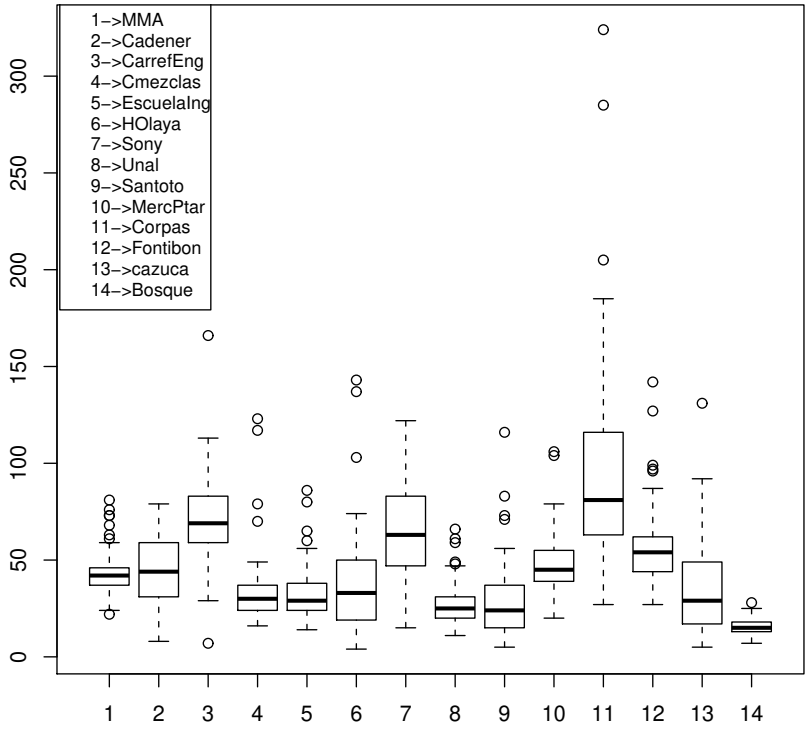


Figura 7.2. PM10 Bogotá.

ambiental (s). *De esta forma, si* $Y(s,t)$ *es la variable que resulta de la transformación Box Cox, esta variable puede ser descompuesta como:*

$$Y(s,t) = \mu(s,t) + Z(s,t). \tag{7.8}$$

y el modelo usado para la tendencia es

$$\mu(s,t) = \nu + \alpha_t + \beta_s. \tag{7.9}$$

donde ν *es la media global, común a todo el proceso; el efecto fila* α_t*, corresponde al efecto de la hora* t *y el efecto columna* β_s *es el efecto correspondiente a la estación ambiental* s*;* $Z(s,t)$ *son los errores del modelo. En el gráfico de Tukey (Figura 7.6), se verifica que no es necesario incluir un término de interacción entre filas y columnas.*

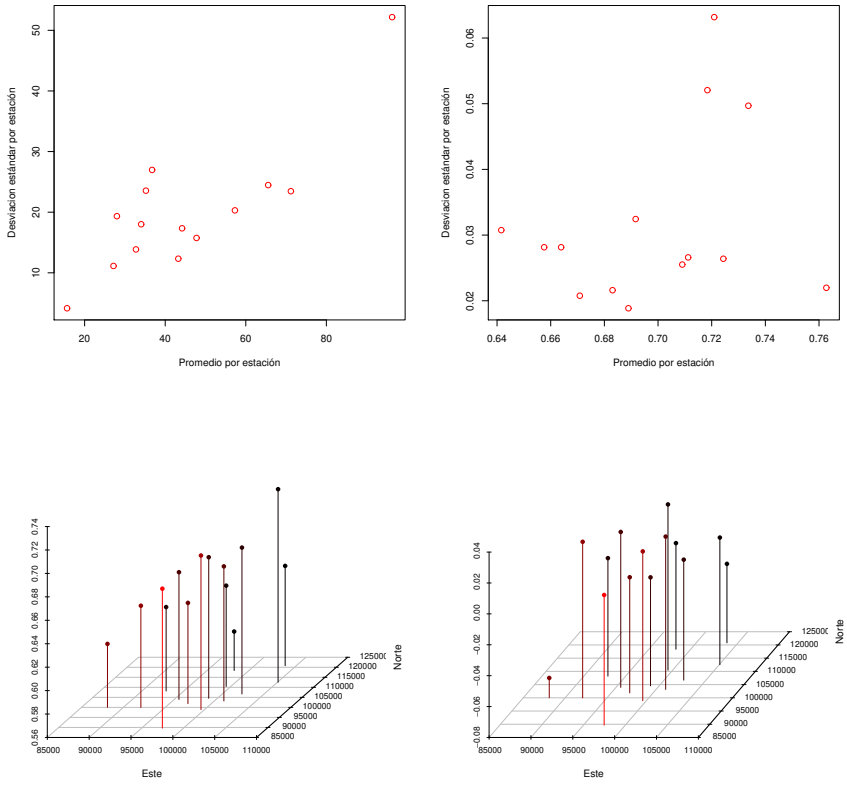


Figura 7.3. a. Promedio vs. Desviación Estándar, b. Promedio vs. Desviación Estándar, datos transformados, c. Dispersión de la variable transformada. d. Dispersión de $\hat{Z}(s,t)$.

Algunas de las series de tiempo resultantes del pulimiento de medianas se muestran en la Figura 7.4. Estas no muestran tendencia ni comportamientos estacionales; espacialmente, tampoco hay indicios de que la media no sea constante.

Se procede al modelamiento de la estructura de dependencia espacio-temporal de $Z(s,t)$. Se lleva a cabo una prueba de separabilidad, [6], para explorar la significancia de la interacción entre espacio y tiempo y verificar si se requiere un modelo de covarianza espacio-temporal no separable. En general, estos procesos muestran fuertes interacciones espacio-tiempo. Nótese que esta prueba se lleva a cabo principalmente con propósitos exploratorios. Se separan los datos tanto temporal como espacialmente en franjas entre las que se pueda garantizar independencia.

Para determinar la mínima separación espacial a la cual se puede suponer independencia, se hace uso de los semivariogramas para cada una de las 73 horas consideradas. En general, la dependencia espacial disminuye en las noches y nuevamente aumenta a medida que el día avanza. En la Figura 7.5 se observa que el alcance práctico de los semivariogramas

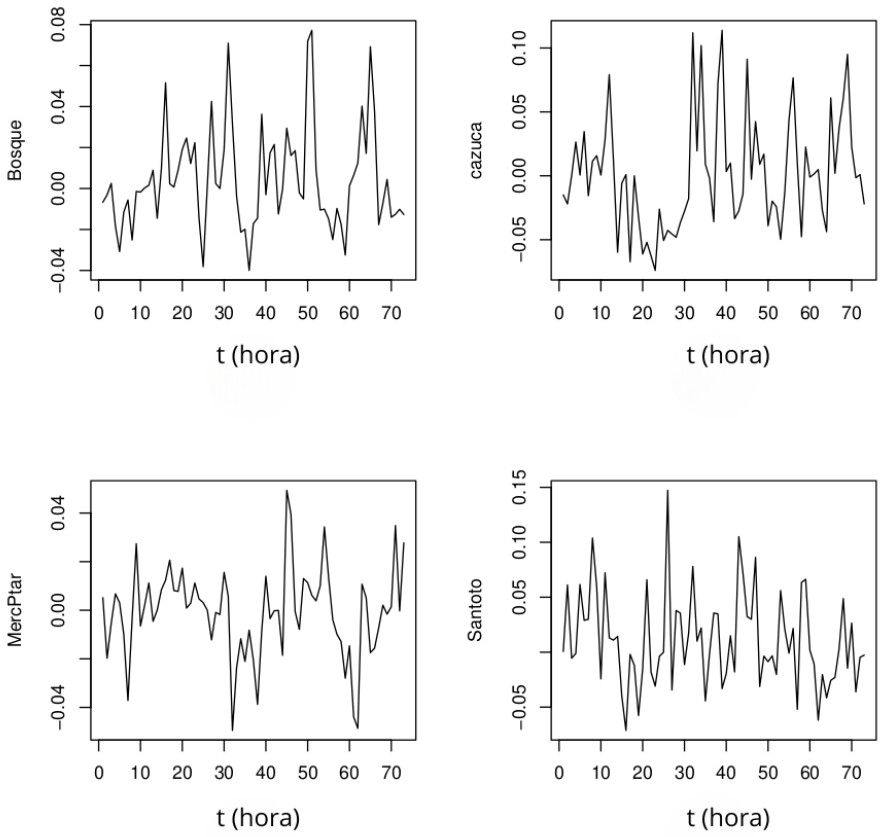


Figura 7.4. Series de tiempo por estación

es a partir de 12.000. Por lo tanto, se pueden establecer tres regiones en Bogotá: norte, centro y sur. Aunque en algunos casos este rango práctico es menor, se toma el mayor para garantizar independencia entre zonas. En este caso, el proceso se asume isotrópico, ya que desafortunadamente no hay suficientes puntos en el espacio que permitan llevar a cabo un análisis satisfactorio de esta propiedad. Se estiman los periodogramas suavizados [13] para un punto de truncamiento $r = 10$*, asignando las ponderaciones a través de la función de Hanning:*

$$\mathbf{W}(t) = \frac{1}{2}[1 + \cos(\pi t/r)] \tag{7.10}$$

con $t = 0, \ldots T - 1$*. Se utiliza el concepto de coherencia que se interpreta como la relación existente entre dos series temporales en distintas frecuencias. Se busca si hay frecuencias en las que se sincronizan las series de tiempo. Para garantizar independencia, se incluyen únicamente las coherencias cada 10 frecuencias; estas se muestran en la Tabla 7.2. En el gráfico de medias de las coherencias estimadas (Figura 7.6) se ilustra la ausencia de*

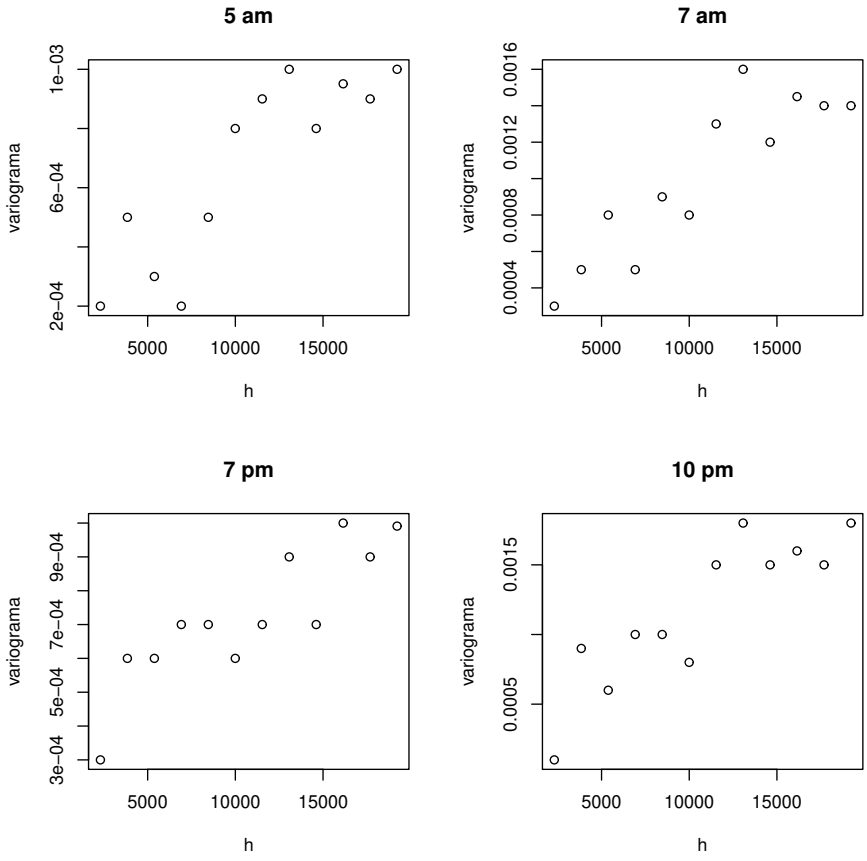


Figura 7.5. Algunos semivariogramas para establecer el alcance espacial.

separabilidad, ya que el valor de la coherencia $\hat{R}_{a,b}(\tau)$ *cambia notoriamente de una a otra frecuencia. Esto se corrobora con un 95% de confianza con el análisis de varianza de la Tabla 7.3, que muestra que la coherencia cambia significativamente con la frecuencia,* ($R^2 = 0,7726$)*. Las series de tiempo más correlacionadas se encuentran ubicadas en las estaciones Cazuca y Sony; las dos se encuentran en zonas con características muy similares, como se detalla en la Tabla 7.1.*

Las coherencias más bajas se encuentran en la zona central, entre las estaciones de Santo Tomás y el Ministerio de Medio Ambiente, que, aunque también se encuentran en zonas de similares características, no son zonas con fuerte presencia de sector comercial o industrial, que son dos de los factores que más influyen en la presencia de PM10.

Frecuencia	Corpas vs. EscuelaIng.	SantoTomas vs. MMA	Cazuca vs. Sony
$\pi/37$	0,71	0,41	0,82
$11\pi/37$	0,71	0,42	0,69
$21\pi/37$	0,68	0,35	0,35
$31\pi/37$	0,53	0,30	0,50
$41\pi/37$	0,60	0,12	0,56
$51\pi/37$	0,66	0,32	0,36
$61\pi/37$	0,32	0,24	0,37
$71\pi/37$	0,46	0,28	0,53

Tabla 7.2. Coherencias ($\hat{R}_{a,b}(\tau)$) PM10 Bogotá

Análisis de separabilidad

Se aplica una prueba de separabilidad utilizando un 95% de confianza para ver si resulta significativa la diferencia de la variable $\phi_{(a,b)_i}(\tau_j)$ entre frecuencias, lo que indicaría que efectivamente se requiere un modelo no separable para la dependencia espacio−tiempo del PM10 en Bogotá. El gráfico de comparación de cuantiles para los residuales se muestra en la Figura 7.6; la transformación hecha para $\phi_{(a,b)_i}(\tau_j)$, ha tenido buenos resultados, [6].

Fuente de variación	gl	Suma cuadrados	Cuadrado medio	Valor F	Pr(>F)
Entre ubicaciones	2	0,24	0,12	13,02	0,0006
Entre frecuencias	7	0,20	0,03	3,07	0,0351
Residuals	14	0,13	0,01		

Tabla 7.3. Prueba de separabilidad PM10 Bogotá, $\hat{\phi}_{(a,b)_i}(\tau_j)$, análisis de varianza bajo supuesto de normalidad

Una alternativa es el uso de la regresión beta. Los mejores resultados se obtuvieron usando la función de enlace log-log, y se encontró un Pseudo R-cuadrado de 0,7806. Se concluye de nuevo ausencia de separabilidad (Tabla 7.4).

Coeficientes	Estimador	Error Std.	valor z	$Pr(>\|z\|)$
(Intercept)	0,38986	0,15112	2,580	0,009884
$\tau.2$	-0,09931	0,18840	-0,527	0,598106
$\tau.3$	-0,53141	0,18380	-2,891	0,003837
$\tau.4$	-0,54012	0,18376	-2,939	0,003291
$\tau.5$	-0,61526	0,18355	-3,352	0,000802
$\tau.6$	-0,54776	0,18374	-2,981	0,002871
$\tau.7$	-0,91930	0,18373	-5,003	5,63e-07
$\tau.8$	-0,59448	0,18360	-3,238	0,001204
$p.2$	-0,74859	0,11081	-6,756	1,42e-11
$p.3$	-0,17610	0,11062	-1,592	0,111380

Tabla 7.4. Prueba de separabilidad PM10 Bogotá, $\hat{R}_{(a,b)_i}(\tau_j)$, regresión beta

Estimación de la función de covarianza espacio-tiempo

Una vez se ha verificado la ausencia de separabilidad, se procede con la estimación de la función de covarianza espacio-tiempo no separable, usando el método de estimación CL. La Tabla 7.5 muestra algunos de los modelos estimados, con sus respectivos criterios de información CLIC.

El caso Gneiting-Matern, mostrado en el Ejemplo 17, genera modelos con igual suavidad en el origen que lejos del origen [9]. Debido a esto, no se realiza estimación del parámetro v; en su lugar, se estima un caso particular para cuando $v = 1/2$. *En este caso, la función queda simplificada según la expresión* (7.11). *Así se obtiene un CLIC de 169794.*

$$C(\|\boldsymbol{h}\|, u) = \frac{\sigma^2}{(au^{2\alpha}+1)^{\beta}} \exp\left(-\frac{c\|\boldsymbol{h}\|}{(au^{2\alpha}+1)^{\beta}}\right) \tag{7.11}$$

Se ajustaron cuatro clases distintas de modelos de [26], así como los de [36] presentados en la Tabla 7.6. Finalmente, el modelo que mejor se ajusta a los datos de material particulado PM10, según el CLIC es

$$C(h, u) = \frac{\sigma^2(a^2u^2+1)}{((a^2u^2+1)^2+b^2h^2)^{3/2}} \tag{7.12}$$

Comparación de modelos

Parámetros	Cressie - Ej. 1	Gneiting Ej. 16	Gneiting Ej. 17, $v = 1/2$
a	1,2824	1,9565	1,2568
s.e. (a)	0,1940	0,3652	0,1456
c	0,3695	0,2989	0,3412
s.e.(c)	0,0523	0,1402	0,0174
α		0,6859	0,7016
s.e.(α)		0,0468	0,0985
β		0,5942	0,5241
s.e.(β)		0,0384	0,0829
γ		0,5124	
s.e.(γ)		0,0193	
σ^2	0,0042	0,0040	0,0036
s.e.(σ^2)	$1,83 \times 10^{-4}$	$2,00 \times 10^{-4}$	$1,92 \times 10^{-4}$
$CLIC$	165841	152523	169794

Tabla 7.5. Desempeño de algunos modelos de covarianza espacio tiempo para PM10 en Bogotá. Parámetros estimados, errores estándar y criterio de información

con un $CLIC$ igual a 150835. La gráfica de suavidad de este modelo presenta buenas características, por lo cual se utilizará como base para algunas predicciones de interés.

Predicción

Por último, se examinan los mapas de predicción para algunas horas. Los datos fueron corregidos por el sesgo generado por la transformación Box Cox, utilizando la expansión de Taylor de segundo orden, kriging transgaussiano, [67]. En las Figuras 7.7 se muestran respectivamente el mapa de predicción del PM10 a las 6 p.m. con sus respectivas varianzas. A esta hora en Suba se alcanzan niveles que superan los $200 \mu g/m^3$. Estos empiezan a descender a las 7 p.m., pero se recuperan al amanecer. En la Figura 7.8 se muestra la predicción espacial entre las 7 p.m. y las 2 a.m.. Se puede ver en los gráficos por hora, que los niveles de $PM10$ están siendo superados, principalmente en las zonas de alto flujo vehicular y de actividad industrial y comercial. Las zonas más afectadas son Suba, Puente Aranda y Fontibón. En la noche, en general, bajan los niveles en toda la ciudad, pero los

Parámetros	a	b	σ^2
Estimaciones	1,11256345	0,49526378	0,00354982
s.e.	0,09894236	0,032519432	$1,76 \times 10^{-4}$

Tabla 7.6. Modelo Cressie-Huang (1999). Ver Ejemplo 4

máximos se desplazan hacia la zona industrial de Fontibón y Puente Aranda. Entre las horas de alto y bajo tráfico vehicular, la diferencia entre concentraciones de PM10 puede variar hasta en $150\mu g/m^3$.

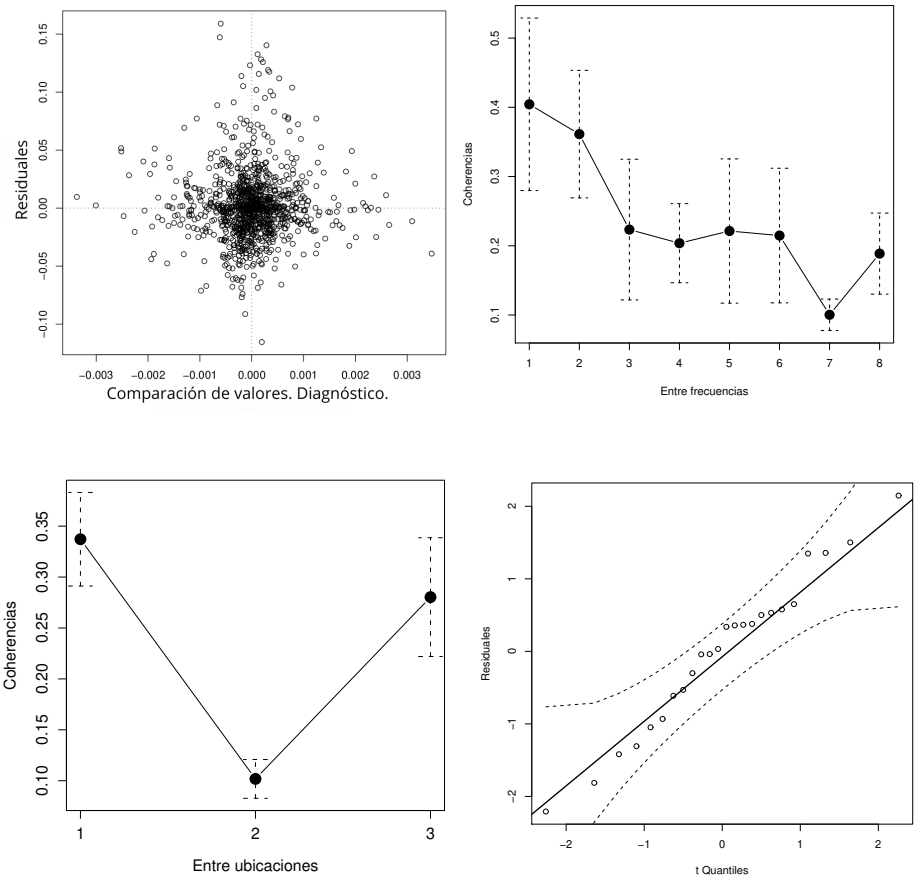


Figura 7.6. Panel superior izquierdo. Gráfico de diagnóstico de aditividad utilizando comparación de cuantiles. Panel superior derecho y Panel inferior izquierdo: coherencias entre frecuencias y entre ubicaciones, respectivamente, $\hat{R}_{a,b}(\tau)$. Panel inferior derecho: gráfico de cuantiles para los residuales de $\hat{\phi}_{(a,b)_i}(\tau_j)$

.

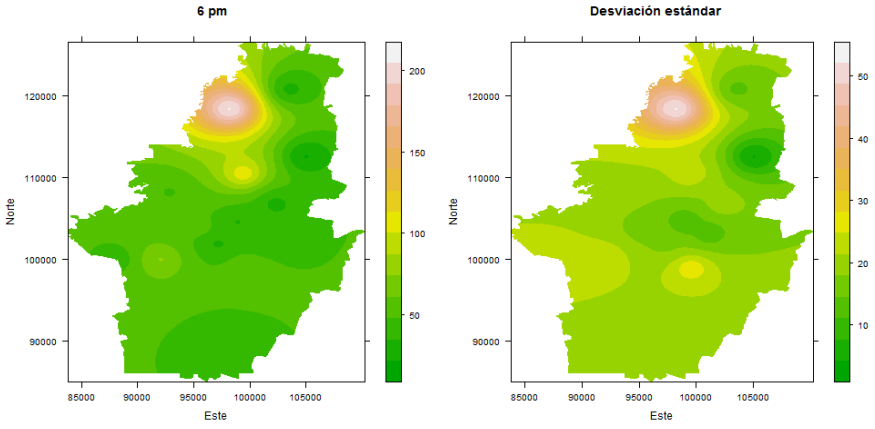


Figura 7.7. Predicción PM10 Bogotá 2024 6 p.m.
a. Kriging de la variable PM10 b. Varianzas estimadas del error predicción 6 p.m.

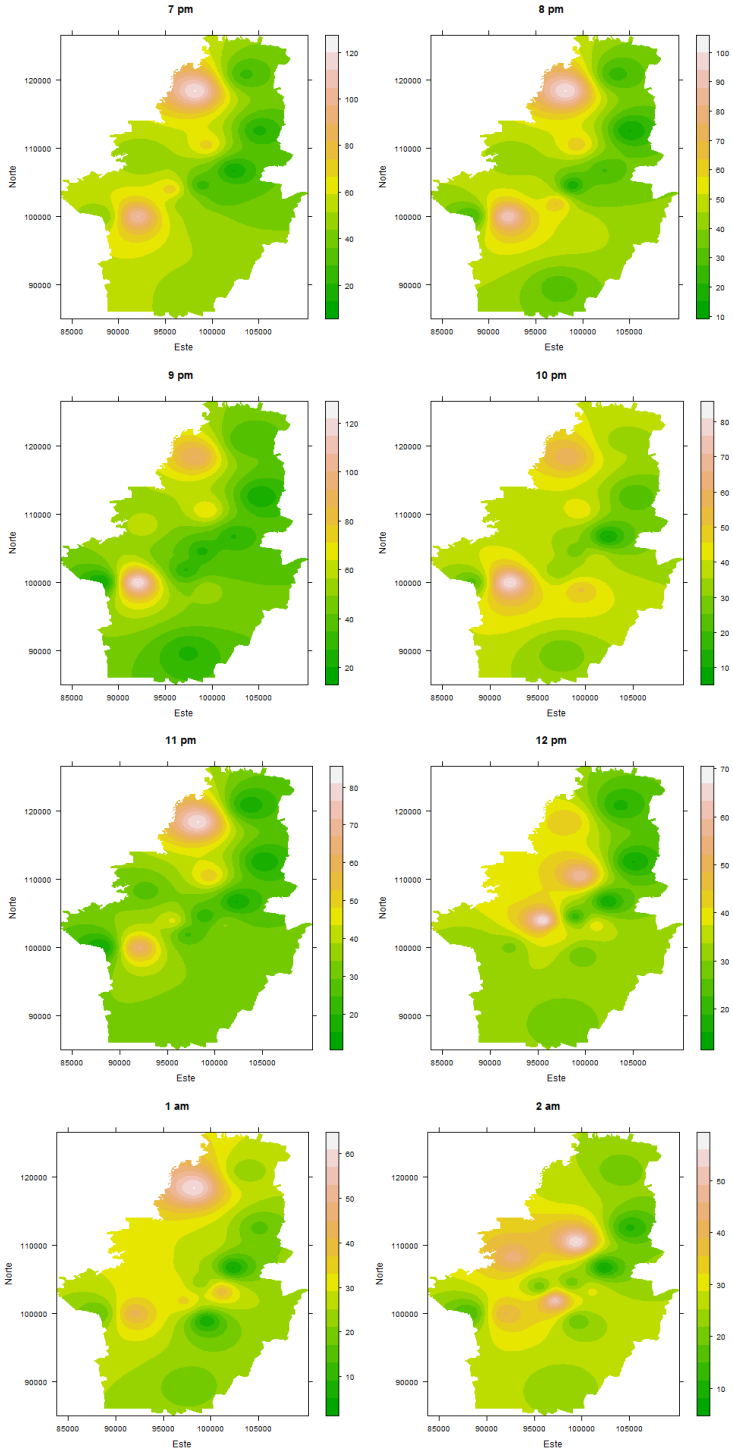


Figura 7.8. Predicción PM10 Bogotá 2024 7 p.m. - 2 a.m.

Parte III

Geoestadística funcional

Capítulo

ocho

Conceptos preliminares de datos funcionales.

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

Se han analizado hasta ahora datos escalares o vectores de datos escalares, correspondientes a realizaciones de variables espaciales aleatorias de valor real. Esto es, las variables que se han considerado son números reales, $Y(\boldsymbol{s})$ o $Y(\boldsymbol{s}, t)$, o vectores de números reales tales como $\mathbb{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))$ y $(\mathbb{Y}_1, \ldots, \mathbb{Y}_P)$ para el caso multivariado.

Los datos funcionales permiten ver los procesos espacio-temporales de una forma más precisa. Específicamente, en este capítulo se considera el caso en el que $T \subset \mathbb{R}$, esto es, la variable aleatoria funcional es una curva. Esta curva o función aleatoria puede tener como argumento el tiempo, la frecuencia o una dimensión espacial como este, norte o profundidad. Si la variación es en dos dimensiones, el argumento es bivariado y la función aleatoria es una superficie, por ejemplo, el argumento puede ser (este,norte). Ejemplos de datos funcionales de superficie son imágenes o mapas. Observe que un valor escalar de temperatura limita el proceso a un punto específico del espacio-tiempo, mientras que hablar de la curva de temperatura, da la visión general del proceso espacial de las series de tiempo de temperatura a través de la zona de interés. En la otra dirección, también es posible hablar de cómo varía el mapa de temperatura de una región a lo largo del tiempo (Ver Figura 8.1). Los datos funcionales constituyen una manera más global de estudiar los procesos espacio-tiempo que varían en dominio continuo.

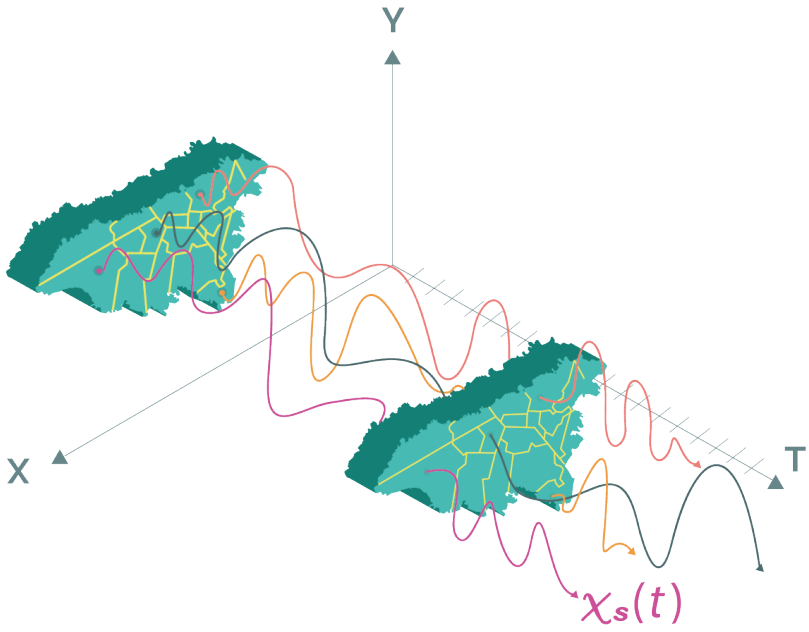


Figura 8.1. Las curvas varían a través del espacio y las superficies espaciales varían a través del tiempo. Usar geoestadística funcional permite predecir las curvas densamente en el espacio para así llenar el volumen y luego extraer superficies de nivel, esto es, mapas en tiempos específicos. También se podrían predecir las superficies densamente en el tiempo y luego extraer curvas en puntos de interés.

Por lo tanto, a partir de este capítulo, la idea principal es que se tiene un objeto aleatorio diferente:

El objeto aleatorio es una función de valor real con uno o más argumentos. Si tiene un argumento, su gráfico es una curva; si tiene dos argumentos, su gráfico es una superficie.

Ejemplo 19 (Sismogramas). *La Figura 8.2 muestra una ilustración de una variable aleatoria funcional. Las curvas suavizadas son sismogramas. Un sismograma es un registro a través del tiempo o de alguna dirección espacial, de movimientos del suelo generados por la energía provocada por un sismo o explosión. La media es la línea gruesa roja, que es calculada de manera usual, se suman las 20 funciones y este resultado se divide en 20.*

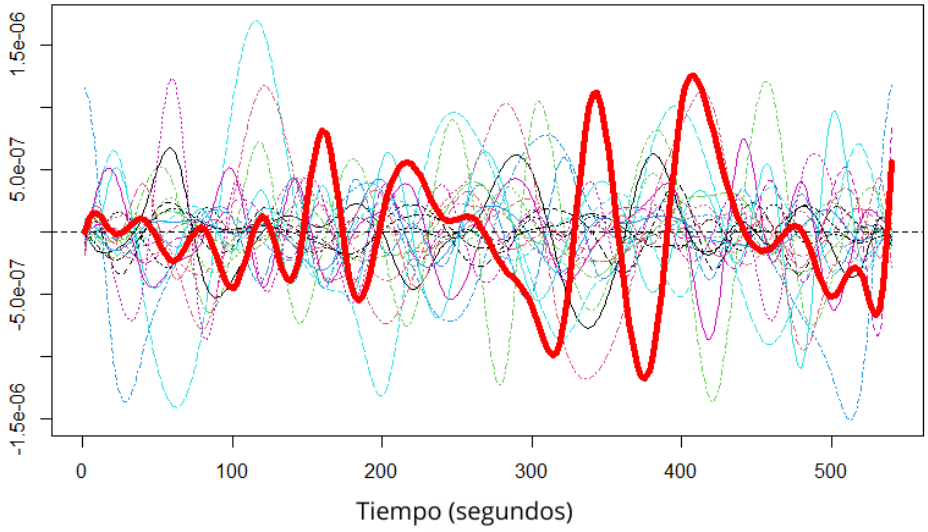


Figura 8.2. Sismogramas. Se presenta el registro del sismógrafo en 20 estaciones de una red sismológica durante los primeros 9 segundos después de ocurrido el sismo. La curva gruesa roja es la mediana funcional.

Ejemplo 20 (Temperatura). *La Figura 8.3 muestra una ilustración de una variable aleatoria funcional. En el panel superior se muestran las observaciones escalares que corresponden a la mediana mensual de la temperatura. En el panel inferior se observan los datos suavizados. La curva roja es la media funcional.*

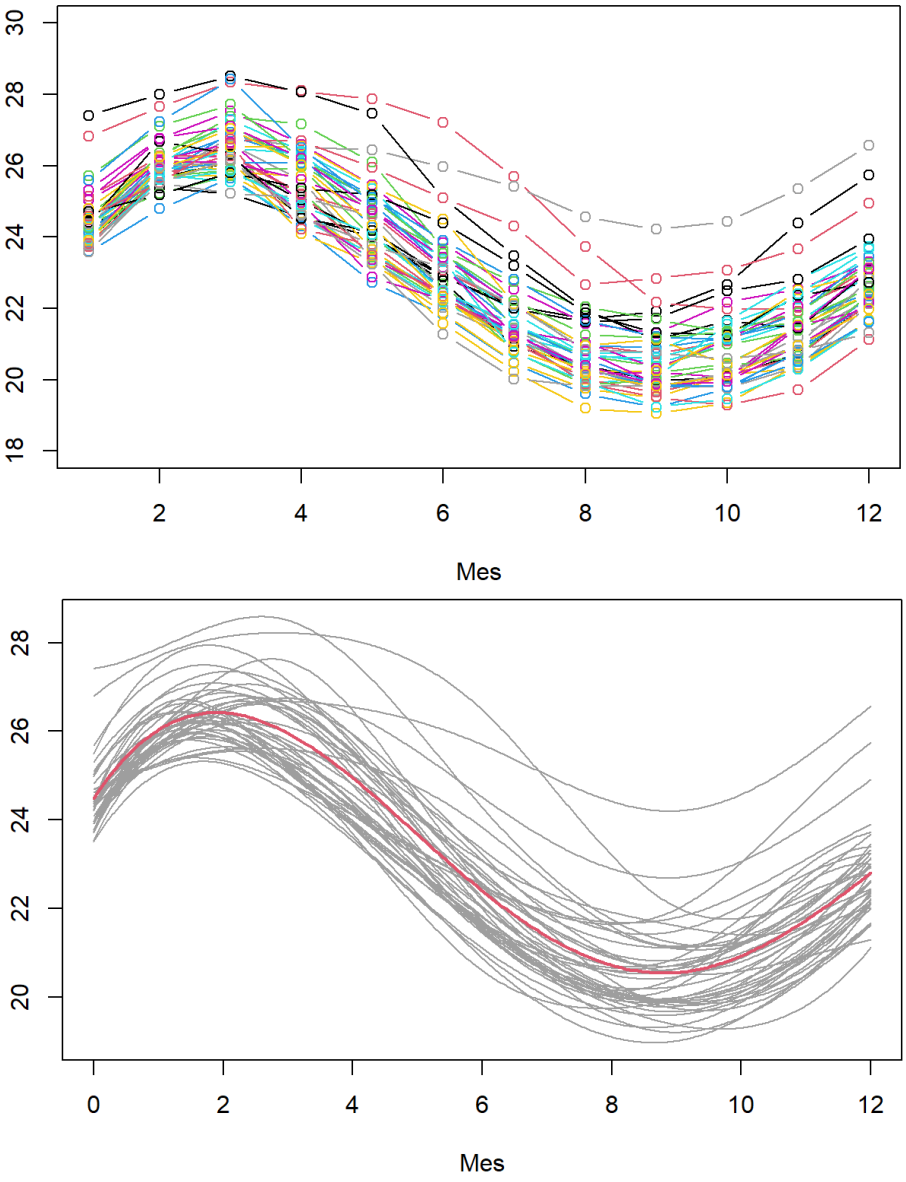


Figura 8.3. Temperatura mensual durante un año. En el panel superior se muestran los datos escalares observados cada mes en cada una de las 40 estaciones. En el panel inferior se muestran las curvas suavizadas y su respectiva media funcional.

A continuación, se presentan los conceptos básicos de los datos funcionales necesarios para desarrollar la geoestadística funcional en los capítulos sucesivos.

8.1. Notación y definiciones básicas

Sea $\boldsymbol{\mathcal{Y}}(t)$ una función definida sobre algún conjunto $\mathbb{T}$. Entonces:

1. Una variable aleatoria $\boldsymbol{\mathcal{Y}}(t)$, $t \in \mathbb{T}$ es llamada variable funcional si toma valores en un espacio infinito dimensional (o espacio funcional) (ver [31] para una exposición más detallada).
2. Una observación $\mathcal{Y}(t)$ de $\boldsymbol{\mathcal{Y}}(t)$, $t \in \mathbb{T}$ es llamada un dato funcional. Se asume que las observaciones están en L^2.
3. Un conjunto de datos funcionales $\mathcal{Y}_1(t), \ldots, \mathcal{Y}_n(t)$ es la realización de n variables aleatorias funcionales, $\boldsymbol{\mathcal{Y}}_1(t), \ldots, \boldsymbol{\mathcal{Y}}_n(t)$ idénticamente distribuidas como $\boldsymbol{\mathcal{Y}}(t)$, la cual es cuadrado integrable.
4. Sea $D_s \subset \mathbb{R}^d$, usualmente $d = 2$, la región espacial en la que es de interés analizar la variable funcional $\boldsymbol{\mathcal{Y}}_s(t)$, $s \in D_s$. Un conjunto de datos funcionales espaciales $\mathcal{Y}_{s_1}(t), \ldots, \mathcal{Y}_{s_n}(t)$ es la observación de n variables funcionales $\boldsymbol{\mathcal{Y}}_{s_1}(t), \ldots, \boldsymbol{\mathcal{Y}}_{s_n}(t)$ en el conjunto $S \subset D_s$, de ubicaciones espaciales
$$S = \{s_1, \ldots, s_n\}$$
Ver Definición 12 para una definición formal de proceso espacial funcional. En este trabajo los objetos funcionales considerados son curvas, así que se considera el caso específico en el que los objetos están en $L^2(\mathcal{B})$, $\mathcal{B} \subset \mathbb{R}$, es decir, los conjuntos $\mathcal{B}$ son los conjuntos borelianos.

El modelo geoestadístico en el caso funcional, se asume aditivo como en el caso escalar. Sea $E\left[\boldsymbol{\mathcal{Y}}_s(t)\right] = \mu(t)$,

$$\boldsymbol{\mathcal{Y}}_s(t) = \mu(t) + \boldsymbol{\chi}_s(t) \tag{8.1}$$

La función media $\mu(t)$ en una primera etapa, es estimada usando la media muestral

$$\hat{\mu}(t) = \overline{\mathcal{Y}}_s(t) = n^{-1} \sum_{i=1}^{n} \mathcal{Y}_{s_i}(t).$$

si se considera constante en la región o posiblemente usando un modelo de regresión adecuado si se observa alguna tendencia. Sin embargo, después de haber estimado la covarianza entre funciones, se puede mejorar la estimación de la media involucrando esta información. Ver Sección 9.2 y expresión 9.10. En este capítulo se usa el proceso espacial centrado $\boldsymbol{\chi}_s(t)$ para las definiciones y el desarrollo de los métodos, dado que al considerar $\mu(t)$ determinístico, se tiene que

$$\text{Var}\left[\boldsymbol{\mathcal{Y}}_s(t)\right] = \text{Var}\left[\boldsymbol{\chi}_s(t)\right]$$

y con este proceso se modela la estructura de covarianza y se encuentra la predicción. La media o su estimación se suma a las predicciones obtenidas.

Sean $\xi, \zeta \in \mathcal{H}$, $\boldsymbol{\xi}, \boldsymbol{\zeta} \in \mathcal{H}^n$, $b \in \mathbb{R}$ y $\langle \boldsymbol{\xi}, \boldsymbol{\zeta} \rangle$ el producto interno usual en $L^2(\mathcal{B})$. La suma, la multiplicación por un escalar y el producto interno para los elementos de $\mathcal{H}^n$ están definidos como

$$\begin{aligned} \boldsymbol{\xi} + \boldsymbol{\zeta} &\equiv (\xi_1 + \zeta_1, \ldots, \xi_n + \zeta_n) \\ b\boldsymbol{\zeta} &\equiv (b\zeta_1, \ldots, b\zeta_n) \\ [\boldsymbol{\xi}, \boldsymbol{\zeta}] &= \langle \xi_1, \zeta_1 \rangle + \ldots + \langle \xi_n, \zeta_n \rangle \end{aligned} \tag{8.2}$$

$$\langle \xi, \zeta \rangle = \int \xi(t)\zeta(t)dt \tag{8.3}$$

8.2. Construcción de los datos funcionales

Las funciones que se consideran aquí varían en tiempo y espacio continuo. Sin embargo, solo pueden ser observadas en un número finito de valores. Usualmente se tiene un conjunto de T observaciones puntuales escalares en cada ubicación espacial $\boldsymbol{s}$. Por lo tanto, es necesario ajustar un modelo para reconstruir la función $\mathcal{Y}_{\boldsymbol{s}_i}(t)$ a partir de sus observaciones en T puntos

$$\{y_{s_i}(t_1), \ldots, y_{s_i}(t_T)\}.$$

Las observaciones son como se ilustra en la expresión 8.4. No es necesario que las observaciones estén regularmente espaciadas ni en el tiempo ni en el espacio y usualmente hay datos faltantes.

Como ilustración, en caso de que no haya datos faltantes y se tenga la misma cantidad T de observaciones puntuales escalares en cada ubicación espacial s_i, la configuración se muestra en la matriz 8.4. Cada elemento de la matriz es un escalar. Cada columna corresponde a una ubicación $s_i, i = 1, \ldots, n$, y cada fila a un tiempo específico t_j, $j = 1, \ldots, T$, $t_j \in \mathbb{T} = [0, T]$ y $T \in \mathbb{R}^+$.

$$\begin{pmatrix} y_{s_1}(t_1) & y_{s_2}(t_1) & \ldots & y_{s_n}(t_1) \\ y_{s_1}(t_2) & y_{s_2}(t_2) & \ldots & y_{s_n}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ y_{s_1}(t_T) & y_{s_2}(t_T) & \ldots & y_{s_n}(t_T) \end{pmatrix} \tag{8.4}$$

Los datos funcionales no requieren mediciones regularmente espaciadas ni en el tiempo ni en el espacio. La Tabla 6.1 muestra las características más comunes de los datos espacio-temporales.

El modelo puede ser paramétrico $\mathcal{Y}(t, \hat{\boldsymbol{\beta}}_i)$ $i = 1, \ldots, n$, como en [38], o no-paramétrico, como en [63], usando bases de funciones. Este último procedimiento es mas adecuado debido a que en presencia de una gran cantidad de datos no es posible encontrar parámetros fijos para todo el intervalo en el que ocurre la función.

Una base de funciones es un conjunto de funciones conocidas $\phi_1, \ldots, \phi_K$ que son matemáticamente independientes dos a dos, y tales que una combinación lineal de un número suficientemente grande K de estas funciones aproxima arbitrariamente bien cualquier curva del espacio, [63].

El conjunto de funciones base aproxima una función $\mathcal{Y}_{\boldsymbol{s}_i}(t)$ usando una expansión truncada así:

$$\mathcal{Y}_{\boldsymbol{s}_i}(t) \approx \sum_{k=1}^{K} a_k(\boldsymbol{s}_i)\phi_k(t) = \boldsymbol{\phi}'(t)\boldsymbol{a}(\boldsymbol{s}_i) \tag{8.5}$$

- $\boldsymbol{\phi}'(t) = (\phi_1(t), \ldots, \phi_K(t))$ es el vector con las K funciones base evaluadas en t.
- Los $a_k(\boldsymbol{s}_i)$, $k = 1, \ldots, K$ son los respectivos coeficientes de las K funciones base $\phi_k(t)$, para el dato $\mathcal{Y}_{\boldsymbol{s}_i}(t)$.
- $\boldsymbol{a}'(\boldsymbol{s}_i) = (a_1(\boldsymbol{s}_i), \ldots, a_K(\boldsymbol{s}_i))$ es el vector con los K coeficientes para reconstruir $\mathcal{Y}_{\boldsymbol{s}_i}(t)$.

Así, el objetivo es utilizar los valores discretos para encontrar la función de la cual provienen. Si se puede asegurar que no hay término de error asociado, se puede interpolar para construir la curva. Sin embargo, si se considera que puede existir algún error en la observación, que es lo usual, entonces el paso de datos escalares a funciones continuas se enmarca en los procesos de suavizamiento.

En consecuencia, se requiere una estrategia eficiente para construir funciones con parámetros sencillos de estimar y con capacidad de ajustarse a las características de las curvas, mientras se mantiene la parsimonia. Por ejemplo, para el caso de la temperatura se utiliza una B-spline de orden 4.

Por lo tanto, asumiendo como modelo

$$\underset{\sim}{y}_{\boldsymbol{s}_i}(t_j) = \mathcal{Y}_{\boldsymbol{s}_i}(t_j) + \upsilon_{\boldsymbol{s}_i}(t_j) = \boldsymbol{\phi}'(t_j)\boldsymbol{a}(\boldsymbol{s}_i) + \upsilon_{\boldsymbol{s}_i}(t_j) \tag{8.6}$$

El objetivo es encontrar el vector $\boldsymbol{a}(\boldsymbol{s}_i)$ que minimice la expresión

$$\sum_{j=1}^{T} \left(\underset{\sim}{y}_{\boldsymbol{s}_i}(t_j) - \boldsymbol{\phi}'(t_j)\boldsymbol{a}(\boldsymbol{s}_i)\right)^2 \tag{8.7}$$

En el proceso de suavizamiento es fundamental tener en cuenta la rugosidad. Por lo tanto, se suele incluir en la minimización un parámetro de penalización para disminuir la variabilidad del ajuste. El número de funciones base K y el coeficiente λ se estiman a través de procedimientos de validación cruzada. Uno de los criterios mas usados para la penalización es el de la segunda derivada como ilustra la

ecuación de validación cruzada generalizada GCV en la expresión 8.8, ver [63]. La Figura 8.4 muestra la optimización del parámetro para los datos de temperatura del ejemplo 8.3. Tomando los 40 años reportados (de 1982 a 2021), se evidencia que el punto que minimiza la estadística GCV (8.8) se encuentra alrededor de 0.017 con valores muy superiores antes de 0.01 y después de 0.04. Este valor minimiza la suma del criterio GCV de todas las 40 curvas que se encuentran dentro de la muestra.

$$\text{GCV} = \sum_{j=1}^{T} \left(y_{s_i}(t_j) - \mathcal{Y}_{s_i}(t_j)\right)^2 + \lambda \int_t \mathcal{Y}''_{s_i}(t)dt \tag{8.8}$$

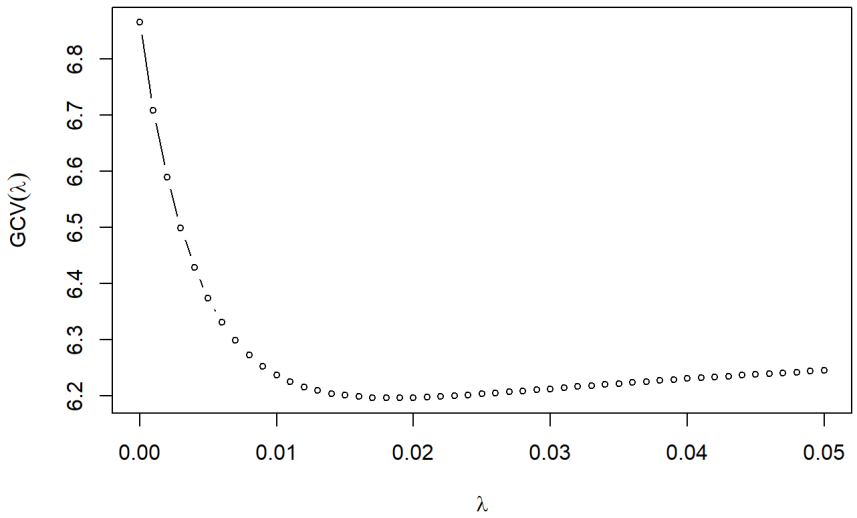


Figura 8.4. Optimización de la expresión (8.8) para la selección del parámetro de suavizamiento λ.

Para datos con comportamientos periódicos, se suele utilizar la base de funciones de Fourier y para datos que no presentan comportamientos periódicos se utilizan otras bases tales como B-splines o polinomios de Legendre por nombrar algunos (ver Figuras 8.5 a 8.8). Cuando los datos presentan comportamientos bruscos o discontinuidades es mas aconsejable usar onduletas, (ver Figura 8.7).

Es importante enfatizar que este paso de suavizamiento es necesario porque no es posible medir directamente las funciones sino una versión discretizada de ellas, ver expresión (8.4). Solo se cuenta hasta el momento con dispositivos que miden punto a punto, es decir, miden datos escalares que luego son usados para reconstruir a través de regresión, ya sea paramétrica o no paramétrica, los datos completos de donde provienen las curvas o datos funcionales. Como criterio de selección se puede minimizar el cuadrado medio residual para no causar sobreajuste.

En algunos casos como en los electrocardiogramas, sismogramas o acelerogramas, el aparato dibuja el gráfico pero no los datos ni ninguna herramienta usada para ajuste, así que es necesario reconstruir la información

desde esta representación gráfica (ver Figura 8.2). Las figuras 8.5 a 8.7 muestran algunos elementos de tres bases de funciones ortonormales: Fourier, Legendre y dos clases de onduletas (ver [20] y [57]). Las onduletas, descomponen señales en sus componentes de frecuencia, a través de traslaciones y dilataciones de funciones específicas, generando familias de bases ortonormales.

En esta etapa en la que se construyen los datos funcionales, no es necesario que las bases sean ortonormales. Posteriormente, para la representación de las funciones en términos de sus componentes principales funcionales, si se requiere que las bases utilizadas cumplan esta propiedad (ver Sección 8.3). Esto no es una restricción ya que es posible convertir cualquier base a una ortonormal aplicando el proceso de Gram-Schmidt, [40]. Por ejemplo, la Figura 8.8 muestra la base de funciones B-spline que aunque no es ortonormal es una de las bases mas usadas para la construcción de los datos funcionales por su versatilidad, Ver [16].

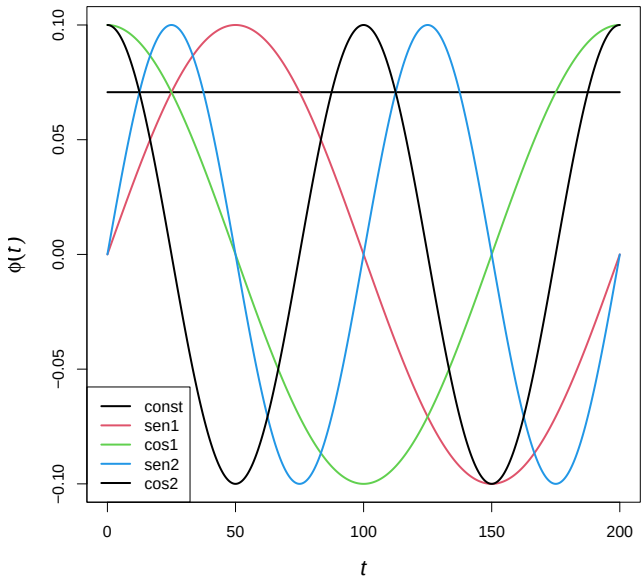


Figura 8.5. Primeras 5 funciones de la base ortonormal de Fourier. Se suele utilizar esta base cuando los datos presentan comportamientos periódicos.

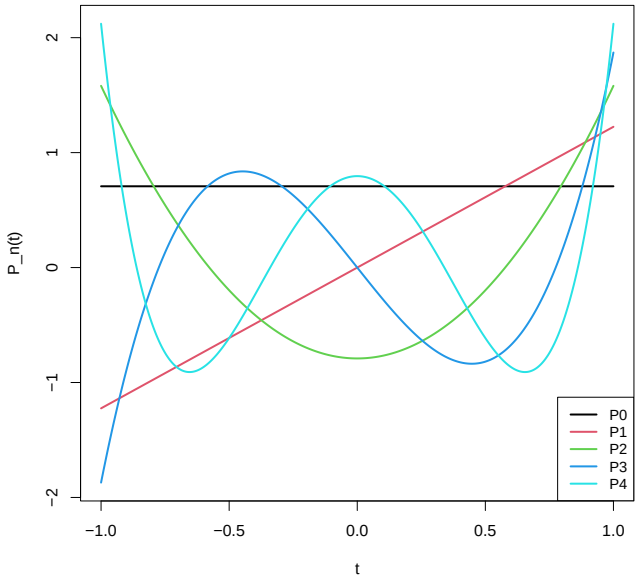


Figura 8.6. Primeras 5 funciones de la base ortonormal de polinomios de Legendre.

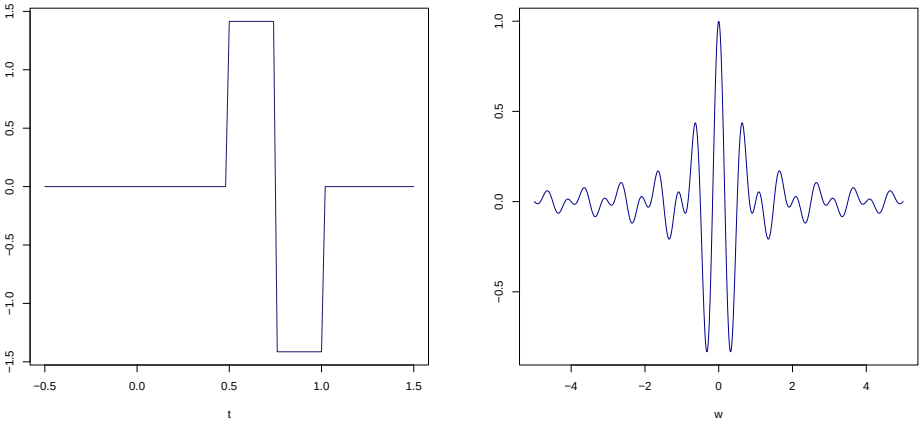


Figura 8.7. Dos ejemplos de onduletas. Panel izquierdo: Onduleta de Haar. Panel derecho: Onduleta de Shannon, parte real.

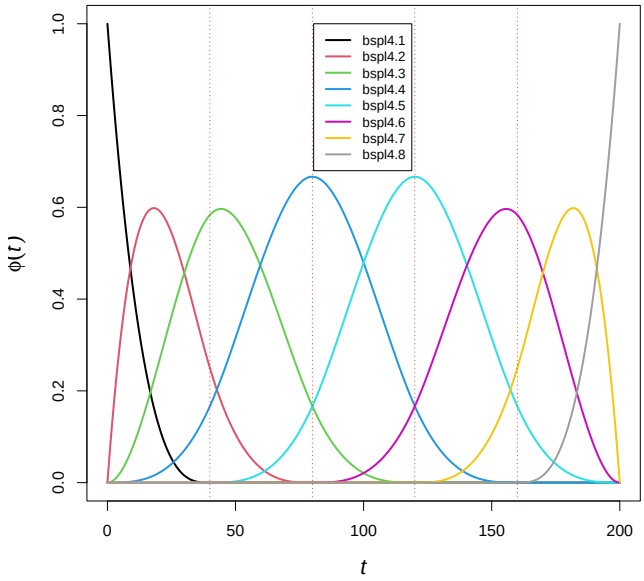


Figura 8.8. Primeras 8 bases B-splines de orden 4. Los nodos son los puntos donde se unen los segmentos polinomiales. Se debe cumplir la relación #bases = orden del polinomio + #nodos - 2. También se usa parámetro de suavizamiento.

Ejemplo 21. Estudio suelos para el metro subterráneo. *Se presenta la aplicación de la metodología propuesta para un estudio de caso en Bogotá. Actualmente, la ciudad se encuentra desarrollando su primera línea de metro. Se han llevado a cabo varios estudios para su diseño y construcción. Uno de ellos fue el proyecto de metro de 2015, que llevó a cabo varios sondeos de exploración del suelo a lo largo de un corredor de línea de metro propuesto (ver Figura 8.3).*

La exploración del suelo realizada incluyó tanto perforaciones tradicionales como pruebas de laboratorio y una colección bastante extensa de pruebas de penetración por cono (CPTu). Los datos recuperados son únicos, ya que son los primeros de su tipo en alcanzar este nivel de detalle a escala regional (ver [59]).

Nótese que en este caso las curvas no varían en función del tiempo, sino en función de la profundidad a la cual se toman los datos de suelo. El interés aquí es predecir las curvas de las características del suelo en profundidad para reconstruir todo el perfil y dar recomendaciones para las obras de ingeniería posteriores.

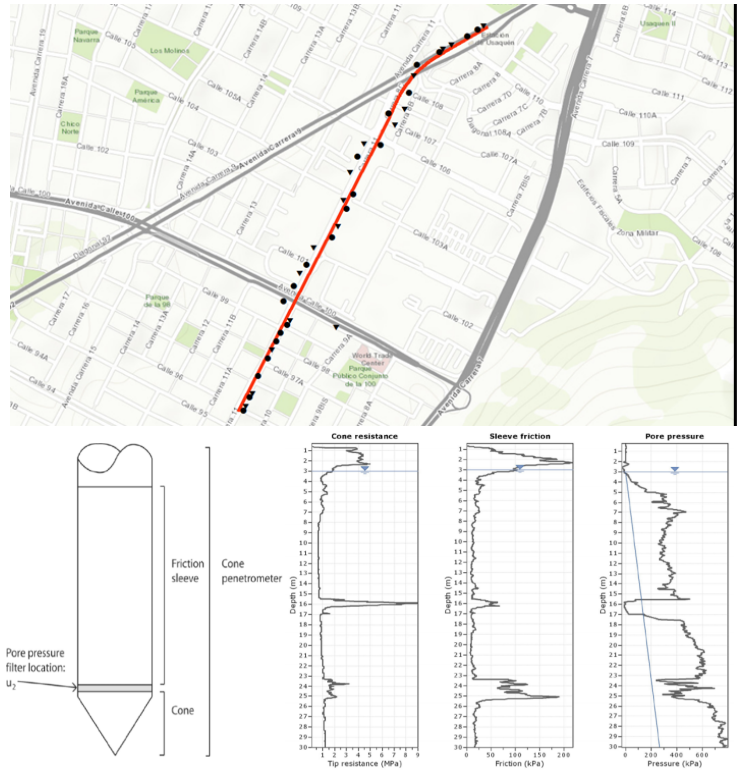


Figura 8.9. Superior: distribución de los dispositivos para las pruebas CPT en el área de estudio en una parte del corredor del Metro. Inferior: curvas suavizadas de los datos medidos en profundidad con el dispositivo.

8.3. Componentes principales funcionales

En esta sección se revisa el marco teórico de los componentes principales funcionales que son una herramienta fundamental para los desarrollos que se proponen en este trabajo en geoestadística funcional. Los conceptos son análogos al caso de componentes principales en el caso multivariado escalar, solo que las variables aleatorias ahora son funciones infinito-dimensionales y se requiere el operador de covarianza. Los valores propios y los puntajes son escalares y los vectores propios son funciones propias. Asumiendo que los campos aleatorios espaciales funcionales son elementos aleatorios de $L^2(\mathcal{B})$ y que $E[\chi_s(t)] = 0$, para casi todo $t \in T$, entonces el operador de covarianza C de $\chi_s(t)$ es

$$C(y) = E[\langle \chi_s, y \rangle \chi_s] \qquad y \in L^2(\mathcal{B}) \tag{8.9}$$

Así,

$$C(y)(t') = \int c(t', t)y(t)dt, \quad \text{donde} \quad c(t', t) = E[\chi_s(t')\chi_s(t)]$$

con estimadores dados por

$$\hat{C}(y) = \frac{1}{n}\sum_{i=1}^{n}\left(\langle \chi_{s_i}, y \rangle \chi_{s_i}\right), \qquad y \in L^2(\mathcal{B})$$

y

$$\hat{C}(y)(t') = \int \hat{c}(t', t)y(t)dt, \quad \text{donde} \quad \hat{c}(t', t) = \frac{1}{n}\sum_{i=1}^{n}\chi_{s_i}(t')\chi_{s_i}(t)$$

Un operador continuo y acotado C sobre $\mathcal{H}$ es un operador de covarianza si y solamente si es simétrico, definido positivo, y sus valores propios η_k satisfacen $\sum_{k=1}^{\infty}\eta_k < \infty$, ver [44]. Los componentes principales funcionales (FPC) se definen como las funciones propias del operador de covarianza $\{\xi_k,\ k = 1, \ldots\}$ (8.9) (Ver [44]). Los estimadores de los FPC son llamados *componentes principales funcionales empíricos* (EFPC). Dado que el principal interés es la reconstrucción de la curva χ_{s_i}, una selección razonable para el sistema de funciones base es la de EFPC formada por las funciones propias $\xi_k,\ k = 1, \ldots, K$ del operador de covarianza C de χ_s con coeficientes dados por los puntajes $f_k(s_i)$ asociados a los componentes principales, y que se definen como

$$f_k(s_i) = \left\langle \chi_{s_i}, \xi_k \right\rangle, \qquad k = 1, \ldots, K, \quad i = 1, \ldots, n \tag{8.10}$$

de acuerdo a [44], la aproximación de esta base es uniformemente óptima en el sentido de minimizar $\hat{\delta}^2$ dado por

$$\hat{\delta}^2 = \sum_{i=1}^{n}\left\|\chi_{s_i}(t) - \sum_{k=1}^{K} f_k(s_i)\xi_k(t)\right\|^2. \tag{8.11}$$

Denotando por η_k el correspondiente valor propio, se selecciona K tal que asegure un porcentaje mínimo de variabilidad acumulada, previamente establecido. El usuario decide el umbral, aunque el más frecuente es el 85 %.

Usando en el modelo geoestadístico la expansión de Karhunen-Loève [12], se asume el siguiente modelo

$$\mathcal{Y}_s(t) = \mu(t) + \chi_s(t) = \mu(t) + \sum_{k=1}^{\infty} f_k(\boldsymbol{s})\xi_k(t), \quad \mathcal{Y}_s(t) \in L^2(\mathcal{B}) \tag{8.12}$$

donde $E\left[\mathcal{Y}_s(t)\right] = \mu(t)$. Es importante notar que se puede aplicar el método de componentes principales funcionales a una sola variable aleatoria funcional, dado que se aplica en t, argumento de la función y en el cual las funciones son infinito-dimensionales. Es decir, la reducción de la dimensionalidad es llevada a cabo en la dimensión temporal, o de frecuencia o de profundidad, según sea el caso. Por lo tanto, existen infinitos componentes principales, y es necesario truncar esta descomposición en los primeros K componentes, según el criterio elegido. En consecuencia, cuando la autocorrelación en el eje t es fuerte, es posible tener una muy buena aproximación usando solo unos pocos EFPC. Además como se observa en la Definición 10, los puntajes resultantes del proceso de los componentes principales funcionales forman un campo aleatorio espacial. En general, aún cuando existan varias variables aleatorias funcionales, una posibilidad es llevar a cabo los FPC por separado y posteriormente analizar las propiedades entre los elementos obtenidos de cada descomposición.

Definición 10 (El campo aleatorio espacial de los puntajes). *Para k, $k = 1, \ldots, K$ y $\boldsymbol{s} \in D_s$, $f_k(\boldsymbol{s})$ es un campo aleatorio escalar espacial, con realizaciones en las ubicaciones $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. Así, el correspondiente vector asociado a cada componente principal k es*

$$\boldsymbol{f}_k(\boldsymbol{s}) = (f_k(\boldsymbol{s}_1), \ldots, f_k(\boldsymbol{s}_n))$$

y si se seleccionan K componentes principales se obtiene el campo aleatorio escalar espacial K-dimensional

$$\mathbb{F} = (\boldsymbol{f}_1(\boldsymbol{s}_1), \ldots, \boldsymbol{f}_K(\boldsymbol{s}_n)) .$$

Además, por construcción los puntajes son campos aleatorios espaciales escalares centrados dado que

$$E\left[f_k(\boldsymbol{s}_i)\right] = E\left\langle \chi_{s_i}, \xi_k \right\rangle = \langle 0, \xi_k \rangle = 0 \tag{8.13}$$

Si bien el supuesto de normalidad no es necesario en este contexto para la predicción, si es fundamental para el proceso de simulación. A continuación se presenta la definición de campo aleatorio conjuntamente gaussiano.

Definición 11 (Campo aleatorio conjuntamente gaussiano). *$\boldsymbol{\Xi}_s \in \mathcal{H}^P$ es un campo aleatorio conjuntamente gaussiano que toma valores en $\mathcal{H}^P$, si la variable real,*

$$[\boldsymbol{\Xi}_s, \boldsymbol{\zeta}] = \langle \boldsymbol{\chi}_s^1, \zeta^1 \rangle + \ldots + \langle \boldsymbol{\chi}_s^P, \zeta^P \rangle \tag{8.14}$$

es gaussiana para todo $\boldsymbol{\zeta} \in \mathcal{H}^P$ *(Ver [11]). Sean* $\xi_1^p, \ldots, \xi_{K_p}^p$, $p = 1, \ldots, P$ *las primeras* K_p *funciones propias del operador de covarianza de* $\chi_{\boldsymbol{s}}^p$, *entonces de acuerdo con la notación en* (8.10), *los puntajes correspondientes son*

$$f_{Kp}^p(\boldsymbol{s}) = \left\langle \chi_{\boldsymbol{s}}^p, \xi_{K_p}^p \right\rangle. \tag{8.15}$$

Para $b_{11}, \ldots, b_{PK_P}$ *números arbitrarios reales, se define el vector* $\boldsymbol{\zeta}$ *como*

$$\boldsymbol{\zeta} = \left(b_{11}\xi_1^1 + \ldots + b_{1K_1}\xi_{K_1}^1, \ldots, b_{P1}\xi_1^P + \ldots + b_{PK_p}\xi_{K_P}^P \right)$$

así, se tiene que

$$[\boldsymbol{\Xi}_{\boldsymbol{s}}, \boldsymbol{\zeta}] = b_{11}f_1^1(\boldsymbol{s}) + \ldots + b_{1K_1}f_{K_1}^1(\boldsymbol{s}) + \ldots + b_{P1}f_1^P(\boldsymbol{s}) + \ldots + b_{PK_P}f_{K_P}^P(\boldsymbol{s}) \tag{8.16}$$

es una variable aleatoria gaussiana real, y por lo tanto el vector

$$\left(f_1^1(\boldsymbol{s}), \ldots, f_{K1}^1(\boldsymbol{s}), \ldots, f_1^P(\boldsymbol{s}), \ldots, f_{K_P}^P(\boldsymbol{s}) \right)$$

es un campo aleatorio gaussiano multivariado en $\mathbb{R}^{K_1+\ldots+K_P}$, *ver [18].*

Ejemplo 22. *En la Ciudad de Bogotá, se registran cada hora varios contaminantes del aire, entre ellos está el material particulado por debajo de 10 micras (PM10). Las estaciones activas suelen ser alrededor de nueve. En algunas épocas pueden estar en funcionamiento algunas más pero como es común en estos casos, pueden descomponerse y dejar de medir o quedar fuera de control en cualquier momento. Sin embargo, en los escasos lugares donde si funcionan las estaciones, los datos a lo largo del tiempo son una cantidad considerable. La Tabla 8.10 panel izquierdo, muestra la red de estaciones y el panel derecho muestra las curvas de la concentración de PM10 para una semana. Nótese lo eficiente que es un dato funcional para mostrar el comportamiento de la variable de interés durante el intervalo de tiempo observado. En cada ubicación durante la semana se registran 168 datos escalares, que son una visión muy parcial del verdadero dato que es la curva en función del tiempo, del PM10 durante la semana.*

- $\mathcal{Y}_{\boldsymbol{s}}(t)$*: Curva de PM10,* D_s *es Bogotá.* $\boldsymbol{s} \in D_s$.
- *Vector aleatorio funcional espacial observado:* $\Upsilon_S = \left(\mathcal{Y}_{\boldsymbol{s}_0}(t), \ldots, \mathcal{Y}_{\boldsymbol{s}_9}(t)\right)$
- *Modelo geoestadístico:* $\mathcal{Y}_{\boldsymbol{s}}(t) = \mu(t) + \chi_{\boldsymbol{s}}(t),\ t \in [0, 168],\ \boldsymbol{s} \in D_s$

Se establece un umbral de varianza acumulada para la selección del número de dimensiones y a continuación se aplica el método de componentes principales funcionales (Ver Sección 8.11). La Figura 8.11 contiene los primeros dos componentes principales funcionales. Estas son las dos primeras funciones propias que son los correspondientes vectores propios en este espacio funcional infinito-dimensional.

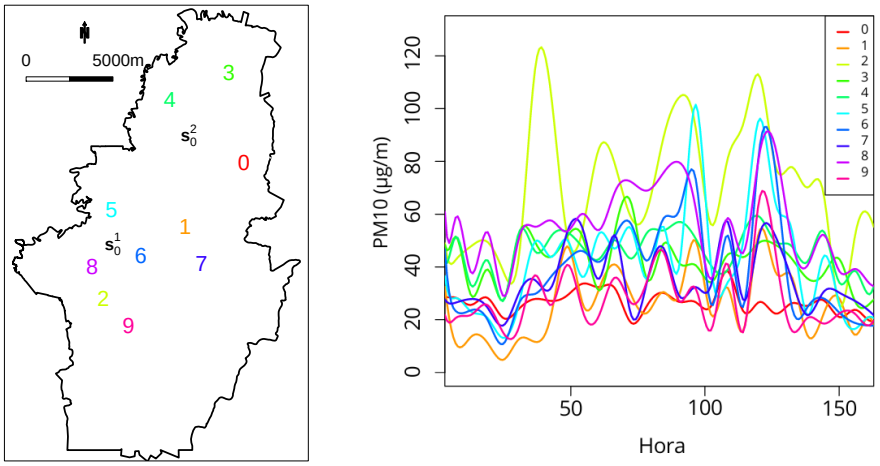


Figura 8.10. **Panel izquierdo**: red de estaciones que miden la calidad del aire en Bogotá. ID: 0: Bosque, 1: IDRD, 2: Sony, 3: Guaymaral, 4: Corpas,5: Fontibón, 6: PteAranda, 7: MAVDT, 8: Kennedy, 9: Tunal **Panel derecho**: curvas temporales de PM10, Mayo 13 a Mayo 20, 2023

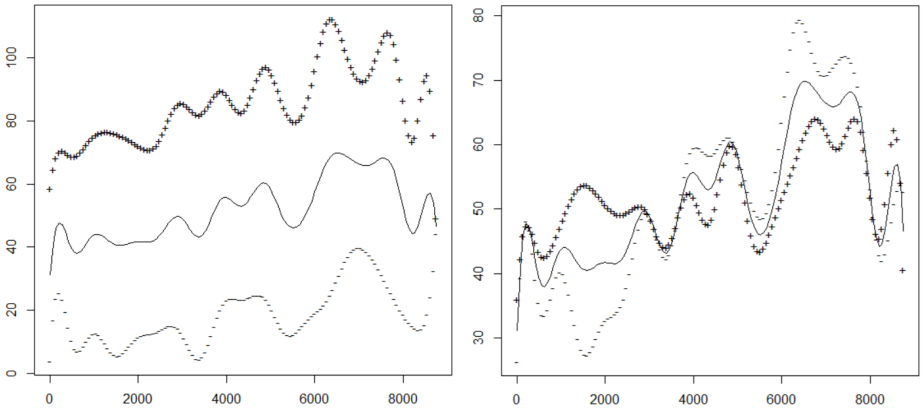


Figura 8.11. La línea sólida es la media funcional de PM10. Se le suman y restan múltiplos de las curvas de los componentes principales para ver el efecto de cada uno. **Panel izquierdo**: El primer componente principal del PM10 acumula una varianza del 94.1 %. **Panel derecho** El segundo componente principal del PM10 acumula una varianza del 2.5 %.

Capítulo

nueve

Predicción espacial funcional univariada. Kriging funcional

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

El principal objetivo de la geoestadística funcional univariada es predecir curvas en lugares no observados, con base en la estimación de la estructura de autocovarianza espacial del proceso.

En este capítulo se extienden los conceptos de campos aleatorios al caso de datos funcionales y se construye el predictor espacial funcional, esto es, se presenta el método *kriging funcional*. Así, los métodos expuestos en los capítulos 2, 3 y 4 para datos escalares, ahora se extienden al caso de curvas que varían espacialmente.

9.1. Campo aleatorio espacial funcional

Definición 12 (Proceso espacial funcional univariado. *Campo aleatorio espacial funcional.*). *Sea* $D_s \subset \mathbb{R}^d$ *el conjunto índice espacial, y sea* $\boldsymbol{\mathcal{Y}}_s(t)$, $s \in D_s$ *una variable aleatoria funcional espacial, ver Sección 8.1. Un conjunto de datos funcionales espaciales* $\mathcal{Y}_{s_1}(t), \ldots, \mathcal{Y}_{s_n}(t)$ *es la observación de n variables aleatorias funcionales cuadrado integrables,* $\boldsymbol{\mathcal{Y}}_{s_1}(t), \ldots, \boldsymbol{\mathcal{Y}}_{s_n}(t)$ *en el conjunto de ubicaciones espaciales* $S \subset D_s$, $S = \{s_1, \ldots, s_n\}$, *(ver [8]).* $\boldsymbol{\mathcal{Y}}_s(t) \in L^2(\mathcal{B})$. *También se llama campo aleatorio funcional espacial y está dado por*

$$\{\boldsymbol{\mathcal{Y}}(s) : s \in D_s \subset \mathbb{R}^d\}$$

El método geoestadístico se basa en el vector aleatorio Υ_S dado por

$$\Upsilon_S = \left(\mathcal{Y}_{s_1}(t), \ldots, \mathcal{Y}_{s_n}(t)\right), \quad s \in D_s, \quad t \in \mathbb{T}$$

que es la realización del vector aleatorio $\boldsymbol{\Upsilon}_S = (\boldsymbol{\mathcal{Y}}_{s_1}(t), \ldots, \boldsymbol{\mathcal{Y}}_{s_n}(t))$. El predictor kriging funcional de $\boldsymbol{\mathcal{Y}}_{s_0}(t)$ se construye con la variable centrada $\boldsymbol{\chi}_s(t)$ con base en el modelo geoestadístico (8.12), $\boldsymbol{\mathcal{Y}}_s(t) = \mu(t) + \boldsymbol{\chi}_s(t)$ donde $E\left[\boldsymbol{\mathcal{Y}}_s(t)\right] = \mu(t)$ y de dónde se obtiene el vector aleatorio centrado

$$\mathcal{X}_S = \left(\boldsymbol{\chi}_{s_1}(t), \ldots, \boldsymbol{\chi}_{s_n}(t)\right)$$

con matriz de covarianza $\boldsymbol{\Omega}$, y vector de covarianza $\boldsymbol{\varsigma}$ con la variable funcional a predecir $\boldsymbol{\mathcal{Y}}_{s_0}(t)$, dados por

$$\boldsymbol{\Omega} = \mathrm{Cov}\left[\boldsymbol{\Upsilon}_S\right] = \mathrm{Cov}\left[\mathcal{X}_S\right] \quad \text{y} \tag{9.1}$$

$$\boldsymbol{\varsigma} = \mathrm{Cov}\left[\boldsymbol{\Upsilon}_S, \boldsymbol{\mathcal{Y}}_{s_0}(t)\right] = \mathrm{Cov}\left[\mathcal{X}_S, \boldsymbol{\chi}_{s_0}(t)\right] \tag{9.2}$$

El predictor final es $\check{\mathcal{Y}}_{s_0}(t) = \mu(t) + \check{\chi}_{s_0}(t)$.

9.2. Predicción espacial funcional univariada. Kriging funcional.

En esta sección se propone una metodología para llevar a cabo la predicción de una variable aleatoria funcional en lugares no observados. La propuesta en este trabajo es modelar la estructura de autocovarianza espacial del campo aleatorio funcional para encontrar un predictor kriging funcional con las mismas características y propiedades del predictor kriging escalar. Es decir, se construye un predictor funcional insesgado y que minimiza el cuadrado de la norma del error de predicción.

El predictor kriging funcional propuesto es la combinación lineal de las curvas observadas, Ver [7].

$$\check{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t)$$

Se construye insesgado. $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ es el vector que contiene las ponderaciones de las n curvas observadas y se selecciona tal que minimice la esperanza de la norma del error de predicción dada por

$$E \left\| \boldsymbol{\chi}_{s_0}(t) - \check{\chi}_{s_0}(t) \right\|^2$$

Desarrollando el cuadrado y tomando esperanzas, se obtiene

$$E\langle \boldsymbol{\chi}_{s_0}, \boldsymbol{\chi}_{s_0} \rangle - 2 \sum_{i=1}^{n} \lambda_i E\langle \chi_{s_i}, \boldsymbol{\chi}_{s_0} \rangle + \sum_{i,i'=1}^{n} \lambda_i \lambda_{i'} E\langle \chi_{s_i}, \chi_{s_{i'}} \rangle \qquad (9.3)$$

9.3. Estimación de la autocovarianza espacial entre curvas

La metodología propuesta consiste en expresar las curvas, en términos de la base de funciones ortogonales que resulta de la descomposición en FPC. Se demuestra que es posible modelar la variabilidad espacial del campo aleatorio funcional usando los puntajes resultantes de los FPC. Específicamente, se usa el campo aleatorio escalar de puntajes obtenidos de la representación de las curvas a través de los EFPC, ver Definición 10, con la ventaja de que en presencia de una fuerte autocorrelación en el argumento de la función t, es posible reducir notoriamente la dimensión. La autocovarianza entre dos funciones resulta ser la suma de todas las autocovarianzas espaciales de los vectores de puntaje resultantes de los componentes principales funcionales.

La predicción espacial funcional presentada, se lleva a cabo bajo el supuesto de estacionariedad de segundo orden. Nótese que las variables espaciales funcionales se asumen en $L^2(\mathcal{B})$.

Para centrar la variable espacial funcional se pueden utilizar los mismos mecanismos que en el caso escalar. Se puede usar el promedio de las curvas observadas o un modelo de regresión adecuado (ver Secciones 1.1.4 y 8.1).

Observe que la expresión (9.3) depende completamente de las covarianzas espaciales entre pares de curvas. El primer término es la varianza, el segundo término contiene las covarianzas entre los lugares observados y el lugar a predecir y el tercer término contiene todas las posibles covarianzas entre lugares observados. A continuación la Sección 9.3 detalla el procedimiento para encontrar estas covarianzas. A continuación se sustituye cada curva por su representación en términos de la base ortonormal obtenida de los FPC. Nótese que este procedimiento solo requiere de las autocovarianzas y no necesita de las covarianzas cruzadas entre los vectores de puntajes.

Sustituyendo cada curva por su representación en términos de la base ortonormal obtenida de los EFPC y aplicando que

$$\langle \xi_k, \xi_{k'} \rangle = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k' \end{cases}$$

se obtiene que la covarianza entre cualquier par de curvas es

$$E\langle \chi_{s_i}, \chi_{s_{i'}} \rangle = \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} E\left[f_k(\boldsymbol{s}_i) f_{k'}(\boldsymbol{s}_{i'})\right] \left(\langle \xi_k, \xi_{k'} \rangle\right) \tag{9.4}$$

$$= \sum_{k=1}^{\infty} E\left[f_k(\boldsymbol{s}_i) f_k(\boldsymbol{s}_{i'})\right] \tag{9.5}$$

En consecuencia, $E\langle \chi_{s_i}, \chi_{s_{i'}} \rangle$, está completamente determinada por la suma de las autocovarianzas de todos los campos aleatorios escalares espaciales de puntajes $f_k(\boldsymbol{s})$ generados a partir de los FPC, para cada par de ubicaciones $(\boldsymbol{s}_i, \boldsymbol{s}_{i'})$ (ver (8.10)). Esto permite estimar la covarianza entre curvas, aplicando los métodos vistos para la covarianza entre escalares.

La esperanza del cuadrado de la norma del error de predicción para un sitio no observado $\boldsymbol{s}_0$, aplicando (9.4) y luego minimizando queda

$$E\left\| \chi_{s_0}(t) - \check{\chi}_{s_0}(t) \right\|^2 = \sum_{k=1}^{\infty} \eta_k - 2\varsigma\boldsymbol{\lambda} + \boldsymbol{\lambda}^T \Omega \boldsymbol{\lambda} \tag{9.6}$$

- ς es el vector de dimensión n que contiene la covarianza entre las curvas observadas y la curva en el lugar a predecir $\boldsymbol{s}_0$ (ver (9.2)). Para $i = 1, \ldots, n$

$$\varsigma = \left(\sum_{k=1}^{\infty} E\left[f_k(\boldsymbol{s}_i) f_k(\boldsymbol{s}_0) \right] \right) \tag{9.7}$$

- Ω es una matriz de dimensión $n \times n$ que contiene la covarianza entre todas las curvas observadas (ver (9.1)) .

$$\Omega = \sum_{k=1}^{\infty} \Sigma_k \quad \text{y} \quad \Sigma_k = \left(E\left[f_k(\boldsymbol{s}_i) f_k(\boldsymbol{s}_{i'}) \right] \right) \tag{9.8}$$

  Entonces, para $i = 1, \ldots, n$, ς (ver (9.2)) y Ω (ver (9.1)) se encuentran sumando las autocovarianzas de los puntajes para cada par de ubicaciones.

El vector solución del predictor kriging funcional simple es: $\boldsymbol{\lambda} = \Omega^{-1}\varsigma$, y reemplazando en (9.6), el cuadrado de la norma del error de predicción es

$$E \left\| \chi_{s_0}(t) - \check{\chi}_{s_0}(t) \right\|^2 = \sum_{k=1}^{\infty} \eta^k - 2\varsigma\boldsymbol{\lambda} + \boldsymbol{\lambda}^T \Omega \boldsymbol{\lambda} = \sum_{k=1}^{\infty} \eta^k - \varsigma\Omega\varsigma \tag{9.9}$$

Tenga en cuenta para la estimación del modelo de autocovarianza en cada dimensión k, que la respectiva varianza o cota superior es conocida e igual a η^k. Finalmente, $\sum_{k=1}^{\infty} \eta^k$ es la varianza total.

Nótese que de nuevo, las funciones de autocovarianza entre curvas que determinan Ω y ς dependen solo de las distancias entre los lugares con observaciones y los lugares donde se requiere la predicción. Para que las expresiones (9.8), (9.3) y (9.9) sean calculables es necesario elegir un valor K de componentes principales funcionales y usar así las sumas truncadas. Al realizar el truncamiento todas las expresiones, se convierten en aproximaciones. Sin embargo, es posible encontrar excelentes aproximaciones con unos pocos componentes cuando la dimensión en la cuál se aplican los FPC muestra una fuerte autocorrelación. Esta dimensión es aquella en la que se construyen las funciones, puedes ser tiempo, frecuencia o profundidad, por ejemplo. El ejemplo 23 muestra como se encuentran los vectores y matrices de covarianzas entre curvas. El valor de K para truncar la representación en términos de los EFPC es flexible y se pueden incluir suficientes términos hasta que la varianza acumulada alcance algún umbral prefijado. Este método no usa las covarianzas cruzadas entre los puntajes; solo requiere el ajuste de las autocovarianzas para cada vector de puntajes. Dado que estos puntajes son variables escalares espaciales su autocovarianza se modela con los métodos vistos en los Capítulos 2 y 3.

Ejemplo 23. Uno y dos FPC *A manera de ilustración, si se elige solo el primer componente principal funcional, se tiene que la silla del semivariograma es* η_1*,*

$$\varsigma = \left(E\left[f_1(s_1)f_1(s_0)\right], \ldots, E\left[f_1(s_n)f_1(s_0)\right]\right)$$

$$y$$

$$\Omega = \Sigma_1 = \mathrm{Cov}[f_1] = \Sigma_{f_1}$$

dónde Σ_1 *es una matriz definida positiva que se construye utilizando un modelo de covarianza espacial válido acotado superiormente por* η_1 *(ver Sección 2.5).* Σ_1 *es*

$$\Sigma_1 = \begin{pmatrix} \eta_1 & E\left[f_1(s_1)f_1(s_2)\right] & E\left[f_1(s_1)f_1(s_3)\right] & \ldots & E\left[f_1(s_1)f_1(s_n)\right] \\ E\left[f_1(s_2)f_1(s_1)\right] & \eta_1 & E\left[f_1(s_2)f_1(s_3)\right] & \ldots & E\left[f_1(s_2)f_1(s_n)\right] \\ E\left[f_1(s_3)f_1(s_1)\right] & E\left[f_1(s_3)f_1(s_2)\right] & \eta_1 & \ldots & E\left[f_1(s_3)f_1(s_n)\right] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E\left[f_1(s_n)f_1(s_1)\right] & E\left[f_1(s_n)f_1(s_2)\right] & E\left[f_1(s_n)f_1(s_3)\right] & \ldots & \eta_1 \end{pmatrix}$$

Si se seleccionan dos componentes principales funcionales, el vector ς *y la matriz de covarianza* Ω *están dados por*

$$\varsigma = \left(E\left[f_1(s_1)f_1(s_0)\right] + E\left[f_2(s_1)f_2(s_0)\right], \ldots, E\left[f_1(s_n)f_1(s_0)\right] + E\left[f_2(s_n)f_2(s_0)\right]\right)$$

$$\Omega = \Sigma_1 + \Sigma_2 = \Sigma_{f_1} + \Sigma_{f_2} = \left(E\left[f_1(s_i)f_1(s_j)\right] + E\left[f_2(s_i)f_2(s_j)\right]\right), \quad i, j = 1, \ldots, n$$

La diagonal de Ω *es la suma de los valores propios.*

$$E\left[f_1(s_i)f_1(s_i)\right] + E\left[f_2(s_i)f_2(s_i)\right] = \eta_1 + \eta_2$$

La predicción se lleva a cabo con la variable centrada

$$E\left[\chi_s(t)\right] = 0.$$

Si la media se asume constante, en la primera etapa se estima con el promedio para centrar la variable. Al final del método geoestadístico se re-estima la media funcional usando la estructura de covarianza encontrada, esto es,

$$\mu_\Omega(t) = \frac{\mathbf{1}'\Omega^{-1}\Upsilon_S}{\mathbf{1}'\Omega^{-1}\mathbf{1}} \tag{9.10}$$

y esta nueva estimación se le suma a la predicción obtenida. La predicción final queda

$$\mathcal{Y}_{s_0}(t) = \check{\chi}_{s_0}(t) + \mu_\Omega(t)$$

Ejemplo 24. Señales cerebrales de lenguaje silencioso para manejo de prótesis. *Se lleva a cabo un experimento para medir señales emitidas por el cerebro en habla silenciosa. Esto es, se piensa en la palabra o letra, pero no se emite ningún sonido. En este caso se pide a la persona que piense en cada una de las cinco vocales, una a la vez. Los 21 electrodos para tomar los datos se ubican en el hemisferio izquierdo y en zonas específicas que manejan el lenguaje (ver Figuras 9.1 y 9.2). El EEG extrae las medidas de señal que están más relacionadas con el pensamiento de la vocal, que forma parte de cómo el cerebro humano procesa el lenguaje. La Figura 9.2 muestra las curvas de las señales generadas por el cerebro al pensar cada vocal. Nótese que las curvas no varían en función del tiempo, sino en función de la frecuencia. Se pretende generar una sucesión de imágenes del cerebro para cada vocal. Se aplica kriging para predecir las curvas de las señales del cerebro y posteriormente se hacen cortes transversales para obtener las imágenes. Las Figuras 9.3 y 9.4 muestran los modelos de semivariogramas y una imagen específica extraída para cada una de las 5 vocales.*

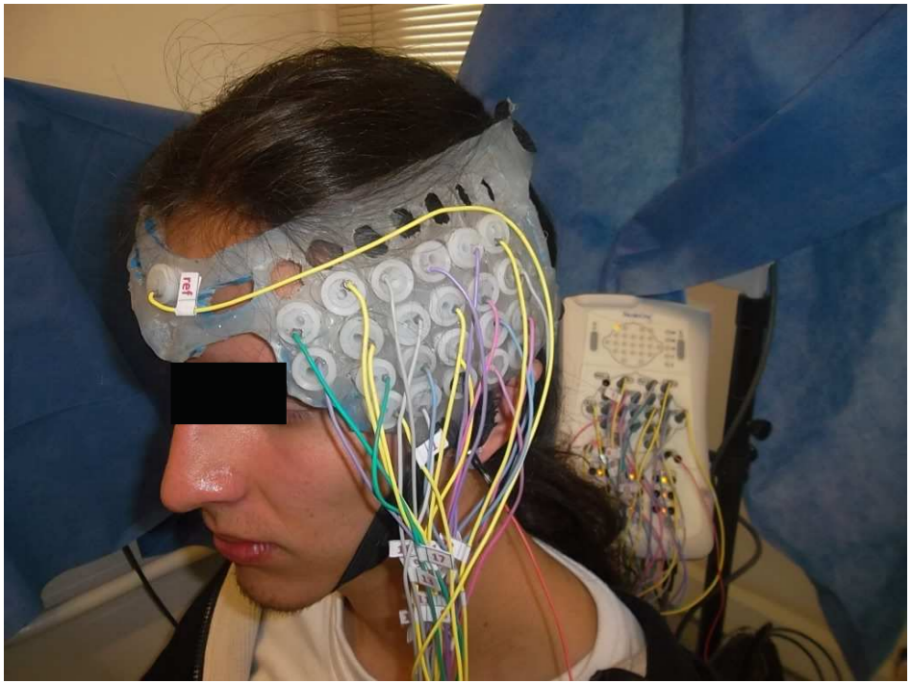


Figura 9.1. Persona que utiliza el neuroauricular EEG en la experimentación de vocales del habla silenciosa. El dispositivo permite colocar 21 electrodos en una disposición matricial con una distancia euclidiana de 16 mm entre ellos, tanto horizontal como verticalmente. Esta configuración incluye un electrodo de referencia ubicado en el centro de la frente y un electrodo de tierra (GND) ubicado en el lóbulo de la oreja izquierda. La posición de los electrodos se diseñó para cubrir las áreas de Wernicke y Broca en el hemisferio izquierdo, dado que son las mas relacionados con las funciones de lenguage en el cerebro.

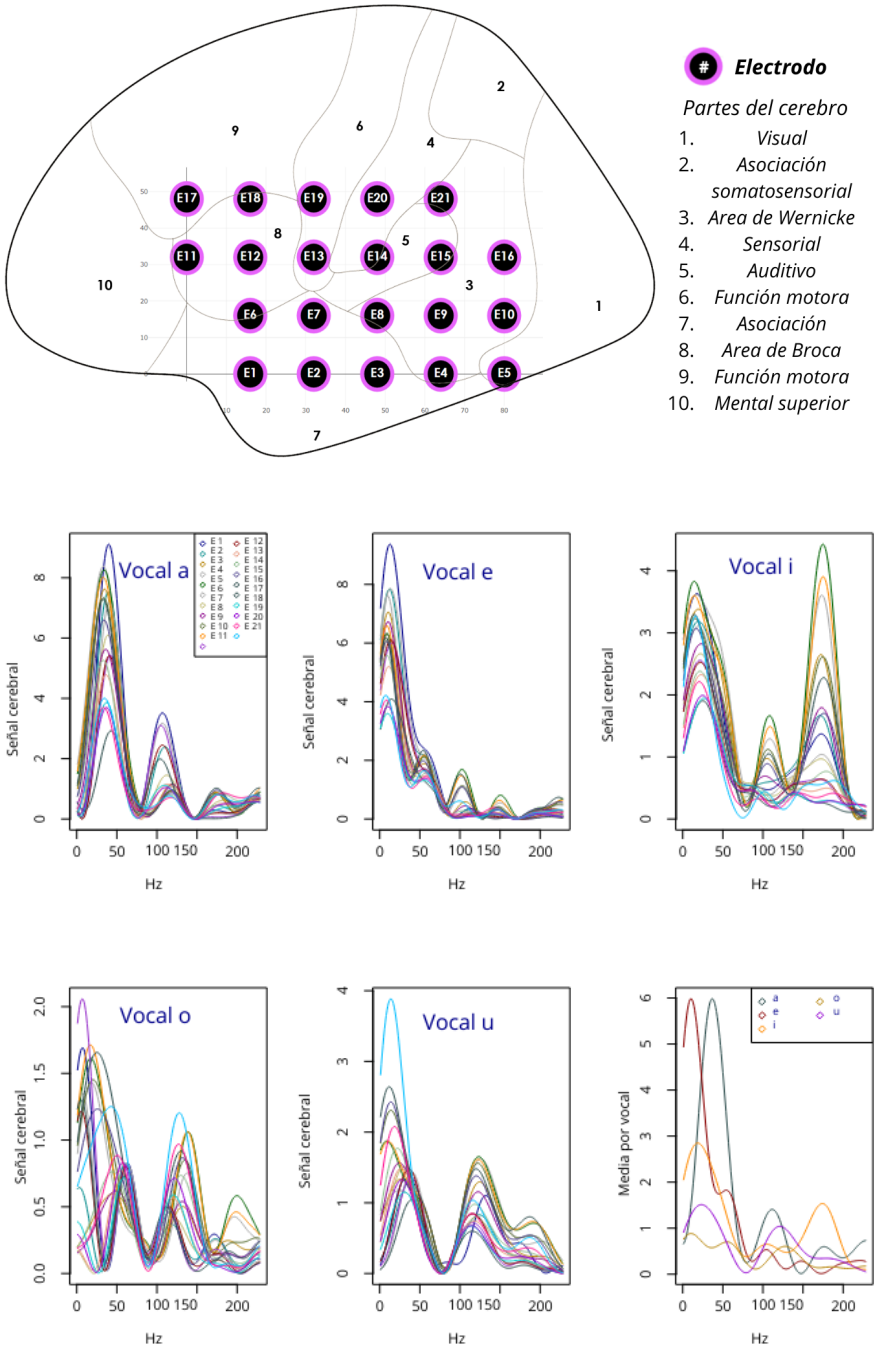


Figura 9.2. **Panel superior**: Configuración espacial de los 21 electrodos en el área del lenguaje del cerebro en el hemisferio izquierdo. **Panel inferior**: Curva de la señal emitida por el cerebro cuando piensa cada vocal, en cada una de las 21 ubicaciones donde se encuentran conectados los electrodos. Nótese que las curvas representan la señal del cerebro como función de la frecuencia en el intervalo [0,228].

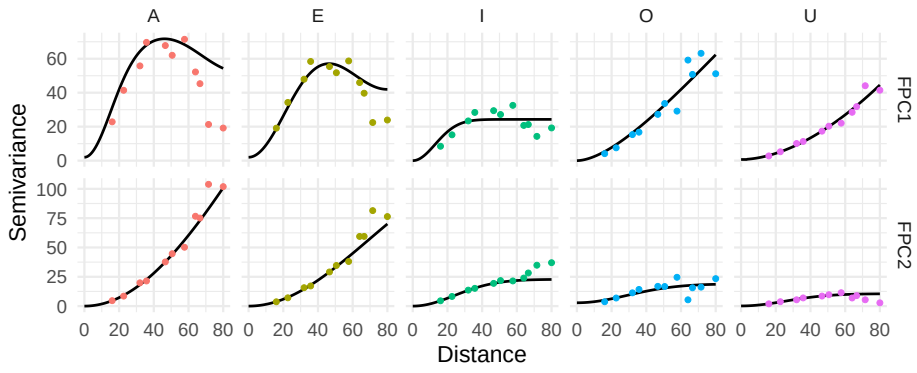


Figura 9.3. Semivariogramas (puntos) y modelo (línea negra) para los puntajes de los dos primeros componentes principales funcionales para cada una de las cinco vocales pensadas por el individuo 13^{th} individual.

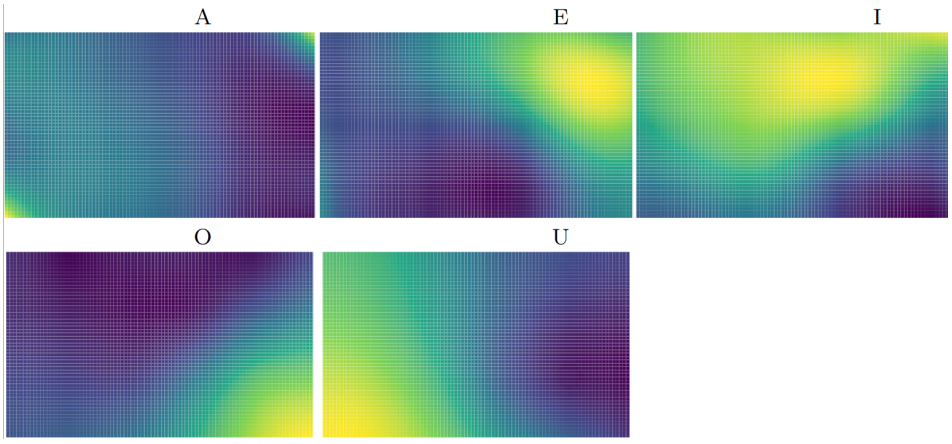


Figura 9.4. Imagen en un punto específico de la frecuencia: 500, obtenida del kriging funcional para las cinco vocales del individuo 13^{th}. En este caso, el interés es generar suficientes imágenes que permitan definir un modelo de clasificación a partir del cual se pueda identificar con alto nivel de precisión, la vocal que el sujeto piensa.

Figura 9.5. Prótesis mioeléctrica de mano. Usando los resultados del modelo de clasificación, cada vocal se traduce en una acción que debe llevar a cabo la prótesis. Por ejemplo, si se identifica que el sujeto piensa la vocal a, se traduce en la orden abrir, y así para cada señal identificada.

Ejemplo 25. Calidad del aire. *El material particulado por debajo de 10 micras de diámetro (PM10), fue analizado usando kriging espacio-tiempo en el Capítulo 6.*

En esta ocasión se quiere hacer uso de series de tiempo más largas del contaminante, para tomar datos históricos más lejanos y porque de esta manera, se puede medir más densamente y hacer un seguimiento mas preciso a la concentración del PM10.

Así, la cantidad de datos de interés crece rápidamente y los procesos de optimización que usan la matriz de covarianza espacio−tiempo que fueron descritos en el Capítulo 6 van a ser muy restrictivos. Por otro lado, hay que notar que el PM10 varía de forma continua tanto en el tiempo como en el espacio, así que el fenómeno se puede interpretar de una forma diferente que vaya mas acorde a las características de este fenómeno en la naturaleza. Las curvas de la Figura 8.10 panel derecho, son suavizadas a partir de los datos observados en cada punto del tiempo. Así, cada curva es un dato funcional georreferenciado utilizando bases de funciones como se describe en la Sección 8.1. En lugar de ver el dato como el valor numérico del material particulado en una ubicación espacio-tiempo específica, el dato es la curva de PM10 que existe en cada ubicación espacial de Bogotá (ver Figura 8.10).

La función de covarianza que presenta mejor ajuste para ambos scores resulta ser el modelo de Gneiting-Matern, (ver [34]), con el vector de parámetros

$$\boldsymbol{\Theta} = \left(\sigma^2, \phi, \kappa_1, \kappa_2\right).$$

reportados para cada caso en la Figura 9.6.

$$C(h|\Theta) = \frac{1}{2^{\kappa_1-1}}\Gamma(\kappa_1)\left(\frac{h}{\phi}\right)^{\kappa_1} K_{\kappa_1}\left(\frac{h}{\phi}\right)$$
$$\sigma^2\left(1+8\left(\iota\frac{h}{\phi\kappa_2}\right)+25\left(\iota\frac{h}{\phi\kappa_2}\right)^2+32\left(\iota\frac{h}{\phi\kappa_2}\right)^3\left(1-\left(\frac{\iota h}{\phi\kappa_2}\right)\right)^8\right)$$

donde $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_{i'}$. $K_{\kappa_1}(.)$ *es la función de Bessel modificada de la tercera clase, de orden* κ_1, $\iota = 10/47$. $\iota\frac{h}{\phi\kappa_2} \leq 1$, *o en caso contrario* $C(h|\Theta) = 0$. *La Figura 9.7 muestra los resultados del kriging funcional con sus respectivas varianzas para la predicción de curvas de PM10 en 28 ubicaciones de interés en Bogotá.*

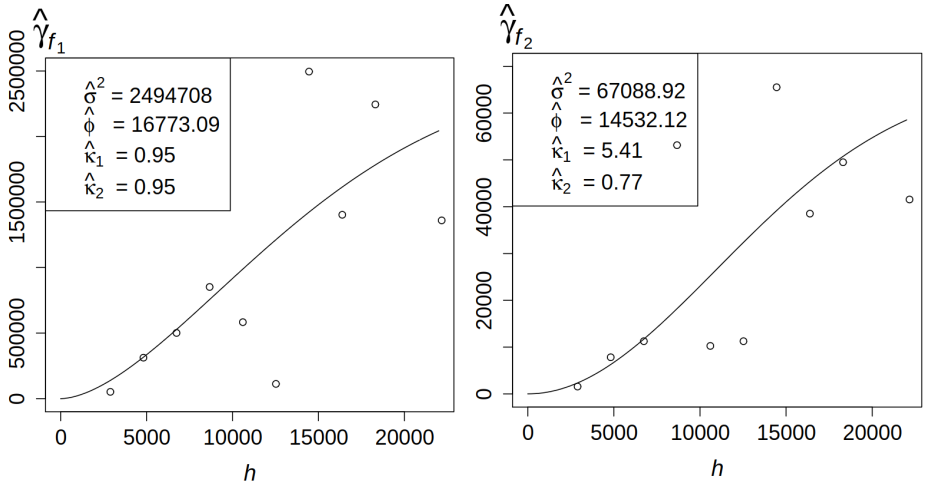


Figura 9.6. Izquierda: variograma de f_1. Derecha: variograma de f_2.

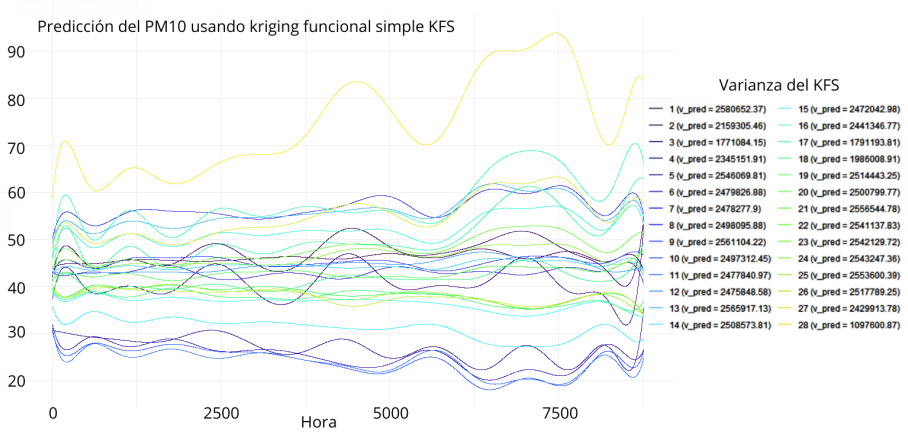


Figura 9.7. Predicción y varianza del error de predicción de curvas de PM10 en 28 lugares de la ciudad de Bogotá usando kriging funcional simple.

Nota 9 (Aspectos prácticos).

- *El predictor kriging funcional, esto es el método presentado en la Sección 9 garantiza directamente el insesgamiento de la función predicha y minimiza directamente el cuadrado de la norma de la diferencia entre la función a predecir y el predictor funcional.*

- *Una alternativa es aplicar kriging simple a los scores como se presenta en la Sección 10.3. Es decir, predecir cada score en el lugar $\boldsymbol{s}_0$, y reconstruir los datos funcionales usando la base EFPC. De tal manera, que no se modela la covarianza de las curvas sino solamente la covarianza de cada uno de los puntajes.*

  Los desarrollos relacionados se encuentran en la sección 10.3. Si se incluyen suficientes puntajes, este método también puede dar buenos resultados, pero es importante tener en cuenta que el que realmente hace una optimización usando datos funcionales es el kriging funcional. Ambas metodologías están implementadas en el paquete SpatFD [10] del software libre R cran [62].

Ejemplo 26. Material particulado.

En este ejemplo se aplica el predictor kriging funcional a los datos de material particulado PM10 en Bogotá. En contraste con el Ejemplo 25, aquí se usan mas funciones en el suavizamiento. La base de funciones usada para la construcción de los datos es B-splines cúbico con 191 bases para conservar los detalles de los datos.

En este caso se evitó suavizar mucho. Además, se usaron 9 dimensiones que completan 99.9% de variabilidad acumulada. Es decir, se modelaron los semivariogramas de las primeras 9 dimensiones de la descomposición en FPC.

La Figura 9.1 muestra los resultados del kriging funcional para predecir las funciones en los lugares observados, utilizando validación cruzada, esto es, lleva a cabo la predicción en cada sitio utilizando los $n-1$ sitios restantes, para verificar la calidad de las predicciones.

Nótese que la mayoría de los residuos se encuentran entre -20 y 20, solo el dato atípico identificado desde el comienzo tiene unos errores considerables. Se seleccionan los dos primeros componentes principales y se modelan las respectivas funciones de covarianza de sus scores.

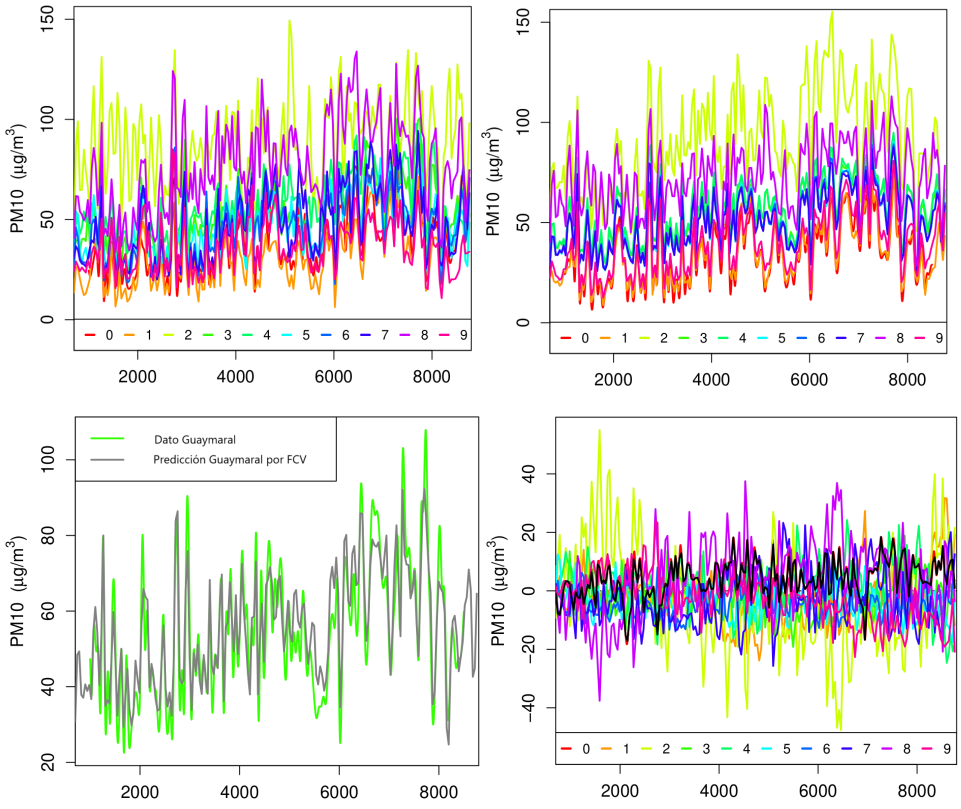


Tabla 9.1. Datos del material particulado en Bogotá medidos cada hora, entre el 13 de mayo de 2023 y el 13 de mayo de 2024. **Panel superior izquierdo.** Datos originales suavizados con B-splines cúbicos y 191 funciones base, buscando conservar bastantes detalles de los datos. Nótese la rugosidad de las curvas. **Panel superior derecho:** predicción funcional espacial para los 10 sitios originales mediante validación cruzada. **Panel inferior izquierdo**. Dato original Vs. kriging funcional para la estación Guaymaral. **Panel inferior derecho.** Residuos.

9.4. Ejercicios

1. Encuentre la predicción espacial, espacio-temporal y funcional con su respectiva varianza para el caso en el que no existe ni dependencia espacial ni dependencia temporal. Interprete los resultados obtenidos.
2. Suponga un campo aleatorio espacial funcional multivariado conjuntamente gaussiano, con función de media y estructura de covarianza, ambas conocidas. Se tienen observaciones funcionales en n ubicaciones. Encuentre la esperanza condicional y la respectiva varianza para $\chi_{s_0}(t)$, dado el vector $\left(\chi_{s_1}(t), \ldots, \chi_{s_n}(t)\right)$. Compare las expresiones del predictor encontrado y de la respectiva varianza del error de predicción, con las del kriging funcional simple.
3. Ilustre el cumplimiento de las propiedades de forma cuadrática de una función de covarianza definida positiva y de una función de semivarianza condicionalmente definida negativa para algunos casos particulares de datos funcionales y de modelos.
4. Encuentre la estimación de la función media, bajo el supuesto de que es constante en toda la zona y para un campo aleatorio funcional espacial gaussiano.

Capítulo

diez

Predicción espacial funcional multivariada. Cokriging funcional

$(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s$
$D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in \mathbb{R}^d \times \mathbb{R}$ $(s,t) \in D_s \times D_T \in$

En este capítulo, se presenta la predicción espacial de una variable aleatoria funcional en lugares no observados, usando su propia información y además la de otras covariables espaciales funcionales disponibles. Los métodos expuestos en el capítulo 5 para datos escalares, ahora se extienden al caso de procesos multivariados de curvas que varían espacialmente.

10.1. Campo aleatorio espacial funcional multivariado.

Definición 13 (Proceso espacial funcional multivariado o Campo aleatorio espacial funcional multivariado.). *Sea $D_s \subset \mathbb{R}^d$ el conjunto índice espacial, y sean*

$$\chi_s^1(t), \ldots, \chi_s^P(t), \ s \in D_s$$

P campos aleatorios funcionales espaciales cuadrado integrables, tales que

$$\chi_s^p(t) \in L^2(\mathcal{B})$$

$p = 1, \ldots, P$. Se considera aquí el caso en el que $t \in \mathcal{B} \subset \mathbb{R}$, esto es, la variable aleatoria funcional es una curva. Nótese que $L^2(\mathcal{B})$ es un espacio real separable de Hilbert, $\mathcal{H}$. El campo aleatorio multivariado espacial funcional está dado por

$$\{\Xi_s : s \in D_s \subset \mathbb{R}^d\}$$

donde

$$\Xi_s = \left(\chi_s^1(t), \ldots, \chi_s^P(t)\right).$$

Un conjunto de datos funcionales espaciales multivariados es una observación de Ξ_s en un conjunto de sitios particulares, $S \subset D_s$. Si los P campos aleatorios funcionales pueden ser medidos en el mismo conjunto de ubicaciones $S = \{s_1, \ldots, s_n\}$, se tiene que en cada ubicación s_i se observan los P procesos funcionales. Esto es,

$$\left(\chi_{s_i}^1(t), \ldots, \chi_{s_i}^P(t)\right) \quad i = 1, \ldots, n$$

En otro caso, cada campo aleatorio funcional espacial $\chi_s^p(t)$ es observado en un conjunto S_p con n_p ubicaciones espaciales $p = 1, \ldots, P$. Así, las n_p realizaciones de la p-ésima variable aleatoria espacial funcional están contenidas en el siguiente vector aleatorio:

$$\left(\chi_{s_1}^p(t), \ldots, \chi_{s_{n_p}}^p(t)\right), \ p = 1, \ldots, P$$

Usualmente, al menos algunas ubicaciones son comunes para varias variables.

Ahora, sea

$$\mathcal{H}^P = \mathcal{H} \oplus \ldots \oplus \mathcal{H}$$

La suma directa de los P espacios reales separables de Hilbert, entonces $\boldsymbol{\Xi}_{\boldsymbol{s}} \in \mathscr{H}^P$ ([64] y [11]). Sean $\boldsymbol{\xi}, \boldsymbol{\zeta} \in \mathscr{H}^P$, $b \in \mathbb{R}$ y $\langle \xi^p, \zeta^p \rangle$ es el producto interno usual en $L^2(\mathscr{B})$. Con la suma, la multiplicación por un escalar y el producto interno definidos como en el Capítulo 8.

Se usa la teoría de componentes principales funcionales para la representación de cada función. Esta representación permite encontrar tanto la autocovarianza como la covarianza cruzada, es decir, la covarianza entre curvas de diferentes variables funcionales en diferentes lugares, usando los puntajes asociados. Estos puntajes a su vez son campos aleatorios escalares espaciales multivariados (ver Sección 8.3).

El método cokriging funcional es desarrollado primero para el caso de dos campos aleatorios funcionales para mostrar detalladamente su construcción y después se extiende al caso general de P campos aleatorios espaciales funcionales.

Si bien las variables aleatorias funcionales espaciales son los objetos aleatorios en este contexto, las ideas generales son similares a los métodos análogos del caso espacial escalar (ver Sección 5 y la Definición 13).

10.2. Cokriging para dos campos aleatorios espaciales funcionales

Sean dos campos aleatorios espaciales funcionales

$$\boldsymbol{\chi}_{\boldsymbol{s}}^1(t) \quad \text{y} \quad \boldsymbol{\chi}_{\boldsymbol{s}}^2(t)$$

Se asume que ya han sido centrados, usando el promedio de las curvas observadas o algún modelo de regresión apropiado. Esto es,

$$E\left[\boldsymbol{\chi}_{\boldsymbol{s}}^1(t)\right] = 0 \quad \text{y} \quad E\left[\boldsymbol{\chi}_{\boldsymbol{s}}^2(t)\right] = 0.$$

El predictor cokriging de $\boldsymbol{\chi}_{\boldsymbol{s}_0}^1(t)$ para un lugar no observado $\boldsymbol{s}_0$, usando tanto la información propia de $\boldsymbol{\chi}_{\boldsymbol{s}}^1$ como la de la covariable espacial $\boldsymbol{\chi}_{\boldsymbol{s}}^2$, está dado por

$$\check{\chi}_{s_0}^1(t) = \sum_{i=1}^{n_1} \lambda_i^{11} \chi_{s_i}^1(t) + \sum_{j=1}^{n_2} \lambda_j^{12} \chi_{s_j}^2(t)$$

donde λ_i^{11} $i = 1, \ldots, n_1$ son las n_1 ponderaciones de las observaciones de $\chi_{\boldsymbol{s}}^1(t)$ y λ_j^{12} $j = 1, \ldots, n_2$ son las n_2 ponderaciones de las observaciones de $\chi_{\boldsymbol{s}}^2(t)$. Note que no se requiere que ambos procesos sean medidos en los mismos lugares. Además, el predictor es insesgado. Note que

$$E\left[\check{\chi}_{s_0}^1(t) - \boldsymbol{\chi}_{\boldsymbol{s}_0}^1(t)\right] = E\left[\sum_{i=1}^{n_1} \lambda_i^{11} \chi_{s_i}^1(t) + \sum_{j=1}^{n_2} \lambda_j^{12} \chi_{s_j}^2(t)\right] = 0$$

y $\boldsymbol{\lambda} = \left(\lambda_i^{11}\right)\ i = 1, \ldots, n_1$ y $\boldsymbol{\beta} = \left(\lambda_j^{12}\right)\ j = 1, \ldots, n_2$ son constantes que minimizan

$$\boldsymbol{Q} = E\|\chi_{s_0}^1(t) - \check{\chi}_{s_0}^1(t)\|^2$$

Ahora, se minimiza $\boldsymbol{Q}$ para obtener el sistema de ecuaciones de cokriging

$$\begin{aligned}
\boldsymbol{Q} &= E\|\chi_{s_0}^1(t) - \check{\chi}_{s_0}^1(t)\|^2 \\
&= E\langle \chi_{s_0}^1, \chi_{s_0}^1\rangle - 2E\langle \chi_{s_0}^1, \check{\chi}_{s_0}^1\rangle + E\langle \check{\chi}_{s_0}^1, \check{\chi}_{s_0}^1\rangle \\
&= E\langle \chi_{s_0}^1, \chi_{s_0}^1\rangle - 2\sum_{i=1}^{n_1} \lambda_i^{11} E\langle \chi_{s_i}^1, \chi_{s_0}^1\rangle - 2\sum_{j=1}^{n_2} \lambda_j^{12} E\langle \chi_{s_j}^2, \chi_{s_0}^1\rangle \\
&\quad + \sum_{i=1}^{n_1}\sum_{i'=1}^{n_1} \lambda_i^{11}\lambda_{i'}^{11} E\langle \chi_{s_i}^1, \chi_{s_{i'}}^1\rangle + 2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \lambda_i^{11}\lambda_j^{12} E\langle \chi_{s_j}^2, \chi_{s_i}^1\rangle \\
&\quad + \sum_{j=1}^{n_2}\sum_{j'=1}^{n_2} \lambda_j^{12}\lambda_{j'}^{12} E\langle \chi_{s_j}^2, \chi_{s_{j'}}^2\rangle
\end{aligned} \tag{10.1}$$

Así, para $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$ las derivadas parciales están dadas por

$$\frac{\partial dQ}{d\lambda_i^{11}} = -2E\langle \chi_{s_i}^1, \chi_{s_0}^1\rangle + 2\sum_{i'=1}^{n_1} \lambda_{i'}^{11} E\langle \chi_{s_i}^1, \chi_{s_{i'}}^1\rangle + 2\sum_{j=1}^{n_2} \lambda_j^{12} E\langle \chi_{s_j}^2, \chi_{s_i}^1\rangle$$

y

$$\frac{\partial dQ}{d\lambda_j^{12}} = -2E\langle \chi_{s_j}^2, \chi_{s_0}^1\rangle + 2\sum_{i=1}^{n_1} \lambda_i^{11} E\langle \chi_{s_j}^2, \chi_{s_i}^1\rangle + 2\sum_{j'=1}^{n_2} \lambda_{j'}^{12} E\langle \chi_{s_j}^2, \chi_{s_{j'}}^2\rangle$$

por lo tanto, las ecuaciones del cokriging son

$$\sum_{i'=1}^{n_1} \lambda_{i'}^{11} E\langle \chi_{s_i}^1, \chi_{s_{i'}}^1\rangle = E\langle \chi_{s_i}^1, \chi_{s_0}^1\rangle - \sum_{j=1}^{n_2} \lambda_j^{12} E\langle \chi_{s_j}^2, \chi_{s_i}^1\rangle \tag{10.2}$$

y

$$\sum_{j'=1}^{n_2} \lambda_{j'}^{12} E\langle \chi_{s_j}^2, \chi_{s_{j'}}^2\rangle = E\langle \chi_{s_j}^2, \chi_{s_0}^1\rangle - \sum_{i=1}^{n_1} \lambda_i^{11} E\langle \chi_{s_j}^2, \chi_{s_i}^1\rangle \tag{10.3}$$

Reemplazando (10.2) y (10.3) en (10.1) obtenemos

$$E\|\chi_{s_0}^1(t) - \check{\chi}_{s_0}^1(t)\|^2 = E\langle \chi_{s_0}^1, \chi_{s_0}^1\rangle - \sum_{i=1}^{n_1} \lambda_i^{11} E\langle \chi_{s_i}^1, \chi_{s_0}^1\rangle - \sum_{j=1}^{n_2} \lambda_j^{12} E\langle \chi_{s_j}^2, \chi_{s_0}^1\rangle \tag{10.4}$$

Sea $f_k^1(\boldsymbol{s}),\ k = 1, \ldots, K$ y $f_l^2(\boldsymbol{s}),\ l = 1, \ldots, L$ el campo aleatorio espacial escalar formado por los puntajes asociados a los EFPC de las variables aleatorias $\chi_s^1(t)$ y $\chi_s^2(t)$, respectivamente. Expresando cada función en (10.4) en términos de sus componentes principales funcionales, se obtiene

$$E\langle \chi_{s_0}^1, \chi_{s_0}^1 \rangle = \sum_{k=1}^{K} \sum_{k'=1}^{K} E\left[f_k^1(\boldsymbol{s}_0) f_{k'}^1(\boldsymbol{s}_0)\right] \langle \xi_k^1, \xi_{k'}^1 \rangle$$

$$\sum_{i=1}^{n_1} \lambda_i^{11} E\langle \chi_{s_i}^1, \chi_{s_0}^1 \rangle = \sum_{i=1}^{n_1} \sum_{k=1}^{K} \sum_{k'=1}^{K} \lambda_i^{11} E\left[f_k^1(\boldsymbol{s}_i) f_{k'}^1(\boldsymbol{s}_0)\right] \langle \xi_k^1, \xi_{k'}^1 \rangle$$

$$\sum_{j=1}^{n_2} \lambda_j^{12} E\langle \chi_{s_j}^2, \chi_{s_0}^1 \rangle = \sum_{j=1}^{n_2} \sum_{l=1}^{L} \sum_{k=1}^{K} \lambda_j^{12} E\left[f_l^2(\boldsymbol{s}_i) f_k^1(\boldsymbol{s}_0)\right] \langle \xi_l^2, \xi_k^1 \rangle$$

donde $\xi_k^1,\ k = 1, \ldots, K$ son las funciones propias del operador de covarianza de $\chi_s^1(t)$ y $\xi_l^2,\ l = 1, \ldots, L$ son las funciones propias del operador de covarianza de $\chi_s^2(t)$. Dada la ortonormalidad de los EFPC, se tiene que

$$\langle \xi_k^1, \xi_{k'}^1 \rangle = \begin{cases} 0 & \text{if } k \neq k' \\ 1 & \text{if } k = k' \end{cases}$$

Notese que

$$\sum_{k=1}^{K} E\left[f_k^1(\boldsymbol{s}_0) f_k^1(\boldsymbol{s}_0)\right] = \sum_{k=1}^{K} \eta_k^1$$

En consecuencia,

El cuadrado de la norma del error de predicción $E\|\chi_{s_0}^1(t) - \check{\chi}_{s_0}^1(t)\|^2$ es

$$\sum_{k=1}^{K} \eta^{1k} - \sum_{i=1}^{n_1} \sum_{k=1}^{K} \lambda_i^{11} E\left[f_{s_i}^{1k} f_{s_0}^{1k}\right] - \sum_{j=1}^{n_2} \sum_{l=1}^{L} \sum_{k=1}^{K} \lambda_j^{12} c_{12}^{lk} E\left[f_{s_i}^{2l} f_{s_0}^{1k}\right] \qquad (10.5)$$

donde $c_{lk}^{12} = \langle \xi_l^2, \xi_k^1 \rangle$.

Observe que los productos internos cuyo resultado es 1 o 0, ocurren solo para pares de funciones propias correspondientes a la misma variable. Cuando las funciones propias corresponde a bases ortonormales de diferentes variables, el resultado puede ser un escalar cualquiera.

En conclusión, la varianza y el sistema de ecuaciones del cokriging funcional depende solamente de las autocovarianzas y de las covarianzas cruzadas de los vectores de puntajes, que a su vez son procesos espaciales escalares multivariados. Ver Sección 3.6.

10.3. Cokriging con P campos aleatorios funcionales

El objetivo es la optimización de la predicción espacial funcional de

$$\chi^r_{s_0}(t), \quad 1 \leq r \leq P$$

en un lugar no observado $\boldsymbol{s}_0$ basada en las P variables espaciales funcionales,

$$\check{\chi}^r_{s_0}(t) = \sum_{p=1}^{P} \sum_{i=1}^{n_p} \lambda_i^{rp} \chi^p_{s_i}(t)$$

El interés es la minimización del cuadrado de la norma del error de predicción dado por

$$Q = E\|\chi^r_{s_0}(t) - \check{\chi}^r_{s_0}(t)\|^2 \tag{10.6}$$

Para $m = 1, \ldots, P$, las derivadas y las ecuaciones del cokriging toman la forma

$$\frac{\partial Q}{d\lambda_i^{rm}} = -2E\langle \chi^m_{s_j}, \chi^r_{s_0}\rangle + 2\sum_{p=1}^{P}\sum_{i=1}^{n_p} \lambda_i^{rp} E\langle \chi^m_{s_j}, \chi^p_{s_i}\rangle, \quad j = 1, \ldots, n_m \tag{10.7}$$

y

$$E\langle \chi^m_{s_j}, \chi^r_{s_0}\rangle = \sum_{p=1}^{P}\sum_{i=1}^{n_p} \lambda_i^{rp} E\langle \chi^m_{s_j}, \chi^p_{s_i}\rangle, \quad j = 1, \ldots, n_m \tag{10.8}$$

respectivamente. Reemplazando (10.7) y (10.8) en (10.6), se obtiene

$$E\|\chi^r_{s_0}(t) - \check{\chi}^r_{s_0}(t)\|^2 = E\langle \chi^r_{s_0}, \chi^r_{s_0}\rangle - \sum_{p=1}^{P}\sum_{i=1}^{n_p} \lambda_i^{rp} E\langle \chi^p_{s_i}, \chi^r_{s_0}\rangle \tag{10.9}$$

Ahora, usando la representación en componentes principales funcionales de todas las variables involucradas, se muestra que (10.9) depende solamente de las autocovarianzas y de las covarianzas cruzadas entre los vectores de puntajes.

$$E\|\chi^r_{s_0}(t) - \check{\chi}^r_{s_0}(t)\|^2 = \sum_{k=1}^{K} E\left[f_k^r(\boldsymbol{s}_0) f_k^r(\boldsymbol{s}_0)\right] - \sum_{i=1}^{n_p}\sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{p=1}^{P} \lambda_i^{rp} c_{kl}^{rp} E\left[f_k^p(\boldsymbol{s}_i) f_l^r(\boldsymbol{s}_0)\right] \tag{10.10}$$

donde, como antes, denotando η_k^r, $k = 1, \ldots, K$ a los valores propios de χ^r_s, se tiene que

$$\sum_{k=1}^{K} E\left[f_k^r(\boldsymbol{s}_0) f_k^r(\boldsymbol{s}_0)\right] = \sum_{k=1}^{K} \eta_k^r$$

y

$$c_{kl}^{rp} = \begin{cases} 1 & \text{Si } p = r \text{ y } k = l \\ 0 & \text{Si } p = r \text{ y } k \neq l \\ \langle \xi_k^p, \xi_l^r \rangle & \text{Si } p \neq r \end{cases}$$

En la mayoría de los casos es suficiente con unos pocos componentes principales funcionales, tal vez uno o dos, ver Sección 8.3. El valor propio correspondiente al primer componente principal es mucha mayor que el resto $\eta^{1r} \gg \eta^{2r}$. Esto simplifica el uso del MLC. Note que todo proceso espacial de puntajes considerado tiene media constante, varianza finita y una estructura de covarianza que depende únicamente de la distancia entre ubicaciones espaciales, ver (10.15), esto es, todos los vectores de puntaje son procesos estacionarios de segundo orden, para mas detalles ver [8]. Finalmente, una medida global de la calidad de la predicción óptima del campo aleatorio funcional $\chi_{s_0}^r(t)$ en B sitios no observados es

$$\sum_{b=1}^{B} \left(E \| \chi_{s_0^b}^r(t) - \check{\chi}_{s_0^b}^r(t) \|^2 \right). \tag{10.11}$$

Ejemplo 27 (Ilustración simulación. Datos funcionales multivariados correlacionados). *Con base en la media y el modelo de covarianza del proceso funcional, es posible encontrar los predictores kriging o cokriging o simular. La simulación de procesos funcionales espaciales, consiste en generar vectores de puntajes espacialmente correlacionados para que sean los coeficientes de combinaciones lineales de funciones propias. Por ejemplo, para simular un proceso bivariado con una versión truncada con* $K = 2$ *y* $K = 3$ *componentes principales para la primera y segunda variable, respectivamente, se tienen*

$$\chi_s^1 = \sum_{k=1}^{2} f_{s_i}^{1k} \xi^{1k}(t); \qquad \chi_s^2 = \sum_{k=1}^{3} f_{s_i}^{2k} \zeta^{2k}(t), \quad s \in \mathbb{R}^2 \tag{10.12}$$

se toman las funciones $\xi^{1k}(t)$ *del conjunto* $\{\xi^{11}(t) = sin(\pi t),\ \xi^{12}(t) = cos(\pi t)\}$ *y* $\zeta^{2k}(t)$ *del conjunto*

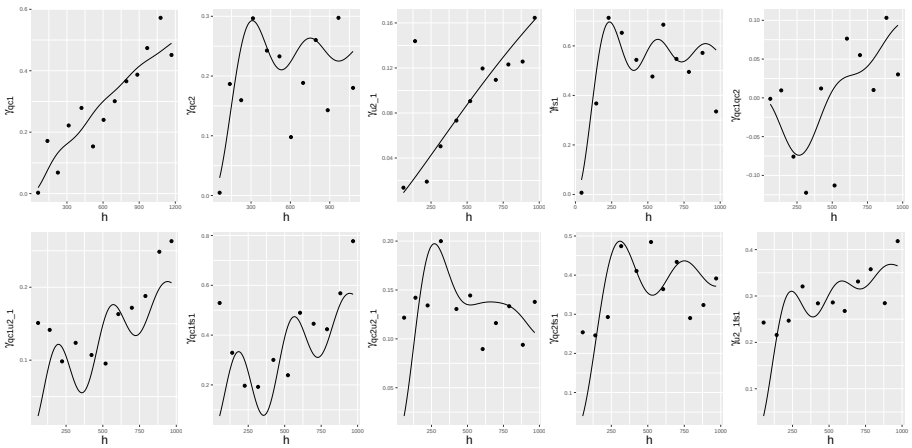
$$\left\{ \zeta^{21}(t) = \sqrt{\frac{3}{2}} t,\ \zeta^{22}(t) = \frac{\sqrt{5}(3t^2 - 1)}{2\sqrt{2}},\ \zeta^{23}(t) = \frac{5}{2}\sqrt{\frac{7}{2}} \left(t^3 - \frac{3}{5} t \right) \right\}.$$

Ambos son conjuntos de funciones ortonormales en $L^2[-1, 1]$. *Estas funciones pueden seleccionarse de diferentes bases o de una base común. Con respecto a los puntajes, sean* $\gamma_1(h; \Theta_1)$, $\gamma_2(h; \Theta_2)$, $\gamma_3((h); \Theta_3)$ *modelos válidos de semivarianza. El vector de puntuaciones* $(f_s^{11}, f_s^{12}, f_s^{13}, f_s^{21}, f_s^{22})$ *es una realización de una distribución gaussiana multivariada con estructura de correlación dada por un modelo válido de corregionalizacíón con base en los modelos* $\gamma_1(h; \Theta_1)$, $\gamma_2(h; \Theta_2)$, $\gamma_3((h); \Theta_3)$. *Ver Definición 11. Los vectores de puntajes se simulan de un procesos normal multivariado y con base en estos y en las bases de funciones ortonormales se construyen los datos funcionales simulados. En el paquete SpatFD [10] se pueden simular procesos funcionales espaciales tanto univariados como multivariados.*

Ejemplo 28 (...Estudio de suelos para el metro subterráneo). *Con el fin de construir los perfiles verticales de los suelos de Bogotá, se aplica cokriging funcional para predecir las tres variables aleatorias observadas en profundidad en el estudio para la viabilidad del metro subterráneo presentado en el Ejemplo 21.*

Se tienen tres variables aleatorias funcionales espaciales: $\chi_s^1(t)$*: la resistencia del cono a la penetración qc,* $\chi_s^2(t)$*:la fuerza de resistencia por fricción* fs *y* $\chi_s^3(t)$*: la presión de poro* $u2$*.*

Se lleva a cabo el cokriging funcional usando los dos primeros componentes de la variable qc, y el primero de cada una de las otras, $u2$ *y* fs*. Se modela la correlación cruzada para estos cuatro vectores de puntajes, es decir, para el vector aleatorio escalar espacial* $\left(f^{qc_1}(\boldsymbol{s}), f^{qc_2}(\boldsymbol{s}), f^{u2_1}(\boldsymbol{s}), f^{fs_1}(\boldsymbol{s})\right)$ *y con base en este modelo se lleva a cabo la predicción espacial funcional con covariables. Se realiza la predicción usando cokriging funcional en 5000 puntos en profundidad, repartidos en todo el corredor del metro. Posteriormente se hacen cortes verticales para determinar perfiles del suelo. La Figura 10.1 muestra cortes verticales de cada una de las tres variables funcionales. Los resultados fueron comparados con los estudios existentes de suelos y la coincidencia es casi perfecta.*



Modelo lineal de corregionalización ajustado a los puntajes vectoriales $\left(f^{qc_1}(\boldsymbol{s}), f^{qc_2}(\boldsymbol{s}), f^{u2_1}(\boldsymbol{s}), f^{fs_1}(\boldsymbol{s})\right)$, correspondientes a los componentes principales con la mayor variación explicada y la covarianza cruzada espacial más fuerte.

En primer lugar se realiza la estimación empírica de los semivariogramas tanto individuales como cruzados. Luego se ajustan los semivariogramas individuales y posteriormente con base en estos modelos se ajustan las matrices de corregionalización procurando capturar lo mas posible de los cruzados.

Para lograr una buena estimación, es importante complementar los resultados iniciales arrojados automáticamente por los paquetes como gstat de R cran project, [62] con un ajuste visual.

Esto se debe a que los paquetes deben imponer muchas restricciones incluso mas de las que el MLC tiene, para lograr estimar matrices de covarianza definidas positivas. Esto lleva a que los resultados no siempre son los mejores. Por esta razón, es importante

procurar parsimonia. Lograr una buena estimación de un MLC para muchas variables es extremadamente difícil.

Los cuatro modelos univariados usados fueron los siguientes 4 modelos acotados.

- $\gamma_{f^{qc_1}}(\boldsymbol{h}) = \text{Matern}(\Theta = (1, 1306), \kappa = 0.54)$,
- $\gamma_{f^{qc_2}}(\boldsymbol{h}) = \text{Wave}(\Theta = (1, 68.87))$,
- $\gamma_{f^{u2_1}}(\boldsymbol{h}) = \text{Circular}(\Theta = (1, 1899.45))$ *y*
- $\gamma_{f^{fs_1}}(\boldsymbol{h}) = \text{Wave}(\Theta = (1, 71.6))$

*A continuación se presentan las matrices de corregionalización, B*1*, B*2*, B*3 *y B*4*, para completar así los elementos del MLC. Esta metodología es descrita en detalle en el Capítulo 5 en el que se presentan los detalles del cokriging escalar y el modelamiento de la covarianza correspondiente.*

$$B1 = \begin{pmatrix} 0,6856198 & 0,23642266 & 0,3342175 & 0,23758051 \\ 0,2364227 & 0,08324265 & 0,1152632 & 0,08016074 \\ 0,3342175 & 0,11526318 & 0,1629204 & 0,11579767 \\ 0,2375805 & 0,08016074 & 0,1157977 & 0,08413924 \end{pmatrix}$$

$$B2 = \begin{pmatrix} 0,28111581 & -0,21372481 & -0,07689685 & -0,30865572 \\ -0,21372481 & 0,28657380 & 0,08562907 & 0,27273587 \\ -0,07689685 & 0,08562907 & 0,02698218 & 0,09276572 \\ -0,30865572 & 0,27273587 & 0,09276572 & 0,35057575 \end{pmatrix}$$

$$B3 = \begin{pmatrix} 0,19228074 & 0,03119251 & -0,08014018 & 0,16640880 \\ 0,03119251 & 0,08194164 & -0,14365412 & -0,02389135 \\ -0,08014018 & -0,14365412 & 0,25543582 & 0,01712071 \\ 0,16640880 & -0,02389135 & 0,01712071 & 0,17769931 \end{pmatrix}$$

$$B4 = \begin{pmatrix} 0,35486255 & 0,008982880 & 0,16888729 & 0,501219343 \\ 0,00898288 & 0,028879397 & 0,02038903 & 0,009839113 \\ 0,16888729 & 0,020389034 & 0,08943982 & 0,236939829 \\ 0,50121934 & 0,009839113 & 0,23693983 & 0,708221660 \end{pmatrix}$$

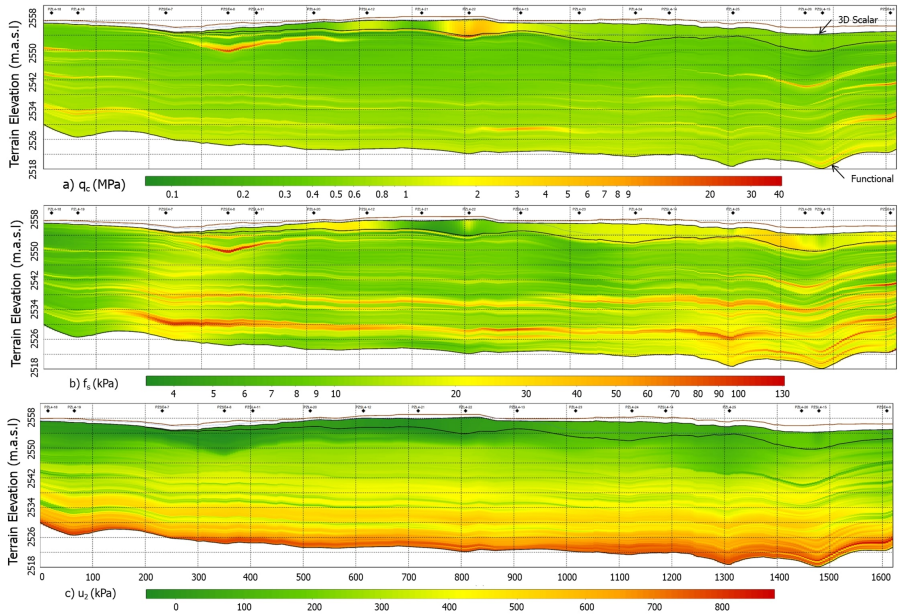


Figura 10.1. Predicciones de perfil para: a. Resistencia de la punta del cono, b. Fricción y c. Presión de poro.

La construcción del modelo completo en R cran project consiste en hacer la combinación lineal entre los modelos de semivarianza o covarianza seleccionados y las matrices de corregionalización. Esto es, los cuatro modelos de covarianza seleccionados, cada uno con su respectivo rango y parámetros adicionales ya definidos, excepto la silla que se fija en 1, dado que la sillas de todos los modelos vienen en las matrices B_1, B_2, B_3, B_4 son:

```
Modelo_qc_1=function(x){cov.spatial(x,cov.model="matern",cov.pars=c(1,1306),kappa = 0.54)}
Modelo_qc_2=function(x){cov.spatial(x,cov.model="wave",cov.pars=c(1,68.87))}
Modelo_u2_1=function(x){cov.spatial(x,cov.model="circular",cov.pars=c(1,1899.45))}
Modelo_fs_1=function(x){cov.spatial(x,cov.model="wave",cov.pars=c(1,71.6))}}

################################################
###Modelo lineal de corregionalización
################################################


qc_1=function(B1,B2,B3,B4,x){B1[1,1]*Modelo_qc_1(x)+B2[1,1]*Modelo_qc_2(x)+B3[1,1]*Modelo_u2_1(x)+B4[1,1]*Modelo_fs_1(x)}
qc_2=function(B1,B2,B3,B4,x){B1[2,2]*Modelo_qc_1(x)+B2[2,2]*Modelo_qc_2(x)+B3[2,2]*Modelo_u2_1(x)+B4[2,2]*Modelo_fs_1(x)}
u2_1=function(B1,B2,B3,B4,x){B1[3,3]*Modelo_qc_1(x)+B2[3,3]*Modelo_qc_2(x)+B3[3,3]*Modelo_u2_1(x)+B4[3,3]*Modelo_fs_1(x)}
fs_1=function(B1,B2,B3,B4,x){B1[4,4]*Modelo_qc_1(x)+B2[4,4]*Modelo_qc_2(x)+B3[4,4]*Modelo_u2_1(x)+B4[4,4]*Modelo_fs_1(x)}
qc_1_qc_2=function(B1,B2,B3,B4,x){B1[1,2]*Modelo_qc_1(x)+B2[1,2]*Modelo_qc_2(x)+B3[1,2]*Modelo_u2_1(x)+B4[1,2]*Modelo_fs_1(x)}
qc_1_u2_1=function(B1,B2,B3,B4,x){B1[1,3]*Modelo_qc_1(x)+B2[1,3]*Modelo_qc_2(x)+B3[1,3]*Modelo_u2_1(x)+B4[1,3]*Modelo_fs_1(x)}
qc_1_fs_1=function(B1,B2,B3,B4,x){B1[1,4]*Modelo_qc_1(x)+B2[1,4]*Modelo_qc_2(x)+B3[1,4]*Modelo_u2_1(x)+B4[1,4]*Modelo_fs_1(x)}
qc_2_u2_1=function(B1,B2,B3,B4,x){B1[2,3]*Modelo_qc_1(x)+B2[2,3]*Modelo_qc_2(x)+B3[2,3]*Modelo_u2_1(x)+B4[2,3]*Modelo_fs_1(x)}
qc_2_fs_1=function(B1,B2,B3,B4,x){B1[2,4]*Modelo_qc_1(x)+B2[2,4]*Modelo_qc_2(x)+B3[2,4]*Modelo_u2_1(x)+B4[2,4]*Modelo_fs_1(x)}
u2_1_fs_1=function(B1,B2,B3,B4,x){B1[3,4]*Modelo_qc_1(x)+B2[3,4]*Modelo_qc_2(x)+B3[3,4]*Modelo_u2_1(x)+B4[3,4]*Modelo_fs_1(x)}
```

Ejemplo 29 (Calidad de aire de la ciudad de México). *Se analizan a continuación datos de calidad del aire de la ciudad de México durante la temporada seca de 2015 debido a que en la temporada lluviosa la concentración de los contaminantes en el aire disminuye.*

Los datos corresponden a horas consecutivas desde el $1.^o$ *de enero a la 1:00 a.m. hasta el 30 de mayo a las 12:00 a.m., en 23 estaciones ambientales (ver Figura 10.2 panel izquierdo).*

Las estaciones en la red de de la calidad del aire RAMA monitorean cada hora el material particulado (PM10) que tenga un tamaño hasta de 10 micrómetros y el dióxido de nitrógeno, (NO2) entre otros. Ver [8] para más detalles acerca de la red y de los efectos adversos del PM10 y el NO2 en la salud de las personas y en el deterioro de las cosas expuestas. La temperatura (Temp) es tomada de la red meteorológica REDMET. RAMA y REDMET tienen 15 estaciones en común.

La secretaria de medio ambiente de Ciudad de México actualmente opera la red de calidad del aire para obtener, procesar y divulgar la calidad del aire, así como para evaluar el cumplimiento de las normas y definir políticas de control de la contaminación. Los datos son obtenidos del sistema de monitoreo automático [69].

Para construir las curvas, se usan bases de funciones B-splines de orden 4 con knots igualmente espaciados y un parámetro de suavizamiento de $0,00001$. *Para el conjunto de datos de PM10 se usa un conjunto de* 163 *B-splines, para NO2, se usan* 157 *B-splines, y para el conjunto de datos de Temp se usan* 121 *B-splines. La Figura 10.2 muestra el mapa de México con las redes RAMA y REDMET así como las curvas observadas de estos procesos espaciales para la última semana del período seleccionado.*

El primer componente principal explica el $75\,\%$*;* $84,6\,\%$ *y* $85,7\,\%$ *de la variabilidad para PM10, NO2 y Temp respectivamente, y el segundo componente principal solo explica* $13,9\,\%$*;* $13,8\,\%$ *y* $13,1\,\%$. *Así, usando* $85\,\%$ *como umbral, se incluyen dos vectores de puntajes para PM10, mientras que solo el primer vector de puntajes es incluido para NO2 y Temp.*

El interés principal es la predicción espacial funcional del PM10 usando NO2 y Temp como covariables funcionales. Siguiendo la notación de la Sección 10, PM10 es $\chi_s^1(t)$*, NO2 es* $\chi_s^2(t)$ *y Temp es* $\chi_s^3(t)$. *La Figura 10.3 muestra los variogramas empíricos y teóricos ajustados según el modelo lineal de corregionalización.*

Se utilizan dos estructuras Matern anidadas usando el MLC, esto es, combinadas linealmente, con parámetros de suavizamiento 0,1 y 5; y rangos 3.000 y 13.000. Por lo tanto, γ_f^{11} *y* γ_f^{12} *son los semivariogramas para los dos primeros componentes principales de PM10,* γ_f^{21} *y* γ_f^{31} *son los variogramas para los vectores de puntaje correspondientes al primer componente principal de NO2 y Temp, respectivamente. Todos los semivariogramas tanto empíricos como ajustados se muestran en la Figura 10.3. La tabla 10.1 contiene las sillas correspondientes a los semivariogramas entre todos los pares posibles de vectores de puntaje.*

De acuerdo con los variogramas empíricos, no hay razón para suponer discontinuidad en el origen, ya que no hay salto en $\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\| = 0$. *Por este motivo, se ajustan modelos sin efecto pepita.*

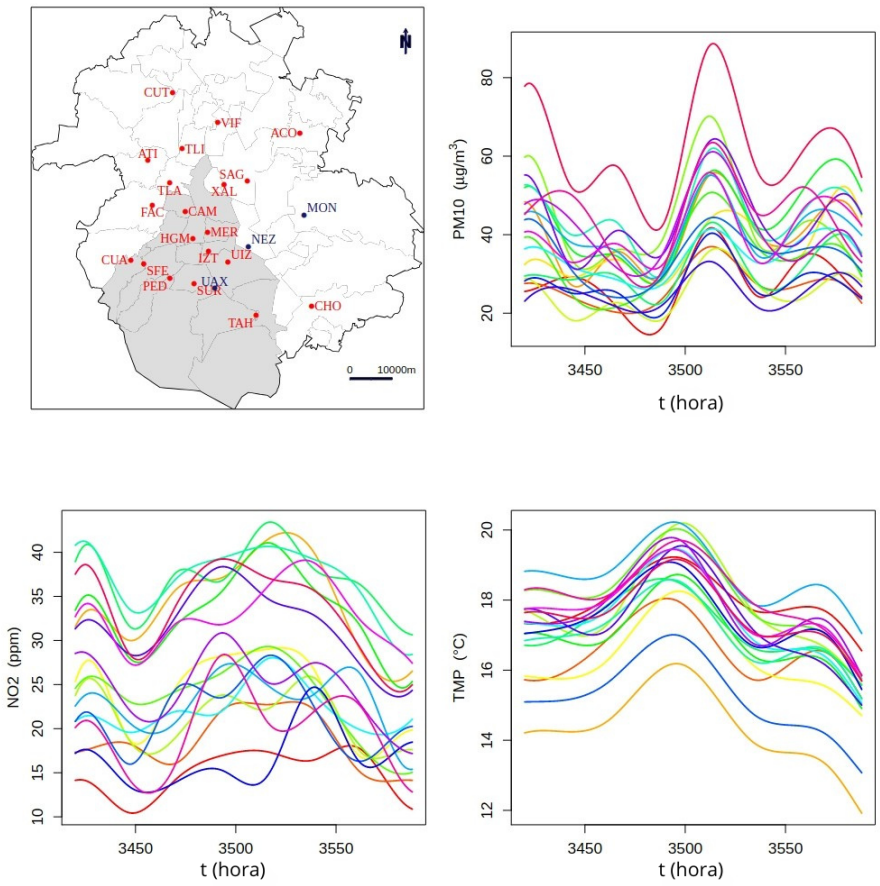


Figura 10.2. *Panel arriba izquierda*: ciudad de México. Red de calidad del aire RAMA (estaciones en rojo). Los puntos azules son las estaciones que además de las de RAMA miden la temperatura pero que pertenecen a REDMET. *Panel arriba derechal*: PM10, Mayo 23 a Mayo 30, 2015. *Panel abajo izquierda*: NO2, Mayo 23 a Mayo 30, 2015.*Panel abajo derecha*: temperatura, Mayo 23 a Mayo 30, 2015.

Modelo	$\hat{\sigma}^2$									
$\hat{\gamma}_f^{pk}$	$\hat{\gamma}_{f^{11}}$	$\hat{\gamma}_{f^{21}}$	$\hat{\gamma}_{f^{31}}$	$\hat{\gamma}_{f^{11}f^{12}}$	$\hat{\gamma}_{f^{11}f^{21}}$	$\hat{\gamma}_{f^{11}f^{31}}$	$\hat{\gamma}_{f^{12}}$	$\hat{\gamma}_{f^{12}f^{21}}$	$\hat{\gamma}_{f^{12}f^{31}}$	$\hat{\gamma}_{f^{21}f^{31}}$
v =0,1	216778,0	24415,3	56381,7	-11931	71160,1	104764,3	-11931,2	31932,8	-3508,696	-6856,4
v =5	1923456,9	386064,8	1181738,7	202568	-680077,6	-92372,9	202568,5	388685,0	3251,8	164083,5

Tabla 10.1. Modelo lineal de corregionalización usando dos estructuras de Matern.

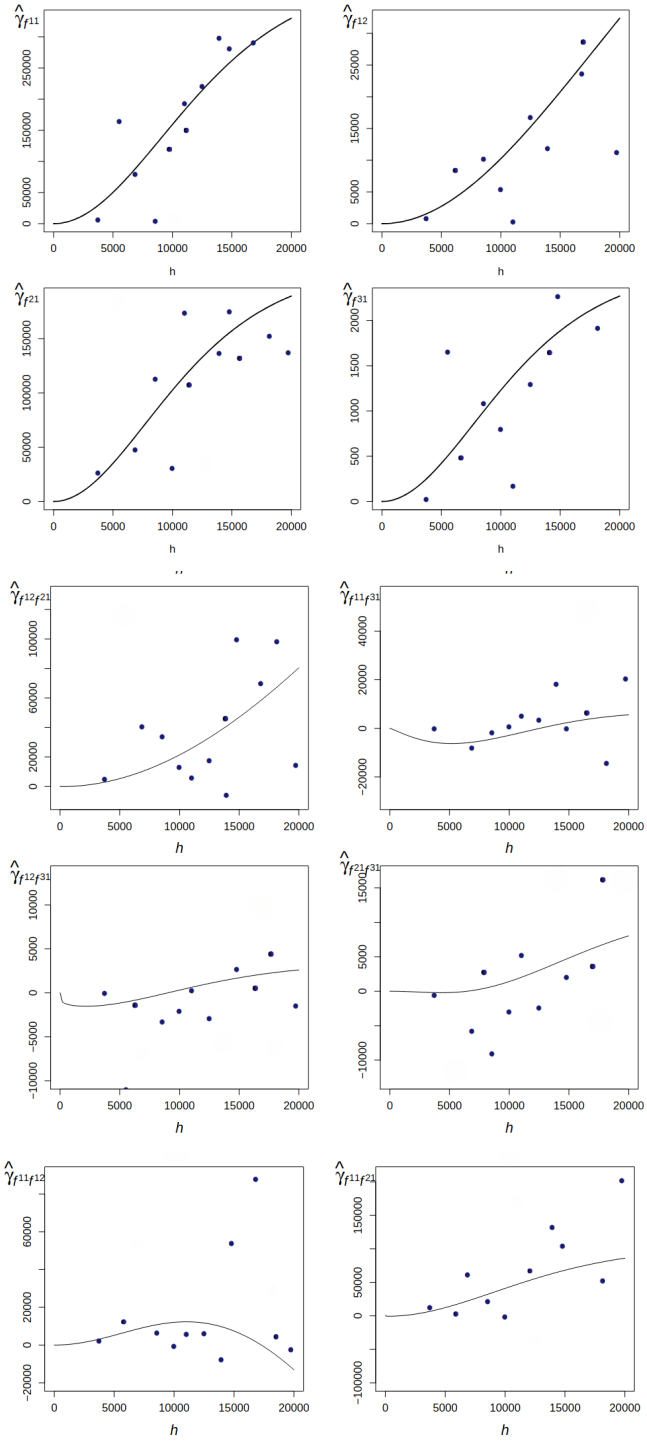


Figura 10.3. Variogramas empíricos y ajustados usando el MLC.

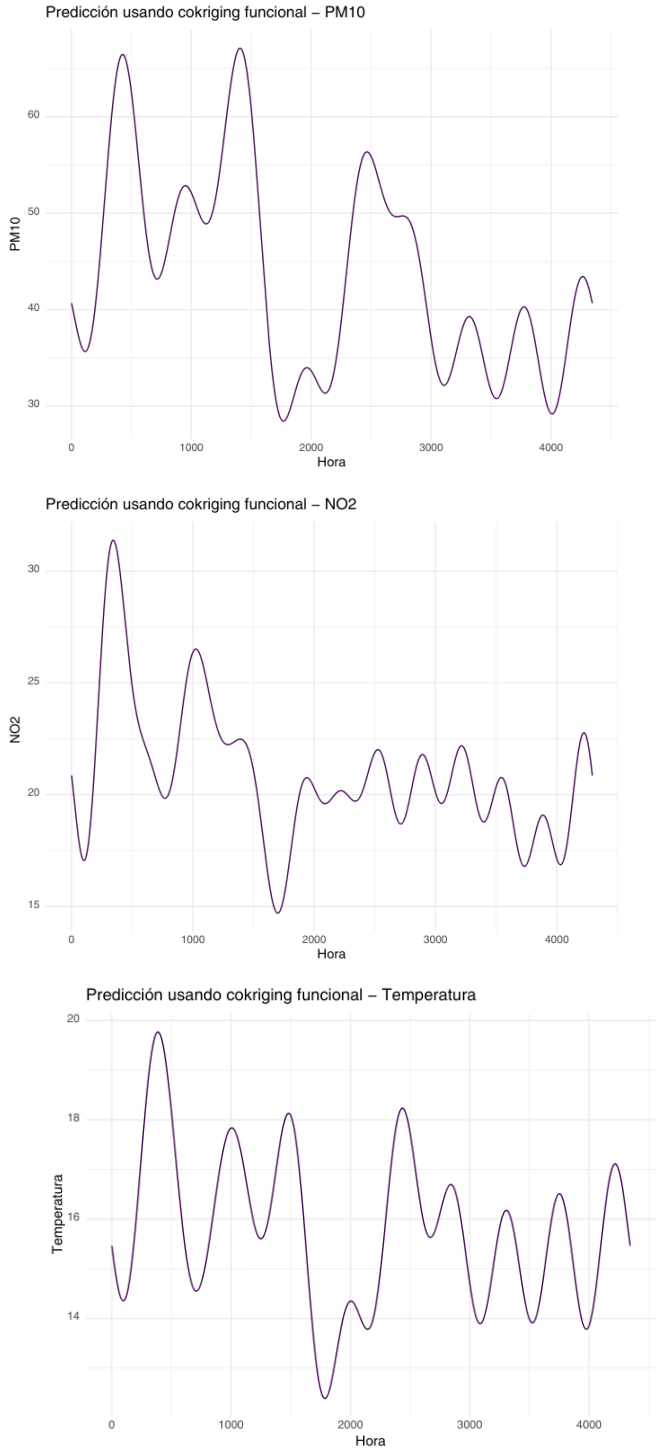


Figura 10.4. Predicción usando cokriging funcional para el PM10, el NO2 y la temperatura de México en la ubicacion espacial (509926,2179149) usando el paquete SpatFD, [10].

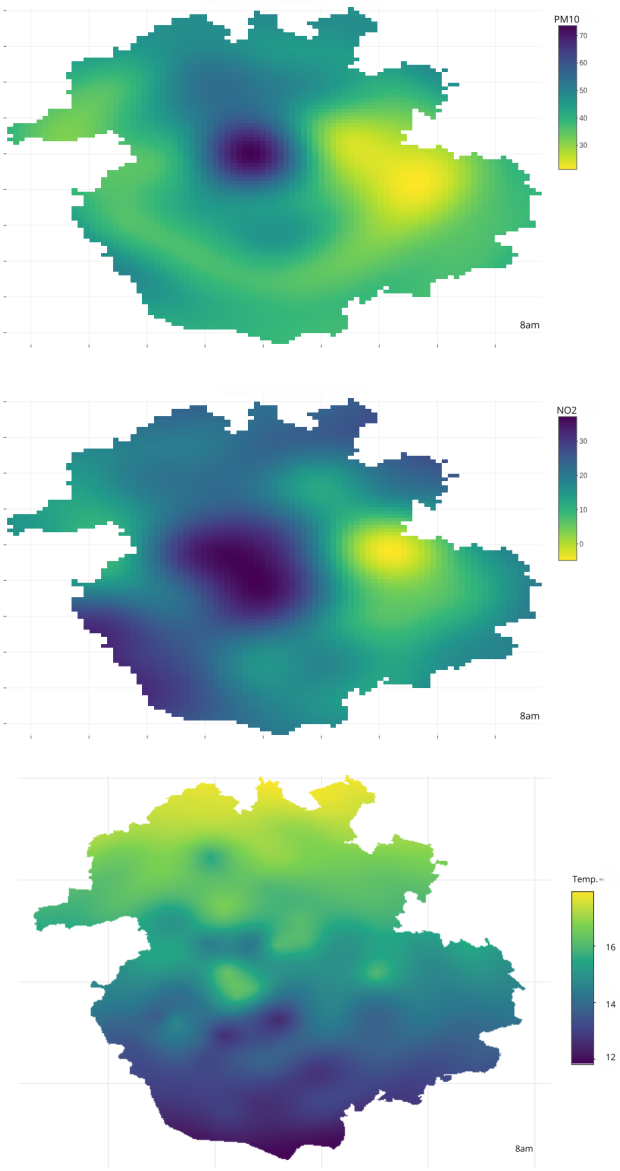


Figura 10.5. Mapas de las tres variables en un mismo momento del tiempo, generados haciendo los respectivos cortes del volumen obtenido aplicando cokriging funcional en muchos puntos en el espacio.

Predictor alternativo. Kriging o cokriging simple escalar de los puntajes

Finalmente, se presenta un predictor alternativo que no optimiza directamente la predicción de los datos funcionales, sino que usa kriging o cokriging escalar sobre los scores y luego lleva a cabo la combinación lineal de estos predictores escalares con las funciones propias.

Una alternativa es usar un predictor que no minimiza directamente el cuadrado de la norma del error de predicción funcional sino que aplica geoestadística escalar a los vectores de puntaje y reemplaza las predicciones encontradas en la expansión de Karhunen-Loève [12]. Esta alternativa es una buena opción, pero es importante enfatizar en que en este caso, no se optimiza una medida de variabilidad funcional, mientras que las propuestas presentadas en los capítulos 9 y 10 están construidas directamente con objetos funcionales.

[33] y [56] aproximan cada función del conjunto de datos usando bases de K funciones y utilizan cokriging escalar del proceso espacial formado por los puntajes para predecir los coeficientes correspondientes a las bases de funciones en el sitio no muestreado $\boldsymbol{s}_0$. Sin embargo, cuando el número de bases aumenta, también aumenta la dimensión del campo aleatorio multivariado de coeficientes, y el modelo lineal de corregionalización usado en el cokriging para construir una matriz de covarianza válida se vuelve intratable computacionalmente.

Por lo tanto, para que esta metodología se pueda utilizar, se requiere el uso de un sistema de funciones base que asegure un número reducido de dimensiones. Los coeficientes involucrados en la reconstrucción de cada función con los EFPC pueden ser dos o tres en la mayoría de casos, lo que hace más factible la utilización del MLC.

Así, una opción muy práctica es el uso de los *componentes principales funcionales empíricos*. De acuerdo con la Sección 8.3, se tiene que

$$E\left[\chi_s(t)\right] = 0, \quad E\left[f_k(\boldsymbol{s}_i)\right] = 0, \quad i = 1, \ldots, n \ \text{ y } \ k = 1, \ldots, K$$

por lo tanto, se puede usar cokriging escalar simple (ver Sección 5) para predecir el vector

$$\boldsymbol{f}(\boldsymbol{s}_0) = (f_1(\boldsymbol{s}_0), \ldots, f_K(\boldsymbol{s}_0))'$$

en la ubicación no muestreada $\boldsymbol{s}_0$ (ver Sección 5). El predictor cokriging simple del vector de puntajes en $\boldsymbol{s}_0$ se encuentra con los métodos descritos en la Sección 5 con

$$\boldsymbol{f}(\boldsymbol{s}_i) = (f_1(\boldsymbol{s}_i), \ldots, f_K(\boldsymbol{s}_i))'$$

$$\boldsymbol{\Sigma}_{(\boldsymbol{s}_i, \boldsymbol{s}_{i'})} = \left(Cov\left(f_k(\boldsymbol{s}_i), f_{k'}(\boldsymbol{s}_{i'})\right)\right), \ k, k' = 1, \ldots, K.$$

Aunque se usa una base ortonormal, las covarianzas cruzadas entre los respectivos vectores de puntajes representan la covarianza cruzada entre pares de funciones observadas. En general, no existe ninguna razón para asumir independencia entre los vectores de puntaje en distintos sitios. Nótese que

$$
\begin{aligned}
E\left[f_k(\boldsymbol{s}_i)f_{k'}(\boldsymbol{s}_{i'})\right] &= E\left[\langle \chi_{\boldsymbol{s}_i}, \xi_k\rangle\langle \chi_{\boldsymbol{s}_{i'}}, \xi_{k'}\rangle\right] \\
&= E\left(\int \chi_{\boldsymbol{s}_i}(t)\xi_k(t)dt \int \chi_{\boldsymbol{s}_{i'}}(r)\xi_{k'}(r)dr\right) \\
&= E\left(\int\int \chi_{\boldsymbol{s}_i}(t)\xi_k(t)\,\chi_{\boldsymbol{s}_{i'}}(r)\xi_{k'}(r)dtdr\right) \\
&= \int \xi_{k'}(r)\left(\int E\left(\chi_{\boldsymbol{s}_i}(t)\,\chi_{\boldsymbol{s}_{i'}}(r)\right)\xi_k(t)dt\right)dr \\
&= \int \xi_{k'}(r)\left(\int c_{\boldsymbol{s}_i,\boldsymbol{s}_{i'}}(t,r)\xi_k(t)dt\right)dr \\
&= \int \xi_{k'}(r)C_{\boldsymbol{s}_i,\boldsymbol{s}_{i'}}(\xi_k)dr \\
&= \langle C_{\boldsymbol{s}_i,\boldsymbol{s}_{i'}}(\xi_k), \xi_{k'}\rangle
\end{aligned} \tag{10.13}
$$

Si $i = i'$, la covarianza para diferentes puntajes es 0. Es decir, la ortogonalidad se verifica para diferentes vectores de puntaje en un mismo lugar. Entonces,

$$
\begin{aligned}
E\left(f_k(\boldsymbol{s}_i)f_{k'}(\boldsymbol{s}_i)\right) &= \int \xi_{k'}(r)\left(\int c_{\boldsymbol{s}_i,\boldsymbol{s}_i}(t,r)\xi_k(t)dt\right)dr \\
&= \int \xi_{k'}(r)C_{\boldsymbol{s}_i,\boldsymbol{s}_i}(\xi_k)dr \\
&= \int \xi_{k'}(r)\eta_k\xi_k(r)dr \\
&= \eta_k\langle \xi_{k'}, \xi_k\rangle \\
&= \begin{cases} \eta_k, & if \quad k = k' \\ 0 & if \quad k \neq k' \end{cases}
\end{aligned} \tag{10.14}
$$

Sea $\boldsymbol{\xi}'(t)$ el vector que contiene las primeras K funciones propias seleccionadas. La representación de las funciones en términos de sus componentes principales funcionales está dada por

$$\chi_{\boldsymbol{s}_i}(t) = \boldsymbol{\xi}'(t)\boldsymbol{f}(\boldsymbol{s}_i),\ i = 1, \ldots, n,$$

y la predicción de la curva $\chi_{\boldsymbol{s}_0}(t)$ es

$$\chi^*_{\boldsymbol{s}_0}(t) = \boldsymbol{\xi}'(t)\boldsymbol{f}^*(\boldsymbol{s}_0), \qquad i = 1, \ldots, n.$$

$f^*(\boldsymbol{s}_0)$ es el predictor cokriging simple del vector de puntajes en $\boldsymbol{s}_0$.

La función de autocovarianza espacial para cada vector de puntajes k es

$$
E\left(f_k(\boldsymbol{s}_i)f_k(\boldsymbol{s}_{i'})\right) = \begin{cases} \eta_k, & if \quad i = i' \\ \eta_k\rho_k\left(\|\boldsymbol{s}_i - \boldsymbol{s}_{i'}\|; \boldsymbol{\Theta}\right) & if \quad i \neq i' \end{cases} \tag{10.15}
$$

donde $\rho_k(.)$ es la función de autocorrelación del campo aleatorio escalar $f_k(s)$. Las funciones válidas utilizadas aquí, así como los métodos para su estimación y predicción son los mismos revisados en los capítulos 2, 3, 4 y 5.

Así, (10.15) muestra que el campo aleatorio es estacionario de segundo orden; la silla de la función de covarianza es la varianza de cada vector de puntajes, es decir, el valor propio como máximo. Sin embargo, para cada modelo el valor de su parámetro de rango muestra de una manera diferente el alcance de la silla. Por ejemplo, en los modelos de soporte compacto como el esférico la silla se alcanza perfectamente a partir del valor del rango, mientras que en modelos como el exponencial, el valor del rango indica la distancia a la cual empieza a disminuir la velocidad de crecimiento del semivariograma.

En consecuencia, el modelo de autocovarianza de cada $f_k(s_i)$ tiene silla conocida y finita. Las matrices en la diagonal principal de (5.7) pueden ser denotadas en forma más general como $\mathbf{\Sigma}_0$, esto es,

$$\mathbf{\Sigma}_0 = \mathbf{\Sigma}_{(s_i,s_i)} = \left(Cov\left(f_k(s_i), f_{k'}(s_i)\right)\right), \; k, k' = 1, \ldots, K.$$

Por lo tanto, su traza es constante y está dada por

$$Tr\left(\mathbf{\Sigma}_0\right) = \sum_{k=1}^{K} \eta_k \tag{10.16}$$

La varianza del error de predicción puede ser obtenida usando la representación de la variable aleatoria funcional en términos de su descomposición usando FPC, esto es,

$$\begin{aligned} Var\left(\chi_{s_0}(t) - \chi^*_{s_0}(t)\right) &= Var\left(\boldsymbol{\xi}'(t)\boldsymbol{f}_{s_0} - \boldsymbol{\xi}'(t)\boldsymbol{f}^*(s_0)\right) \\ &= \boldsymbol{\xi}'(t) Var\left(\boldsymbol{f}(s_0) - \boldsymbol{f}^*(s_0)\right)\boldsymbol{\xi}(t) \\ &= \boldsymbol{\xi}'(t)\left(Tr\left(\Sigma_0\right) - Tr\left(\sum_{i=1}^{n}\left(\mathbf{\Sigma}_{(s_0,s_i)}\Gamma_i\right)\right)\right)\boldsymbol{\xi}(t) \\ &= \sigma^2_{f(s_0)-f^*(s_0)}\boldsymbol{\xi}'(t)\boldsymbol{\xi}(t) \end{aligned} \tag{10.17}$$

donde

$$\sigma^2_{f(s_0)-f^*(s_0)} = Tr\left(\mathbf{\Sigma}_0\right) - Tr\left(\sum_{i=1}^{n}\left(\mathbf{\Sigma}_{(s_0,s_i)}\Gamma_i\right)\right)$$

es la varianza acumulada del predictor cokriging escalar del vector

$$\boldsymbol{f}(s_0)$$

donde

$$Tr\left(\mathbf{\Sigma}_0\right)$$

es constante (ver (10.16)).

10.4. Ejercicios

1. Compare la calidad de las predicciones para un campo aleatorio espacial funcional bivariado, usando kriging funcional para cada variable por separado y usando cokriging funcional.
2. Sean los procesos funcionales espaciales $\chi_s^1(t)$ y $\chi_s^2(t)$, los cuales son observados en n_1 y una (1) ubicaciones, respectivamente. Se requiere predecir $\chi_{s_0}^1(t)$. Encuentre la ponderación que le corresponde a la observación que se tiene de $\chi_s^2(t)$ si se usa el predictor cokriging

   - Si $\mu_2(t)$ es conocida y constante,
   - Si $\mu_2(t)$ es desconocida pero constante.
   - Influye si la ubicación donde $\chi_s^2(t)$ es observada, coincide (o no) con alguna de las n_1 ubicaciones donde $\chi_s^1(t)$ es observada?
   - Sería necesario tener una realización de $\chi_s^2(t)$ en s_0? Justifique.

3. Verifique que los conjuntos de funciones usadas en el Ejemplo 27 son ortonormales.
4. Construya modelos de corregionalización utilizando las funciones de semivarianza propuestas en el Ejemplo 27.
5. Utilice el paquete SpatFD de R cran project ([10]) para simular campos aleatorios funcionales, tanto univariados como multivariados con diferentes estructuras de covarianza y usando diferentes familias de componentes principales. Realice una descripción de cada conjunto simulado. Compare los predictores vistos para conjuntos de ubicaciones no observadas. Justifique en todos los casos.

Capítulo

once

Diseños de muestreo óptimo para la predicción espacial

Los diseños de muestreo para procesos en dominio continuo, y espacialmente correlacionados, encuentran la configuración espacial ideal de ubicaciones dónde deben observarse las variables de interés con fines de estimación de las funciones de covarianza o semivarianza espacial y predicciones óptimas en lugares no observados. Se conocen como diseños de muestre óptimos, [54].

Nótese que, dado que estos procesos ocurren en infinitos lugares, existen infinitas muestras posibles y, además, lo que ocurre en un sitio da información sobre lo que ocurre a su alrededor. En consecuencia, una muestra de tamaño n de datos independientes contiene más información que una muestra del mismo tamaño de datos correlacionados.

En contraste, los diseños de muestreo tales como el aleatorio simple, el sistemático, el estratificado, ppt, πpt por nombrar algunos, son específicos para poblaciones finitas de individuos independientes, y su fin en general, es la estimación de un parámetro, como la media, un total o una proporción. Por lo tanto, este enfoque no es el apropiado en el contexto espacial.

11.1. Diseños óptimos de muestreo espacial. Definición y generalidades.

Un diseño óptimo de muestreo es aquel que encuentra la mejor combinación predictor-diseño o estimador-diseño, de acuerdo con la optimización de un criterio previamente establecido. Por lo tanto, el criterio de diseño óptimo se define con base en los objetivos de cada estudio. Así, por ejemplo si el objetivo principal es la predicción óptima, el diseño determina las ubicaciones que permiten minimizar la varianza del error de predicción.

Definición 14 (Diseño óptimo de muestreo). *Un diseño óptimo S_n^* es definido como la configuración de sitios espaciales tal que*

$$S_n^* = \arg \max_{S_n \in \Xi_n} \Phi(\boldsymbol{\Theta}, S_n). \tag{11.1}$$

donde $\Phi(\boldsymbol{\Theta}, S_n)$ es el criterio de diseño, y puede ser cualquier medida escalar de información o calidad, obtenida a partir de la configuración S_n y que depende del vector de parámetros $\boldsymbol{\Theta}$.

El criterio de diseño en el contexto de muestreo espacial depende de una medida de incertidumbre asociada al objetivo principal del estudio. En el caso usual de que el objetivo sea la predicción óptima, hay que tener en cuenta que esta a su vez depende de la estructura de covarianza espacial ([7] y [8]). Hay que tener en cuenta que D_s es un conjunto continuo, así que existen infinitas

opciones para definir nuevos sitios de observación, lo que hace el proceso complejo computacionalmente.

En la práctica, una opción es llevar a cabo la optimización sobre un conjunto $D'_s \subset D_s$, que contiene un número finito de configuraciones posibles previamente determinadas. D'_s debe ser construido de acuerdo con el conocimiento de las condiciones de la región, las condiciones de acceso y las restricciones económicas. Hay sitios dónde por sus características no es posible instalar un aparato de medición. Adicionalmente, es importante notar que no tiene sentido tomar observaciones en sitios extremadamente cerca, ya que la correlación espacial lleva a información redundante y, por lo tanto, a un desperdicio de recursos. Otra opción, es definir el criterio sobre una grilla regular fina de la región de interés.

11.2. Diseño y rediseño de una red de muestreo

En ocasiones se toman los datos por una única vez en los lugares seleccionados. Pero, en muchos casos, para tomar series de observaciones espacio-temporales se instalan dispositivos en los sitios de muestreo, y estos dispositivos registran de manera automática los datos. Por ejemplo, las redes ambientales o meteorológicas. A estos conjuntos se les suele llamar redes de muestreo. Si es la primera vez que se va a establecer la configuración, se le llama diseño de la red. Si la red ya existe pero se va a revisar, aumentar o disminuir se le llama rediseño. Conservar el histórico de datos a través del tiempo es importante, por lo tanto, no es usual mover estaciones que llevan mucho tiempo funcionando bien, a menos que haya varias muy cerca y la alta correlación entre ellas genere redundancia, mostrando que se puede prescindir de alguna estación. En cada caso lo importante es definir el objetivo y determinar la función a optimizar. Se presentan a continuación algunos casos para ilustrar la metodología.

11.3. Diseños de muestreo óptimo para la predicción espacial escalar

A continuación se presenta en detalle el procedimiento para el rediseño de una red en caso de que el objetivo sea la predicción óptima y se quieran instalar nuevas estaciones y mantener las existentes.

Se requiere agregar m estaciones de medición de una variable espacial y el objetivo es su predicción óptima en el conjunto S_0 de B ubicaciones, en las cuales no hay observaciones

$$S_0 = \{\boldsymbol{s}_0^1, \ldots, \boldsymbol{s}_0^B\}$$

Sea

$$S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}.$$

el conjunto actual de ubicaciones de muestreo, y sea

$$S_m = \{\boldsymbol{s}_{n+1}, \ldots, \boldsymbol{s}_{n+m}\}.$$

el conjunto de nuevas ubicaciones que debe ser determinado. La red de muestreo aumentada es,

$$\overline{S} = S \cup S_m = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{s}_{n+1}, \ldots, \boldsymbol{s}_{n+m}\} \tag{11.2}$$

Por lo tanto, entre todos los posibles subconjuntos con m ubicaciones espaciales tales que $S_m \subset D'_s$, se selecciona aquel conjunto S_m tal que $\overline{S}$ minimiza la suma de las B varianzas del error del predictor kriging en los s_0^b, $b = 1, \ldots, B$ lugares de interés. Si ya se conocen o se tiene una buena estimación de los parámetros de la función de covarianza, el criterio de diseño está dado por $\Phi\left(\overline{S}\right)$. Nótese que para cada configuración evaluada en el criterio, la varianza del predictor kriging simple se puede calcular usando la función de covarianza (ver expression (4.9)).

$$\Phi\left(\overline{S}\right) = \sum_{b=1}^{B} \text{Var}\left[Z^*\left(s_0^b\right) - Z\left(s_0^b\right)\right]. \tag{11.3}$$

El conjunto de ubicaciones $\overline{S}$ que minimice $\Phi\left(\overline{S}\right)$, contiene la configuración espacial en la cual deben tomarse los datos para garantizar que la suma de las varianzas del error de predicción sea mínima y es el diseño de muestreo óptimo (ver (11.2)). De manera explícita la varianza del kriging simple en términos de la función de covarianza es,

$$\Phi\left(\overline{S}\right) = \sum_{b=1}^{B} \left(C\left(0\right) - \overline{\boldsymbol{\sigma}}_0^b \overline{\Sigma}^{-1} \overline{\boldsymbol{\sigma}}_0^b\right). \tag{11.4}$$

$C\left(0\right)$ es la silla del modelo, $\overline{\Sigma}$ es la $(n+m) \times (n+m)$- matriz de covarianza del vector

$$\mathbb{Y} = \left(Y(\boldsymbol{s_1}), \ldots, Y(\boldsymbol{s_n}), Y(\boldsymbol{s_{n+1}}) \ldots, , Y(\boldsymbol{s_{n+m}})\right)$$

$\overline{\sigma}^b$ es el $(n+m)$-vector de covarianza entre el proceso en el lugar a predecir s_0^b y el vector $\mathbb{Y}$. Dado que la silla es constante minimizar el criterio (11.4) es equivalente a maximizar

$$\Phi\left(\overline{S}\right) = \sum_{b=1}^{B} \overline{\boldsymbol{\sigma}}_0^b \overline{\Sigma}^{-1} \overline{\boldsymbol{\sigma}}_0^b \tag{11.5}$$

Esto es, el criterio de diseño queda

$$\overline{S}^* = \arg\max_{\overline{S} \in \wp_{n+m}} \Phi\left(\overline{S}\right). \tag{11.6}$$

$\wp_{n+m}$ es el conjunto de todas las configuraciones $\overline{S}$ de $n+m$ lugares en D_s, sobre el cual se lleva a cabo la optimización. Ver por ejemplo [74] y [54].

En general, cualquier criterio de diseño, como por ejemplo la expresión (11.4), puede desarrollarse en términos del semivariograma. De hecho, si no hay estacionariedad de segundo orden, pero si intrínseca, es necesario plantear la varianza del kriging en términos de $\gamma(\boldsymbol{h})$. Un objetivo alternativo puede ser encontrar el diseño de muestreo que permita optimizar la estimación del vector de parámetros Θ de un modelo de semivariograma. El criterio de diseño cambia de acuerdo al objetivo del estudio.

11.4. Diseños de muestreo óptimo para la predicción espacial de datos funcionales.

En esta sección se derivan los criterios de diseño para optimizar la predicción espacial de las curvas utilizando los predictores discutidos en los Capítulos 9 y 10. La idea es la misma, determinar el objetivo y la medida de calidad asociada será el criterio de diseño. En este caso el objetivo es optimizar la predicción espacial de curvas.

11.4.1. Optimización de la predicción espacial de curvas. Una variable aleatoria funcional.

Extendiendo la Sección 11.3 al caso funcional, supongamos que se requiere encontrar los m sitios óptimos para ubicar m nuevas estaciones. Dado que se pretende encontrar la predicción óptima para un sitio no muestreado $\boldsymbol{s}_0$, la expresión que da el criterio de diseño, es la varianza del error de predicción del kriging simple con base en la red ampliada

$$\overline{S} = S \cup S_m$$

aplicando (9.4), esta varianza se expresa como

$$\sigma^2_{SK} = \sum_{k=1}^{\infty} \eta_j - 2\varsigma\lambda + \lambda^T \Omega \lambda \tag{11.7}$$

donde ς es el $(n+m)$-vector formado por las autocovarianzas entre las variables funcionales observadas y la variable funcional a predecir en el sitio $\boldsymbol{s}_0$,

$$\varsigma = \left(\sum_{k=1}^{\infty} E\left[f_k(\boldsymbol{s}_i) f_k(\boldsymbol{s}_0) \right] \right), \quad i = 1, \ldots, n \tag{11.8}$$

y $\boldsymbol{\Omega}$ es la $(n+m)\times(n+m)$-matriz formada por las autocovarianzas entre las variables funcionales observadas

$$\boldsymbol{\Omega} = \sum_{k=1}^{\infty} \Sigma_k \quad \text{y} \quad \Sigma_k = E\left[f_k(\boldsymbol{s}_i) f_k(\boldsymbol{s}_{i'})\right], \quad i, i' = 1, \ldots, n \tag{11.9}$$

Nótese que tanto la matriz como el vector son calculables porque son la suma de las autocovarianzas espaciales de los puntajes resultantes de los FPCA, las cuales son funciones de la distancia entre ubicaciones. Esto es lo que permite computar el criterio involucrando la ubicación a predecir aunque allí no se cuente con observación (11.8) y (11.9).

Por otra parte, el vector solución usando kriging funcional simple es $\boldsymbol{\lambda} = \boldsymbol{\Omega}^{-1}\boldsymbol{\varsigma}$, por lo tanto, sustituyendo en (11.7) la expresión para la varianza del error de predicción queda

$$\sigma^2_{SK} = \sum_{k=1}^{\infty} \eta_k - \boldsymbol{\varsigma}\boldsymbol{\Omega}\boldsymbol{\varsigma}$$

lo que reduce el criterio de diseño para la predicción óptima de las B curvas a

$$\arg\max_{S \in \wp_{n+m}} \sum_{b=1}^{B} \overline{\boldsymbol{\varsigma}}_0^b \overline{\boldsymbol{\Omega}} \overline{\boldsymbol{\varsigma}}_0^b \tag{11.10}$$

donde para $i = 1, \ldots, n$ y para cada b se tiene el vector

$$\overline{\boldsymbol{\varsigma}}_0^b = \left(\sum_{k=1}^{\infty} E\left[f_k(\boldsymbol{s}_i) f_k(\boldsymbol{s}_0^b)\right]\right), \; b = 1, \ldots, B.$$

Del mismo modo visto en el capítulo 9, para que sea posible llevar los métodos a la práctica se requiere truncar la descomposición en componentes principales funcionales y seleccionar solo K componentes. La función de covarianza que determina el criterio de diseño, depende únicamente de las distancias entre las observaciones y los sitios de predicción. En este planteamiento se asumen conocidos tanto el modelo como los parámetros. El valor de K que trunca la representación en términos de *componentes principales funcionales empíricos* puede ser flexible y se pueden incluir tantos términos como se requiera, hasta que la varianza acumulada alcance un umbral prefijado, aprovechando que este método no utiliza covarianzas cruzadas entre vectores de puntajes y así el ajuste es más sencillo.

Ejemplo 30 (Muestreo óptimo para la calidad de aire de la ciudad de Bogotá). *Para ilustrar la metodología de la Sección 11.4.1, primero se asume que todas las estaciones pueden moverse. Se eligen* $\boldsymbol{s}_0^1$ *y* $\boldsymbol{s}_0^2$ *como puntos de interés para la predicción. Estos puntos se ubican en zonas con alta densidad de población, por lo que la contaminación tiene un fuerte impacto.*

En cuanto al conjunto $D'_{\boldsymbol{s}}$*, se toma la cuadrícula de muestreo de* 300 *puntos espaciales separados por* 1 *km,* 10 *puntos de oeste a este y* 30 *de sur a norte. La figura 11.1 (izquierda) muestra las ubicaciones de interés* $\boldsymbol{s}_0^1$ *y* $\boldsymbol{s}_0^2$*, así como la cuadrícula de muestreo óptima*

obtenida mediante el criterio global (11.13)*, que es equivalente en este caso particular a* (11.10) *truncado en el segundo vector de puntuación,*

$$\Omega = \Sigma_1 + \Sigma_2.$$

Actualmente, esta red cuenta con una estación móvil de monitoreo de la calidad del aire para mejorar la información obtenida con la red, y dado que esta red no puede moverse con frecuencia, nuestra metodología puede utilizarse para optimizar la ubicación de esta estación móvil.

La figura 11.1 (derecha) muestra la ubicación óptima de muestreo para esta estación móvil manteniendo fija la red actual y utilizando (11.13).

Para el procedimiento de optimización, utilizamos el recocido simulado (Simulated annealing) [14], cuyo estado se define mediante el diseño de muestreo espacial aplicado en cada iteración. La función de energía se define mediante los criterios correspondientes.

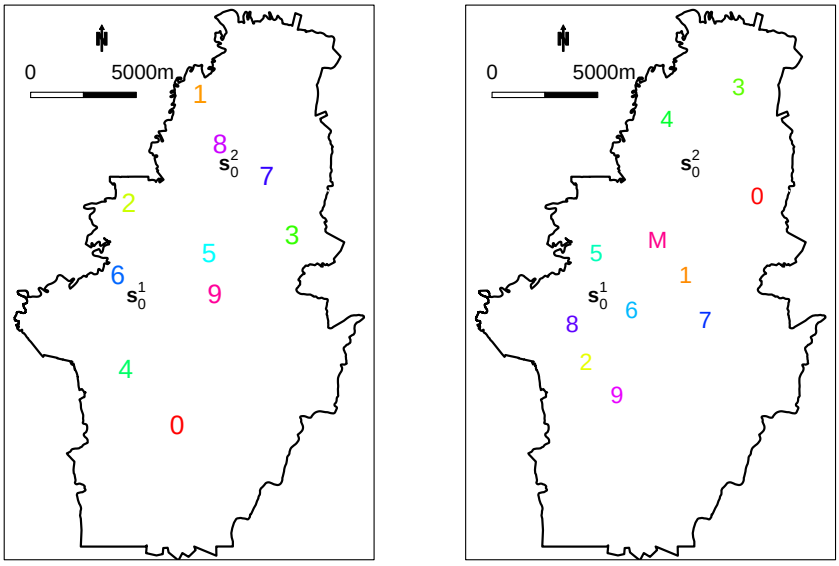


Figura 11.1. Izquierda: Red de muestreo óptima suponiendo que todas las estaciones se pueden mover. Derecha: Ubicación óptima para la estación móvil manteniendo fija la red actual.

11.4.2. Optimización de la predicción espacial de vectores. P variables aleatorias funcionales.

Se requiere diseñar o re-diseñar los p conjuntos

$$S_p = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n_p}\}, \quad p = 1, \ldots, P.$$

o al menos aquellos que puedan ser modificados, para asegurar una predicción funcional espacial óptima $\chi^r_{\boldsymbol{s}}(t)$ en un conjunto de ubicaciones de interés $S_0 = \{\boldsymbol{s}_0^1, \ldots, \boldsymbol{s}_0^B\}$ con base en P campos aleatorios funcionales espacialmente correlacionados. El procedimiento consiste en seleccionar la configuración óptima en el sentido de que garantice la mínima norma del error de predicción del cokriging funcional, ver (10.11). Suponga primero que m_p estaciones, $p = 1, \ldots, P$, pueden ser agregadas para la observación de cada uno de los campos aleatorios $\chi^p_{\boldsymbol{s}}(t)$. Así, la red aumentada para cada caso es

$$\overline{S}_p = S_p \cup S_{m_p} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n_p}, \boldsymbol{s}_{n_p+1}, \ldots, \boldsymbol{s}_{n_p+m_p}\}, \ p = 1, \ldots, P$$

Sea $\bigcup_{p=1}^P \overline{S}_{m_p} = S_{m_1} \cup \ldots \cup S_{m_P} \subset \wp_{m_p}$ el conjunto de las nuevas ubicaciones que deben ser determinadas. Así, entre todos los posibles subconjuntos de nuevos sitios posibles de observación para agregar a la red $\bigcup_{p=1}^P \overline{S}_{m_p}$, se debe elegir aquel que minimiza el cuadrado de la norma del error de predicción del cokriging funcional. Así, de acuerdo con (10.10) y (10.11) el criterio de diseño para predecir la variable aleatoria funcional χ^r en B sitios está dado por

$$\arg\min_{\overline{S}_p \subset \wp_{n_p+m_p}} \sum_{b=1}^{B} E\| \chi^r_{\boldsymbol{s}_0^b}(t) - \check{\chi}^r_{\boldsymbol{s}_0^b}(t)\|^2 \tag{11.11}$$

donde $\sum_{k=1}^{K} \eta^{rk}$ es constante y por lo tanto el criterio (11.11) se reduce a

$$\arg\max_{\overline{S}_p \subset \wp_{n_p+m_p} \subset D'_s} \sum_{b=1}^{B} \sum_{i=1}^{n_p+m_p} \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{p=1}^{P} \lambda_i^{rp} c_{lk}^{rp} E\left(f_k^p(\boldsymbol{s}_i) f_l^r(\boldsymbol{s}_0^b)\right). \tag{11.12}$$

El criterio (11.12) establece el caso general, pero frecuentemente todos los campos aleatorios son medidos en el mismo conjunto $S = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ de ubicaciones. Si este es el caso, la optimización es sobre los posibles conjuntos $\overline{S} \subset \wp_{n+m}$ y existe solo una red aumentada $\overline{S} = S \cup S_m \subset \wp_{n+m}$ para todos los campos aleatorios.

De nuevo, gracias a la estacionariedad de segundo orden, el LMC y el criterio (11.12) dependen solo de la distancia entre observaciones y lugares de predicción. Los números K y L de componentes principales usualmente son pequeños, 1 or 2; así, el cálculo iterativo de la inversa de la matriz de covarianza no representa un costo computacional demasiado alto, a menos que haya demasiados puntos posibles en el espacio involucrados en el proceso de optimización.

También es posible plantear el criterio de diseño para el predictor alternativo que consiste en llevar a cabo cokriging sobre los vectores de puntajes y usar las predicciones obtenidas para la combinación lineal con las funciones propias.

Dado que $\boldsymbol{\xi}(t)$ es conocida, la incertidumbre de la predicción basada en la red aumentada $\overline{S} = S \cup S_m$ solo depende de $\sigma^2_{f(s_0)-f^*(s_0)}$, ver (10.17),

$$\sigma^2_{f_{s_0}-f^*_{s_0}} = Tr\left(\boldsymbol{\Sigma}_0\right) - Tr\left(\sum_{i=1}^{n+m}\left(\boldsymbol{\Sigma}_{(s_0,s_i)}\Gamma_i\right)\right) = \sum_{k=1}^{K}\eta^k - Tr\left(\sum_{i=1}^{n+m}\left(\boldsymbol{\Sigma}_{(s_0,s_i)}\Gamma_i\right)\right)$$

El criterio de diseño para la predicción óptima de B curvas en un conjunto de lugares $S_0 = \{s_0^1, \ldots, s_0^B\}$ de interés usando la varianza del error de predicción, está dado por

$$\arg\max_{S_m \subset D'_s} \sum_{b=1}^{B}\left(Tr\left(\sum_{i=1}^{n+m}\left(\boldsymbol{\Sigma}_{(s_0^b,s_i)}\Gamma_i\right)\right)\right) \tag{11.13}$$

Denotando por $\Delta_0^b = \left(\boldsymbol{\Sigma}_{(s_0^b,s_i)}\right)$ $i = 1, \ldots, n+m$, la solución para el cokriging de los vectores de puntaje es $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}^{-1}\Delta_0^b$. Así, debido a la estacionariedad de segundo orden, el LMC y el criterio (11.13) dependen solo de la distancia entre observaciones y lugares a predecir.

Para la estimación de las estructuras de auto-covarianza y LMC dados en las secciones 9.2 y 10, se usan los estimadores clásicos de los variogramas univariados y cruzados, y los parámetros del modelo pueden ser ajustados por mínimos cuadrados ponderados si no quiere recurrir a supuestos distribucionales, o si se prefiere como en el caso escalar se pueden usar métodos basados en verosimilitud. En la Sección 8.3 se mostró que si el campo aleatorio funcional espacial es conjuntamente Gaussiano, los puntajes son un proceso Gaussiano multivariado.

Finalmente, en los modelos de semivarianza o covarianza se usa el método plug-in para llevar a cabo la optimización de los criterios de muestreo. Esto es, en cada lugar donde los términos $\eta^k\rho^k\left(\|s_i - s_{i'}\|;\Theta\right)$ or $\gamma(\|s_i - s_{i'}\|;\Theta)$ aparecen, se reemplazan con $\eta^k\rho^k\left(\|s_i - s_{i'}\|;\widehat{\Theta}\right)$ y $\gamma(\|s_i - s_{i'}\|;\widehat{\Theta})$. [41] proponen una corrección a la varianza del kriging para incorporar la incertidumbre debida al desconocimiento de $\boldsymbol{\Theta}$. [75] encuentra que esta corrección es importante solo cuando la auto-correlación espacial es débil. Sin embargo, esta corrección se basa en el supuesto distribución Gaussiana y en la estimación usando los métodos de máxima verosimilitud y máxima verosimilitud restringida y la expresión depende de Θ. Así, la mejor opción es usar el método plug-in siempre que la auto-correlación espacial sea moderada o fuerte. Ver [67].

Todos los criterios de diseño mostrados en este capítulo dependen del parámetro Θ. Si este parámetro es desconocido y debe ser estimado, tiene incertidumbre, por lo que el diseño no es óptimo pero aún es ***localmente óptimo***. Así (11.6) se convierte en

$$S_n^* = \arg\max_{S_n \in \Xi_n} \Phi(\widehat{\Theta}, S_n).$$

Además, estos criterios estadísticos se pueden usar para decidir el número de sitios de observación n. En este caso, se determina previamente una varianza de predicción máxima, y la optimización se lleva a cabo para $n = 1$. Luego, manteniendo esta ubicación fija, la segunda ubicación se optimiza, y así sucesivamente hasta encontrar un n que alcance el umbral establecido. Si se usa el supuesto distribucional de normalidad multivariada, el criterio puede ser alguna medida de la matriz de información de Fisher. Ver [54]

El diseño de muestreo espacial óptimo para datos funcionales espaciales es una extensión natural de su contraparte con variables escalares. Los criterios utilizados aquí son útiles al mover o adicionar una ubicación o varias ubicaciones. Una vez que se ha modelado la covarianza entre las curvas, el proceso de optimización para encontrar un diseño óptimo tiene los mismos requerimientos computacionales que en el caso escalar. Su rendimiento depende de la calidad de los estimadores de los parámetros de covarianza y del algoritmo de optimización utilizado.

Ejemplo 31 (Muestreo óptimo para la calidad de aire de la ciudad de México). *Para ilustrar la metodología de los diseños óptimos de muestreo, escogemos dos ubicaciones de interés para predecir el PM10,* $\boldsymbol{s}_0^1$ *y* $\boldsymbol{s}_0^2$*, Ver Figura 11.2 (panel izquierdo). Como conjunto* $D'_{\boldsymbol{s}}$*, se toma una grilla de muestreo de* 375 *ubicaciones espaciales separadas cada* 2*km,* 25 *puntos de oeste a este y* 15 *de sur a norte, restringidos al área con estaciones. La Figura 11.2 (panel izquierdo) muestra las ubicaciones de interés* $\boldsymbol{s}_0^1$ *y* $\boldsymbol{s}_0^2$ *y la ubicación óptima (OL) para agregar una nueva estación manteniendo fija la red actual y usando* (11.12). *Para el procedimiento de optimización, se usa simulated annealing [14] con el estado dado por el criterio de diseño de muestreo espacial aplicado en cada iteración. La función de energía viene dada por el criterio* (11.12). *Con el fin de evaluar la calidad de la predicción espacial basada en el cokriging funcional, utilizamos el método de validación cruzada funcional, se deja una observación por fuera y se predice con el resto [53]. Aunque hay algunos residuos grandes al comienzo de la temporada debido a la variación de los contaminantes en este período, el rendimiento es bueno; la función media residual varía cerca de cero, de -20 a 20 en la mayoría de los casos, ver Figura 11.2 (Panel derecho).*

11.5. Diseño de muestreo óptimo en un tiempo futuro. Modelo dinámico.

En caso de que se puedan cambiar de manera recurrente una o varias estaciones de medición como es el caso de las estaciones móviles ambientales, se puede considerar el comportamiento dinámico del proceso espacio-temporal a través de la evolución de sus parámetros del modelo de covarianza espacial. Ver Figura 11.3.

Es posible que la variación temporal de los parámetros presente algún tipo de autocorrelación. Para que este análisis se pueda llevar a cabo se necesitaría la estimación del parámetro de covarianza espacial en cada punto del tiempo. Esta metodología es sencilla y combina modelos de series de tiempo con análisis geoestadístico.

A continuación se presenta una formulación para este caso.

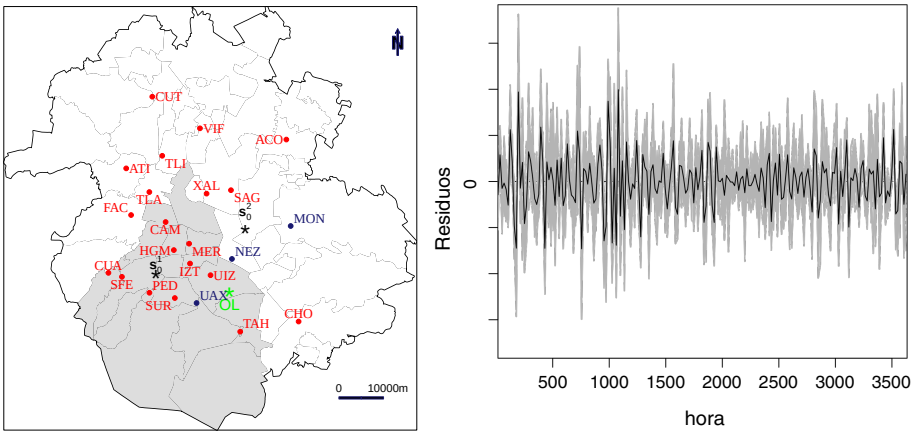


Figura 11.2. *Panel izquierdo*: En verde la ubicación óptima para una ubicación adicional de la red de medio ambiente, REDMA con el fin de predecir el PM10 de manera óptima en las ubicaciones s_0^1 y s_0^2. *Panel derecho*: Residuales de la validación cruzada con su respectiva media

Sea $Y_t(\boldsymbol{s})$ el proceso espacial en el instante t, donde $t \in D_t \subset \mathbb{R}$, $\boldsymbol{s} \in D_s \subset \mathbb{R}^2$, donde el componente espacial varía en $\mathbb{R}^2$ y el temporal en $\mathbb{R}$. Asumiendo media constante, considere el modelo:

$$\boldsymbol{Y}_t = \mathbf{1}\mu_t + \boldsymbol{\delta}_t, \qquad t = 1, ..., T \tag{11.14}$$

donde

$$\boldsymbol{Y}_t = \left(Y_t(\boldsymbol{s}_1), ..., Y_t(\boldsymbol{s}_{n_t})\right)', \ t = 1, .., T$$

es el vector aleatorio en n_t ubicaciones espaciales en cada punto de tiempo t, observado en

$$S_t = \{\boldsymbol{s}_1, ..., \boldsymbol{s}_{n_t}\}.$$

Además se tiene que

$$\mu_t = E\left[Y_t(\boldsymbol{s})\right], \quad \mathbf{1} = (1, ..., 1)'$$

y

$$\boldsymbol{\delta}_t = (\delta_t(\boldsymbol{s}_1), ..., \delta_t(\boldsymbol{s}_{n_t}))'$$

es un proceso estacionario de segundo orden correlacionado tal que

$$E\left[\boldsymbol{\delta}_t\right] = \ 0 \ \text{ y } \ \text{Var}\left[\boldsymbol{\delta}_t\right] = \text{Var}\left[\boldsymbol{Y}_t\right] = \Sigma_t$$

Σ_t es la matriz de covarianza del proceso espacial en el instante t en el conjunto de ubicaciones S_t. Es decir, la entrada i, j-ésima de Σ_t es

$$(\Sigma_t)_{ij} = \text{Cov}\left[Y_t(\boldsymbol{s}_i), Y_t(\boldsymbol{s}_j)\right] \quad i, j = 1, \ldots, n_t \quad t = 1, \ldots, T$$

$\text{Cov}\left[Y_t(\boldsymbol{s}_i), Y_t(\boldsymbol{s}_j)\right]$ está dado por una función de covarianza definida positiva conocida $C(.|\Theta_t)$ con el vector de parámetros Θ_t (ver Sección 6.3). Tenemos las siguientes posibilidades:

a. $\text{Cov}\left(Y_t(\boldsymbol{s}_i), Y_t(\boldsymbol{s}_j)|\Theta\right)$; la función C y el parámetro son constantes, $\forall t \in D_t$.

b. $\text{Cov}\left(Y_t(\boldsymbol{s}_i), Y_t(\boldsymbol{s}_j)|\Theta_t\right)$; la función C es constante, pero el parámetro varía con el tiempo. Entonces, $\Theta_t \; t = 1, \ldots, T$, proviene de la serie temporal $\{\Theta_t, \quad t \in D_t\}$.

El caso $C_t\left(\boldsymbol{s}_i - \boldsymbol{s}_j|\Theta_t\right)$, en el que, además del vector de parámetros, la función de covarianza espacial también varía en cada t, no se considera aquí. Este supuesto implica un proceso muy inestable y no es realista.

Para el pronóstico se pueden utilizar modelos ARIMA o cualquier otra alternativa que permita pronosticar estos parámetros para reemplazarlos en el criterio de diseño en el tiempo futuro en el cual se requiere mover las estaciones. Estas series temporales son estimaciones de parámetros de covarianza y, por lo tanto, se utilizan modelos para datos con error de medición. El enfoque más general consiste en utilizar modelos de espacio de estados para extraer la señal y generar predicciones. Se asume que el término de error es un proceso aditivo de ruido blanco, y el sesgo de los estimadores $(\hat{\theta}_t)_k$, $k = 1, \ldots, K$ no es significativo. Por ejemplo, el modelo de nivel local ([30]) para algún parámetro de covarianza $(\theta_t)_k$ para tiempo discreto y continuo toma la forma, respectivamente:

Tiempo discreto $t = 1, \ldots, T$

$$\begin{aligned}(\hat{\theta}_t)_k &= (\theta_t)_k + (\epsilon_t)_k \qquad (\epsilon_t)_k \sim N(0, \sigma^2_{\epsilon_k}) \\ (\theta_{t+1})_k &= (\theta_t)_k + (\eta_t)_k \qquad (\eta_t)_k \sim N(0, \sigma^2_{\eta_k})\end{aligned}$$

$(\epsilon_t)_k$ y $(\eta_t)_k$ son mutuamente independientes e independientes de $(\theta_t)_k$.

Tiempo continuo $0 \leq t \leq T$

$$\begin{aligned}&(\hat{\theta}(t))_k = (\theta(t))_k + (\epsilon(t))_k, \qquad t = t_1, \ldots, t_T, \qquad (\epsilon(t_i))_k \sim N(0, \sigma^2_{\epsilon_k}(t_i)), \\ &(\theta(t))_k = (\theta(0))_k + \sigma_{\eta_k} w(t)\end{aligned}$$

$w(t)$ es el proceso de movimiento browniano, y $\sigma_{\eta_k} > 0$ es un parámetro de escala. $\sigma^2_{\epsilon_k}(t)$ es una función no estocástica que puede depender de parámetros desconocidos y está significativamente acotada desde cero. Para la estimación de parámetros utilizamos los modelos reducidos

$$\begin{aligned}&(\hat{\theta}(t_i))_k = (\theta(t_i))_k + (\epsilon(t_i))_k, \qquad t = t_1, \ldots, t_T, \qquad (\epsilon(t_i))_k \sim N(0, \sigma^2_{\epsilon_k}(t_i)), \\ &(\theta(t_i + 1))_k = (\theta(t_i))_k + (\eta_{t_i})_k\end{aligned}$$

Donde se supone que $(\epsilon(t))_k$ son independientes y $(\eta_{t_i})_k = \sigma_{\eta_k}\,[w(t_i+1) - w(t_i)]$. Este es un modelo de nivel discreto que permite la variación uniforme de $(\epsilon(t))_k$. Esta es la única diferencia con el modelo de nivel local discreto.

Los pasos a seguir para usar esta metodología se resumen a continuación:

1. Determinar el modelo de covarianza espacial C.
2. Estimar el parámetro de covarianza espacial Θ_t, en cada momento del tiempo t, $t = 1, \ldots, T$ con base en $\boldsymbol{\delta}_t$.
3. Modelar la serie temporal formada por las estimaciones del parámetro de covarianza espacial $\widehat{\Theta}_1, \ldots, \widehat{\Theta}_T$.
4. Encontrar el pronóstico $\widetilde{\Theta}_{T+\ell}$ del parámetro de covarianza espacial para el tiempo futuro $T + \ell$, cuando las ubicaciones de observación se pueden cambiar, $\ell \in \mathbb{N}$.
5. Establecer el nuevo conjunto $S_{T+\ell}$ de covarianza espacial ubicaciones de observación en el punto temporal $T+\ell$ optimizando el criterio especificado.

Ejemplo 32 (Optimización de la estimación de la media espacial). *Si el objetivo es la estimación de la media espacial en un instante de tiempo futuro $T + \ell$, el estimador de mínimos cuadrados generalizados (GLS) que tiene en cuenta la correlación espacial existente en este instante temporal es $\hat{\mu}_{T+\ell}$,*

$$\hat{\mu}_{T+\ell} = \frac{\mathbf{1}'\widehat{\Sigma}^{-1}_{T+\ell}\boldsymbol{\delta}_{T+\ell}}{\mathbf{1}'\widehat{\Sigma}^{-1}_{T+\ell}\mathbf{1}}. \tag{11.15}$$

y el estimador de la varianza de $\hat{\mu}_{T+\ell}$ y que por lo tanto es el criterio con base en el cual se puede encontrar el diseño espacial óptimo en el tiempo $T + \ell$ está dado por la varianza pronosticada $\widetilde{Var}(\hat{\mu}_{T+\ell})$, que viene dada por

$$\widetilde{Var}(\hat{\mu}_{T+\ell}) = \left(\mathbf{1}'\widetilde{\Sigma}^{-1}_{T+\ell}\mathbf{1}\right)^{-1} = \left(\sum_{i=1}^{n_{T+\ell}}\sum_{j=1}^{n_{T+\ell}} \mathscr{C}_{\boldsymbol{s}_i\boldsymbol{s}_j}\right)^{-1} \tag{11.16}$$

con $\widetilde{\Sigma}_{T+\ell} = \left(C(\boldsymbol{s}_i - \boldsymbol{s}_j|\widetilde{\Theta}_{T+\ell})\right)$, $i, j = 1, \ldots, n_{T+\ell}$. El criterio de diseño resultante es la suma de las entradas de la inversa de la matriz de covarianza espacial en el instante de tiempo de interés, $\widetilde{\Sigma}^{-1}_{T+\ell}$. Entonces, la configuración espacial óptima $S_{T+\ell}$ para estimar $\mu_{T+\ell}$ está dada por el diseño muestral $n_{T+\ell}$ que minimiza

$$\widetilde{Var}(\hat{\mu}_{T+\ell}) \tag{11.17}$$

Este enfoque puede utilizarse cuando no es necesario o posible mover todas las ubicaciones de muestreo, sino solo algunas. Es ideal para determinar la ubicación de estaciones móviles.

Tenga en cuenta que una ubicación de observación $\boldsymbol{s}_i$ en el tiempo T puede variar a $\boldsymbol{s}_i$ en el siguiente tiempo $T + \ell$ y posteriormente. Además, los tamaños de muestreo espacial pueden ser diferentes para cada instante temporal, $\boldsymbol{s}_i, \boldsymbol{s}_i \in \mathbb{R}^d$.

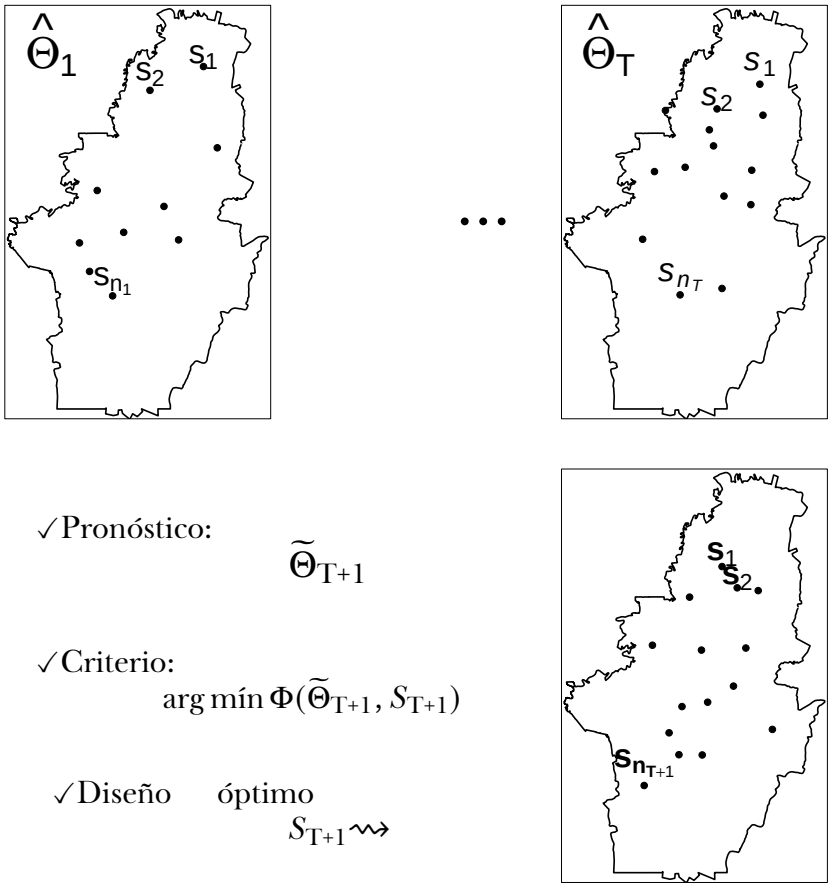


Figura 11.3. Diseño óptimo en el tiempo futuro $T + 1$. $\widetilde{\Theta}_{T+1}$ es el vector de predicciones de los parámetros de la función de covarianza en el tiempo $T+1$ basado en la serie temporal de estimaciones de estos parámetros en los puntos temporales 1, ..., T. Nótese que los tamaños de muestra espacial n_t pueden ser diferentes para cada instante temporal. Además, una ubicación de observación $\boldsymbol{s}_i$ en el tiempo t puede variar a s_i en el siguiente tiempo $t+1$ y así sucesivamente. $\mathbf{s}_i$, $\boldsymbol{s}_i$, $\mathbf{s}_i \in \mathbb{R}^d$. En general, $d = 2$

En conclusión, teniendo claridad sobre las especificidades de caso, se puede llevar a cabo la adaptación de la metodología. Si la variación de las series de tiempo es a una frecuencia muy corta o las series son muy largas puede ser mejor recurrir a los datos funcionales como se observa en la Sección 11.2.

Elemento	Escalar espacial	Escalar espacio-temporal	Funcional
Campo aleatorio	$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D_s \subset \mathbb{R}^d\}$	$\{Y(\boldsymbol{s}, t) : (\boldsymbol{s}, t) \in D_s \times D_t \subset \mathbb{R}^d \times \mathbb{R}\}$	$\{\boldsymbol{\mathcal{Y}}_s(t) : \boldsymbol{s} \in D_s \subset \mathbb{R}^d\}$
Dominio de la variable	$Y \in \mathbb{R}$	$Y \in \mathbb{R}$	$\boldsymbol{\mathcal{Y}}(t) \in \mathbb{R}^\infty$
Vector espacial aleatorio	$(Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))$	$(Y(\boldsymbol{s}_i, 1), \ldots, Y(\boldsymbol{s}_i, t_i)) \quad i = 1, \ldots, n$	$\left(\boldsymbol{\mathcal{Y}}_{s_1}(t), \ldots, \boldsymbol{\mathcal{Y}}_{s_n}(t)\right)$
Realización del proceso	$y(\boldsymbol{s})$	$y(\boldsymbol{s}, t)$	$\mathcal{Y}_s(t)$
Datos observados	$(\mathcal{Y}(\boldsymbol{s}_1), \ldots, \mathcal{Y}(\boldsymbol{s}_n))$	$(y(\boldsymbol{s}_i, 1), \ldots, y(\boldsymbol{s}_i, t_i)) \quad i = 1, \ldots, n$	$(\mathcal{Y}(\boldsymbol{s}_1), \ldots, \mathcal{Y}(\boldsymbol{s}_n))$
Modelo geoestadístico	$Y(\boldsymbol{s}) = \mu(\boldsymbol{s}) + Z(\boldsymbol{s})$	$Y(\boldsymbol{s}, t) = \mu(\boldsymbol{s}, t) + Z(\boldsymbol{s}, t)$	$\boldsymbol{\mathcal{Y}}_s(t) = \mu(t) + \chi_s(t)$
Media del proceso	$E[Y(\boldsymbol{s})] = \mu$	$E(Y(\boldsymbol{s}, t)) = \mu$	$E\left[\boldsymbol{\mathcal{Y}}_s(t)\right] = \mu(t)$
Proceso centrado	$Z(\boldsymbol{s})$	$Z(\boldsymbol{s}, t)$	$\chi_s(t)$
Autocovarianza (par)	$E\left[Z(\boldsymbol{s}_i), Z(\boldsymbol{s}_j)\right]$	$E\left[Z(\boldsymbol{s}_i, t), Z(\boldsymbol{s}_j, t')\right]; \; t, t' = 1, \ldots, t_i, \; i, j = 1, \ldots, n$	$E\langle \chi_{s_i}, \chi_{s_j} \rangle$
Predictor kriging (KP)	$Z^*(\boldsymbol{s}_0) = \sum_{i=1}^n \lambda_i z(\boldsymbol{s}_i)$	$Z^*(\boldsymbol{s}_0, t_0) = \sum_{i=1}^n \lambda_i z(\boldsymbol{s}_i, t) \quad t = 1, \ldots, t_i$	$\check{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi_{s_i}(t)$
Varianza del KP	$E[Z(\boldsymbol{s}_0) - Z^*(\boldsymbol{s}_0)]^2$	$E[Z(\boldsymbol{s}_0, t_0) - Z^*(\boldsymbol{s}_0, t_0)]^2$	$E\left\|\boldsymbol{\chi}_{s_0}(t) - \check{\chi}_{s_0}(t)\right\|^2$

Tabla 11.1. Paralelo entre los elementos fundamentales, de la geoestadística escalar espacial, la geoestadística escalar espacio-temporal y la geoestadística funcional. El problema es similar en el dominio espacial, pero la diferencia radica en el dominio de ocurrencia de la variable de interés. En el caso escalar $Z \in \mathbb{R}$, $Y \in \mathbb{R}$, mientras que en el caso funcional $\mathcal{Y} \in \mathbb{R}^\infty$, $\chi \in \mathbb{R}^\infty$.

.

La Tabla 11.1 hace un paralelo entre los tres grandes métodos presentados para enfatizar la similitud en cuanto a objetivos y fundamentos teóricos, así como la diferencia entre los objetos aleatorios, en cada caso.

1. El elemento aleatorio que varía espacialmente es el principal cambio entre la geoestadística escalar y la funcional. En el primer caso, es una variable de valor real la que varía a través del espacio, mientras que en el segundo es una curva o función.
2. En muchos fenómenos no es posible tener mediciones de los datos espacio-temporales, ni en todos los tiempos ni en todas las ubicaciones debido a su variación continua en estas dimensiones.
3. Es usual que la longitud de las series de tiempo sea diferente en cada ubicación, aun cuando son observadas en el mismo intervalo de tiempo.
4. En el contexto geoestadístico no se requiere que los datos sean observados a una frecuencia temporal o espacial regular.
5. En el modelo geoestadístico se asume media constante. Sin embargo, lo mas común es que exista algún tipo de tendencia, ya sea espacial o temporal. Esta tendencia se suele estimar usando un modelo polinómico paramétrico, a partir del cual se obtiene un proceso residual de media cero con la misma estructura de covarianza que el proceso original. La ventaja del modelo de media paramétrico es que se puede estimar la media en cualquier ubicación de interés.

El análisis geoestadístico de datos funcionales espaciales es un enfoque alternativo al modelamiento espacio-temporal cuando las curvas son series temporales que varían espacialmente. Específicamente, la geoestadística funcional permite llevar a cabo la predicción espacial óptima de toda la curva de interés en lugares no muestreados. La limitación de la dimensión para el MLC es el problema más crítico del método de cokriging. Sin embargo, esta dificultad se resuelve por el hecho de que la representación EFPC generalmente no necesita demasiadas funciones propias, incluso con una o dos podría ser suficiente en la mayoría de los casos, lo que hace posible utilizar este tipo de modelo de covarianza.

El proceso espacio-temporal y el proceso funcional se refieren al mismo fenómeno de la naturaleza. Nótese que si tanto el conjunto índice espacial como el temporal son continuos, estamos en el contexto geoestadístico y es posible aplicar kriging espacio-temporal a la variable escalar o geoestadística funcional espacial a la función del tiempo que varía espacialmente. El enfoque dependerá de la cantidad de observaciones existentes, pero sin duda el enfoque más acorde a estas clases de fenómenos es el funcional espacial. Si hay un número moderado de observaciones tanto en el espacio como en el tiempo, pero suficiente que permita una estimación adecuada de la función de covarianza espacio temporal, es posible usar directamente la variable escalar. Sin embargo, si en alguna de las dimensiones existe una gran cantidad de mediciones, los métodos geoestadísticos para variable escalar no se pueden aplicar de manera eficiente, debido a las dimensiones que alcanza la matriz de covarianza del vector aleatorio. En este caso, es mucho más

eficiente, construir el dato funcional usando métodos de suavizamiento, en la dimensión que más densamente se encuentra observada. Así, el nuevo elemento aleatorio es una función del tiempo que varía espacialmente o una función de la ubicación espacial que varía temporalmente. Si el proceso está densamente muestreado se pueden obtener superficies o curvas suavizadas, pero esto impide profundizar en las características y la comprensión de los fenómenos, así como estimar medidas de incertidumbre.

11.6. Ejercicios

1. Seleccione 10 puntos dentro de una zona arbitraria y simule un proceso espacial univariado en esos 10 lugares. Suponga que estas son las realizaciones de la variable. Determine los lugares donde se deben instalar dos nuevos puntos de medición de tal manera que la predicción cumpla algún criterio de optimalidad. Detalle el procedimiento. Diseñe algoritmo y código.
2. Repita el ejercicio anterior para el caso bivariado. Plantee un criterio que recorra toda la región de interés y otro criterio con optimización discreta sobre un conjunto definido de ubicaciones posibles.
3. Use el paquete SpatFD de R cran. Utilice bases ortonormales diferentes, por ejemplo Fourier y Legendre para la reconstrucción de los datos funcionales. Con estas funciones propias, repita los ejercicios anteriores de simulación ahora para el caso funcional tanto univariado como bivariado.

Índice

Referencias

[1] Luc Anselin, *Spatial econometrics: methods and models*, vol. 4, Springer Science & Business Media, 2013.

[2] C. Berg, E. Porcu, and J. Mateu, *The Dagum family of isotropic correlation functions.*, Bernoulli **14** (2008), no. 4, 1134 – 1149.

[3] Moreno Bevilacqua, Jorge Mateu, Emilio Porcu, Hui Zhang, and Armand Zini, *Weighted composite likelihood-based tests for space-time separability of covariance functions*, Statistics and Computing **20** (2010), no. 3, 283–293.

[4] Moreno Bevilacqua, Víctor Morales-Oñate, and Christian Caamaño-Carrillo, *Geomodels: Procedures for gaussian and non gaussian geostatistical (large) data analysis*, 2024, R package version 2.0.1.

[5] Lindsay BG., *Composite likelihood methods. statistical inference from stochastic processes*, Contemp. Math., 80, Amer. Math. Soc. (1988), 221–239.

[6] M. Bohorquez, *Diferenciabilidad de funciones de covarianza espacio temporal no separables*, Master's thesis, Universidad Nacional de Colombia, 2010.

[7] M. Bohorquez, R. Giraldo, and J. Mateu, *Optimal sampling for spatial prediction of functional data*, Statistical Methods & Applications **25** (2016), no. 1, 39–54.

[8] ———, *Multivariate functional random fields: prediction and optimal sampling*, Stochastic Environmental Research and Risk Assessment **31** (2017), no. 1, 53–70.

[9] M. Bohorquez, J. Mateu, and L. Diaz, *A note on smoothness measures for space–time surfaces*, Stochastic Environmental Research and Risk Assessment **28** (2014), no. 4, 1011–1022.

[10] M Bohorquez, D Sandoval, A Villamil, S Sanchez, R Guevara, R Giraldo, and J Mateu, *Spatfd: Functional geostatistics: Univariate and multivariate functional spatial prediction*, 2024, R package version 0.0.1.

[11] EG Bongiorno, Ernesto Salinelli, A Goia, and P Vieu, *Contributions in infinite-dimensional statistics and related topics*, Societa Editrice Esculapio, 2014.

[12] D. Bosq, *Linear processes in function spaces: Theory and applications*, vol. 149, Springer, 2000.

[13] A.E. Brockwell and P.J. Brockwell, *A class of non-embeddable arma processes*, Journal of Time Series Analysis **20** (1999), no. 5, 483–486.

[14] S.P. Brooks and B.J.T. Morgan, *Optimization using simulated annealing*, Journal of the Royal Statistical Society: Series D (The Statistician) **44** (1995), no. 2, 241–257.

[15] J Buescu and AC Paixao, *Real and complex variable positive definite functions*, Sao Paulo Journal of Mathematical Sciences **6** (2012), no. 2, 155–169.

[16] Jana Burkotová, Ivana Pavluu, Hiba Nassar, Jitka Machalová, and Karel Hron, *Efficient spline orthogonal basis for representation of density functions*, Journal of Applied Statistics (2025), 1–37.

[17] George Casella and Roger Berger, *Statistical inference*, Chapman and Hall/CRC, 2024.

[18] Liliana Blanco Castañeda, *Probabilidad*, Univ. Nacional de Colombia, 2004.

[19] J.P. Chiles and P. Delfiner, *Geostatistics: modeling spatial uncertainty*, Wiley Series in probability and statistics, 1999.

[20] John B Conway, *A course in functional analysis*, vol. 96, Springer, 2019.

[21] N. Cressie, *Fitting variogram models by weighted least squares*, Mathematical Geology **17** (1985), no. 5.

[22] ———, *Fitting variogram models by weighted least squares*, Mathematical Geology **17** (1985), no. 5.

[23] ———, *Statistics for spatial data*, Wiley Series in Probability and Mathematical Statistics, 1993.

[24] N Cressie, *Statistics for spatial data: Revised edition*, John Wiley & Sons, 1993.

[25] N. Cressie and H. C. Huang, *Classes of nonseparable, spatio-temporal stationary covariance functions*, JASA, 94 (1999).

[26] N. Cressie and H.C. Huang, *Classes of nonseparable, spatio-temporal stationary covariance functions*, Journal of the American Statistical Association (1999), 1330–1340.

[27] FC Curriero and S Lele, *A composite likelihood approach to semivariogram estimation*, Journal of Agricultural, biological, and Environmental statistics (1999), 9–28.

[28] S. De Iaco, DE Myers, and D. Posa, *Nonseparable space-time covariance models: some parametric families*, Mathematical Geology **34** (2002), no. 1, 23–42.

[29] P J Diggle, Jonathan A Tawn, and R A Moyeed, *Model-based geostatistics*, Journal of the Royal Statistical Society: Series C (Applied Statistics) **47** (1998), no. 3, 299–350.

[30] James Durbin and Siem Jan Koopman, *Time series analysis by state space methods*, Oxford university press, 2012.

[31] F. Ferraty and P. Vieu, *Nonparametric functional data analysis: Theory and practice*, Springer Verlag, 2006.

[32] X Gao and P X-K Song, *Composite likelihood bayesian information criteria for model selection in high-dimensional data*, Journal of the American Statistical Association **105** (2010), no. 492, 1531–1540.

[33] R. Giraldo, *Cokriging based on curves, prediction and estimation of the prediction variance*, InterStat **2** (2014), 1–30.

[34] T. Gneiting, *Correlation functions for atmospheric data analysis*, Quarterly Journal of the Royal Meteorological Society **125** (1999), no. 559, 2449–2464.

[35] _____, *Nonseparable, stationary covariance functions for space-time data*, JASA (2002).

[36] _____, *Nonseparable, stationary covariance functions for space-time data*, Journal of the American Statistical Association **97** (2002), no. 458, 590–600.

[37] Pierre Goovaerts, *Geostatistics for natural resources evaluation*, Oxford University Press on Demand, 1997.

[38] M. Goulard and M. Voltz, *Geostatistical interpolation of curves: A case study in soil science*, In A. Soares (Ed.) Geostatistics Tróia'92 **2** (1993), no. 2, 805–816.

[39] D.S. Greenbaum, J.D. Bachmann, D. Krewski, J.M. Samet, Ro. White, and R.E. Wyzga, *Particulate air pollution standards and morbidity and mortality: case study*, American journal of epidemiology **154** (2001), no. 12, 78–90.

[40] Stanley Grossman, I Stanley, et al., *Álgebra lineal*, Biblioteca Hernán Malo González, 2019.

[41] D.A. Harville and D.R. Jeske, *Mean squared error of estimation or prediction under a general linear model*, Journal of the American Statistical Association **87** (1992), no. 419, 724–731 (English).

[42] PJ Heagerty and SR Lele, *A composite likelihood approach to binary spatial data*, Journal of the American Statistical Association **93** (1998), no. 443, 1099–1111.

[43] D.C. Hoaglin, F. Mosteller, and J.W. Tukey, *Understanding robust and exploratory data analysis*, vol. 3, Wiley New York, 1983.

[44] L. Horvath and P. Kokoszka, *Inference for functional data with applications*, Springer, 2012.

[45] Danie G Krige and DG Krige, *Lognormal-de wijsian geostatistics for ore evaluation*, South African Institute of mining and metallurgy Johannesburg, 1981.

[46] P.C. Kyriakidis and A. G. Journel, *Geostatitical space time models: A review*, Mathematical Geology (1999).

[47] Christian Lantuéjoul, *Geostatistical simulation: models and algorithms*, no. 1139, Springer Science & Business Media, 2001.

[48] B. Lindsay, *Composite likelihood methods*, Contemporary Mathematics **80** (1988).

[49] Chunsheng Ma, *Spatio-temporal stationary covariance models*, Journal of Multivariate Analysis **86** (2003), no. 1, 97–107.

[50] J. R Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*, John Wiley & Sons, 1988.

[51] J.E. Marsden and A.J. Tromba, *Cálculo vectorial*, Addison Wesley Longman, 1998.

[52] Georges Matheron, *The intrinsic random functions and their applications*, Advances in applied probability **5** (1973), no. 3, 439–468.

[53] J.M. Montero, G. Fernandez-Aviles, and J Mateu, *Spatial and spatio-temporal geostatistical modeling and kriging*, Wiley, 2015.

[54] W.G. Müller, *Collecting spatial data: Optimum design of experiments for random fields*, Springer Verlag, 2007.

[55] D.E. Myers and A. Journel, *Variograms with zonal anisotropies and noninvertible kriging systems*, Mathematical Geology **22** (1990), no. 7, 779–785.

[56] D. Nerini, P. Monestiez, and C. Manté, *Cokriging for spatial functional data*, Journal of Multivariate Analysis **101** (2010), no. 2, 409–418.

[57] Peter V O'neil et al., *Matemáticas avanzadas para ingeniería*, México Cengage Learning, 2008.

[58] A Paldy, J Bobvos, M Lustigova, H Moshammer, E M Niciu, P Otorepec, V Puklova, K Szafraniec, T Zagargale, M Neuberger, et al., *Health impact assessment of pm10 on mortality and morbidity in children in central-eastern european cities*, Epidemiology **17** (2006), no. 6, S131.

[59] Luis Jose Parra Gómez, *Spatial geotechnical modeling of a lacustrine deposit using functional geostatistical analysis of cptu tests*, Ph.D. thesis, 2019.

[60] HA Peters, *Neighbour-regulated mortality: the influence of positive and negative density dependence on tree populations in species-rich tropical forests*, Ecology letters **6** (2003), no. 8, 757–765.

[61] E. Porcu, J. Mateu, A. Zini, and R. Pini, *Modelling spatio-temporal data: A new variogram and covariance structure proposal*, Statistics and Probability Letters **77** (2007), no. 1, 83–89.

[62] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2025.

[63] J.O. Ramsay and B.W. Silverman, *Functional data analysis*, Springer, New York, 2005.

[64] M Reed and B Simon, *Methods of modern mathematical physics i: Functional analysis*, Academic Press, Inc. San Diego., 1980.

[65] S. Rouhani and D.E. Myers, *Problems in space-time kriging of geohydrological data*, Mathematical Geology **22** (1990), no. 5, 611–623.

[66] J.A. Royle and L.M. Berliner, *A hierarchical approach to multivariate spatial modeling and prediction*, Journal of Agricultural, Biological, and Environmental Statistics (1999), 29–56.

[67] O Schabenberger and CA Gotway, *Statistical methods for spatial data analysis*, Chapman and Hall/CRC, 2005.

[68] Martin Schlather and Olga Moreva, *A parametric model bridging between bounded and unbounded variograms*, Stat **6** (2017), no. 1, 47–52.

[69] SEDEMA. Secretaría del Medio Ambiente México, 2015.

[70] M.L. Stein, *Space-time covariance functions*, JASA (2005).

[71] C Varin, *On composite marginal likelihoods*, AStA Advances in Statistical Analysis **92** (2008), no. 1, 1.

[72] J.M. Ver Hoef and N. Cressie, *Multivariable spatial prediction*, Mathematical Geology **25** (1993), no. 2, 219–240.

[73] H. Wackernagel, *Multivariate geostatistics: An introduction with applications*, Springer Verlag, 1998.

[74] AW Warrick and DE Myers, *Optimization of sampling locations for variogram calculations*, Water Resources Research **23** (1987), no. 3, 496–500.

[75] Z. Zhu and M.L. Stein, *Spatial sampling design for prediction with estimated parameters*, Journal of Agricultural, Biological, and Environmental Statistics **11** (2006), no. 1, 24–44.

[76] DL Zimmerman and N Cressie, *Mean squared prediction error in the spatial linear model with estimated covariance parameters*, Annals of the Institute of Statistical Mathematics **44** (1992), no. 1, 27–43.

Colección
Textos

Este libro presenta de forma clara, aplicada y didáctica los elementos fundamentales de la geoestadística desde sus herramientas más básicas hasta su estado del arte actual, el cual permite manejar grandes volúmenes de datos, así como datos con estructuras más complejas. Se procura profundizar en todos los conceptos a través de diversas ilustraciones y de ejemplos con datos reales en una gran variedad de áreas del conocimiento. Se presentan los predictores y estimadores existentes, así como nuevas propuestas para el análisis eficiente de grandes volúmenes de datos espacio temporales. Además, se discuten aspectos prácticos importantes a tener en cuenta en la aplicación de todas las herramientas. Los métodos presentados en el libro cuentan con todo el código, software y ejemplos necesarios para su reproducción.

Estadística