

Estimación Paramétrica y No Paramétrica de la
Tendencia en Datos con Dependencia Espacial. Un
Estudio de Simulación.

Miguel Ángel Reyes Cortés

30 de junio de 2010

Índice general

1. Conceptos de Geoestadística	9
1.1. Introducción	9
1.2. Procesos Espaciales	10
1.2.1. Procesos Espaciales Estacionarios	11
1.3. Función Variograma y Covarianza	12
1.3.1. Variograma y Covarianza Teóricos	13
1.3.2. Anisotropía	15
1.3.3. Estimación del Variograma	15
1.3.4. Modelos de Variograma	17
1.3.5. Métodos de Ajuste del Variograma Experimental, Criterios de Selección y Validación	19
2. Estimación de la tendencia	23
2.1. El Modelo General	23
2.2. Estimador paramétrico	24
2.3. Estimador No Paramétrico	25
2.3.1. Estimador local lineal	25
2.3.2. Estimador local lineal en datos espaciales	28
2.3.3. Selección de la matriz ventana H	28
2.4. Predicción Espacial Lineal Óptima	29
3. Descripción del Estudio de Simulación	35
3.1. Comparación de las Estimaciones de la Tendencia	35
3.2. Simulación de una muestra	36
3.3. Estimadores	37
3.4. Definición de los Parámetros de la Simulación	39
3.5. Plan de Simulación	41
4. Resultados	43
4.1. Especificación Adecuada del Modelo	43
4.1.1. Variación de la proporción del nugget	43
4.1.2. Variación del rango	47
4.2. Especificación Inadecuada del Modelo	48

4.2.1. Variación de la proporción de nugget	48
4.2.2. Variación del rango	51
4.3. Estimación No Paramétrica	52
5. Conclusiones	55

Resumen

El presente estudio de simulación tiene por objetivo estimar la tendencia de un conjunto de datos con dependencia espacial de manera paramétrica y no paramétrica. Para ello se efectúa un plan de simulación consistente en diferentes escenarios en los que varía la dependencia espacial por medio del rango y del nugget, parámetros ambos de la función de covarianza. Las estimaciones de la tendencia por los diferentes métodos y en los diferentes escenarios se compara mediante el error cuadrático medio integrado (MISE), lo que permite determinar no sólo su exactitud y su precisión, sino también comparar el efecto de variar las condiciones de dependencia espacial. Finalmente, la evidencia sugiere que la estimación no paramétrica puede ser una buena opción en casos en los que nada se sabe sobre el modelo de tendencia de los datos observados.

Introducción

Una problema importante en la mayoría de las investigaciones científicas es la formulación de modelos estadísticos que representen adecuadamente el fenómeno aleatorio en estudio. En muchos casos, la utilización de modelos sencillos que supongan que las observaciones de dicho fenómeno han sido tomadas bajo condiciones idénticas e independientes unas de otras no es adecuado. La falta de homogeneidad en los datos suele ser modelada a través de la suposición de una tendencia no constante. Por otra parte, si existe evidencia para pensar que los datos cercanos en el espacio o en el tiempo son más semejantes que aquellos que están alejados, es decir, que los datos puedan presentar dependencia espacial y/o temporal, resulta más conveniente emplear modelos que exploten adecuadamente dicha componente espacial o espacio-temporal. Si bien estos modelos han recibido mucha atención por la comunidad estadística en los últimos años debido sobre todo a su interés práctico, resulta llamativo el escaso número de aportaciones donde el principal objetivo sea la realización de inferencia estadística en dichos modelos cuando se supone la existencia de una función media o tendencia. En este trabajo nos centraremos en la estimación de la tendencia de modelos con dependencia espacial y/o espacio-temporal utilizando técnicas no paramétricas de tipo polinómico local. Estas técnicas no paramétricas no presuponen un modelo funcional paramétrico para dicha función tendencia. Por otra parte, en la práctica, para poder realizar estas estimaciones será necesario estimar también de alguna manera la dependencia entre las observaciones. Para ello, será necesario formular modelos adecuados. Con el objeto de realizar un estudio lo más completo posible, utilizaremos varios modelos de dependencia paramétricos y los compararemos con un modelo no paramétricos. La utilización de estas técnicas no paramétricas, tanto para estimar la tendencia como para estimar la dependencia espacial o espacio-temporal, proporcionará una mayor flexibilidad en la estimación y esperamos unos mejores resultados en la práctica. Otros problemas inherentes a los métodos de estimación no paramétricas como la selección del parámetro de suavizado serán también tratados. Por último, se aplicarán estas y otras técnicas a conjuntos de datos mediante estudios de simulación, lo que conllevará que será necesario la elaboración de software adecuado para poder llevar a cabo este cometido.

Capítulo 1

Conceptos de Geoestadística

1.1. Introducción

De manera general, *datos espaciales* son aquellos que tienen asociada una localización en el espacio. Consecuentemente, y ampliando la idea anterior, un dato espacio-temporal es simplemente la observación de una variable en cierta localización espacial considerando al tiempo como coordenada adicional. Este tipo de datos es muy común en las ciencias de la tierra (geología, meteorología, oceanografía, paleontología...), pero también en otras áreas del conocimiento tales como epidemiología, geografía, ecología, astronomía, ciencias agrícolas, etc.

Las variables espacio-temporales de interés pueden ser *continuas* o *discretas*, y se presupone que existe correlación entre dos variables que tengan asociada distinta referencia geográfica. La idea fundamental con este tipo de datos es que las observaciones más cercanas son más parecidas entre sí, y conforme éstas se distancian, la correlación entre las variables tiende a disminuir, anulándose en algún momento.

Se le llama *estadística espacial* al conjunto de modelos y métodos que tienen por objetivo el análisis de datos espacialmente referenciados. Dentro de la estadística espacial misma, se le llama *geoestadística* al conjunto de métodos y modelos para el estudio de datos espaciales con las siguientes características: 1) Los valores de la variable se observan en un conjunto discreto de localizaciones muestrales dentro de alguna región espacial. 2) Cada observación es una medida de (o está relacionada directamente con) un fenómeno espacial continuo en las correspondientes localizaciones muestrales [1].

La geoestadística puede considerarse una rama de las matemáticas aplicadas cuyos orígenes se remontan a principios de la década de 1950. Las primeras aportaciones las hizo el ingeniero de minas D.G. Krige y el estadístico H.S. Sichel, en Sudáfrica, en un intento por mejorar los métodos para calcular las reservas en las minas de oro. Posteriormente, sus técnicas

llamaron la atención de los ingenieros de minas franceses, en especial de Georges Matheron, quien formalizó los conceptos de Krige en el marco de la *Teoría de las Variables Regionalizadas* [2].

1.2. Procesos Espaciales

Los datos geoestadísticos pueden tener dependencia espacial, temporal, o ambas. En muchos casos, además, este tipo de datos suelen ser multivariantes.

Sea $D \subset \mathbb{R}^d$ un dominio espacial d -dimensional. Puede pensarse para fines prácticos en $d = 3$. Para un conjunto de localizaciones $\mathbf{x}_\alpha \in D$, a un tiempo t_α , con $\alpha = 1, 2, 3, \dots, n$, se pueden realizar mediciones de varios atributos físicos z_α^i , donde $i = 1, 2, 3, \dots, N$. No obstante, considérese por un momento que la componente temporal no se tiene en cuenta y que sólo se mide un atributo físico en cada localización espacial (que corresponde al caso $N = 1$). De esta manera se tiene una observación z en cada localización \mathbf{x}_α , lo que se representa por medio de $z(\mathbf{x}_\alpha)$. Si z está definida en cada punto de D , entonces decimos que z es una *variable regionalizada* (VR). Más aún, puede pensarse que cada valor $z(\mathbf{x}_\alpha)$ es una realización de una variable aleatoria (VA) $Z(\mathbf{x}_\alpha)$, con lo que el conjunto $\{z(\mathbf{x}), \mathbf{x} \in D\}$ es una realización de un conjunto infinito de variables aleatorias. A este conjunto infinito de variables aleatorias $\{Z(\mathbf{x}), \mathbf{x} \in D\}$ se le denomina *función aleatoria* o *proceso espacial* (PE).

Considérese un vector de posición arbitrario \mathbf{x}_0 en el dominio D (que puede o no formar parte de una muestra). La variable aleatoria correspondiente, $Z(\mathbf{x}_0)$, genera valores $z(\mathbf{x}_0)$ de acuerdo con su función de distribución

$$F_{\mathbf{x}_0}(z) = P(Z(\mathbf{x}_0) \leq z).$$

Así mismo, en el caso de n variables aleatorias asociadas a n vectores de posición, la función de distribución se expresa como

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(z_1, \dots, z_n) = P(Z(\mathbf{x}_1) \leq z_1, \dots, Z(\mathbf{x}_n) \leq z_n). \quad (1.2.1)$$

Este modelo general constituye la ley espacial de probabilidad de $Z(\mathbf{x})$ y permite describir multiplicidad de procesos en la naturaleza. No obstante, en la práctica es imposible determinarla, pues al considerar muestras finitas de n posiciones espaciales sólo se tiene una realización parcial $\{z(\mathbf{x}_1), \dots, z(\mathbf{x}_n)\}$ del PE $Z(\mathbf{x})$. Resulta necesario entonces hacer algunos supuestos que permitan realizar inferencias [2].

Por lo general, en las aplicaciones geoestadísticas resulta suficiente estimar los momentos de $Z(\mathbf{x})$ hasta de segundo orden, además de que en la mayoría de los casos la información disponible no permite inferir momentos de orden superior [3].

El momento de primer orden de $Z(\mathbf{x})$ es la esperanza, definida como

$$m(\mathbf{x}) = E[Z(\mathbf{x})]. \quad (1.2.2)$$

Los momentos de segundo orden que se consideran son

- La varianza de $Z(\mathbf{x})$

$$\sigma^2(\mathbf{x}) = \text{Var}(Z(\mathbf{x})) = E\left[\{Z(\mathbf{x}) - m(\mathbf{x})\}^2\right] \quad (1.2.3)$$

- La covarianza entre $Z(\mathbf{x}_i)$ y $Z(\mathbf{x}_j)$

$$\text{Cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) = E\left[\{Z(\mathbf{x}_i) - m(\mathbf{x}_i)\}\{Z(\mathbf{x}_j) - m(\mathbf{x}_j)\}\right] \quad (1.2.4)$$

- El semivariograma

$$\gamma(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) = \frac{1}{2}\text{Var}[Z(\mathbf{x}_i) - Z(\mathbf{x}_j)]. \quad (1.2.5)$$

1.2.1. Procesos Espaciales Estacionarios

Se dice que el PE $Z(\mathbf{x})$ es *estrictamente estacionario* en D si la función de distribución (1.2.1) es invariante ante traslaciones; es decir, que dado un conjunto de n puntos $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ y para algún vector \mathbf{h} , se cumple que

$$F_{\mathbf{x}_1, \dots, \mathbf{x}_n}(z_1, \dots, z_n) = F_{\mathbf{x}_1 + \mathbf{h}, \dots, \mathbf{x}_n + \mathbf{h}}(z_1, \dots, z_n),$$

lo que significa que si se cambia la configuración de puntos en una determinada dirección, la distribución múltiple no se altera. Dado que usualmente se trabaja con los momentos hasta de segundo orden, tiene sentido práctico acotar y relajar la hipótesis de estacionariedad estricta a estos primeros momentos.

Se dice que $Z(\mathbf{x})$ es *estacionaria de segundo orden* en D si

1. Su valor esperado existe y es el mismo para cualquier \mathbf{x}

$$E[Z(\mathbf{x})] = m \quad (1.2.6)$$

2. Para cualquier par de variables, $Z(\mathbf{x})$, $Z(\mathbf{x} + \mathbf{h})$, existe la covarianza y sólo depende del vector de separación \mathbf{h}

$$\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) = C(\mathbf{h}). \quad (1.2.7)$$

La condición (1.2.7) implica que la varianza existe, es finita y no depende de \mathbf{x} , pues

$$\sigma^2 = C(\mathbf{0}) = \text{Var}(Z(\mathbf{x})). \quad (1.2.8)$$

Adicionalmente, el semivariograma también es estacionario y cumple la siguiente relación con la función de covarianza

$$\begin{aligned}\gamma(\mathbf{h}) \equiv \gamma(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) &= \frac{1}{2}E \left[\{Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\}^2 \right] \\ &= C(\mathbf{0}) - C(\mathbf{h}),\end{aligned}\tag{1.2.9}$$

con lo que cualquiera de ambas funciones permite caracterizar la dependencia espacial del proceso.

Existen algunos PE que representan a fenómenos físicos con una capacidad de variación prácticamente ilimitada, por lo que para estas funciones no está definida la varianza ni la covarianza. Sin embargo, es posible que sus diferencias $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$ tengan varianza limitada; es decir, que las diferencias sean estacionarias de segundo orden.

Así, se dice que un PE es *estacionario intrínseco* en D si

1. El valor esperado de las diferencias cumple que

$$E [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0\tag{1.2.10}$$

2. La varianza de las diferencias es finita y está dada por

$$Var [Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h})\tag{1.2.11}$$

Por lo anterior, si un PE tiene varianza finita, eso implica que sus diferencias también tendrán varianza finita; es decir, que si un PE es estacionario de segundo orden también será estacionario intrínseco. En contraparte, si un PE es estacionario intrínseco, no podemos asegurar que también sea estacionario de segundo orden [2][3].

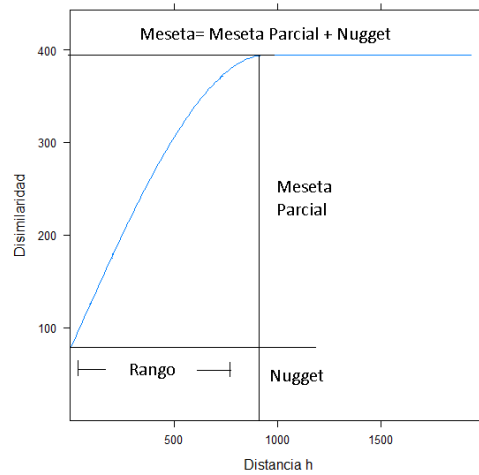
En este capítulo nos centraremos en el caso de media constante. El caso general se comentará más adelante.

1.3. Función Variograma y Covarianza

El variograma¹ es una de las herramientas más importantes en la geostatística, pues es la base para caracterizar la estructura de dependencia espacial de algún PE.

¹Aunque estrictamente hablando no son lo mismo, es común referirse al semivariograma como variograma.

Figura 1.3.1: Semivariograma .



1.3.1. Variograma y Covarianza Teóricos

La descripción espacial de un PE, $Z(\mathbf{x})$, puede hacerse por medio de sus diferencias o incrementos $Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})$. En este sentido, puede decirse que el variograma teórico se define a partir de la condición de estacionariedad intrínseca, dada por las ecuaciones (1.2.10) y (1.2.11).

Puede observarse, por definición, [ecs. (1.2.5), (1.2.9)], que el variograma γ cumple que

- Es siempre positivo o cero, si es que se evalúa en el origen

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

- Es una función par

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

- El variograma aumenta más lentamente que $\|\mathbf{h}\|^2$,

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

La forma típica de un variograma acotado se muestra en la Figura 1.3.1.

- Al valor que acota superiormente al variograma se le denomina *sill* o *meseta*. Si la variable $Z(\mathbf{x})$ es estacionaria de segundo orden, entonces la meseta coincide con $Var(Z(\mathbf{x}))$.

- Al valor de $\|\mathbf{h}\|$ a partir del cual el valor del variograma es constante e igual a la meseta, se le denomina *rango*. Para $\|\mathbf{h}\|$ mayores que el rango, $Z(\mathbf{x})$ y $Z(\mathbf{x} + \mathbf{h})$ son incorrelacionadas.
- Teóricamente, $\gamma(0) = 0$. Sin embargo, en la práctica suele ocurrir que $\lim_{\|\mathbf{h}\| \rightarrow 0} \gamma(\mathbf{h}) \neq 0$. A esta desigualdad en el origen se le denomina efecto *nugget* o *pepita*. Su existencia se debe a variaciones espaciales a distancias menores que el intervalo de muestreo más pequeño.
- A la diferencia entre la meseta y el nugget se le denomina *sill parcial* (*psill*) o *meseta parcial*.

Por su parte, la función de covarianza o covariograma se define a partir de la condición de estacionariedad de segundo orden, dada por las ecs. (1.2.6) y (1.2.7). Con respecto a esta función, puede decirse que

- Ésta es una función acotada superiormente por la varianza, es decir,

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

- Al igual que con el variograma, la covarianza es una función par,

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

pero, a diferencia de la primera, ésta sí puede adquirir valores negativos.

- En ocasiones suele dividirse la covarianza entre la varianza, dando lugar a la *función de correlación* o *correlograma*,

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

expresión que está acotada entre 1 y -1.

Es interesante mencionar que a partir de una función de covarianza es posible deducir una función variograma por medio de la ecuación (1.2.9); no obstante, lo contrario no siempre es posible debido a que el variograma no siempre es acotado [2]. De cualquier manera, se prefiere usar el estimador del variograma por poseer menor sesgo en comparación con el estimador del covariograma [4]. Además, si la media no es constante y no se elimina adecuadamente, pueden producirse malas estimaciones del covariograma.

La función de covarianza debe ser *semidefinida positiva*; es decir, que la función $C(\mathbf{h})$ involucrada en el cálculo de la varianza de una combinación lineal de $n + 1$ VA $Z(\mathbf{x}_\alpha)$ estacionarias debe ser positiva:

$$\text{Var} \left\{ \sum_{\alpha=0}^n w_\alpha Z(\mathbf{x}_\alpha) \right\} = \sum_{\alpha=0}^n \sum_{\beta=0}^n w_\alpha w_\beta C(\mathbf{x}_\alpha - \mathbf{x}_\beta) \geq 0, \quad (1.3.1)$$

para cualesquiera vectores \mathbf{x}_α y pesos w_α . Lo que permite garantizar que la varianza de la combinación lineal de cualquier subconjunto de VA, de una PE, sea positiva, es que la suma de los pesos w_α sea igual a cero. Así, si $\sum_{\alpha=0}^n w_\alpha = 0$, la varianza de una combinación lineal de VA intrínsecamente estacionarias se define por

$$\text{Var} \left\{ \sum_{\alpha=0}^n w_\alpha Z(\mathbf{x}_\alpha) \right\} = - \sum_{\alpha=0}^n \sum_{\beta=0}^n w_\alpha w_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) \geq 0, \quad (1.3.2)$$

pues, evidentemente, al no existir la covarianza para PE intrínsecos, la expresión (1.3.1) debe quedar en términos del variograma γ . Por lo anterior, decimos que $-\gamma$ es *condicionalmente semidefinido positivo*, o bien, que γ es *condicionalmente semidefinido negativo* [2].

1.3.2. Anisotropía

Si la estructura espacial sólo depende de la norma de \mathbf{h} y, por tanto, es independiente de su dirección, se dice que $Z(\mathbf{x})$ es *isotrópica*. Cuando no sólo depende de la norma de \mathbf{h} , sino que el variograma es diferente considerando diferentes direcciones, $Z(\mathbf{x})$ es *anisotrópica*.

Se le conoce como *anisotropía geométrica* a aquella situación en la que, al variar la dirección en el espacio, varía también el rango del variograma. Este tipo de anisotropía puede corregirse mediante un cambio de coordenadas adecuado.

Por otra parte, la anisotropía zonal es aquella en la que al variar la dirección varía también el valor de la meseta. En estos casos, es común considerar una combinación de variograma isotrópico en cierta dirección, γ_1 , y agregar un variograma anisotrópico geométrico, γ_2 , de tal manera que no influya en la dirección de γ_1 . De esta manera, el PE resultante sobrepone dos procesos incorrelacionados, $Z(\mathbf{x}) = Z_1(\mathbf{x}) + Z_2(\mathbf{x})$ [2].

1.3.3. Estimación del Variograma

Para obtener un modelo de variograma que describa adecuadamente la dependencia espacial de los datos, habitualmente se procede de acuerdo a lo siguiente:

1. Estimación preliminar del semivariograma usando algún tipo de estimador experimental.
2. Ajustar un modelo válido de semivariograma a las observaciones preliminares, ya que éstas, normalmente, no pueden usarse directamente en la predicción espacial.
3. Validación del modelo.

Considérese una variable regionalizada $z(\mathbf{x})$ en el dominio D . Sean z_α y z_β los valores de la VR en los vectores de posición \mathbf{x}_α y \mathbf{x}_β , respectivamente. Se define la *disimilaridad*, γ^* , entre estas dos cantidades, mediante

$$\gamma_{\alpha\beta}^* = \frac{1}{2}(z_\alpha - z_\beta)^2. \quad (1.3.3)$$

Más aún, puede pensarse en el vector resultante de la diferencia de ambos vectores, $\mathbf{h} = \mathbf{x}_\alpha - \mathbf{x}_\beta$, y expresar la disimilaridad (1.3.3) en términos de éste, obteniéndose

$$\gamma^*(\mathbf{h}) = \frac{1}{2}(z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha))^2,$$

es decir, que la disimilaridad depende de la orientación dada por \mathbf{h} . Evidentemente, la disimilaridad entre z_α y z_β debe ser la misma que entre z_β y z_α ; o bien, en otras palabras, que la disimilaridad es independiente de considerar el vector \mathbf{h} o $-\mathbf{h}$. Decimos entonces que la disimilaridad es simétrica:

$$\gamma^*(-\mathbf{h}) = \gamma^*(\mathbf{h}).$$

La idea de disimilaridad entre dos valores z_α y z_β puede extenderse a todos los pares de valores de $z(\mathbf{x})$ en las diferentes posiciones muestrales. Esta idea generalizada consiste en calcular la disimilaridad promedio entre todos los pares de valores de $z(\mathbf{x})$ cuya diferencia de sus correspondientes vectores de posición sea un vector \mathbf{h} en particular. Formalmente, se dice que se calcula el promedio de la disimilaridad $\gamma^*(\mathbf{h})$ para una clase dada de vectores ϖ . Los vectores pertenecientes a la clase ϖ se caracterizan por tener la misma norma y dirección dentro de los márgenes de cierta tolerancia establecida.

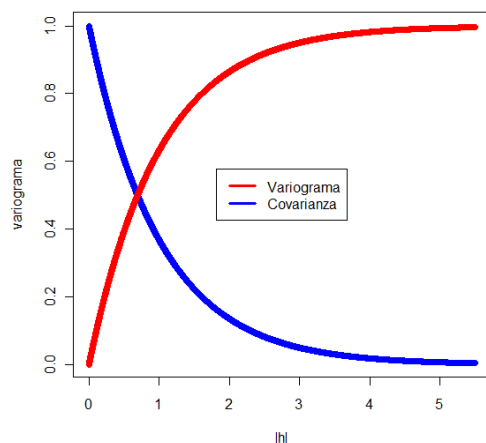
De esta manera, se define el *variograma experimental* como

$$\gamma^*(\varpi_q) = \frac{1}{2n_c} \sum_{\alpha=1}^{n_c} \{z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)\}^2, \quad (1.3.4)$$

siendo $\mathbf{h} \in \varpi_q$, y n_c el número de pares de vectores de posición cuya diferencia es \mathbf{h} . La ecuación (1.3.4), propuesta por Matheron en 1962, se considera el estimador clásico del variograma teórico. No obstante que es insesgado, tiene poca robustez, lo que es entendible debido a que es esencialmente una media muestral. Por esta razón, Cressie y Hawkins propusieron en 1980 un estimador alternativo más robusto [2][5].

En la práctica, lo usual es considerar vectores \mathbf{h} cuya norma sea menor que la mitad del diámetro del dominio en cuestión. Con respecto a la región de tolerancia para \mathbf{h} , esta debe ser lo suficientemente grande como para que no aparezcan inestabilidades; en este sentido, se recomienda que el número de aportaciones a la estimación en un salto de \mathbf{h} sea al menos de 30 [6].

Figura 1.3.2: Variograma exponencial, con $a = b = 1$. El rango efectivo se alcanza en $|h| = 3$.



1.3.4. Modelos de Variograma

Los semivariogramas experimentales no siempre satisfacen las propiedades de una función de semivariograma, razón por la que se suelen ajustar semivariogramas teóricos válidos.

El variograma se puede caracterizar mediante una función de covarianza continua. Si el variograma es finito y limitado por un valor $\gamma(\infty)$, podemos expresar entonces una función de covarianza por medio de

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

En la práctica es común encontrar variogramas experimentales para los que, a partir de cierto valor de $\|\mathbf{h}\|$, sus estimaciones parece fluctuar alrededor de un valor constante. A este tipo de variogramas experimentales se les puede ajustar un variograma teórico como el de la figura 1.3.1, lo que en esencia es una función de covarianza:

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

donde $b = C(\mathbf{0})$ es el valor de la función de covarianza en el origen.

A continuación se presentan algunos modelos de covarianza utilizados comúnmente en casos de PE isotrópicos.

- *Efecto nugget.*

$$C_{nug}(\mathbf{h}) = \begin{cases} b & \|\mathbf{h}\| = 0 \\ 0 & \|\mathbf{h}\| > 0 \end{cases},$$

con $b > 0$. El correspondiente variograma es $b - C_{nug}(\mathbf{h})$, cuyo valor es 0 en el origen y b fuera de él.

- *Exponencial*

$$C_{exp}(\mathbf{h}) = b \exp\left(-\frac{1}{a}\|\mathbf{h}\|\right),$$

con $a, b > 0$. El variograma correspondiente se aproxima asintóticamente a b , por lo que en la práctica se considera como *rango efectivo* aquella norma $\|\mathbf{h}\|$ para la cual se ha alcanzado el 95% de $C(\mathbf{0})$. El valor del rango efectivo es $3a$. La figura 1.3.2 representa tanto la función de covarianza como el variograma, en el caso particular de $a = b = 1$.

- *Esférico*

$$C_{esf}(\mathbf{h}) = \begin{cases} b\left(1 - \frac{3}{2}\frac{\|\mathbf{h}\|}{a} + \frac{1}{2}\frac{\|\mathbf{h}\|^3}{a^3}\right) & 0 \leq \|\mathbf{h}\| \leq a \\ 0 & \|\mathbf{h}\| > a \end{cases}.$$

En este modelo, a representa el rango y b el máximo valor de la covarianza. Así, el variograma correspondiente crece hasta llegar al rango, a partir del cual cambia de valor abruptamente haciéndose constante en b . La figura 1.3.3 representa la función de covarianza y el variograma para el caso $a = b = 1$.

- Modelo Matern o K -Bessel

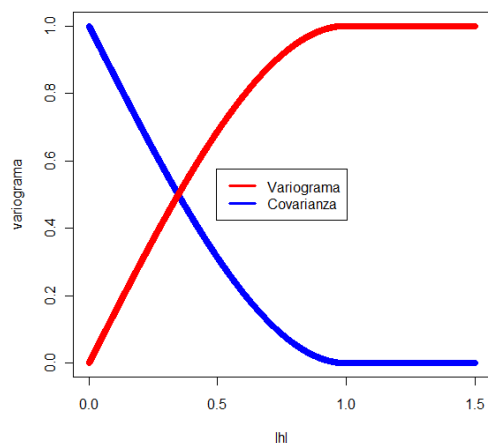
$$C_{Matern}(\mathbf{h}) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{\|\mathbf{h}\|}{a}\right)^{\kappa} K_{\kappa}\left(\frac{\|\mathbf{h}\|}{a}\right),$$

donde $a > 0$, $\kappa \geq 0$, Γ es la función gamma, y K_{κ} es la función de Bessel modificada de segundo tipo de orden κ .

Este modelo resulta particularmente interesante y muy flexible para modelar variogramas que tiendan más o menos rápido hacia la meseta, lo que se consigue variando el valor de κ , que se conoce como *parámetro de suavizado*. En particular, si $\kappa = 0.5$ se obtiene el modelo de covarianza exponencial. Con valores de κ mayores que 0.5 se consiguen variogramas que tienden más lentamente hacia la meseta, es decir, son modelos que consideran un mayor rango de dependencia. Contrariamente, para valores de κ menores que 0.5 se consiguen variogramas que tienden más rápidamente hacia la meseta, lo que significa que se consiguen modelos que consideran menores rangos de dependencia [7].

Existen muchos otros modelos de variograma, que en general pueden dividirse en *acotados* y *no acotados*. El modelo exponencial y esférico son ejemplos de variogramas acotados, pero dentro de esta categoría puede mencionarse también al *modelo lineal con meseta*, al *modelo circular*, *modelo*

Figura 1.3.3: Variograma esférico con $a = b = 1$. A partir de $|h| = 1$, el variograma es constante e igual a b .



gaussiano, y al *modelo de efecto agujero*. Hay otros casos en los que la varianza aparenta incrementarse sin límite. Incluso, si se toman intervalos más pequeños siempre queda variación sin explicar. Un ejemplo de esto es el modelo nugget, ya que toda la varianza se encuentra dentro de un intervalo de muestro más pequeño. Para ver estos y otros modelos, véase, por ejemplo, la obra de J.P. Chiles [7].

1.3.5. Métodos de Ajuste del Variograma Experimental, Criterios de Selección y Validación

Para ajustar un modelo de variograma suele utilizarse la aproximación por mínimos cuadrados. Sea θ el vector de parámetros del variograma, y sean $\hat{\gamma} = (\hat{\gamma}(\mathbf{h}_1), \dots, \hat{\gamma}(\mathbf{h}_k))^t$ y $\gamma(\theta) = (\gamma(\mathbf{h}_1, \theta), \dots, \gamma(\mathbf{h}_k, \theta))^t$ los vectores que representan el variograma experimental y el variograma teórico en los vectores \mathbf{h}_j , respectivamente. Entonces, el método de mínimos cuadrados consiste en los siguientes:

- *Mínimos Cuadrados Ordinarios.*

Este criterio supone que los residuos están normalmente distribuidos, que son independientes y que las semivarianzas estimadas poseen igual varianza. Así, se elige θ de tal manera que se minimice la expresión

$$(\hat{\gamma} - \gamma(\theta))^t (\hat{\gamma} - \gamma(\theta)).$$

- *Mínimos Cuadrados Generalizados*

En el caso de variables dependientes, este criterio escoge θ de tal ma-

nera que minimice

$$(\hat{\gamma} - \gamma(\theta))^t \mathbf{V}^{-1}(\theta) (\hat{\gamma} - \gamma(\theta)),$$

donde $\mathbf{V}(\theta)$ es la matriz de covarianza de $\hat{\gamma}$. No obstante, en la práctica es difícil de obtener o aproximar.

■ *Mínimos Cuadrados Ponderados*

Suponiendo que las varianzas de $\hat{\gamma}$ no son iguales, este criterio busca aquél θ que minimice

$$(\hat{\gamma} - \gamma(\theta))^t \mathbf{W}^{-1}(\theta) (\hat{\gamma} - \gamma(\theta)),$$

donde $\mathbf{W}(\theta)$ es una matriz diagonal cuyos valores son las varianzas de $\hat{\gamma}$. Explícitamente, se minimiza

$$\sum_j \frac{[\hat{\gamma}(\mathbf{h}_j) - \gamma(\mathbf{h}_j, \theta)]^2}{Var[\hat{\gamma}(\mathbf{h}_j)]}; \quad (1.3.5)$$

sin embargo, el problema es que $Var[\hat{\gamma}(\mathbf{h}_j)]$ casi siempre se desconoce.

Cressie propone una alternativa al estimador de mínimos cuadrados ponderados, demostrando que una buena aproximación consiste en reemplazar $Var[\hat{\gamma}(\mathbf{h}_j)]$ por $[\gamma(\mathbf{h}_j, \theta)]^2 / N(\mathbf{h}_j)$, dando así mayor peso a las semivarianzas que se esperan menores [8]. Si $\hat{\gamma}$ se evalúa en un conjunto finito de vectores \mathbf{h}_j , se escoge θ tal que minimice

$$\sum_j w_j(\theta) [\hat{\gamma}(\mathbf{h}_j) - \gamma(\mathbf{h}_j, \theta)]^2,$$

siendo $w_j(\theta) = |N(\mathbf{h}_j)| / \gamma(\mathbf{h}_j, \theta)^2$.²

Otra forma de proceder en cuanto al ajuste de un modelo válido consiste en la estimación por máxima verosimilitud. Este método es muy conocido en la inferencia estadística paramétrica, aunque su empleo en geoestadística ha sido relativamente reciente (mediados de 1980). Si suponemos que la distribución de los datos es normal, se puede deducir fácilmente la expresión de la función de verosimilitud y obtener las estimaciones de los parámetros buscando los valores que la maximizan. Para más detalles, véase la obra de Cressie[5].

La validación del modelo de variograma puede hacerse de varias maneras. Una de ellas es usando el método *leave one out*, que consiste en quitar un elemento de la muestra y predecir el valor en ese punto (mediante kriging) con el modelo de variograma obtenido, procediendo de la misma manera

²Otra alternativa es considerar los pesos $w_j(\theta)$ como $|N(\mathbf{h}_j)| / \hat{\gamma}(\mathbf{h}_j)^2$, ya que en la forma primeramente mencionada los pesos dependen de θ , con lo que al minimizar la ecuación (1.3.5) se maximizan también las varianzas.

para los demás puntos de la muestra. Al final del procedimiento se obtiene un mapa de las diferencias entre los valores reales y estimados. Si el modelo de variograma reproduce correctamente la estructura espacial de los datos, entonces dichas diferencias deberán ser cercanas a cero [3].

Capítulo 2

Estimación de la tendencia

En este capítulo se tratará la estimación de la tendencia en un conjunto de datos espaciales. Una primera aproximación al problema consiste en proponer modelos paramétricos de la tendencia y estimar los coeficientes de la misma planteando un problema de regresión lineal. Otra forma de proceder, no obstante, es de manera no paramétrica, estimando la tendencia en cada punto del dominio de definición por medio de algún estimador local.

2.1. El Modelo General

Considérese una variable respuesta $Y = Z(\mathbf{X})$ cuya dependencia radica en un vector de coordenadas $\mathbf{X} \in \mathbb{R}^d$, mismas que pueden representar una línea si $d = 1$, o un plano si $d = 2$. De manera general, el modelo para Y se descompone en una parte determinística y otra residual de la forma

$$Y_i = m(\mathbf{X}_i) + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (2.1.1)$$

donde $m(\mathbf{X})$, conocida como función de tendencia o deriva, da cuenta de la variabilidad a gran escala. A su vez, $\epsilon(\mathbf{x})$ representa el comportamiento local o evolución a pequeña escala, siendo un proceso de media cero, varianza constante y con función de covarianza que depende de la diferencia entre las posiciones en cuestión; es decir,

$$E[\epsilon_i | \mathbf{X}_i] = 0, \quad (2.1.2)$$

$$Var(\epsilon_i | \mathbf{X}_i) = \sigma^2, \quad (2.1.3)$$

$$Cov(\epsilon_i, \epsilon_j | \mathbf{X}_i, \mathbf{X}_j) = \sigma^2 \rho(\mathbf{X}_i - \mathbf{X}_j). \quad (2.1.4)$$

Denominaremos al modelo 2.1.1 *modelo general*.

2.2. Estimador paramétrico

Uno de los modelos más utilizados en estadística para el caso de datos no homogéneos es el modelo clásico de regresión lineal. Sea $Z(\mathbf{x})$ una función aleatoria definida en un dominio D , y exprese de la siguiente manera:

$$Z(\mathbf{x}) = \sum_{j=0}^p x_j(\mathbf{x})\beta_j + \epsilon(\mathbf{x});$$

es decir, que expresamos la función aleatoria $Z(\mathbf{x})$ según (2.1.1), considerando que $m(\mathbf{x}) = \sum_{j=0}^p x_j(\mathbf{x})\beta_j$, con $\beta = \{\beta_0, \dots, \beta_p\}^t$ un vector desconocido de coeficientes, $\{x_j, j = 0, \dots, p\}$ un conjunto de variables explicativas y ϵ es un proceso de media cero y, en principio, incorrelacionado y de varianza constante.

Supóngase que el objetivo es la estimación eficiente de la tendencia, o bien, la estimación óptima de los parámetros de la variación a gran escala, β , a partir de las observaciones disponibles en un conjunto de posiciones espaciales $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Puede entonces escribirse el sistema de ecuaciones correspondiente de manera matricial como

$$\mathbf{Z} = \mathbf{X}\beta + \epsilon, \quad (2.2.1)$$

siendo $\mathbf{Z} = \{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)\}$, \mathbf{X} la matriz de diseño y $\epsilon = \{\epsilon(\mathbf{x}_1), \dots, \epsilon(\mathbf{x}_n)\}^t$.

Suponiendo que los errores son incorrelacionados, de media cero y varianza constante σ^2 , el mejor estimador lineal insesgado de β es

$$\hat{\beta}_{ols} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Z}; \quad (2.2.2)$$

no obstante, si los errores no tienen varianza constante o no son incorrelacionados, como ocurre en una situación de dependencia espacial, es posible encontrar un mejor estimador de β . Supóngase entonces que los errores tienen matriz de covarianza $\sigma^2\mathbf{\Sigma}$, siendo $\mathbf{\Sigma}$ la matriz de correlación de los errores. En lugar de la ecuación (2.2.2), se tiene que la mejor estimación de β es

$$\hat{\beta}_{gls} = (\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{Z}, \quad (2.2.3)$$

que se conoce como el estimador por mínimos cuadrados generalizados.

Factorización Cholesky. En efecto, en el contexto de regresión por mínimos cuadrados generalizados, la ecuación (2.2.3) surge como solución de minimizar

$$(\mathbf{Y} - \mathbf{X}\beta)^t \mathbf{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\beta). \quad (2.2.4)$$

Nótese que en el caso en el que se supone que los errores son independientes e incorrelacionados, $\mathbf{\Sigma} = I$, dando lugar a la solución

(2.2.2). Ahora, puesto que $\sigma^2\Sigma$ es la matriz de covarianza de los errores, ésta debe ser no singular y definida positiva. Así, es posible encontrar mediante factorización Cholesky una matriz triangular S , tal que $\Sigma = SS^t$. Sustituyendo esta expresión en la expresión (2.2.4) y desarrollando, se obtiene que

$$(S^{-1}\mathbf{Y} - S^{-1}\mathbf{X}\beta)^t (S^{-1}\mathbf{Y} - S^{-1}\mathbf{X}\beta).$$

Llamando a $S^{-1}\mathbf{Y} = \mathbf{Y}'$ y $S^{-1}\mathbf{X} = \mathbf{X}'$, se tiene que

$$(\mathbf{Y}' - \mathbf{X}'\beta)^t (\mathbf{Y}' - \mathbf{X}'\beta),$$

es decir, que ahora se realiza la regresión sobre las nuevas variables \mathbf{Y}' y \mathbf{X}' . Es fácil ver que los nuevos errores $\epsilon' = \mathbf{Y}' - \mathbf{X}'\beta$ son incorrelacionados y de varianza constante:

$$Var(\epsilon') = Var(\mathbf{Y}' - \mathbf{X}'\beta) = Var(S^{-1}(\mathbf{Y} - \mathbf{X}\beta)) = Var(S^{-1}\epsilon),$$

y como $\Sigma = SS^t$ es la matriz de correlación de los errores ϵ ,

$$Var(S^{-1}\epsilon) = S^{-1}Var(\epsilon)S^{-t} = S^{-1}\sigma^2SS^tS^{-t} = \sigma^2I. \quad (2.2.5)$$

De esta manera, la factorización Cholesky resulta de interés en el caso de estimación bajo dependencia espacial porque permite hacer una transformación de los datos originales a datos independientes, como quedó demostrado en (2.2.5). Esto será de mucha utilidad más adelante para los propósitos de este trabajo.

2.3. Estimador No Paramétrico

El estimador no paramétrico de la tendencia consiste en una generalización de la estimación de la tendencia en un punto x considerando un promedio ponderado de las respuestas Y , dando mayor peso a las respuestas cercanas y menor peso a las lejanas. El estimador no paramétrico a utilizar en este trabajo es el llamado *local lineal*. Este estimador es una generalización del estimador de Nadaraya-Watson [9].

2.3.1. Estimador local lineal

Considerando por simplicidad el caso unidimensional, la estimación de la tendencia $\hat{m}(x)$ en un punto x , se calcula por medio de un promedio ponderado de las observaciones Y_i ,

$$\hat{m}(x) = \sum_{i=1}^n W_i^x Y_i,$$

donde

$$W_i^x = K_h(X_i - x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right)}.$$

Asimismo, K es una función núcleo, simétrica y definida positiva, y h es una ventana de suavización que determina el tamaño del vecindario para realizar la respuesta media. La función núcleo actúa como una función de peso que le otorga mayor importancia a los puntos cercanos y menor importancia a los puntos alejados de x .

La función K debe cumplir las siguientes propiedades:

- $K(x) \geq 0$ para todo x .
- $\int K(t)dt=1$
- $K(-x) = K(x)$

Hay varias funciones que se pueden considerar como núcleo. Algunas de ellas son

- Núcleo tricubo

$$K(x) = \begin{cases} (1 - |x|^3)^3 & x \in [-1, 1] \\ 0 & x \notin [-1, 1] \end{cases}$$

- Núcleo Gaussiano

$$K(x) = \frac{1}{2\pi} \exp[-0.5x^2]$$

- Núcleo bicuadrado

$$K(x) = \begin{cases} (1 - x^2)^2 & x \in [-1, 1] \\ 0 & x \notin [-1, 1] \end{cases}$$

- Núcleo de Epanechnikov

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & x \in [-1, 1] \\ 0 & x \notin [-1, 1] \end{cases}$$

El grado de suavidad de las estimaciones se determina mediante la elección del parámetro h . En este sentido, la elección del núcleo no es tan importante, sino más bien importa elegir un tamaño de ventana adecuado. Si h es muy grande, la estimación resulta muy suave; si es muy pequeña, prácticamente se interpolan los datos.

El estimador de Nadaraya-Watson puede pensarse como un ajuste local constante entorno a un punto x . De esta manera, una generalización de este estimador consiste en considerar un ajuste local polinómico de grado p , usando los datos alrededor de x . Así, se tiene que

$\mathbf{p}=0$ → Estimador de Nadaraya-Watson

$\mathbf{p}=1$ → Estimador local lineal

$\mathbf{p}=2$ → Estimador local cuadrático

$\mathbf{p}=3$ → Estimador local cúbico

El estimador polinómico local se basa en la idea de que la función de regresión m se puede aproximar por un polinomio de grado p por medio del desarrollo de Taylor. De esta manera se justifica el ajuste polinómico local y en cada x se minimiza la expresión siguiente:

$$\sum_{i=1}^n K_h(X_i - x) (Y - [\beta_0 + \beta_1(X_i - x) + \dots + \beta_p(X_i - x)^p])^2. \quad (2.3.1)$$

La solución del problema de minimización viene dado por

$$\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^t = (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{Z}$$

siendo

$$X = \begin{bmatrix} 1 & X_1 & \cdot & \cdot & \cdot & X_1^1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_n & \cdot & \cdot & \cdot & X_n^p \end{bmatrix}$$

y

$$\mathbf{W} = \text{diag} [K_h(X_1 - x), \dots, K_h(X_n - x)].$$

De esta manera, si interesa la estimación de m en el punto $x_0 = x$, habrá que evaluar la expresión

$$\hat{\beta}_0 + \hat{\beta}_1(x - x_0) + \dots + \hat{\beta}_p(x - x_0)^p,$$

obteniéndose

$$\hat{m}(x_0) = \hat{\beta}_0. \quad (2.3.2)$$

Otro de los aspectos por determinar es el grado p del polinomio. En este sentido, algunos resultados teóricos demuestran que el comportamiento

asintótico de los estimadores mejora conforme aumenta p [10]. No obstante, en la práctica no queda del todo claro el beneficio de utilizar polinomios de grado elevado puesto que a mayor p , mayor varianza en la estimaciones. Es frecuente que se necesite de tamaños muestrales considerables para tener una mejoría notable en el comportamiento del estimador.

Adicionalmente, se sabe que las estimaciones que se basan en polinomios de grado impar presentan mejores propiedades teóricas que aquellas basadas en polinomios de grado par. Por tal razón, en la práctica es suficiente utilizar estimadores locales cúbicos ($p = 3$) o *estimadores locales lineales* ($p = 1$).

2.3.2. Estimador local lineal en datos espaciales

En el caso de datos espaciales se considera un conjunto de vectores aleatorios $\{(\mathbf{X}_i, Y_i)\}_{i=1}^n$, donde las Y_i son respuestas escalares y \mathbf{X}_i es, en general, un vector en \mathbb{R}^d . El problema multivariante de regresión no paramétrica consiste en estimar $m(\mathbf{x})$ para algún \mathbf{x} en un subdominio de \mathbb{R}^d . Desde luego, no se hace supuesto alguno sobre la forma funcional paramétrica de m .

Como ya se ha mencionado, el modelo que se asume es aquél dado por la ecuación (2.1.1), junto con las condiciones (2.1.2), (2.1.3) y (2.1.4). En analogía con las ecuaciones (2.3.1) y (2.3.2), el estimador local lineal para m en \mathbf{x} es la solución para β_0 de la minimización por mínimos cuadrados siguiente:

$$\min_{\beta_0, \beta} \sum_{i=1}^n \{K_H(\mathbf{X}_i - \mathbf{x}) \{Y_i - \beta_0 - \beta^t (\mathbf{X}_i - \mathbf{x})\}^2\},$$

donde H es una matriz simétrica definida positiva, K es un kernel d -variado y $K_H(\mathbf{u}) = |H|^{-1}K(H^{-1}\mathbf{u})$. La matriz H , llamada *matriz ventana*, es la que controla tanto la forma como el tamaño del vecindario que se usará en la estimación de $m(\mathbf{x})$. Explícitamente, el estimador local lineal toma la forma

$$\hat{m}(\mathbf{x}, H) = e_1^t (\mathbf{X}_x^t \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^t \mathbf{W}_x \mathbf{X}_x \mathbf{Y} = \mathbf{s}_x^t \mathbf{Y}, \quad (2.3.3)$$

donde e_1 es un vector de longitud $d + 1$ cuya primera coordenada es 1 y las coordenadas subsecuentes 0; \mathbf{X}_x es una matriz cuya fila i -ésima es $(1, (\mathbf{X}_i - \mathbf{x})^t)$, $\mathbf{W}_x = \text{diag}\{K_H(\mathbf{X}_1 - \mathbf{x}), \dots, K_H(\mathbf{X}_n - \mathbf{x})\}$, e \mathbf{Y} es el vector de respuestas escalares $(Y_1, \dots, Y_n)^t$.

2.3.3. Selección de la matriz ventana H

Tal como ocurría con h en el caso unidimensional, en este caso resulta de especial importancia seleccionar adecuadamente la matriz H . La selección de H puede hacerse mediante el criterio de *validación cruzada generalizada* (GCV), que consiste en elegir aquella H que minimice la expresión

$$GCV(H) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \hat{m}_i}{1 - \frac{1}{n} \text{tr} \mathbf{S}} \right)^2, \quad (2.3.4)$$

donde \mathbf{S} es la matriz cuya i -ésima fila está dada por $\mathbf{s}_{\mathbf{X}_i}^t$; es decir, el vector de suavización para $\mathbf{x} = \mathbf{X}_i$. No obstante, hay evidencia de que no es correcto usar el criterio (2.3.4) en presencia de errores correlacionados, ya que su esperanza se ve afectada por dicha correlación [11]. En realidad, la mayoría de los métodos de selección de la matriz H suelen no funcionar del todo bien en presencia de correlación espacial [12]. En ese sentido, resulta necesario introducir alguna corrección en la expresión (2.3.4) a fin de que tome en cuenta adecuadamente la correlación espacial y corrija el sesgo.

Al respecto, M. Francisco-Fernández y J. Opsomer [13] proponen la siguiente corrección para la ecuación (2.3.4):

$$GCV_c(H) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \hat{m}_i}{1 - \frac{1}{n} \text{tr} \mathbf{S} \mathbf{R}} \right)^2, \quad (2.3.5)$$

donde \mathbf{R} es la matriz de correlación de las observaciones. Aún así, en la práctica se suele desconocer la matriz de correlación, por lo que estos autores proponen asumir un modelo paramétrico de covarianza y estimar \mathbf{R} [13]:

$$GCV_{ce}(H) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \hat{m}_i}{1 - \frac{1}{n} \text{tr} \mathbf{S} \hat{\mathbf{R}}} \right)^2. \quad (2.3.6)$$

Evidentemente, (2.3.6) se aproximará mejor a (2.3.5) en la medida en que el modelo de covarianza asumido sea el adecuado y que las estimaciones de los parámetros de covarianza sean lo suficientemente buenas.

Más aún, considérese la expresión para el promedio del error cuadrático medio (*mean average squared error*, MASE), dado por la siguiente expresión:

$$MASE(H) = \frac{1}{n} (\mathbf{S} \mathbf{m} - \mathbf{m})^t (\mathbf{S} \mathbf{m} - \mathbf{m}) + \frac{1}{n} \sigma^2 \text{tr} (\mathbf{S} \mathbf{R} \mathbf{S}^t), \quad (2.3.7)$$

donde \mathbf{m} es el vector de tendencia $(m(\mathbf{X}_1), \dots, m(\mathbf{X}_n))^t$, y \mathbf{R} es la matriz de correlación teórica de los errores. Denomínese a aquella matriz H que minimice la expresión (2.3.7) como H óptima, H_{opt} .

2.4. Predicción Espacial Lineal Óptima

Aunque en el presente trabajo el objetivo no se centra en la predicción del proceso sino en la estimación de la tendencia m , se ha decidido incorporar una subsección relativa al kriging dada su importancia en el análisis geoestadístico.

En muchos casos, el objetivo final es la predicción del proceso en una posición espacial \mathbf{x} . Si se sigue la aproximación tradicional, podría pensarse en utilizar el estimador de la tendencia como predictor; sin embargo, en un contexto de dependencia espacial, la tendencia no es el predictor lineal óptimo. Es en estos casos que resulta de especial importancia el Kriging.

Supóngase que se tiene información sobre cierto atributo físico z en diferentes posiciones de un dominio D . Un problema típico en esta situación es tratar de estimar el valor de z en aquellas posiciones de D en donde no hubo medición, teniendo en cuenta la estructura de covarianza de las variables aleatorias $Z(\mathbf{x})$ definidas en las posiciones en cuestión. El método utilizado es muy similar a una regresión lineal múltiple aplicada a un contexto espacial, en donde las variables aleatorias $Z(\mathbf{x})$ funcionan como variables regresoras, y la variable aleatoria en el punto donde interesa la estimación, digamos, $Z(\mathbf{x}_0)$, funge como la variable dependiente. Al conjunto de algoritmos de regresión lineal cuyo propósito es ese, se le conoce como *kriging*. Esta es una técnica de estimación local que tiene la cualidad de ser el mejor estimador lineal insesgado de z [14].

Kriging

Considérese el problema de estimar el valor de un atributo z en alguna posición \mathbf{x} a partir del conocimiento de los valores de z en n diferentes posiciones del dominio D , digamos, $z(\mathbf{x}_\alpha)$, $\alpha = 1, \dots, n$. Los estimadores kriging son variantes de estimador clásico de regresión lineal $Z^*(\mathbf{x})$,

$$Z^*(\mathbf{x}) - m(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha(\mathbf{x}) [Z(\mathbf{x}_\alpha) - m(\mathbf{x}_\alpha)], \quad (2.4.1)$$

donde $\lambda_\alpha(\mathbf{x})$ es un peso específico que se le asigna a la observación $z(\mathbf{x}_\alpha)$, interpretada como una realización de la variable $Z(\mathbf{x}_\alpha)$, y $m(\mathbf{x})$ y $m(\mathbf{x}_\alpha)$ son los valores esperados de $Z(\mathbf{x})$ y $Z(\mathbf{x}_\alpha)$, respectivamente. Hay ocasiones en las que no suelen considerarse todos los datos disponibles, sino sólo los $n(\mathbf{x})$ datos más cercanos al punto donde interesa la estimación de z ; es decir, se considera un vecindario $W(\mathbf{x})$ centrado en \mathbf{x} [14].

El objetivo de las técnicas kriging es minimizar la varianza del error de predicción $\sigma_E^2(\mathbf{x})$ sujeto a la restricción de insesgaredad del estimador. Es decir, se minimiza

$$\sigma_E^2(\mathbf{x}) = \text{Var}\{Z^*(\mathbf{x}) - Z(\mathbf{x})\},$$

sujeto a

$$E[Z^*(x) - Z(x)] = 0.$$

El estimador kriging depende del modelo que se adopte para la función aleatoria $Z(\mathbf{x})$. Por lo general, $Z(\mathbf{x})$ se suele descomponer en una componente

de tendencia y una componente residual, tal como se expresa en la ecuación (2.1.1), que aquí se reproduce:

$$Z(\mathbf{x}) = m(\mathbf{x}) + \epsilon(\mathbf{x}),$$

donde se supone conocido el variograma o el covariograma de $\epsilon(\mathbf{x})$.

El valor esperado de Z en la posición \mathbf{x} representa el valor de la tendencia en dicha posición:

$$E[Z(\mathbf{x})] = m(\mathbf{x}).$$

Las variantes del kriging dependen del modelo que se adopte para la tendencia $m(\mathbf{x})$.

1. El *kriging simple* (KS) supone

$$m(\mathbf{x}) = m,$$

es decir, que la media $m(\mathbf{x})$ es conocida en todo el dominio D .

2. El *kriging ordinario* (KO) supone que la tendencia $m(\mathbf{x}) = m$ es constante pero desconocida. Además, se ciñe a fluctuaciones locales de la media dentro de un vecindario $W(\mathbf{x})$, dentro del cual se pueda considerar la media estacionaria.
3. El *kriging universal* (KU) considera que la media $m(\mathbf{x})$ es una función que varía suavemente en todo el dominio D . La tendencia se modela por medio de combinaciones lineales de ciertas funciones de las coordenadas, $f_k(\mathbf{x})$:

$$m(\mathbf{x}) = \sum_{k=0}^K \beta_k f_k(\mathbf{x}), \quad (2.4.2)$$

Los coeficientes β_k se consideran desconocidos. Se considera que $f_0(\mathbf{x}) = 1$, de tal manera que cuando $K = 0$, se llega al caso particular del kriging ordinario.

Kriging Universal

El kriging ordinario considera de manera implícita un PE no estacionario en el dominio D ; no obstante, considera que el proceso es estacionario dentro de cierta vecindad $W(\mathbf{x})$. Sin embargo, hay ocasiones en las que no es posible suponer esto último, incluso considerando vecindades relativamente pequeñas. En casos como estos se considera entonces el kriging universal, que modela la tendencia del PE según la ecuación (2.4.2) [14].

Considerando las ecuaciones (2.4.1) y (2.4.2), se tiene que

$$Z^*(\mathbf{x}) = \sum_{k=0}^K \beta_k f(\mathbf{x}) + \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}) \left[Z(\mathbf{x}_{\alpha}) - \sum_{k=0}^K f_k(\mathbf{x}_{\alpha}) \beta_k \right],$$

o bien, reagrupando algunos términos y factorizando,

$$Z^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}) Z(\mathbf{x}_{\alpha}) + \sum_{k=0}^K \beta_k \left[f_k(\mathbf{x}) - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}) f_k(\mathbf{x}_{\alpha}) \right]. \quad (2.4.3)$$

Para que el segundo término de la suma en (2.4.3) quede sin efecto, es necesario que se cumpla que

$$\sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}) f_k(\mathbf{x}_{\alpha}) = f_k(\mathbf{x}), \quad k = 0, \dots, K. \quad (2.4.4)$$

Como ya se mencionó, se considera, por convención, que $f_0(\mathbf{x}) = 1$, de tal manera que si $K = 0$ se llega al caso particular del kriging ordinario, en donde los pesos satisfacen que $\sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}) = 1$.

Las restricciones (2.4.4) permiten entonces expresar el estimador $Z^*(\mathbf{x})$ en términos únicamente de las $n(\mathbf{x})$ variables aleatorias $Z(\mathbf{x}_{\alpha})$,

$$Z_{KU}^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{KU}(\mathbf{x}) Z(\mathbf{x}_{\alpha}), \quad (2.4.5)$$

donde, evidentemente, $\lambda_{\alpha}^{KU}(\mathbf{x})$ satisface que

$$\sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{KU}(\mathbf{x}) f_k(\mathbf{x}_{\alpha}) = f_k(\mathbf{x}), \quad k = 0, \dots, K. \quad (2.4.6)$$

Un poco de álgebra puede mostrar que el estimador (2.4.5) es insesgado,

$$E[Z_{KU}^*(\mathbf{x}) - Z(\mathbf{x})] = 0,$$

debido precisamente a la condición (2.4.6). Por otra parte, la minimización del correspondiente error de la varianza $\sigma_E^2(\mathbf{x})$ requiere utilizar multiplicadores de Lagrange. Así, el resultado de minimizar el error de la varianza es:

$$\sigma_{KU}^2(\mathbf{x}) = C_{\epsilon}(0) - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{KU}(\mathbf{x}) C_{\epsilon}(\mathbf{x}_{\alpha} - \mathbf{x}) - \sum_{k=0}^K \mu_k^{KU}(\mathbf{x}) f_k(\mathbf{x}), \quad (2.4.7)$$

donde los μ_k^{KU} representan los correspondientes multiplicadores de Lagrange y C_{ϵ} representa la función de covarianza de la parte residual ϵ . Cuando

$K = 0$, las ecuaciones (2.4.5) y (2.4.7) representan al estimador y la varianza, respectivamente, en el caso de kriging ordinario (KO).

El KU requiere fijar las funciones de tendencia $f_k(\mathbf{x})$ y la función de covarianza de la componente residual $\epsilon(\mathbf{x})$. Con respecto a las funciones f_k , éstas pueden proponerse de acuerdo a la situación física de turno, aunque en ocasiones no hay elementos suficientes para poder hacerlo. No obstante, es común utilizar polinomios de grado menor o igual a 2 [14].

Estimación en el caso de tendencia no constante

Considérese un proceso Z intrínsecamente estacionario del que se tienen las observaciones $\{Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)\}$. En este caso se cumplen las condiciones dadas por las ecuaciones (1.2.10) y (1.2.11), y es sobre la base de estas condiciones que se plantea el estimador del variograma en la sección 1.3.1. Cuando la media no es constante no se puede emplear el estimador empírico del variograma; por tal razón, antes de usarlo hay que eliminar la tendencia de los datos. En el caso en el que la media sea conocida (KS), se puede eliminar la tendencia y estimar el variograma a partir de los residuos; sin embargo, si la tendencia $m(\mathbf{x})$ se desconoce, como ocurre en el caso KU, habrá que proceder de otra manera.

Si se conocen los coeficientes β_k (como en el caso de KS), se puede estimar el variograma partir del modelo general mediante

$$\epsilon(\mathbf{x}) = Z(\mathbf{x}) - \sum_{k=0}^K \beta_k f_k(\mathbf{x});$$

sin embargo, en el caso del KU se desconocen los coeficientes β_k . Una idea inmediata consiste en estimar los β_k para así obtener los residuos y a partir de éstos obtener una estimación del variograma. No obstante, el problema es que, para estimar eficientemente el conjunto de β_k , es necesario conocer la varianza del proceso $Z(\mathbf{x})$; es decir, se necesita conocer el variograma γ . Esta problemática circular se considera el principal inconveniente del KU [15].

Sea $\mathbf{P}_{gls} = \mathbf{X}(\mathbf{X}^t \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \boldsymbol{\Sigma}^{-1}$. Sustituyendo (2.2.3) en (2.2.1) se obtiene la estimación de ϵ ,

$$\hat{\epsilon} = \mathbf{Z} - \mathbf{X} \hat{\boldsymbol{\beta}}_{gls} = (\mathbf{I} - \mathbf{P}_{gls}) \mathbf{Z},$$

cuya matriz de covarianza es

$$Var(\hat{\epsilon}) = (\mathbf{I} - \mathbf{P}_{gls}) \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{P}_{gls})^t = \boldsymbol{\Sigma} - \mathbf{X}(\mathbf{X}^t \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^t.$$

Así, aún procediendo de la manera más eficiente, el uso de los residuos introduce un sesgo en la estimación de la dependencia espacial, lo que implica una estimación sesgada del variograma teórico. Por lo general, el sesgo del estimador del variograma es pequeño en los saltos cercanos al origen y se torna considerable en saltos grandes [16].

Además de lo anterior, en la práctica se tiene el problema adicional de no conocer Σ . En 1984, Neuman y Jacobson propusieron una solución consistente en iniciar con el estimador $\hat{\beta}_{ols}$, estimar el variograma a partir de los residuos, ajustar un variograma teórico, calcular el estimador de mínimos cuadrados generalizados basándose en el modelo ajustado y repetir iterativamente hasta llegar a la convergencia. Aunque este procedimiento suele llegar a convergir en unas cuantas iteraciones, no soluciona el problema del sesgo, pues incluso cuando se conoce Σ , el sesgo es diferente de cero [17].

A pesar de lo anterior en cuanto a la estimación sesgada del variograma en el KU, el efecto que esto tiene en la predicción no es considerable; en primer lugar, porque al ajustar un modelo de variograma por mínimos cuadrados generalizados (o ponderados), son los saltos más pequeños los que reciben mayor peso en el ajuste. Adicionalmente, si la predicción espacial se realiza en una vecindad, el variograma se evalúa justamente en los saltos pequeños, que es en donde se tiene tanto buena estimación como buen ajuste. De hecho, Stein muestra que si se logra captar adecuadamente el comportamiento del variograma en el origen se puede lograr una predicción eficiente [18].

Capítulo 3

Descripción del Estudio de Simulación

El presente estudio de simulación tiene por objetivo realizar una comparación de distintos estimadores de la tendencia considerando datos con dependencia espacial de acuerdo al modelo general (2.1.1). La comparación se realizará entre estimadores de la tendencia paramétricos y no paramétricos, vistos en las secciones 2.2 y 2.3.3.

A partir de una muestra de tamaño n , $\{\mathbf{X}_i, Y_i\}_{i=1}^n$, el objetivo entonces es estimar la tendencia m y comparar si el estimador en cuestión, \hat{m} , se encuentra “cerca” de la tendencia que se quiere estimar. Más precisamente, lo que se quiere comparar entre los diversos estimadores usados es, en promedio, qué tan cerca se encuentran de la tendencia verdadera. Evidentemente, esto deberá hacerse considerando un número de muestras suficientemente grande, B . En general, los pasos a seguir serán los siguientes:

- Simular una muestra $\{\mathbf{X}_i, Y_i\}$ de tamaño n .
- Estimar la tendencia m por medio del estimador \hat{m} .
- Repetir este proceso un número grande de veces.
- Considerar una medida de discrepancia entre el estimador \hat{m} y la tendencia m , a fin de poder comparar entre diferentes estimadores.

3.1. Comparación de las Estimaciones de la Tendencia

Para comparar las estimaciones obtenidas con los estimadores, será necesario tener una medida que permita, por una parte, comparar en promedio qué tan cerca se encuentran dichas estimaciones de la tendencia verdadera

y, por otra parte, tener una idea de la variabilidad promedio de las mismas en todo el dominio de definición de los datos.

Una medida que permite evaluar simultáneamente ambas características es el *error cuadrático medio integrado* (MISE), que se define como

$$MISE(\hat{m}) = E \int [\hat{m}(\mathbf{x}) - m(\mathbf{x})]^2 d\mathbf{x},$$

o bien,

$$MISE(\hat{m}) = \int sesgo^2 [\hat{m}(\mathbf{x})] d\mathbf{x} + \int Var [\hat{m}(\mathbf{x})] d\mathbf{x}. \quad (3.1.1)$$

En el presente caso, puesto que se está en un caso discreto, se aproximará la ecuación (3.1.1) por medio de

$$\begin{aligned} & \int sesgo^2 [\hat{m}(\mathbf{x})] d\mathbf{x} \approx \\ & \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \left[\left(\frac{1}{B} \sum_{i=1}^B \hat{m}^i(t_i, r_j) \right) - m(t_i, r_j) \right]^2 \end{aligned} \quad (3.1.2)$$

y de

$$\begin{aligned} & \int Var [\hat{m}(\mathbf{x})] d\mathbf{x} \approx \\ & \frac{1}{k^2} \sum_{i=1}^k \sum_{j=1}^k \left[\left(\frac{1}{B} \sum_{i=1}^B [\hat{m}^i(t_i, r_j)]^2 \right) - \left(\frac{1}{B} \sum_{i=1}^B \hat{m}^i(t_i, r_j) \right)^2 \right] \end{aligned} \quad (3.1.3)$$

siendo (t_i, r_j) los puntos en la rejilla. En efecto, la ecuación (3.1.2) considera el sesgo cuadrático en cada punto de la rejilla y calcula un promedio de éstos, mientras que con la ecuación (3.1.3) se obtiene un promedio de las varianzas en cada punto de la rejilla. Por lo anterior, el mejor estimador será el que tenga el menor MISE.

3.2. Simulación de una muestra

Las simulaciones se realizarán por medio del uso de paquetes específicos dentro del lenguaje de programación **R**; fundamentalmente, se hará uso de **geoR**. Este paquete está diseñado para realizar análisis de datos geostatísticos y predicción espacial [19]. En particular, dados un modelo y unos parámetros de covarianza, la función 'grf' permite generar simulaciones de

funciones aleatorias gaussianas en el número de localizaciones espaciales indicadas.

Antes de comenzar cada simulación será necesario fijar una serie de parámetros, tales como:

- el tamaño muestral n
- el número de repeticiones B
- definir la función de tendencia m que se desea estimar
- el tipo de dependencia espacial con el que se generarán los datos

En el caso de la estimación no paramétrica será necesaria la definición de otros parámetros, tales como la función kernel, o algún parámetro para determinar la ventana.

Una vez definida la tendencia m y la estructura espacial de los errores (alguna de las funciones de covarianza mostradas en la sección 1.3.4), existen dos posibilidades:

- *Diseño fijo.* Consiste en que el experimentador fija de antemano las variables explicativas y son siempre las mismas; por ejemplo, $\mathbf{X}_i = (X_{i1}, X_{i2})$, $i = 1, 2, \dots, n$, tomando X_{i1} y X_{i2} equiespaciadas en el intervalo $(0,1)$ cada una de ellas. De esta manera, en cada una de las B simulaciones la muestra $\{\mathbf{X}_i\}_{i=1}^n$ es la misma, cambiando sólo el valor de $\{Y_i\}_{i=1}^n$.
- *Diseño aleatorio.* Este diseño consiste en que, en cada una de las B simulaciones, se tiene tanto un nuevo conjunto de posiciones muestrales $\{\mathbf{X}_i\}_{i=1}^n$ (que incluso puede ser no equiespaciado) como un nuevo conjunto de valores $\{Y_i\}_{i=1}^n$.

Una vez generadas las variables \mathbf{X}_i se generan los errores (por medio de 'grf') y se suman a la tendencia m que se haya elegido, obteniéndose un valor de Y para cada \mathbf{X}_i . Por simplicidad, en este trabajo se trabajará con *diseño fijo*.

3.3. Estimadores

Haciendo alusión a lo mencionado en la secciones 2.2 y 2.4, con respecto a la estimación de la tendencia, y en referencia también a la solución propuesta por Neuman y Jacobson, en este trabajo se considerarán cuatro estimadores paramétricos de la tendencia:

1. Mínimos cuadrados ordinarios (OLS).
2. Mínimos cuadrados generalizados con una iteración (GLS1).

3. Mínimos cuadrados generalizados iterativamente hasta la convergencia (GLSe).
4. Mínimos cuadrados generalizados considerando los parámetros teóricos de covarianza (GLSt).

El procedimiento transcurre de la manera siguiente. Una vez simulados los datos:

- Se calcula el variograma experimental mediante la función 'variog'. Esta función permite calcular el variograma experimental ya sea mediante el estimador clásico o el estimador robusto de Cressie (sección 1.3.3). Es posible también indicar algún tipo de tendencia, misma que se estima mediante mínimos cuadrados ordinarios y en cuyo caso el variograma se realiza sobre los residuos. De esta manera, se tiene la estimación de los parámetros de tendencia vía OLS.
- El siguiente paso es ajustar un modelo de semivariograma al variograma experimental. Esto se hace mediante la función 'variofit', que estima los parámetros de covarianza (meseta parcial, rango, nugget) ajustando un modelo paramétrico al variograma empírico. El ajuste del variograma puede hacerse por medio de pesos iguales, pesos de Cressie, o bien, pesos proporcionales al número de puntos en cada clase de distancia.
- Una vez estimados los parámetros de covarianza, se calcula la matriz de covarianza. La función 'varcov.spatial' permite construir la matriz de covarianza para un conjunto de localizaciones espaciales. A partir de la matriz de covarianza se obtiene la matriz de correlación, misma con la que se realizará la factorización Cholesky (sección 2.2).
- Habiendo realizado la factorización Cholesky, se efectúa la regresión con las nuevas variables \mathbf{Y}' y \mathbf{X}' , y sobre los residuos obtenidos se vuelve a calcular el variograma experimental para, nuevamente, ajustar el variograma teórico, obtener los parámetros de covarianza, estimar la matriz de correlación, hacer nuevamente factorización Cholesky, y así sucesivamente hasta llegar a la convergencia de los parámetros de covarianza. En ese momento, se detiene proceso.

En resumen:

1. Las primeras estimaciones obtenidas serán los parámetros de covarianza y los coeficientes de tendencia estimados por OLS.
2. La solución por GLS1 será considerando los resultados únicamente en la primera iteración al hacer factorización Cholesky.
3. La solución por GLSe serán los últimos parámetros de covarianza estimados y los últimos coeficientes de tendencia estimados al llegar el algoritmo a la convergencia.

4. El resultado por GLSt se obtiene, en lugar de estimando los parámetros de covarianza, construyendo la matriz de correlación con los parámetros teóricos; es decir, se hace una vez la factorización Cholesky usando la matriz de correlación teórica.

Paralelamente a los estimadores anteriores, se considerará el *estimador no paramétrico* mostrado en la ecuación (2.3.3). En este caso, el procedimiento consiste en una vez generada la muestra, estimar la matriz óptima H_{opt} , considerar un núcleo d -variado y obtener la estimaciones de acuerdo con la ecuación ya mencionada.

3.4. Definición de los Parámetros de la Simulación

Como ya se mencionó, previo a la simulación se deberán establecer parámetros necesarios tales como el tamaño muestral, el tipo de dominio o rejilla donde se simularán los datos, la función o funciones de tendencia tanto para la simulación de los datos como para la estimación de la tendencia (es decir, la especificación del modelo en los casos paramétricos), los parámetros de covarianza, entre otras.

Tipo de rejilla. Se considerará una partición equidistante del cuadrado $(0,1) \times (0,1)$. Cada lado del cuadrado se dividirá en 20 segmentos, con lo que se tendrá una rejilla de 400 puntos.

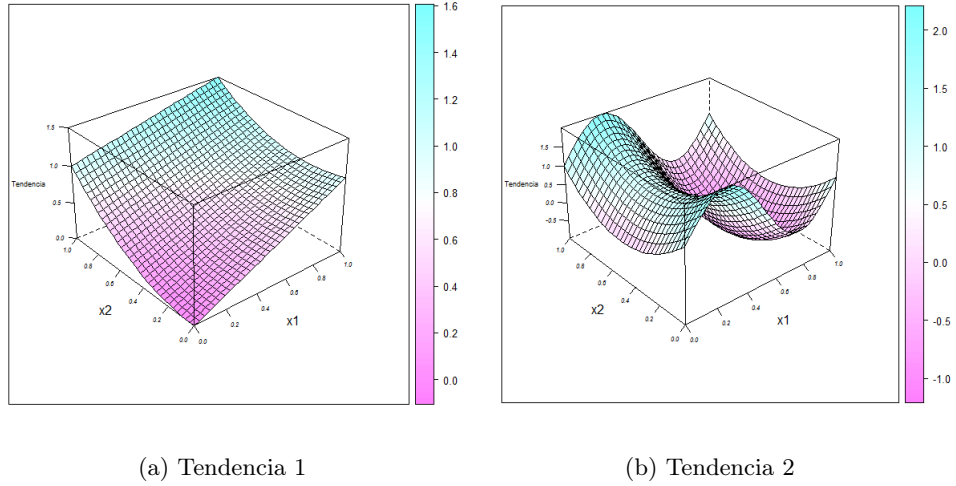
Modelo y parámetros del variograma. Se considerará, tanto en la simulación de los datos como en la estimación de la tendencia, el modelo Matern. Los parámetros teóricos de covarianza que se establecerán por defecto son

$$(0.8, 0.25, 0.2, 0.5), \quad (3.4.1)$$

que corresponden a la *meseta parcial*, *rango*, *nugget*, y κ , respectivamente. Debe notarse que el modelo Matern con $\kappa = 0.5$ equivale a considerar el modelo exponencial de covarianza. Asimismo, el rango que se considera es el propio de la función de covarianza exponencial, con lo que el rango efectivo en este caso es 3 veces este valor; es decir, 0.75.

Estimación del variograma. La estimación del variograma se realizará considerando el modelo Matern, con $\kappa = 0.5$. La máxima distancia hasta donde se calculará la disimilitud será $0.5\sqrt{2}$; es decir, la mitad de la diagonal de la rejilla. Con respecto al cálculo del variograma experimental, se considerará el estimador clásico de Matheron, y

Figura 3.4.1: Tendencias usadas en la simulación



en lo referente al ajuste de variograma teórico, se considerarán pesos iguales.¹

Tendencia. Se considerarán dos tipos de tendencia:

- Tendencia 1 (T1). $m[(x_1, x_2)] = x_1 + x_2^2 - 0.5x_1x_2$
- Tendencia 2 (T2). $m[(x_1, x_2)] = \sin(2\pi x_1) + 4(x_2 - 0.5)^2$
Nótese que los coeficientes de tendencia de T2 son (1,4).

La figura 3.4.1 muestra la apariencia de las tendencias T1 y T2.

Estimador no paramétrico. El estimador no paramétrico (2.3.3) requiere de una matriz H y de un kernel d -variado. Con respecto a la primera, en este trabajo se usará la venta óptima H_{opt} , misma que se obtiene al minimizar la ecuación (2.3.7). En cuanto al kernel, se utilizará el kernel de Epanechnikov, que en su versión multivariante se define como

$$K_{Epa}(\mathbf{x}) = \frac{2}{\pi} \max\{(1 - \|\mathbf{x}\|^2), 0\}.$$

¹Si bien es posible elegir los pesos de Cressie, que en cierta forma se prefieren sobre otros pesos, en el presente estudio de simulación se encontró la dificultad de que al implementar el algoritmo, en algunas ocasiones éste no convergía, dando lugar a valores de los coeficientes de dependencia exorbitantes. Una forma en que se corrigió este problema fue, en primer lugar, considerando pesos iguales en lugar de los de Cressie, y en segundo lugar, implementando límites razonables a los valores estimados de los coeficientes de dependencia. De hecho, se pudo detectar que la función 'variofit' de geoR tiene un error de código, toda vez que los límites para la meseta parcial los considera para el rango, y viceversa. Hay que mencionar que parte importante del tiempo dedicado a este trabajo se invirtió en detectar con precisión y corregir este tipo de errores de código.

Número de repeticiones. El número de repeticiones en cada simulación será $B = 1000$.

3.5. Plan de Simulación

El plan de simulación consiste en comparar por medio del MISE las estimaciones de la tendencia obtenidas con los estimadores paramétricos y no paramétricos descritos en la sección 3.3, teniendo en cuenta diferentes situaciones de dependencia espacial consistentes en:

- *Variar el nugget.*

Manteniendo fijos los demás parámetros de covarianza definidos en (3.4.1), se considerarán diferentes valores del nugget: 0.0, 0.2, 0.5.

En este caso el objetivo es observar el efecto de aumentar o disminuir la variación que se encuentra en intervalos de muestreo menores a los considerados. Esta es una primera forma de considerar variaciones en la dependencia de los datos, pues conforme el nugget aumenta, los datos se acercan a la independencia.

Es importante mencionar que los valores de la meseta parcial serán tales que la suma de ésta con el nugget deberá ser 1; es decir, que lo que se está variando en cada caso es la proporción de varianza nugget.

- *Variar el rango.*

Otra forma de modificar la dependencia espacial es variando el rango de la función de covarianza. A diferencia del caso de variación del nugget, aquí no se trata de que aumente o disminuya la variación en escalas menores a las consideradas en el muestreo, sino que, directamente, se trata de un aumento o disminución en el alcance de la dependencia espacial.

Al igual que en el caso anterior, se mantienen fijos los demás parámetros de covarianza en (3.4.1) mientras se consideran diferentes valores del rango: 0.1, 0.25, 0.5.

Estas situaciones de dependencia espacial se considerarán de acuerdo a cada una de los siguientes casos:

1. *Especificación adecuada del modelo (T2-T2).*

Consiste en simular los datos con el modelo paramétrico T2, y estimar la tendencia con el modelo paramétrico T2. De esta manera se estima la tendencia de la mejor forma en que puede hacerse paramétricamente; es decir, estimamos con el mismo modelo del cual provienen los datos.

2. *Especificación inadecuada del modelo (T2-T1).*

Consiste en simular los datos con el modelo paramétrico T2 y estimar

la tendencia con el modelo paramétrico T1. Así se estima la tendencia de los datos simulados mediante un modelo de tendencia diferente, como podría ocurrir en una situación real.

3. *Estimación no paramétrica (T2-NP)*. Consiste en calcular la H_{opt} para el modelo paramétrico T2 y con ella estimar la tendencia mediante (2.3.3).

Esta es una situación parecida al caso 1, sólo que en el caso no paramétrico. En condiciones habituales, con datos reales de los que no se sabe de qué modelo de tendencia provienen, se tendría que emplear aquella H que minimice la expresión (2.3.6). No obstante, en este trabajo se usará la H_{opt} por cuestiones de optimización de tiempo.

En cualquiera de los casos mencionados, lo que se quiere observar es el efecto de las diferentes situaciones de dependencia espacial en el MISE de las tendencias estimadas, así como el efecto en las estimaciones de los coeficientes de dependencia espacial.

Capítulo 4

Resultados

A continuación se muestran los resultados obtenidos en las simulaciones considerando las diferentes situaciones de dependencia espacial y los diferentes casos de especificación de la tendencia planteados en la sección 3.5.

4.1. Especificación Adecuada del Modelo

4.1.1. Variación de la proporción del nugget

MISE

Los resultados del cuadro 4.1.1 muestran que el sesgo cuadrático integrado de la estimación de la tendencia es prácticamente nulo en cualquiera de las situaciones de dependencia dadas por el valor del nugget, aunque ligeramente mayor ahí cuando se tiene nugget 0, que es cuando se está más lejos de la independencia (en este contexto de variación del nugget). Por lo anterior, el MISE representa prácticamente la varianza integrada de la estimación de la tendencia. Lo que se observa es que el método OLS tiene el mayor MISE de todos los métodos, y que va descendiendo conforme se avanza de OLS a GLSt, lo que representa una ganancia el uso de los métodos diferentes al OLS. Por otra parte, cuando se observan las distintas situaciones de depen-

Figura 4.1.1: $Sesgo^2$ integrado, Varianza integrada y MISE. T2-T2. Variación del nugget.

Método	$Sesgo^2$ Int.	Varianza Int.	MISE
	N00 N02 N05	N00 N02 N05	N00 N02 N05
OLS	0.0004 0.0001 0.0000	0.1670 0.1330 0.0851	0.1674 0.1331 0.0851
GLS1	0.0000 0.0000 0.0000	0.0748 0.0650 0.0473	0.0748 0.0650 0.0473
GLSe	0.0000 0.0000 0.0000	0.0745 0.0646 0.0476	0.0745 0.0646 0.0476
GLSt	0.0000 0.0000 0.0000	0.0697 0.0592 0.0426	0.0697 0.0592 0.0426

Cuadro 4.1: Estadísticos descriptivos de los coeficientes de tendencia estimados. Nugget 0.0

(a) $\beta_0 = 1$

Método	Media	Mediana	Error Estándar
OLS	1.0243	1.0303	0.3971
GLS1	1.0069	0.9965	0.2843
GLSe	1.0054	1.0016	0.2800
GLSt	1.0057	1.0063	0.2752

(b) $\beta_1 = 4$

Método	Media	Mediana	Error Estándar
OLS	3.9311	3.9337	2.4626
GLS1	3.9860	3.9640	1.5480
GLSe	3.9911	3.9797	1.5663
GLSt	3.9907	4.0132	1.4894

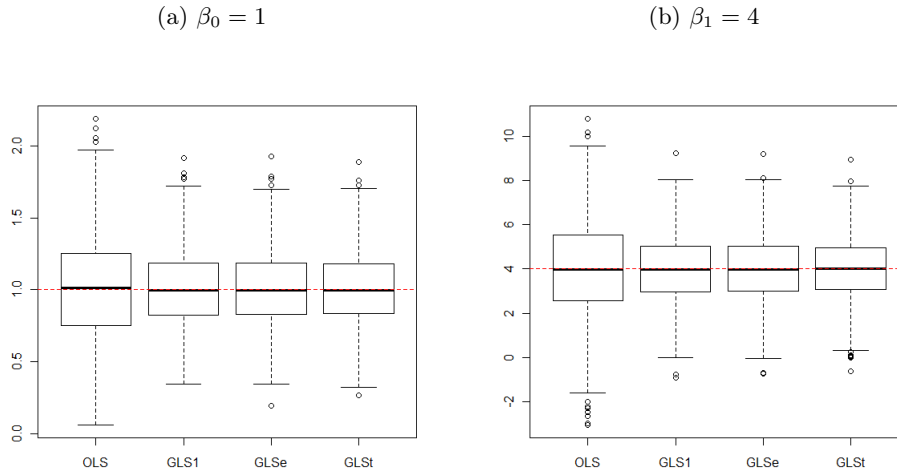
dencia dadas por el nugget se observa que conforme éste aumenta, tiende a disminuir el MISE, lo cual es entendible, pues al aumentar el nugget, los datos se encuentran más cerca de la independencia.

COEFICIENTES DE TENDENCIA

La correcta especificación del modelo permite en este caso tener las estimaciones de los coeficientes de tendencia, de las cuales se pueden observar algunos estadísticos descriptivos en el cuadro 4.1, relativo a nugget 0.0. Se puede observar cómo todos los métodos arrojan estimaciones de los coeficientes de tendencia prácticamente insesgadas, además, considerando la cercanía entre la media y la mediana, se trata de distribuciones simétricas.

Puede observarse también cómo al pasar del método OLS a GLSt, la estimación promedio se acerca más al valor del parámetro, al mismo tiempo que se gana precisión al obtener menores errores estándar. Concretamente, se tiene mayor precisión en la estimación de β_0 que de β_1 en cualquiera de los métodos, situación entendible debido a que la función de la que es coeficiente β_1 es más compleja que la función de la que es coeficiente β_0 .

Al pasar de nugget 0.0 a 0.2, y después a 0.5, en general, se observa el mismo fenómeno que en el cuadro anterior; a saber, que hay mayor precisión en la estimación de β_0 que de β_1 , y que la precisión mejora conforme se pasa del método OLS al GLSt, pero además, se observa que al aumentar el nugget se provoca una disminución en el sesgo del estimador y una mayor precisión en la estimación de ambos coeficientes en cualquiera de los métodos. De hecho, al pasar de nugget 0.2 a 0.5 se observó el mismo fenómeno: menor

Figura 4.1.2: Gráficos de caja de las estimaciones $\hat{\beta}_0$ y $\hat{\beta}_1$. N02

diferencia entre la media de las estimaciones y el valor del coeficiente, y menor error estándar en general, mismo que disminuye al pasar de OLS a GLSt. Para tener una idea de las distribuciones, se muestran los gráficos de caja de la estimaciones de los coeficientes de tendencia en el caso de nugget 0.2.

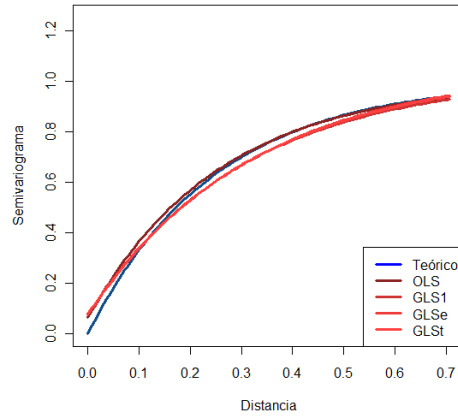
COEFICIENTES DE DEPENDENCIA

En cuanto a los coeficientes de dependencia sólo se tiene estimación de la meseta parcial, del rango y del nugget, toda vez que el parámetro de suavización κ ha permanecido constante por considerar un modelo Matern con $\kappa = 0.5$ tanto en la simulación de los datos como en el proceso de estimación. De manera general, se encuentra que la distribución de las estimaciones de estos parámetros es asimétrica, pudiendo distinguirse fácilmente entre la media y la mediana de las estimaciones. No obstante, las estimaciones promedio de los parámetros, ya sea en nugget 0.0, 0.2 o 0.5, son relativamente cercanas a los parámetros teóricos.

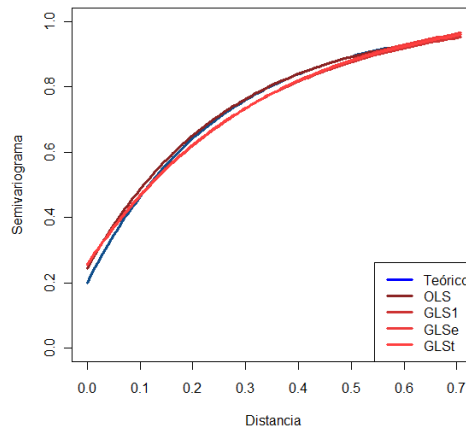
Lo anterior se verifica al observar los variogramas obtenidos con los parámetros de covarianza medios comparados con el variograma teórico en la figura 4.1.3. En general, puede observarse que en la región cercana al nugget hay una ligera sobreestimación, razón por la que el variograma en esa zona se observa por encima del teórico. Después, aproximadamente en la parte media del variograma, parece haber una subestimación del mismo, y por último se observa una región cercana al final del variograma en la que parece haber una cierta sobreestimación del mismo. Eso acusa que, en promedio, tiende a haber una sobre estimación tanto del nugget como del rango y la meseta.

Figura 4.1.3: Comparación del semivariograma teórico con los variogramas obtenidos mediante el valor medio de los parámetros de covarianza, según método.

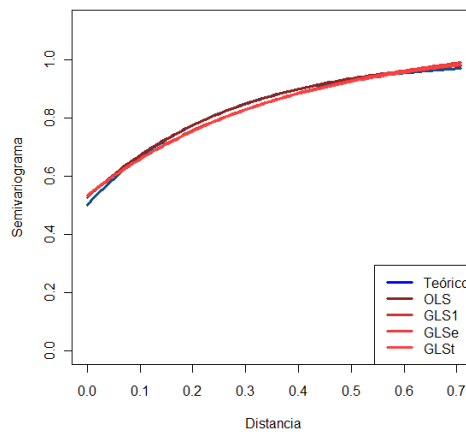
(a) Nugget 0.0



(b) Nugget 0.2



(c) Nugget 0.5



Cuadro 4.2: $Sesgo^2$ integrado, Varianza integrada y MISE. T2-T2. Variación del rango.

Método	$Sesgo^2_{int.}$			Varianza int.			MISE		
	R01	R025	R05	R01	R025	R05	R01	R025	R05
OLS	0.0000	0.0001	0.0001	0.0509	0.1330	0.2156	0.0509	0.1331	0.2157
GLS1	0.0000	0.0000	0.0000	0.0418	0.0650	0.0573	0.0418	0.0650	0.0573
GLSe	0.0000	0.0000	0.0000	0.0424	0.0646	0.0572	0.0424	0.0642	0.0572
GLSt	0.0000	0.0000	0.0000	0.0404	0.0592	0.0497	0.0404	0.0592	0.0497

4.1.2. Variación del rango

MISE

Los resultados del $sesgo^2$ integrado, varianza integrada y MISE considerando variación en el rango se encuentran en el cuadro 4.2. En general, se observa que el sesgo cuadrado integrado en cualquiera de los métodos y en cualquiera de los valores del rango es cero o prácticamente despreciable, aunque quizá, en el caso OLS, si se consideraran valores mayores del rango pudieran observarse valores algo mayores del sesgo. En cuanto a la varianza integrada puede observarse que conforme aumenta el rango tiende a aumentar la varianza integrada de manera notable, básicamente en el método OLS, pues lo que se observa en los otros métodos hace pensar que al aumentar el rango no hay cambios apreciables en la varianza integrada.

Ya que se observó que el sesgo es despreciable, el MISE en cada uno de los métodos es una consecuencia del valor de la varianza integrada y se tendrían los mismo comentarios que para la varianza.

En cuanto a la variación del MISE según método, puede observarse que éste disminuye notablemente al pasar de OLS a GLSt, siendo el MISE en los tres últimos métodos bastante similares y ya sin cambios importantes al pasar de uno a otro; en realidad, el cambio notorio se observa al pasar de OLS a GLS1.

COEFICIENTES DE TENDENCIA

En cuanto a la estimación de los coeficientes de tendencia, se pudo observar que en todos los métodos se obtienen distribuciones en las que la media y la mediana son bastante simétricas. Además, se observó también que usando el método OLS se obtiene mayor error que en los demás métodos. En cuanto al cambio en la precisión de las estimaciones, se pudo observar que al aumentar el rango aumentó la imprecisión de las mismas, al observarse un incremento en el error estándar.

Cuadro 4.3: Estadísticos descriptivos de las estimaciones de $\beta_1 = 4$. Rango 0.1 y 0.5.(a) $\beta_1 = 4$. Rango 0.1

Método	Media	Mediana	Error Estándar
OLS	3.9959	3.9670	1.2294
GLS1	3.9905	3.9838	1.0934
GLSe	3.9907	3.9980	1.1024
GLSt	3.9902	3.9553	1.0680

(b) $\beta_1 = 4$. Rango 0.5

Método	Media	Mediana	Error Estándar
OLS	4.0026	3.9662	3.1480
GLS1	4.0008	4.0367	1.4926
GLSe	4.0030	4.0270	1.5219
GLSt	3.9938	4.0167	1.3707

COEFICIENTES DE DEPENDENCIA

En cuanto a la estimación de los coeficientes de dependencia, en el caso de menor rango, 0.1, se encontró que en promedio se sobreestimó el efecto nugget, observándose los variogramas con los coeficientes promedio por encima del variograma teórico justo en el origen. Después, en la parte media de la gráfica, se observa que los variogramas estimados se encuentran por debajo del teórico, lo que significa que en promedio se sobreestima ligeramente el rango; no obstante, la meseta de los variogramas estimados parece coincidir con el variograma teórico, con lo que se puede decir que la estimación de la meseta fue aceptable.

En cambio, cuando se observan los variogramas en el caso de rango 0.5 se puede constatar que la estimación media del nugget fue aceptablemente correcta en prácticamente todos los métodos; sin embargo, se observa en la parte media que los variogramas estimados con los parámetros medios de dependencia sobrepasan al variograma teórico, lo que implica que en este caso se está subestimando el rango de la dependencia.

4.2. Especificación Inadecuada del Modelo

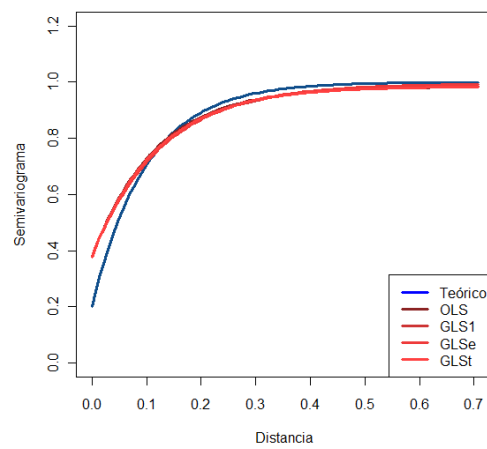
4.2.1. Variación de la proporción de nugget

MISE

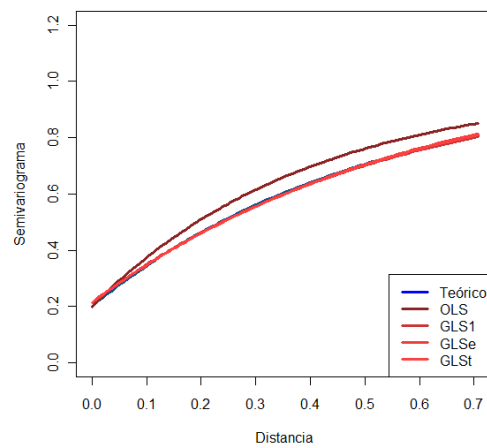
Los resultados del sesgo cuadrático integrado, varianza integrada y MISE se observan en el cuadro 4.4. A diferencia de lo que ocurrió en el caso donde

Figura 4.1.4: Comparación del variograma teórico con los variogramas obtenidos mediante el valor medio de los parámetros de covarianza, según método.

(a) Rango 0.1



(b) Rango 0.5



Cuadro 4.4: $Sesgo^2$ integrado, Varianza integrada y MISE. T2-T1. Variación del nugget.

Método	$Sesgo^2_{int}$			Varianza int			MISE		
	N00 N02 N05	N00 N02 N05	N00 N02 N05	N00 N02 N05	N00 N02 N05	N00 N02 N05	N00 N02 N05	N00 N02 N05	
OLS	0.3163 0.3162 0.3162	0.3917 0.3150 0.2003	0.7079 0.6312 0.5165						
GLS1	0.4704 0.4516 0.4377	0.2887 0.2401 0.1609	0.7590 0.6918 0.5986						
GLSe	0.6615 0.6628 0.6829	0.2942 0.2548 0.1846	0.9557 0.9176 0.8675						
GLSt	0.5788 0.5391 0.4840	0.2757 0.2309 0.1556	0.8545 0.7700 0.6396						

Cuadro 4.5: Promedio de los coeficientes de covarianza, según método. Nugget 0.0

Método	meseta parcial	rango	nugget
OLS	1.0760	0.1687	0.0155
GLS1	1.4488	0.2893	0.0570
GLSe	1.7540	0.3605	0.0765
GLSt	1.6471	0.3538	0.0773

se consideraba especificación adecuada del modelo, en el presente caso resulta difícil en primera instancia distinguir si hay algún patrón en cuanto al comportamiento del MISE y sus componentes en los diferentes métodos usados.

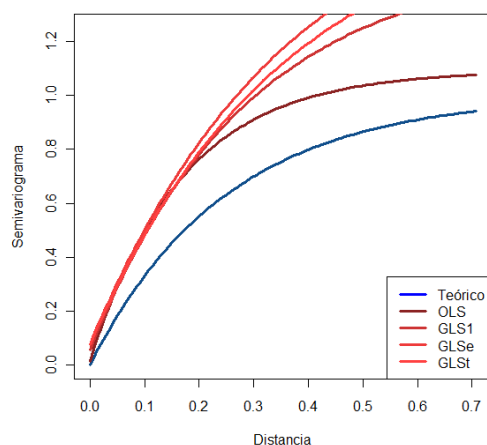
Hablando de manera general, puede decirse que aumentar el valor del nugget tiene como consecuencia un decremente en el valor del MISE; es decir, que los datos, al acercarse más a la independencia, favorecen estimaciones globales con menor error cuadrático medio integrado. Asimismo, se observa que la varianza disminuye al aumentar el nugget y el sesgo cuadrático no parece experimentar grandes cambios. Eso significa que los cambios observados en el MISE al aumentar el rango se deben fundamentalmente al aumento de la varianza integrada. En cualquier caso, es evidente que en el presente caso de mala especificación de la tendencia las estimaciones son mucho peores comparadas con el caso de especificación adecuada del modelo, en el que se tienen sesgos despreciables y varianzas integradas bastante menores a las obtenidas cuando no se especifica bien la tendencia.

COEFICIENTES DE DEPENDENCIA

En relación con los coeficientes de dependencia, se encontró que la mala especificación del modelo condujo a estimaciones bastante inexactas de los parámetros de covarianza, tal como se puede observar en el cuadro 4.5.

Como resultado de las malas estimaciones promedio de los coeficientes de covarianza, los gráficos de los variogramas estimados resultan bastante diferentes al variograma teórico, de tal manera que representan casi por

Figura 4.2.1: Comparación del variograma teórico con los variogramas obtenidos mediante el valor medio de los parámetros de covarianza, según método. Nugget 0.0



completo una estructura espacial de los datos diferente de la que realmente fueron simulados. A manera de ejemplo, puede verse en la gráfica 4.2.1 que si bien el nugget no fue problema grave en cuanto a su estimación media, si lo fueron de alguna manera el rango y, en especial, la meseta parcial. Eso significa que al haber especificado mal la tendencia, las estimaciones sobreestiman la varianza de los datos principalmente al alejarse del origen. En relación con los casos de variación de nugget, se observó igualmente que la estimación del nugget no fue el problema principal, sino el rango (incluso en mayor medida que en el caso de nugget 0.0) y la meseta, cuyos valores fueron todavía mayores que los observados en el cuadro 4.5, teniéndose de esa manera variogramas con mesetas aún más elevadas que las mostradas en la figura 4.2.1.

4.2.2. Variación del rango

MISE

Los comentarios sobre el MISE variando el rango son bastante similares a los obtenidos anteriormente en el caso de variación del nugget en el sentido de que se observan a simple vista valores del sesgo cuadrático integrado y de la varianza integrada que están lejos de considerarse propios de una buena estimación de la tendencia. No obstante, el aumento del rango implicó un aumento del MISE en todos los métodos, a diferencia de lo que ocurría cuando aumentaba el nugget.

La gráfica 4.2.2 permite observar de manera global los sesgos y varianzas

Cuadro 4.6: $Sesgo^2$ integrado, Varianza integrada y MISE.

(a) Variando el nugget

	$Sesgo^2$ int	Varianza int	MISE
Método	N00 N02 N05	N00 N02 N05	N00 N02 N05
NP	0.0523 0.0439 0.0312	0.4686 0.3867 0.2608	0.5209 0.4306 0.2926

(b) Variando el rango

	$Sesgo^2$ int	Varianza int	MISE
Método	R01 R025 R05	R01 R025 R05	R01 R025 R05
NP	0.0393 0.0439 0.0350	0.1853 0.3867 0.5475	0.2246 0.4306 0.5825

en la rejilla considerada.

Dada la forma de las superficies de tendencia, observables en la gráfica 3.4.1, puede entenderse que en el gráfico de $sesgo^2$ se observen específicamente esas zonas en donde el sesgo es casi nulo: es en estas zonas donde coinciden en buena medida ambas tendencias. La varianza, en cambio, tiende a aumentar conforme las posiciones espaciales se alejan del centro de la rejilla.

COEFICIENTES DE DEPENDENCIA

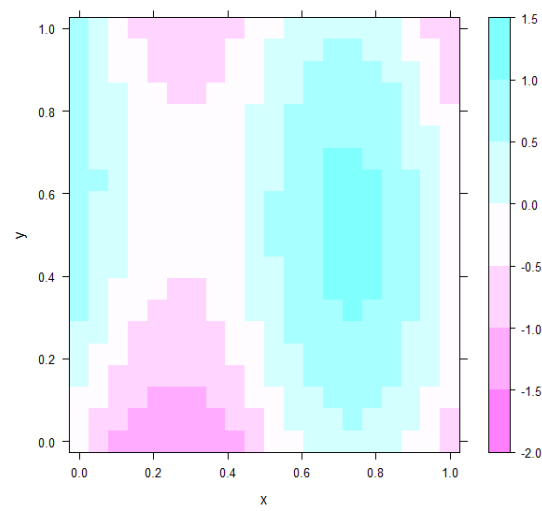
En cuanto a la estimación de los coeficientes de dependencia, la situación es básicamente la misma que en el caso de variación del nugget; es decir, que se tienen muy malas estimaciones en promedio de los parámetros de covarianza. En este caso, las estimaciones del nugget resultan un poco peor en el caso de rango 0.1 que en el caso de rango 0.5; a su vez, las estimaciones de la meseta parcial se tornan peores al aumentar el rango.

4.3. Estimación No Paramétrica

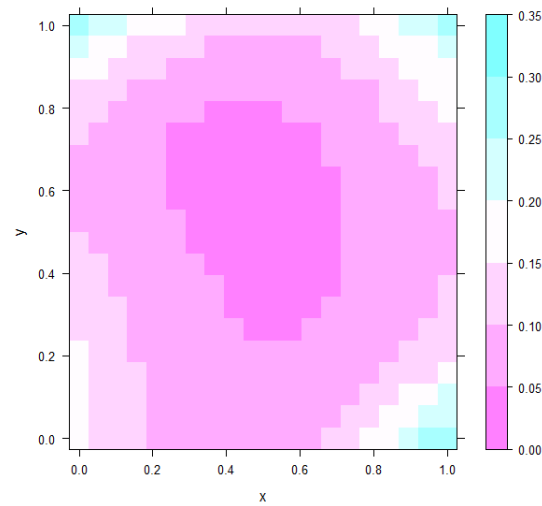
Los resultados de la estimación no paramétrica de la tendencia se muestran en el cuadro 4.6.

Escencialmente, se observa, al igual que en los casos anteriores, que el MISE aumenta conforme el rango aumenta, y disminuye conforme aumenta el nugget. Asimismo, al aumentar el nugget se observa una disminución tanto en el sesgo cuadrático como en la varianza; no así con el rango, que al aumentar su valor el sesgo cuadrático se mantiene aproximadamente igual mientras que es la varianza la que aumenta y la causante de que aumente el MISE.

Es interesante mencionar que los resultados del MISE obtenidos bajo la estimación paramétrica de la tendencia se encuentran en alguna zona intermedia entre los encontrados bajo estimación adecuada e inadecuada de la tendencia en los casos paramétricos.

Figura 4.2.2: $Sesgo^2$ y varianza obtenidos por el método GLSt. Rango 0.0(a) $Sesgo^2$ 

(b) Varianza



Cuadro 4.7: Matriz ventana H_{opt} , según situación de dependencia

(a) Nugget 0.0	(b) Nugget 0.2								
<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">0.287</td><td style="padding: 2px 10px;">0.000</td></tr> <tr><td style="padding: 2px 10px;">0.000</td><td style="padding: 2px 10px;">0.579</td></tr> </table>	0.287	0.000	0.000	0.579	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">0.273</td><td style="padding: 2px 10px;">0.000</td></tr> <tr><td style="padding: 2px 10px;">0.000</td><td style="padding: 2px 10px;">0.546</td></tr> </table>	0.273	0.000	0.000	0.546
0.287	0.000								
0.000	0.579								
0.273	0.000								
0.000	0.546								
(c) Nugget 0.5	(d) Phi 0.1								
<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">0.239</td><td style="padding: 2px 10px;">0.000</td></tr> <tr><td style="padding: 2px 10px;">0.000</td><td style="padding: 2px 10px;">0.505</td></tr> </table>	0.239	0.000	0.000	0.505	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">0.261</td><td style="padding: 2px 10px;">0.000</td></tr> <tr><td style="padding: 2px 10px;">0.000</td><td style="padding: 2px 10px;">0.533</td></tr> </table>	0.261	0.000	0.000	0.533
0.239	0.000								
0.000	0.505								
0.261	0.000								
0.000	0.533								
(e) Phi 0.5									
<table style="border-collapse: collapse; margin: auto;"> <tr><td style="padding: 2px 10px;">0.247</td><td style="padding: 2px 10px;">0.000</td></tr> <tr><td style="padding: 2px 10px;">0.000</td><td style="padding: 2px 10px;">0.529</td></tr> </table>		0.247	0.000	0.000	0.529				
0.247	0.000								
0.000	0.529								

Con respecto a la matriz ventana H_{opt} utilizada en estos casos, éstas se muestran en el cuadro 4.7. Puede observarse cómo al pasar de una situación de mayor dependencia a menor dependencia, como cuando aumenta el nugget, los valores de la diagonal de la matriz H_{opt} tienden a disminuir.

Este efecto resulta interesante debido a que en situaciones de menor dependencia, cada dato aporta más información por sí mismo, no siendo requerido abarcar mayor espacio en la rejilla para calcular el estimador local lineal. En cambio, cuando hay mayor dependencia, cada dato aporta menos información por sí mismo, con lo que se hace necesario abarcar una mayor área para poder estimar adecuadamente la tendencia.

Capítulo 5

Conclusiones

En este trabajo se consideraron diferentes escenarios de dependencia espacial por medio de la variación de parámetros de covarianza tales como el rango y el nugget. En cada caso, se consideraron diferentes formas de especificación de la tendencia a fin de comparar las estimaciones de la tendencia por diferentes métodos paramétricos y no paramétricos.

De manera resumida, el MISE de las estimaciones no paramétricas indica que éstas no fueron tan buenas como en los casos de especificación adecuada de la tendencia de manera paramétrica, pero se obtuvieron mejores resultados que en el caso en el que no se especificaba la tendencia correctamente. Considerando que en la práctica se suele desconocer la tendencia verdadera de los datos, se concluye que a menos que se especifique la tendencia de manera adecuada mediante una función similar a la tendencia verdadera, no se puede usar con buenos resultados la estimación paramétrica de la tendencia. Para la mayoría de los casos en los que nada se sabe sobre la tendencia verdadera, los resultados sugieren apoyarse en la estimación no paramétrica de la tendencia.

Este trabajo deja abiertas varias líneas futuras. Una de ellas consiste en generalizar y realizar estimaciones no paramétricas del variograma, así como trabajar también en la corrección del sesgo en las estimaciones del mismo, obtenidas a partir de los residuos. Se deja abierta la puerta para la búsqueda de métodos más eficientes y menos costosos computacionalmente para seleccionar la matriz ventana H , por ejemplo de tipo plug-in.

Por último, se deja abierta la posibilidad de trabajar en la elaboración de un paquete estadístico que permita utilizar tanto métodos paramétricos como no paramétricos en la estimación de la tendencia, que sea además lo suficientemente flexible para emplearlo incluso cuando el propósito principal no es la estimación de la tendencia, sino la predicción vía kriging.

Bibliografía

- [1] P.J. Diggle y P.J. Ribeiro Jr. *Model-Based Geostatistics*. New York. Springer. 2007.
- [2] H. Wackernagel. *Multivariate Geostatistics. An introduction with applications*. New York. Springer. 2003.
- [3] M.A. Diaz Viera. “Geoestadística Aplicada”. Notas de curso. Instituto de Geofísica, Universidad Nacional Autónoma de México (UNAM). Instituto de Geofísica y Astronomía, CITMA. Cuba. 2002.
- [4] N. Cressie y M.O. Grondona. “A comparison of variogram estimation with covariogram estimation”, en *The art of statistical science*. KV. Mardia (Ed.), Wiley. 191-208. 1992.
- [5] N. Cressie. *Statistics for Spatial Data*. New York. Wiley. 1991.
- [6] A.G. Journel y C.J. Huijbregts. *Mining Geostatistics*. New York. Academic Press. 1978.
- [7] J.P. Chiles y P. Delfiner. *Geostatistics. Modeling Spatial Uncertainty*. New York. Wiley. 1999.
- [8] N. Cressie. “Fitting variogram models using weighted least squares.” *J. Math. Geol.* 17, 693-702. 1985
- [9] L. Wasserman. *All of Nonparametric Statistics*. New York. Springer. 2006.
- [10] M.P. Wand y M.C. Jones. *Kernel Smoothing*. London. Chapman & Hall. 1995
- [11] X.Liu. “Kernel Smoothing for spatially correlated data”. PhD. Thesis. Department of Statistics, Iowa State University. 2001.
- [12] J.D. Opsomer, Y. Wang e Y. Wang. “Nonparametric regression with correlated errors”. *Statistical Science*, 16, 134-153.

- [13] M. Francisco-Fernández y J.D. Opsomer. “Smoothing parameter selection methods for nonparametric regression with spatially correlated errors”. *The Canadian Journal of Statistics*. Vol 33, No.2. p279-295. 2005.
- [14] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford. Oxford University Press. 1997.
- [15] M. Armstrong. “Problems with universal kriging”. *Mathematical Geology*, 16, 101-108. 1984.
- [16] G. Matheron. “The theory of regionalized variables and its applications”. *Cahiers du Centre de Morphologie Mathématique*, 5. Fontainebleau. France. 1971.
- [17] S.P. Neuman y E.A. Jacobson. “Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels”. *Mathematical Geology*, 16, 499-521. 1984.
- [18] M.L. Stein. “Asymptotically efficient prediction of a random field with a misspecified covariance function”. *Annals of Statistics*, 16, 55-63. 1988.
- [19] P. J. Ribeiro Jr & P. J. Diggle. “geoR: a package for geostatistical analysis”. *R-NEWS*, 1(2):15-18. June, 2001.