



Universidade de Vigo

Trabajo Fin de Máster

---

# Estimación de tipo núcleo de la tendencia espacial

---

Natalia Pérez Veiga

Máster en Técnicas Estadísticas

Curso 2018-2019



## Propuesta de Trabajo Fin de Máster

<b>Título en galego:</b> Estimación de tipo núcleo da tendencia espacial
<b>Título en español:</b> Estimación de tipo núcleo de la tendencia espacial
<b>English title:</b> Kernel-type estimation of the spatial trend
<b>Modalidad:</b> Modalidad A
<b>Autor/a:</b> Natalia Pérez Veiga, Universidad de Santiago de Compostela
<b>Director/a:</b> María del Pilar García Soidán, Universidad de Vigo; Tomás R. Cotos Yáñez, Universidad de Vigo
<b>Breve resumen del trabajo:</b> Este trabajo analiza el comportamiento de una herramienta no paramétrica consistente para la estimación de la tendencia de un proceso espacial, tanto desde un punto de vista teórico como a través de un estudio de simulación.



Doña María del Pilar García Soidán, Catedrática de Escuela Universitaria de la Universidad de Vigo, don Tomás R. Cotos Yáñez, Profesor Contratado Doctor de la Universidad de Vigo, informan que el Trabajo Fin de Máster titulado

**Estimación de tipo núcleo de la tendencia espacial**

fue realizado bajo su dirección por don/doña Natalia Pérez Veiga para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 31 de Enero de 2019.

La directora:

El director:

Doña María del Pilar García Soidán

Don Tomás R. Cotos Yáñez

La autora:

Doña Natalia Pérez Veiga



# Índice general

<b>Resumen</b>	<b>IX</b>
<b>Prefacio</b>	<b>XI</b>
<b>1. Introducción a la estadística espacial</b>	<b>1</b>
1.1. Proceso espacial estacionario . . . . .	1
1.1.1. Tipos de estacionariedad e isotropía . . . . .	2
1.1.2. Variograma y covariograma . . . . .	3
1.2. Modelización de la dependencia espacial . . . . .	4
1.2.1. Estimación piloto del variograma . . . . .	4
1.2.2. Selección y ajuste de un modelo de variograma válido . . . . .	5
1.2.3. Diagnósis . . . . .	7
1.3. Predicción espacial . . . . .	8
1.3.1. Kriging simple . . . . .	8
1.3.2. Kriging ordinario . . . . .	9
1.3.3. Kriging universal . . . . .	10
<b>2. Estimación de la tendencia</b>	<b>13</b>
2.1. Introducción . . . . .	13
2.2. Estimador NW de la tendencia . . . . .	14
2.3. Estimador de tipo núcleo en dos etapas . . . . .	16
2.4. Selección de la ventana . . . . .	19
<b>3. Estudio de simulación</b>	<b>21</b>
3.1. Datos simulados . . . . .	21
3.2. Selección de las ventanas . . . . .	22
3.3. Comparación con el estimador de NW . . . . .	29
3.4. Predicción . . . . .	32
3.5. Aplicación a datos reales . . . . .	42
<b>A. Código de R</b>	<b>45</b>
<b>Bibliografía</b>	<b>55</b>





# Resumen

## Resumen en español

La aproximación de la tendencia espacial se puede llevar a cabo utilizando técnicas paramétricas aunque, en ese caso, la mala especificación del modelo seleccionado podría afectar a la eficiencia del estimador resultante. Este riesgo se puede evitar recurriendo a algún mecanismo no paramétrico para caracterizar la tendencia, adaptado al contexto espacial, como es la metodología de tipo núcleo.

En este trabajo se propone un estimador no paramétrico de la tendencia de un proceso espacial, que es asintóticamente insesgado y consistente. Se trata de un método en dos etapas que permite aproximar la tendencia, en cada posición de la región de observación, mediante una media ponderada de las tendencias estimadas en las posiciones observadas, cuyos pesos tienen en cuenta las distancias entre las localizaciones muestreadas y la localización objetivo. Por una parte, se prueban las propiedades teóricas del estimador, bajo determinadas hipótesis. Por otra, se investiga su comportamiento en la práctica, mediante estudios de simulación, en los que también se analiza el problema de selección de las ventanas y su utilidad para llevar a cabo la predicción del proceso en localizaciones no muestreadas. Finalmente se presenta una aplicación a datos reales, que pone de manifiesto el potencial del nuevo estimador.

## English abstract

The approximation of the spatial trend can be addressed through parametric techniques although, in that case, the misspecification of the selected model could affect the efficiency of the resulting estimator. This risk can be avoided by using a nonparametric mechanism to characterize the trend, adapted to the spatial setting, such as the kernel-type methodology.

In this work, a nonparametric estimator is proposed to approximate the trend of a spatial process, which is asymptotically unbiased and consistent. It is a two-step method that allows us to approximate the trend, at each location of the observation region, through a weighted average of the estimated trends at the observed sites, whose weights take into account the distances between the sampled locations and the target location. On the one hand, the theoretical properties of the estimator are proved, under certain hypotheses. On the other, its practical performance is checked, through simulation studies, in which the problem of bandwidth selection is analyzed, as well as its usefulness to perform the prediction of the process at unsampled locations. Finally, an application to real data is presented to show the potential of the new estimator.



# Prefacio

Cuando las observaciones del fenómeno de interés están asociadas a una posición del espacio es necesario emplear técnicas estadísticas específicas que utilicen de manera adecuada la información que proporciona dicha posición, ya que, generalmente, no se satisface la hipótesis de que las observaciones del fenómeno han sido tomadas bajo idénticas condiciones y de manera independiente. En este contexto se hace necesaria la implementación de técnicas particulares que presuponen que datos próximos presentarán propiedades parecidas y entre datos alejados no habrá prácticamente correlación.

La *estadística espacial* o *geoestadística* es el conjunto de técnicas que tienen como objetivo la determinación de la estructura de autocorrelación entre datos asociados a una posición del espacio y la predicción en nuevas localizaciones. Las primeras muestras de estudio de datos espaciales datan del siglo XVII a través de mapas de datos, pero no sería hasta la década de 1980 cuando Matheron acuña el término geoestadística para referirse al conjunto de aplicaciones de la teoría de las variables regionalizadas a la estimación de depósitos minerales. Así, Matheron concebía la geoestadística como una ciencia híbrida entre las matemáticas, geología, ingeniería de minas y estadística, cuyos modelos estadísticos incluían la dependencia a pequeña y gran escala que presentan los datos espaciales.

Como ya se ha mencionado, uno de los objetivos principales de la estadística espacial es reconstruir sobre toda la región de observación un fenómeno que se ha observado en un conjunto finito de localizaciones. Es de nuevo Matheron quien formaliza y generaliza el uso de las técnicas de predicción *kriging* para este propósito a partir de una serie de metodologías desarrolladas por el geólogo D. G. Krige con el fin de hacer predicciones en minas de oro en Sudáfrica en el año 1951. En dichas técnicas *kriging* una mala especificación de la distribución, de la tendencia o de la estructura de segundo orden pueden conducir a predicciones pobres. Pero este no es el único punto donde es importante el papel de la estimación de la tendencia en estadística espacial.

Teniendo esto en cuenta, en este Trabajo de Fin de Máster se propone un método para estimar la tendencia a partir de técnicas de estimación no paramétrica. En particular, se recurre a métodos tipo núcleo, que han sido ampliamente estudiados en distintos contextos de la literatura estadística, en problemas habituales como la estimación de la función de densidad o de regresión.

El Capítulo 1 recoge una breve introducción de los fundamentos básicos de la estadística espacial, tanto definiciones y resultados teóricos como metodología para realizar predicciones o modelar la dependencia espacial. El Capítulo 2 se centra en presentar diferentes estimadores de la tendencia espacial de tipo núcleo y aborda el problema habitual en este tipo de estimadores de la selección del parámetro ventana. Por último, los Capítulos 3 y 4 recogen los resultados de los estudios numéricos para comparar las técnicas introducidas en un estudio de simulación y su aplicación a un caso real.



# Capítulo 1

## Introducción a la estadística espacial

En este capítulo se introducen los conceptos básicos en estadística espacial, así como las técnicas para el modelado de la estructura de dependencia y para la predicción espacial.

### 1.1. Proceso espacial estacionario

Se define un *proceso estocástico espacial* (también llamado variable aleatoria georeferenciada o variable regionalizada) como un proceso aleatorio distribuido en el espacio. Se representará por

$$\{Z(s)/s \in D\},$$

donde  $s$  denota una localización y  $D$  la región de observación en el espacio euclídeo  $d$ -dimensional,  $\mathbb{R}^d$ . Por tanto, para cada  $s$ ,  $Z(s)$  es una variable aleatoria unidimensional o multidimensional.

Sobre la región de observación,  $D$ , se consideran  $n$  localizaciones o posiciones muestrales,  $\{s_1, \dots, s_n\} \subset D$ , para las cuales se ha observado el proceso. Así,  $\{Z(s_1), \dots, Z(s_n)\}$  representa al conjunto de variables aleatorias y  $\{z(s_1), \dots, z(s_n)\}$  una realización.

Lo más habitual es que un proceso aleatorio presente una parte determinista y otra aleatoria,  $Z(s) = \mu(s) + Y(s)$ , donde  $\mu(s)$  es la parte determinista (no aleatoria) y representa los cambios o evolución a gran escala e  $Y(s)$  es la componente aleatoria (errática) y representa el comportamiento local o evolución a pequeña escala. Generalmente se supone que  $\mu(s) = \mathbb{E}[Z(s)] \quad \forall s \in D$ . Una vez la tendencia ha sido eliminada, el residuo resultante es la variación a pequeña escala, en la que se pueden encontrar patrones de comportamiento que obedecen al entorno en el que se encuentran las localizaciones.

Como se verá más adelante, para la estimación del variograma o la obtención de predicciones se suele suponer alguna de las siguientes condiciones para la componente determinista: conocida, constante, se puede caracterizar a partir de los datos, se puede ajustar mediante una función polinómica o utilizando otros procedimientos (como promediar los datos más próximos).

En adelante usaremos la siguiente notación:

- $\mathbb{E}[Z(s)] = \mu(s)$  que representa la tendencia del proceso,
- $Var[Z(s)] = \sigma^2(s)$  que representa su varianza y
- $\mathbb{P}(Z(s_1) \leq z_1, \dots, Z(s_n) \leq z_n) = F_{s_1, \dots, s_n}(z_1, \dots, z_n)$  su función de distribución.

Se dice que un proceso estocástico  $\{Z(s)/s \in D\}$  es *gaussiano* si para cada  $n \in \mathbb{N}$  y para cada conjunto de  $n$  localizaciones  $\{s_1, \dots, s_n\}$  se verifica que su distribución  $F_{s_1, \dots, s_n}$  es normal o gaussiana.

Dado que se dispone de una única realización discreta de la variable regionalizada  $Z(s)$ , deben asumirse ciertas hipótesis para poder realizar inferencia sobre los datos, en particular suele recurrirse a hipótesis de estacionariedad.

### 1.1.1. Tipos de estacionariedad e isotropía

Una función aleatoria es *estrictamente estacionaria* si es invariante ante traslaciones, es decir, si la distribución asociada a las localizaciones se mantiene al desplazar todas las localizaciones en la misma dirección la misma distancia:

$$F_{s_1, \dots, s_k}(z_1, \dots, z_k) = F_{s_1+t, \dots, s_k+t}(z_1, \dots, z_k), \forall t \in \mathbb{R}^d, \forall k \in \mathbb{N}.$$

La estacionariedad estricta es poco habitual en el contexto de datos reales. Además, en los procesos espaciales se suele suponer que existen los momentos de primer y segundo orden. Por tanto, es frecuente asumir otros tipos de estacionariedad menos restrictivos como la estacionariedad de segundo orden.

Un proceso es *estacionario de segundo orden* si se verifica que la tendencia del proceso es constante a lo largo de toda la región de observación y la covarianza entre dos observaciones es función de la separación (dirección y distancia) entre sus localizaciones, esto es, si se cumplen

- $\mathbb{E}[Z(s)] = \mu \quad \forall s \in D$  y
- $Cov[Z(s), Z(s')] = C(s - s') \quad \forall s, s' \in D.$

La función  $C$  se denomina *covariograma* o *función de covarianza*. Si la segunda condición se sustituye por que la covarianza entre dos observaciones solo es función de la distancia entre sus localizaciones y no de la dirección ( $Cov(Z(s), Z(s')) = C(\|s - s'\|)$ , donde  $\|\cdot\|$  es la norma euclídea), se dice que el proceso estacionario de segundo orden es *isotrópico*. En caso contrario, el proceso estacionario de segundo orden es *anisotrópico*.

Bajo las hipótesis de estacionariedad de segundo orden se verifica que  $\sigma^2(s) = Var[Z(s)] = Cov[Z(s), Z(s)] = C(0) = \sigma^2 \quad \forall s \in D$ , luego la varianza del proceso estacionario de segundo orden es independiente de la localización espacial  $s$  y finita. Pero muchos procesos disponen de la capacidad infinita de dispersión, de modo que no existe la varianza de la correspondiente variable regionalizada. En este contexto resulta de utilidad la condición de estacionariedad intrínseca.

Un proceso es *intrínsecamente estacionario* si verifica

- $\mathbb{E}[Z(s)] = \mu \quad \forall s \in D$  y
- $Var[Z(s) - Z(s')] = 2\gamma(s - s') \quad \forall s, s' \in D.$

En este caso, la función  $2\gamma$  se denomina *variograma* y  $\gamma$  es el *semivariograma*. De nuevo, el proceso intrínsecamente estacionario será *isotrópico* cuando  $Var(Z(s) - Z(s')) = 2\gamma(\|s - s'\|) \quad \forall s, s' \in D$  y *anisotrópico* en caso contrario.

La estacionariedad intrínseca no requiere que  $Var[Z(s)]$  sea independiente de la localización ni finita, con lo que la clase de procesos intrínsecamente estacionarios es más general que la clase de procesos estacionarios de segundo orden.

En el caso de procesos espaciales gaussianos la estacionariedad estricta y la estacionariedad de segundo orden son equivalentes, ya que todo proceso estrictamente estacionario con momentos de primer y segundo orden finitos es a su vez estacionario de segundo orden. Por otro lado, la estacionariedad de segundo orden y la estacionariedad estricta son equivalentes si  $\sigma^2(s) = \sigma^2 < \infty, \forall s \in D$ .

### 1.1.2. Variograma y covariograma

En la Sección anterior se han definido el variograma y el covariograma, a continuación se estudian sus propiedades y la relación que existe entre ellos.

Si el proceso  $Z(s)$  es estacionario de segundo orden, entonces  $Cov[Z(s), Z(s')] = C(s-s') \quad \forall s, s' \in D$  o, equivalentemente,

$$C(t) = Cov[Z(s), Z(s+t)] \quad \forall s \in D \text{ y } \forall t \in \mathbb{R}^d,$$

de donde se deducen las siguientes propiedades:

- $Var[Z(s)] = \sigma^2(s) = C(0) = \sigma^2, \forall s \in D.$
- $C$  es una función simétrica:  $C(t) = C(-t), \forall t \in \mathbb{R}^d.$
- $C$  es una función definida positiva:  $\sum \sum a_i a_j C(s_i - s_j) \geq 0, \forall s_1, \dots, s_n \in D \text{ y } \forall a_1, \dots, a_n \in \mathbb{R}.$

Relacionado con la función de covarianza está el *correlograma*  $\rho$ , definido como  $\rho(t) = C(t)/C(0).$

Por otro lado, si el proceso  $Z(s)$  es intrínseco, entonces  $Var[Z(s) - Z(s')] = \mathbb{E}[(Z(s) - Z(s'))^2] = 2\gamma(s-s') \quad \forall s, s' \in D$  o, equivalentemente,

$$\gamma(t) = (1/2)\mathbb{E}[(Z(s) - Z(s+t))^2] \quad \forall s \in D \text{ y } \forall t \in \mathbb{R}^d,$$

de donde se deduce que:

- $\gamma(0) = 0$
- $\gamma(t) \geq 0, \forall t \in \mathbb{R}^d$
- $\gamma$  es una función simétrica:  $\gamma(t) = \gamma(-t), \forall t \in \mathbb{R}^d$
- $\gamma$  es una función condicionalmente definida negativa:  
 $\sum \sum a_i a_j \gamma(s_i - s_j) \leq 0, \forall s_1, \dots, s_n \in D \text{ y } \forall a_1, \dots, a_n \in \mathbb{R} / \sum a_i = 0$
- $\lim_{\|t\| \rightarrow \infty} \frac{\gamma(t)}{\|t\|^2} = 0$

La relación entre la función de covarianza y el semivariograma se tiene si el proceso estacionario es intrínseco y  $Var[Z(s)] = \sigma^2 < \infty$ :

$$\begin{aligned} \gamma(t) &= (1/2)Var[Z(s) - Z(s+t)] = (1/2)(Var[Z(s)] + Var[Z(s+t)] - 2Cov[Z(s), Z(s+t)]) = \\ &= (1/2)2(\sigma^2 - Cov[Z(s), Z(s+t)]) = \sigma^2 - C(t) \end{aligned}$$

En cuanto a las características del variograma, si  $\gamma$  está acotado y existe  $\lim_{\|t\| \rightarrow \infty} \gamma(t)$ , dicho límite se denomina *umbral* o *meseta*.

Si ocurre que  $c_0 = \lim_{\|t\| \rightarrow 0} \gamma(t) \neq 0$ , al valor  $c_0$  se le llama *efecto pepita*. La diferencia  $\sigma^2 - c_0$  recibe el nombre de *meseta parcial* o *umbral parcial*.

Por último, el *rango* o *alcance*, si existe, es un valor real  $r$  tal que si  $\|t\| > r$ , entonces  $\gamma(t) = \sigma^2$ . Esto es equivalente a decir que las variables  $Z(s)$  y  $Z(s+t)$  son incorreladas. El rango no siempre existe, pero si el proceso es estacionario de segundo orden, siempre puede definirse el *rango asintótico* o *efectivo* como el valor real  $r'$  tal que si  $\|t\| \geq r'$ , entonces  $\gamma(t) \geq c_0 + 0.95(\sigma^2 - c_0)$ .

## 1.2. Modelización de la dependencia espacial

Para poder estudiar los procesos espaciales es necesario caracterizar la estructura de dependencia. Cuando se trata de procesos que presentan estacionariedad de segundo orden o intrínseca, esta caracterización puede reducirse a la estimación de la función de covarianza o del variograma, respectivamente.

Dado que los procesos intrínsecos son más generales y que existe una relación entre ellos cuando se tiene estacionariedad de segundo orden, en la práctica es más habitual el estudio del variograma. Además, a diferencia de la función de covarianza, la estimación del variograma no requiere del conocimiento de la media del proceso ni de su estimación.

Para la estimación del variograma es frecuente emplear las siguientes tres etapas:

1. Estimación piloto del variograma
2. Selección y ajuste de un modelo de variograma válido
3. Diagnóstico del modelo de variograma ajustado

### 1.2.1. Estimación piloto del variograma

Se realiza una aproximación inicial del variograma mediante algún estimador no paramétrico. Las opciones más frecuentes son:

- El estimador empírico o muestral (debido a Matheron, 1963), basado en el método de los momentos, que no es más que la media de los cuadrados de las diferencias entre observaciones:

$$\hat{\gamma}(t) = \frac{1}{2|N(t)|} \sum_{(i,j) \in N(t)} (Z(s_i) - Z(s_j))^2,$$

siendo  $t \in R^d$  y  $|N(t)|$  el número de pares en  $N(t) = \{(i, j) / s_i - s_j = t\}$ . Cuando existen pocos pares de datos cuya diferencia sea  $t$  puede sustituirse  $N(t)$  por una región de tolerancia  $A(t) = \{(i, j) / s_i - s_j \approx t\}$ .

Si los datos siguen una distribución normal, el método de los momentos coincide con el de máxima verosimilitud y, en ese caso,  $\hat{\gamma}$  proporcionaría una buena aproximación del semivariograma. En caso contrario es aconsejable usar un estimador robusto ya que será menos sensible a la presencia de datos atípicos.

- El estimador empírico robusto de Cressie y Hawkins

$$\hat{\gamma}(t) = \frac{1}{2B(t)} \left( \text{Mediana} \left\{ |Z(s_i) - Z(s_j)|^{1/2} / s_i - s_j \in N(t) \right\} \right)^4, \text{ siendo } B(t) = 0.457 + \frac{0.494}{|N(t)|}$$

El coeficiente  $B(t)$  evita el sesgo que introduce la mediana. El estimador de Cressie y Hawkins se considera óptimo en condiciones de normalidad, sin embargo, infravalora los datos atípicos.

- El estimador robusto calculado como una media de los datos reescalados:

$$\hat{\gamma}(t) = \frac{1}{2 \cdot 0.457} \left( \frac{1}{|N(t)|} \sum_{(i,j) \in N(t)} |Z(s_i) - Z(s_j)|^{1/2} \right)^4$$

Nótese que si el proceso espacial es isotrópico, en los estimadores anteriores basta cambiar  $N(t)$  por  $N^*(d) = \{(i, j) / \|s_i - s_j\| = d\}$ , ya que la variabilidad en los datos solo depende de la distancia entre las localizaciones y no de la dirección.



- Estimadores tipo núcleo, por ejemplo el presentado en García-Soidán et al. (2004) para un proceso intrínseco e isotrópico, que consiste en un promedio ponderado de las diferencias entre los valores observados al cuadrado:

$$\hat{\gamma}(d) = \frac{\sum_{i,j} w_{i,j}(d) (Z(s_i) - Z(s_j))^2}{2 \sum_{i,j} w_{i,j}(d)},$$

con  $w_{i,j}(d) \geq 0, \forall i, j$  y  $\sum w_{i,j}(d) > 0$ . Según los pesos que se consideren se obtiene:

- Estimador empírico o muestral si  $w_{i,j}(d) = I_{\|s_i - s_j\|=d} = \begin{cases} 1 & \text{si } \|s_i - s_j\| = d \\ 0 & \text{en otro caso} \end{cases}$
- Estimador de tipo núcleo de Nadaraya-Watson (véase García-Soidán et al., 2004) si

$$w_{i,j}(d) = K\left(\frac{\|s_i - s_j\| - d}{h}\right)$$

- Lineal local (véase García-Soidán et al., 2003) si

$$w_{i,j}(d) = K\left(\frac{\|s_i - s_j\| - d}{h}\right) \sum_{k,l} K\left(\frac{\|s_k - s_l\| - d}{h}\right) (\|s_k - s_l\| - d) (\|s_k - s_l\| - \|s_i - s_j\|).$$

Tanto el estimador de Nadaraya-Watson como el lineal local son asintóticamente insesgados. La ventaja adicional del estimador lineal local es que proporciona un estimador directo de la primera derivada del variograma, lo cual es particularmente útil para la selección de un modelo paramétrico válido.

- Estimador de vecinos más próximos de Yu et al. (2006), también bajo las hipótesis de estacionariedad intrínseca e isotropía

$$\hat{\gamma}(d) = \frac{\sum_{i<j} \frac{1}{\delta_0 (\|s_i - s_j\|)} K\left(\frac{d - \|s_i - s_j\|}{\delta \delta_0 (\|s_i - s_j\|)}\right) (Z(s_i) - Z(s_j))^2}{\sum_{i<j} \frac{1}{\delta_0 (\|s_i - s_j\|)} K\left(\frac{d - \|s_i - s_j\|}{\delta \delta_0 (\|s_i - s_j\|)}\right)},$$

donde  $\delta > 0$  y  $\delta_0$  son una constante y una función de suavizado.

En general, esta primera aproximación no paramétrica del variograma no será válida dado que el estimador resultante no verifica la propiedad de ser condicionalmente definido negativo. Dicha condición es necesaria para la predicción al garantizar la existencia de solución de las ecuaciones kriging y que no se obtengan estimaciones negativas del error cuadrático medio de predicción.

### 1.2.2. Selección y ajuste de un modelo de variograma válido

En primer lugar se selecciona una familia paramétrica de variogramas válidos para caracterizar la estructura de dependencia. Las familias más comunes son la exponencial, esférica, gaussiana o de Matern, que se recogen a continuación. Aunque existen otros como el modelo de efecto pepita, apropiado en ausencia de correlación espacial; el modelo lineal, que es poco útil en la práctica por no estar acotado; el modelo potencial o exponencial-potencial; el modelo oscilatorio, útil para procesos con un comportamiento periódico donde existe una sucesión entre zonas ricas y pobres. En la notación utilizada en las parametrizaciones  $c_0 > 0$  representa el efecto pepita y, en caso de existir,  $c_1 > 0$  el umbral parcial y  $c_2 > 0$  el rango o alcance.

$$\begin{aligned}
\text{Esférico: } \gamma_{c_0, c_1, c_2}(d) &= \begin{cases} 0, & \text{si } d = 0 \\ c_0 + c_1 \left( 1.5 \frac{d}{c_2} - 0.5 \left( \frac{d}{c_2} \right)^3 \right), & \text{si } 0 < d < c_2 \\ c_0 + c_1, & \text{si } d \geq c_2 \end{cases} \\
\text{Exponencial: } \gamma_{c_0, c_1, c_2}(d) &= \begin{cases} 0, & \text{si } d = 0 \\ c_0 + c_1 \left( 1 - \exp \left( -\frac{3d}{c_2} \right) \right), & \text{si } d > 0 \end{cases} \\
\text{Gaussiano: } \gamma_{c_0, c_1, c_2}(d) &= \begin{cases} 0, & \text{si } d = 0 \\ c_0 + c_1 \left( 1 - \exp \left( -\frac{3d^2}{c_2^2} \right) \right), & \text{si } d > 0 \end{cases} \\
\text{Matern: } \gamma_{c_0, c_1, c_2, \nu}(d) &= \begin{cases} 0, & \text{si } d = 0 \\ c_0 + c_1 \left( 1 - \frac{1}{2^\nu \Gamma(\nu)} \left( \frac{d}{c_2} \right)^\nu K_\nu \left( \frac{d}{c_2} \right) \right), & \text{si } d > 0 \end{cases}
\end{aligned}$$

donde  $K_\nu$  es la función de Bessel de segunda clase y orden  $\nu$  modificada. Para  $\nu = 0.5$  el modelo de Matern coincide con el exponencial y para  $\nu = \infty$  resulta el modelo gaussiano.

Después se ajustan los parámetros a partir de la estimación piloto del variograma, empleando técnicas clásicas adaptadas a este contexto (máxima verosimilitud; máxima verosimilitud restringida; mínimos cuadrados ordinarios, ponderados o generalizados). Para la aplicación en la práctica de estos métodos se suele recurrir a un mecanismo iterativo que requiere unos valores iniciales generalmente seleccionados a ojo a partir del estimador piloto.

A continuación se detalla el procedimiento de mínimos cuadrados ponderados, que es el que se usará en las simulaciones del Capítulo 3.

Se supone que para ciertos valores  $d_i$ ,  $1 \leq i \leq k$ , se han obtenido las estimaciones del semivariograma  $\hat{\gamma}(d_i)$  a partir de un estimador piloto. Siguiendo las recomendaciones sugeridas por Journel y Huijbregts (1978), solo deben considerarse en el ajuste saltos menores o iguales que la mitad del máximo salto (es decir,  $\|d_i\| \leq \frac{1}{2} \max\{\|s_k - s_i\|\}$ ) y tales que el número de aportaciones a cada estimación sea por lo menos de 30, es decir,  $|N^*(d_i)| \geq 30$ , donde  $N^*(d_i)$  representa el número de pares a una distancia  $d_i$ . El método de ajuste por mínimos cuadrados trata de encontrar  $c_0$ ,  $c_1$  y  $c_2$  tales que minimicen

$$(\hat{\gamma} - \gamma_{c_0, c_1, c_2})^T V_{c_0, c_1, c_2} (\hat{\gamma} - \gamma_{c_0, c_1, c_2}), \quad (1.1)$$

donde  $\hat{\gamma} = (\hat{\gamma}(d_1), \dots, \hat{\gamma}(d_k))^T$  y  $\gamma_{c_0, c_1, c_2} = (\gamma_{c_0, c_1, c_2}(d_1), \dots, \gamma_{c_0, c_1, c_2}(d_k))$ . En función de la forma de la matriz  $V_{c_0, c_1, c_2}$  se obtienen distintos criterios para el ajuste:

- Mínimos cuadrados ordinarios, donde la matriz  $V_{c_0, c_1, c_2}$  es igual a la identidad de tamaño  $k \times k$  y el problema se reduce a obtener  $c_0$ ,  $c_1$  y  $c_2$  minimizando  $(\hat{\gamma} - \gamma_{c_0, c_1, c_2})^T (\hat{\gamma} - \gamma_{c_0, c_1, c_2})$ .
- Mínimos cuadrados ponderados, con la matriz  $V_{c_0, c_1, c_2}$  de la forma

$$V_{c_0, c_1, c_2} = \text{diag}(w_1, \dots, w_k),$$

donde los pesos  $w_i$  son iguales o inversamente proporcionales a las varianzas de  $\hat{\gamma}(d_i)$ . Como este último término es desconocido se puede proceder de dos maneras. Bajo la hipótesis de que el proceso es gaussiano y de que las estimaciones son incorreladas se puede aproximar la varianza del estimador como

$$\text{Var}[\hat{\gamma}(d_i)] \approx \frac{\gamma(d_i)}{|N^*(d_i)|},$$

siendo  $|N^*(d_i)|$  el número de elementos a distancia  $d_i$ . Por tanto, se toman los pesos

$$w_i \approx \frac{|N^*(d_i)|}{(\gamma_{c_0, c_1, c_2}(d_i))^2}.$$

La otra alternativa es seguir un proceso iterativo. En un primer paso se obtienen estimaciones iniciales de los parámetros  $(c_0^{(1)}, c_1^{(1)}, c_2^{(1)})$  tomando  $w_i = 1$ ,  $i = 1, \dots, k$ , (es decir, realizando el ajuste por mínimos cuadrados ordinarios). El segundo paso consistiría en construir los nuevos pesos

$$w_i = \frac{|N^*(d_i)|}{\left(\gamma_{c_0^{(1)}, c_1^{(1)}, c_2^{(1)}}(d_i)\right)^2}$$

y minimizar la expresión (1.1) para obtener  $(c_0^{(2)}, c_1^{(2)}, c_2^{(2)})$ . Se repite este proceso hasta alcanzar la convergencia.

- Mínimos cuadrados generalizados, donde, en lugar de considerar una matriz diagonal, se toma  $\gamma_{c_0, c_1, c_2}$  igual o proporcional a la inversa de la matriz de varianzas covarianzas de  $\hat{\gamma}$ , cuyos términos dependen nuevamente de los parámetros a estimar.

### 1.2.3. Diagnosis

Utilizando el método de validación cruzada se lleva a cabo la diagnosis del variograma ajustado,  $\gamma_{c_0, c_1, c_2}$ , para comprobar que los datos originales se adaptan al variograma obtenido. Para esto es necesario utilizar un método de predicción, cuestión que se aborda en la siguiente sección.

El método de validación cruzada consiste en dejar un dato fuera,  $Z(s_i)$ , y utilizar los restantes para construir un predictor de  $Z(s_i)$  y un estimador del error cuadrático medio de predicción, que se denotarán como  $\hat{Z}_{-i}(s_i)$  y  $\hat{\sigma}_{-i}^2(s_i)$ , para  $i = 1, \dots, n$ . La proximidad de los predictores a los valores de la muestra permitirá diagnosticar la validez del modelo ajustado. Para ello se puede proceder de distintas formas:

- Analizar si la media de los errores tipificados

$$\frac{1}{n} \sum_{i=1}^n \frac{(Z(s_i) - \hat{Z}_{-i}(s_i))}{\hat{\sigma}_{-i}(s_i)}$$

es próxima a 0.

- Analizar si error cuadrático medio adimensional

$$\left[ \frac{1}{n} \sum_{i=1}^n \frac{(Z(s_i) - \hat{Z}_{-i}(s_i))^2}{\hat{\sigma}_{-i}^2(s_i)} \right]^{1/2}$$

es próximo a 1 o dista de 1 significativamente.

- Comprobar que el error cuadrático medio

$$\frac{1}{n} \sum_{i=1}^n (Z(s_i) - \hat{Z}_{-i}(s_i))^2$$

es pequeño. El problema de esta medida de error es que asigna igual peso a todos los datos y no tiene en cuenta las posiciones espaciales. En general los errores son mayores en los puntos más alejados del resto de los datos, lo cual puede tener un efecto dominante en la media global.

- Herramientas gráficas como la representación del histograma o boxplot de los residuos tipificados,  $(Z(s_i) - \hat{Z}_{-i}(s_i))\hat{\sigma}_{-i}^{-1}(s_i)$ , donde la presencia de outliers significaría un mal ajuste.

### 1.3. Predicción espacial

Uno de los objetivos principales de la estadística espacial es, dada una muestra formada por un número finito de datos, obtener una estimación de los valores del proceso original en toda la región de observación. Para dar solución a este problema, como ya se ha mencionado previamente, Matheron introduce en 1963 el uso de las técnicas kriging, denominado así como reconocimiento del trabajo desarrollado en el contexto espacial por el geólogo sudafricano Krige.

Los métodos kriging son algoritmos de predicción de mínimo error en media cuadrática que tienen en cuenta la estructura de segundo orden del proceso. Supongamos entonces que  $\{Z(s)/s \in D\}$  es un proceso espacial, al que se le suele imponer algún tipo de estacionariedad (de segundo orden o intrínseca). Si disponemos de las observaciones  $Z(s_1), \dots, Z(s_n)$  se trata de construir un predictor de  $Z(s)$ , que se denotará por  $\hat{Z}(s)$  tal que

- sea insesgado,  $\mathbb{E}[\hat{Z}(s)] = \mu(s)$  y
- minimice el error cuadrático medio de predicción,  $\mathbb{E}[(\hat{Z}(s) - Z(s))^2]$ , también designado como varianza de la predicción,  $\sigma^2$ .

Se puede diferenciar entre kriging lineal y kriging no lineal, en función del tipo de predictor. En el caso del kriging lineal se trata de construir un predictor de tipo

$$\hat{Z}(s) = \sum_{i=1}^n \lambda_i Z(s_i).$$

Se distinguen los siguientes tipos de kriging lineal:

- El kriging simple, que supone que la media del proceso,  $\mu(s)$  es conocida para todo  $s \in D$ .
- El kriging ordinario, en el que se supone que la media del proceso es desconocida pero constante,  $\mu(s) = \mu$  para todo  $s \in D$ .
- El kriging universal, donde se supone que la media del proceso, desconocida y no constante, puede modelizarse como combinación lineal de funciones conocidas,  $\mu(s) = \sum_{k=0}^K f_k(s) a_k \quad \forall s \in D$ , siendo  $a_i \in \mathbb{R}$  parámetros desconocidos.

A continuación se desarrollan los detalles de los tres tipos de kriging lineal. Nótese que el kriging simple presenta una situación utópica ya que la tendencia del proceso no suele ser conocida. Por su parte, el kriging ordinario es un caso particular del kriging universal donde  $K = 0$  y  $f_0(s) \equiv 1 \quad \forall s \in D$ .

#### 1.3.1. Kriging simple

Se supone que el proceso  $Z(s)$  admite una descomposición de la forma  $Z(s) = \mu(s) + Y(s)$  donde la función de tendencia se supone conocida y donde  $Y(s)$  es un proceso estacionario de segundo orden de media 0 y función de covarianza  $C$  conocida (o estimada) de modo que  $C(t) = \text{Cov}[Z(s), Z(s+t)] = \text{Cov}[Y(s), Y(s+t)]$ . Se desea construir un predictor lineal

$$\hat{Y}(s) = \sum_{i=1}^n \lambda_i Y(s_i),$$

que, en este caso, sería equivalente a construir un predictor  $\hat{Z}(s) = \mu(s) + \sum_{i=1}^n \lambda_i (Z(s_i) - \mu(s_i))$  verificando  $\mathbb{E}[\hat{Y}(s)] = 0$ , condición que se verifica de manera trivial y no aporta ninguna restricción, y los valores  $\lambda_i$  se obtienen de modo que minimicen

$$\begin{aligned} \mathbb{E} \left[ \left( \hat{Y}(s) - Y(s) \right)^2 \right] &= \mathbb{E} \left[ \left( \sum_{i=1}^n \lambda_i Y(s_i) - Y(s) \right)^2 \right] = \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s) + C(0) \end{aligned} \quad (1.2)$$

Derivando con respecto a  $\lambda_i$  e igualando a cero en la ecuación (1.2) se obtiene que

$$2 \sum_{j=1}^n \lambda_j C(s_i - s_j) - 2C(s_i - s) = 0 \Leftrightarrow \sum_{j=1}^n \lambda_j C(s_i - s_j) = C(s_i - s)$$

De forma matricial, las expresiones anteriores se pueden resumir en  $\Sigma \lambda = b$ , con solución  $\lambda = \Sigma^{-1} b$  siempre que  $\sigma$  sea no singular, siendo

$$\Sigma = \begin{pmatrix} C(0) & C(s_1 - s_2) & \cdots & C(s_1 - s_n) \\ C(s_2 - s_1) & C(0) & \cdots & C(s_2 - s_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(s_n - s_1) & C(s_n - s_2) & \cdots & C(0) \end{pmatrix}, b = \begin{pmatrix} C(s_1 - s) \\ C(s_2 - s) \\ \vdots \\ C(s_n - s) \end{pmatrix}, \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Además, la varianza de la predicción es  $\sigma_{KS}^2(s) = \sigma^2 - \lambda^T b$ .

### 1.3.2. Kriging ordinario

Supongamos ahora que el proceso  $Z(s)$  admite una descomposición de la forma  $Z(s) = \mu + Y(s)$ , donde la media del proceso  $\mu$  es desconocida pero se supone constante y donde  $Y(s)$  es un proceso estacionario de segundo orden de media 0 y función de covarianza  $C$ . De nuevo, se desea construir un predictor lineal de la forma

$$\hat{Z}(s) = \sum_{i=1}^n \lambda_i Z(s_i),$$

verificando la condición de insesgadez y minimizando el error cuadrático medio de predicción, es decir:

- $\mathbb{E} \left[ \hat{Z}(s) \right] = \mu \Rightarrow \sum_{i=1}^n \lambda_i \mu = \mu \Rightarrow \sum_{i=1}^n \lambda_i = 1$
- Minimizar  $\mathbb{E} \left[ \left( \hat{Z}(s) - Z(s) \right)^2 \right]$  sujeto a  $\sum_{i=1}^n \lambda_i = 1$

$$\begin{aligned} \mathbb{E} \left[ \left( \hat{Z}(s) - Z(s) \right)^2 \right] + m_0 \left( \sum_{i=1}^n \lambda_i - 1 \right) &= \\ \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s) + C(0) + m_0 \left( \sum_{i=1}^n \lambda_i - 1 \right), \end{aligned} \quad (1.3)$$

donde  $m_0$  es el multiplicador de Lagrange. Derivando con respecto a  $\lambda_i$  y  $m_0$  en la ecuación (1.3) e igualando a 0 se obtiene que

$$2 \sum_{j=1}^n \lambda_j C(s_i - s_j) - 2C(s_i - s) + m_0 = 0 \quad \text{y} \quad \sum_{i=1}^n \lambda_i - 1 = 0.$$

Tomando  $m = m_0/2$ , resulta

$$\sum_{j=1}^n \lambda_j C(s_i - s_j) - C(s_i - s) + m = 0 \quad \text{y} \quad \sum_{i=1}^n \lambda_i = 1.$$

De forma matricial las expresiones anteriores se podrían resumir como  $AX = B$ , con solución  $X = A^{-1}B$  siempre que  $A$  sea no singular, y se obtiene como varianza de predicción  $\sigma_{KO}^2(s) = \sigma^2 - X^T B$ , siendo

$$A = \begin{pmatrix} \Sigma & d \\ d^T & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b \\ 1 \end{pmatrix}, \quad X = \begin{pmatrix} \lambda \\ 1 \end{pmatrix} \quad \text{y} \quad d = (1, \dots, 1)^T.$$

La construcción de un predictor bajo la hipótesis de estacionariedad intrínseca es muy similar. Se obtendría un sistema de ecuaciones análogo donde en lugar de la función de covarianza  $C$  en la matriz  $\Sigma$  y en el vector  $b$  aparece el semivariograma  $\gamma$ . La diferencia aparece en la varianza de predicción resultante,  $\sigma_{KO}^2 = X^T B$ .

### 1.3.3. Kriging universal

Supongamos nuevamente que el proceso se puede descomponer como  $Z(s) = \mu(s) + Y(s)$ , donde  $Y(s)$  es un proceso estacionario de segundo orden de media 0 y función de covarianza  $C$  y donde la tendencia del proceso  $\mu(s)$  es desconocida pero se puede expresar como combinación lineal de ciertas funciones conocidas.

Bajo la hipótesis de estacionariedad de segundo orden, la condición de insesgadez establece que  $\mathbb{E}[\hat{Z}(s)] = \mu(s)$ . Como

$$\mathbb{E}[\hat{Z}(s)] = \mathbb{E}\left[\sum_{i=1}^n \lambda_i Z(s_i)\right] = \sum_{i=1}^n \lambda_i \mathbb{E}[Z(s_i)] = \sum_{i=1}^n \lambda_i \sum_{k=0}^K a_k f_k(s_i) \quad \text{y} \quad \mu(s) = \sum_{k=0}^K f_k(s) a_k,$$

basta imponer que  $\sum_{i=1}^n \lambda_i f_k(s_i) = f_k(s)$ .

Por otro lado, los valores de  $\lambda_i$  se obtienen minimizando  $\mathbb{E}\left[\left(\hat{Z}(s) - Z(s)\right)^2\right]$  sujeto a la restricción  $\sum_{i=1}^n \lambda_i f_k(s_i) = f_k(s)$ , lo que es equivalente a obtener los valores  $\lambda_i$  y  $m_k$  tales que minimicen:

$$\begin{aligned} & \mathbb{E}\left[\left(\hat{Z}(s) - Z(s)\right)^2\right] + \sum_{k=0}^K m_k \left(\sum_{i=1}^n \lambda_i f_k(s_i) - f_k(s)\right) = \\ & \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s) + C(0) + \sum_{k=0}^K m_k \left(\sum_{i=1}^n \lambda_i f_k(s_i) - f_k(s)\right) \end{aligned} \quad (1.4)$$

ya que

$$\begin{aligned} & \mathbb{E}\left[\left(\hat{Z}(s) - Z(s)\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^n \lambda_i Z(s_i) - Z(s)\right) \left(\sum_{j=1}^n \lambda_j Z(s_j) - Z(s)\right)\right] = \\ & \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \mathbb{E}[Z(s_i)Z(s_j)] - 2 \sum_{i=1}^n \lambda_i \mathbb{E}[Z(s_i)Z(s)] + \mathbb{E}[Z(s)^2] = \\ & \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_{i=1}^n \lambda_i C(s_i - s) + C(0) \end{aligned}$$

Derivando en la ecuación (1.4) respecto a  $\lambda_i$  y a  $m_k$  e igualando a 0 se tiene que

$$2 \sum_{j=1}^n \lambda_j C(s_i - s_j) - 2C(s_i - s) + \sum_{k=0}^K m_k f_k(s_i) = 0 \quad \text{y} \quad \sum_{i=1}^n \lambda_i f_k(s_i) - f_k(s) = 0.$$

Tomando  $m'_k = m_k/2$ ,

$$\sum_{j=1}^n \lambda_j C(s_i - s_j) + \sum_{k=0}^K m'_k f_k(s_i) = C(s_i - s) \quad \text{y} \quad \sum_{i=1}^n \lambda_i f_k(s_i) = f_k(s)$$

Expresado en forma matricial se trata de resolver el sistema  $AX = B$ , que tiene como solución  $X = A^{-1}B$ , siempre que  $A$  sea no singular. La varianza de predicción resulta  $\sigma_{KU}^2 = \sigma^2 - X^T B$ , donde

$$A = \begin{pmatrix} \Sigma & d_0 & \cdots & d_k \\ d_0^t & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_k^t & 0 & \cdots & 0 \end{pmatrix}, \quad X = \begin{pmatrix} \lambda \\ M \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} b \\ f_0(s) \\ \vdots \\ f_k(s) \end{pmatrix}.$$

De forma similar a la sección anterior, si considerásemos la hipótesis de estacionariedad intrínseca en el proceso  $Y(s)$  se podría construir un predictor kriging universal teniendo en cuenta la relación  $C(t) = \sigma^2 - \gamma(t)$ .

Por último, cabe mencionar algunas consideraciones sobre los predictores kriging, que son los mejores predictores lineales insesgados. Además, si el proceso es gaussiano el predictor lineal coincide con el mejor predictor posible. Por otro lado, son interpoladores exactos, de modo que si  $s = s_i$  para algún  $i$ , entonces  $\hat{Z}(s_i) = Z(s_i)$  y la varianza de predicción es 0.

Para la implementación de los métodos kriging son necesarios estimadores de la función de covarianza y del variograma, que se exigirá que sean válidos, es decir, que verifiquen la propiedad de ser definido positivo y condicionalmente definido negativo, respectivamente. Esto garantizará la existencia de soluciones kriging. En el caso de kriging simple garantiza que la matriz  $\Sigma$  es definida positiva.

La hipótesis de estacionariedad permite caracterizar de forma simple los pesos de las ecuaciones kriging que dependen de la función de covarianza o del variograma evaluados en los saltos (diferencias entre posiciones). Los predictores kriging son predictores puntuales, es decir, para cada posición  $s \in D$  deben calcularse los pesos  $\lambda_i$  específicos. El hecho de estimar la función de covarianza o variograma incrementa el error final de la predicción, así que éste será mayor que el que indica la varianza de predicción.

Es frecuente elegir los datos más próximos a la posición en la que se pretende obtener la predicción en lugar de considerar todas las observaciones, lo cual simplifica considerablemente el problema objeto de estudio. Además con este mecanismo se utilizan estimaciones más eficientes del variograma (o función de covarianza) al estar calculadas sobre distancias pequeñas. Se han sugerido varias alternativas para seleccionar dicho entorno: considerar entre 20 y 40 datos más próximos a la posición de predicción; utilizar los datos que se encuentren dentro de un radio de búsqueda en torno a la posición de interés; dividir la región de observación en bloques y considerar los datos correspondientes a los bloques más próximos a la posición de predicción.





## Capítulo 2

# Estimación de la tendencia

### 2.1. Introducción

Sea  $\{Z(s) \in \mathbb{R} : s \in D \subset \mathbb{R}^d\}$  un proceso estacionario espacial que admite una descomposición como

$$Z(s) = \mu(s) + Y(s),$$

donde  $\mu(s) = \mathbb{E}[Z(s)]$  representa la tendencia determinista y  $\{Y(s) \in \mathbb{R} : s \in D \subset \mathbb{R}^d\}$  es un proceso espacial intrínsecamente estacionario de media 0, es decir,  $\mathbb{E}[Y(s)] = 0$ .

Denotemos por  $Z(s_1), \dots, Z(s_n)$  los  $n$  datos observados del proceso estacionario en sus respectivas localizaciones,  $s_1, \dots, s_n$ . El objetivo es estimar la tendencia  $\mu(s)$  pues, como ya se ha comentado en el Capítulo 1, cuando la tendencia no es constante a lo largo de la región de observación se requiere su estimación para poder obtener las predicciones en localizaciones no muestreadas. Lo mismo puede ocurrir cuando se trata de aproximar características del proceso subyacente, como la función de distribución. En particular, García-Soidán y Menezes (2012) estudian este último caso y ponen de manifiesto la necesidad de estimar la tendencia del proceso para aplicar el kriging indicador cuando ésta no es constante.

En la literatura se han propuesto diferentes técnicas paramétricas para estimar la función de tendencia. Normalmente se asume un modelo paramétrico lineal (ver, por ejemplo, Neuman y Jacobson, 1984; Cressie 1986) o no lineal (ver, por ejemplo, Crujeiras y Van Keilegom, 2010), donde los parámetros de la tendencia se estiman usando procedimientos iterativos basados en mínimos cuadrados generalizados.

El inconveniente de emplear técnicas paramétricas es que la eficiencia del estimador resultante podría verse afectada por una posible mala especificación del modelo. Este problema puede solucionarse usando técnicas no paramétricas adaptadas al contexto de la estadística espacial. En este escenario, Francisco-Fernández y Vilar-Fernández (2001) estudian la estimación de la función de regresión y sus derivadas usando un ajuste polinómico local ponderado bajo la suposición de que el error procede de un proceso estocástico estrictamente estacionario. Además, Francisco-Fernández y Opsomer (2005) proponen un método de selección de ventana que adapta el criterio de validación cruzada generalizada al efecto de la correlación espacial. A partir de las técnicas desarrolladas en los dos artículos anteriormente mencionados, Francisco-Fernández (2012) utiliza el estimador de regresión lineal local para estimar la tendencia y después el variograma de datos de eventos sísmicos. Por otro lado, Menezes et al. (2010) estudian la estimación tipo núcleo de la predicción espacial.

En este trabajo se adapta el estimador clásico de tipo núcleo de Nadaraya-Watson (NW) a la situación de datos con dependencia espacial. Este estimador pondera cada observación del proceso por una cantidad que será mayor cuanto más cerca esté la nueva localización  $s$  de cada  $s_i$ . Una propuesta

alternativa vendrá dada por un método en dos etapas que también incluya la información de la diferencia entre los valores observados y no solo utilice la información de las localizaciones. Esta idea ya fue estudiada en García-Soidán y Menezes (2017) para la distribución del proceso espacial.

Este Capítulo se estructura como sigue: en la secciones 2.2 y 2.3 se presentan cada uno de los dos estimadores y se comprueba que son asintóticamente insesgados pero solo el segundo es asintóticamente consistente; en la sección 2.4 se aborda el problema de selección de los parámetros ventana y en la sección 2.5 se propone una aplicación de la estimación propuesta para la tendencia, que se utilizaría como alternativa a los estimadores paramétricos en la predicción de tipo kriging.

## 2.2. Estimador NW de la tendencia

La primera idea para construir un estimador tipo núcleo para  $\mu(s)$  conduce al siguiente promedio ponderado de los valores observados en las localizaciones muestreadas, de tal manera que cuanto más próximo esté  $s_i$  de  $s$ , más peso se le asigna a  $Z(s_i)$ :

$$\hat{\mu}(s) = \frac{\sum_{i=1}^n K_d\left(\frac{s-s_i}{h}\right) Z(s_i)}{\sum_{i=1}^n K_d\left(\frac{s-s_i}{h}\right)}, \quad (2.1)$$

donde  $K_d$  representa una función simétrica  $d$ -variante de tipo núcleo y  $h$  es el parámetro ventana.

El estimador (2.1) incorpora la información de la correlación entre los datos mediante las distancias entre las localizaciones.

Para obtener los términos dominantes del sesgo y la varianza de  $\hat{\mu}(s)$ , se asumirán las siguientes condiciones:

- (i)  $\mathbb{E}[Z(s)^2] < \infty, \forall s \in D$ .
- (ii)  $f_{t_1, t_2}(x_1, x_2)$  es continuamente diferenciable como función de  $t_i \in \mathbb{R}^d$  y  $x_i \in \mathbb{R}$ , para  $i = 1, 2$ .
- (iii)  $Z(\cdot)$  es  $\alpha$ -mixing, con  $\alpha(r) = O(r^{-a})$ , para  $r > 0$  y alguna constante  $a > 0$ .
- (iv)  $D = \beta D_0$ , para algún  $\beta = \beta_n \xrightarrow{n \rightarrow +\infty} +\infty$  y  $D_0 \subset \mathbb{R}^d$  acotado.
- (v)  $s_i = \beta u_i$ , para  $1 \leq i \leq n$ , donde  $u_1, \dots, u_n$  denota una realización de una muestra aleatoria de tamaño  $n$  obtenida de una función de densidad  $g_0$  considerada en  $D_0$ .
- (vi)  $K_d$  es una función de densidad  $d$ -variante, simétrica y con soporte compacto.
- (vii)  $\{h + \beta^{-1} + n^{-2}h^{-d}\beta^d\} \xrightarrow{n \rightarrow +\infty} 0$ .

Se verifica que el estimador  $\hat{\mu}$  es asintóticamente insesgado, es decir, que el sesgo de  $\hat{\mu}$  tiende a 0 cuando el tamaño muestral aumenta. De hecho, se tiene que:

$$\begin{aligned} \mathbb{E}[\hat{\mu}(s)] &= \mathbb{E}[\mathbb{E}[\hat{\mu}(s)/s_k, \forall k]] = \mathbb{E}\left[\mathbb{E}\left[\frac{\sum_i K_d\left(\frac{s-s_i}{h}\right) Z(s_i)}{\sum_i K_d\left(\frac{s-s_i}{h}\right)} \middle/ s_k, \forall k\right]\right] \\ &= \sum_i \mathbb{E}\left[\frac{K_d\left(\frac{s-s_i}{h}\right) \mu(s_i)}{\sum_i K_d\left(\frac{s-s_i}{h}\right)}\right] \end{aligned}$$

Para probar esta propiedad y las posteriores que aparezcan en este capítulo se van a tomar como referencia las demostraciones del estimador de la distribución que aparecen en el trabajo de García-Soidán

y Menezes (2017), que a su vez siguen un mecanismo similar al utilizado para probar el Teorema 3.1 de Hall, Fisher y Hoffman (1994). En adelante, el símbolo  $\approx$  indica que sólo se presentará la parte dominante del término posterior.

En base a ello se verifica que:

$$\mathbb{E} [\hat{\mu}(s)] \approx \frac{A_1(s)}{A_2(s)} \approx \mu(s),$$

donde

$$\begin{aligned} A_1(s) &= \int K_d \left( \frac{s-\beta u}{h} \right) \mu(\beta u) g_0(u) du \\ &\stackrel{(a)}{\approx} h^d \beta^{-d} \int K_d(v) g_0 \left( \frac{s-hv}{\beta} \right) \mu(s-hv) dv \approx h^d \beta^{-d} g_0(0) \mu(s) \\ A_2(s) &= \int K_d \left( \frac{s-\beta u}{h} \right) g_0(u) du \\ &\stackrel{(a)}{\approx} h^d \beta^{-d} \int K_d(v) g_0 \left( \frac{s-hv}{\beta} \right) dv \approx h^d \beta^{-d} g_0(0). \end{aligned}$$

En las igualdades (a) se realizó el cambio de variable  $v = \frac{s-\beta u}{h}$ . También se usó que  $g_0$  es la densidad de las  $u_i$  y que  $\int K_d(v) dv = 1$  por la condición (vi).

Luego, el sesgo de  $\hat{\mu}(s)$  es asintóticamente nulo ya que Sesgo  $[\hat{\mu}(s)] = \mathbb{E} [\hat{\mu}(s)] - \mu(s) \approx 0$ .

En cuanto a la varianza, teniendo en cuenta que  $\mathbb{E} [Z(s_i)^2 / s_k, \forall k] = \mu(s_i)^2 + C(0)$  y que  $\mathbb{E} [Z(s_i)Z(s_{i'}) / s_k, \forall k] = C(s_i - s_{i'}) + \mu(s_i)\mu(s_{i'})$  se puede probar que

$$\begin{aligned} \mathbb{E} [\hat{\mu}(s)^2] &= \mathbb{E} [\mathbb{E} [\hat{\mu}(s)^2 / s_k, \forall k]] \approx \\ &\approx \sum_i \mathbb{E} \left[ \frac{K_d \left( \frac{s-s_i}{h} \right)^2 (\mu(s_i)^2 + C(0))}{\left( \sum_i K_d \left( \frac{s-s_i}{h} \right) \right)^2} \right] + \sum_i \sum_{i'} \mathbb{E} \left[ \frac{K_d \left( \frac{s-s_i}{h} \right) K_d \left( \frac{s-s_{i'}}{h} \right) (C(s_i - s_{i'}) + \mu(s_i)\mu(s_{i'}))}{\left( \sum_i K_d \left( \frac{s-s_i}{h} \right) \right)^2} \right] \\ &\approx \frac{B_1(s)}{nA_2(s)^2} + \frac{B_2(s)}{A_2(s)^2} \approx \frac{g_0(0) h^d \beta^{-d} (\mu(s)^2 + C(0)) \int K_d(v)^2 dv}{n(h^d \beta^{-d} g_0(0))^2} + \frac{h^{2d} \beta^{-2d} g_0(0)^2 (C(0) + \mu(s)^2)}{(h^d \beta^{-d} g_0(0))^2} \approx C(0) + \mu(s)^2 \end{aligned}$$

donde

$$\begin{aligned} B_1(s) &= \int K_d \left( \frac{s-\beta u}{h} \right)^2 (\mu(\beta u)^2 + C(0)) g_0(u) du \\ &= h^d \beta^{-d} \int K_d(v)^2 g_0 \left( \frac{s-hv}{\beta} \right) (\mu(s-hv)^2 + C(0)) dv \approx g_0(0) h^d \beta^{-d} (\mu(s)^2 + C(0)) \int K_d(v)^2 dv \\ B_2(s) &= \int \int K_d \left( \frac{s-\beta u_1}{h} \right) K_d \left( \frac{s-\beta u_2}{h} \right) (C(\beta u_1 - \beta u_2) + \mu(\beta u_1) + \mu(\beta u_2)) g_0(u_1) g_0(u_2) du_1 du_2 \\ &\approx h^{2d} \beta^{-2d} g_0(0)^2 (C(0) + \mu(s)^2). \end{aligned}$$

Por tanto,  $\hat{\mu}(s)$  no proporciona un estimador consistente de la tendencia, ya que su varianza no tiende necesariamente a 0 cuando el tamaño muestral crece puesto que:

$$\text{Var} [\hat{\mu}(s)] = \mathbb{E} [\hat{\mu}(s)^2] - (\mathbb{E} [\hat{\mu}(s)])^2 \approx C(0) \neq 0.$$

### 2.3. Estimador de tipo núcleo en dos etapas

Para resolver el problema de consistencia que plantea el estimador de NW de la tendencia, estudiado en la sección anterior, se propone como alternativa la aplicación de un método de tipo núcleo que consta de dos etapas. Así, se comienza por utilizar los datos disponibles para estimar la tendencia en las localizaciones observadas. Los valores resultantes se usan a continuación para obtener un estimador de la tendencia aplicable a todas las localizaciones.

En primer lugar se aproxima la tendencia en cada localización muestreada,  $s_i$ :

$$\tilde{\mu}_1(s_i) = \frac{\sum_{j=1}^n L_1 \left( \frac{Z(s_i) - Z(s_j)}{h_i} \right) Z(s_j)}{\sum_{j=1}^n L_1 \left( \frac{Z(s_i) - Z(s_j)}{h_i} \right)}$$

donde  $L_1$  es una función simétrica de tipo núcleo univariante y  $h_i$  es el parámetro ventana, que puede ser diferente para cada localización.

En la segunda etapa se obtiene una media ponderada de los valores resultantes de la aproximación de la tendencia en cada localización, donde los pesos incorporan la dependencia espacial:

$$\tilde{\mu}(s) = \frac{\sum_{i=1}^n K_d \left( \frac{s - s_i}{h} \right) \tilde{\mu}_1(s_i)}{\sum_{i=1}^n K_d \left( \frac{s - s_i}{h} \right)} = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{K_d \left( \frac{s - s_i}{h} \right) L_1 \left( \frac{Z(s_i) - Z(s_j)}{h_i} \right) Z(s_j)}{\sum_{i'=1}^n K_d \left( \frac{s - s_{i'}}{h} \right) \sum_{j'=1}^n L_1 \left( \frac{Z(s_{i'}) - Z(s_{j'})}{h_{i'}} \right)}, \quad (2.2)$$

donde  $K_d$  es una función núcleo  $d$ -variante y  $h$  es el parámetro ventana global.

La idea del estimador (2.2) es sustituir en la ecuación (2.1) el término  $Z(s_i)$  por un estimador consistente de  $\mu(s_i)$ . El resultado es una media ponderada de los datos disponibles del proceso, cuyos pesos tiene en cuenta la separación de la localización objetivo de cada posición muestreada  $s_i$  y, a su vez, la diferencia entre cada valor observado y  $Z(s_i)$ .

Con un procedimiento similar al de la sección anterior trataremos de demostrar que esta nueva versión sí proporciona un estimador consistente, bajo las condiciones (i), (ii'), (iii)-(vi), (vii') y (viii), donde

(ii')  $f_{t_1, t_2, t_3, t_4}(x_1, x_2, x_3, x_4)$  es continuamente diferenciable como función de  $t_i \in \mathbb{R}^d$  y  $x_i \in \mathbb{R}$ , para todo  $i \leq 4$ .

(vii')  $L_1$  es una función de densidad univariante simétrica y con soporte compacto.

(viii)  $\{h + h_i + \beta^{-1} + n^{-3} (\sum_i h_i)^{-1} h^{-d} \beta^d\} \xrightarrow{n \rightarrow +\infty} 0$ .

Empecemos comprobando que el sesgo de  $\tilde{\mu}$  tiende a 0 cuando el tamaño muestral crece, es decir, Sesgo  $[\tilde{\mu}(s)] = \mathbb{E}[\tilde{\mu}(s)] - \mu(s)$ .

$$\begin{aligned} \mathbb{E}[\tilde{\mu}(s)] &= \mathbb{E}[\mathbb{E}[\tilde{\mu}(s) / s_k, \forall k]] = \\ &= \sum_i \sum_j \mathbb{E} \left[ \frac{K_d \left( \frac{s - s_i}{h} \right)}{\sum_i K_d \left( \frac{s - s_i}{h} \right)} \mathbb{E} \left[ \frac{L_1 \left( \frac{Z(s_i) - Z(s_j)}{h_i} \right) Z(s_j)}{\sum_j L_1 \left( \frac{Z(s_i) - Z(s_j)}{h_i} \right)} \middle/ s_k, \forall k \right] \right] \\ &\approx \sum_i \sum_j \mathbb{E} \left[ \frac{K_d \left( \frac{s - s_i}{h} \right)}{\sum_i K_d \left( \frac{s - s_i}{h} \right)} \frac{C_1(s_i, s_j)}{nC_2(s_i, s_j)} \right] \approx \frac{C_3(s)}{C_4(s)} \approx \int z \frac{C_5(s, z)}{C_6(s)} dz, \end{aligned}$$

donde

$$\begin{aligned}
C_1(s_i, s_j) &= \int \int L_1\left(\frac{z_1 - z_2}{h_i}\right) z_2 f_{s_i, s_j}(z_1, z_2) dz_1 dz_2 \approx h_i \int z f_{s_i, s_j}(z, z) dz \\
C_2(s_i, s_j) &= \int \int L_1\left(\frac{z_1 - z_2}{h_i}\right) f_{s_i, s_j}(z_1, z_2) z_1 dz_2 \approx h_i \int f_{s_i, s_j}(z, z) dz \\
C_3(s) &= (\sum_i h_i) \int \int \int K_d\left(\frac{s - \beta u_1}{h}\right) z f_{\beta u_1, \beta u_2}(z, z) g_0(u_1) g_0(u_2) dz du_1 du_2 \\
&\approx (\sum_i h_i) h^d \beta^{-d} g_0(0) \int \int z f_{s, \beta u}(z, z) g_0(u) dz du = (\sum_i h_i) h^d \beta^{-d} g_0(0) \int z C_5(s, z) dz \\
C_4(s) &= (\sum_i h_i) \int \int \int K_d\left(\frac{s - \beta u_1}{h}\right) f_{s, \beta u_2}(z, z) g_0(u_1) g_0(u_2) dz du_1 du_2 \approx \\
&\approx (\sum_i h_i) h^d \beta^{-d} g_0(0) \int \int f_{s, \beta u}(z, z) g_0(u) dz du = (\sum_i h_i) h^d \beta^{-d} g_0(0) C_6(s) \\
C_5(s, z) &= \int f_{s, \beta u}(z, z) g_0(u) du \\
C_6(s) &= \int \int f_{s, \beta u}(z, z) g_0(u) dz du
\end{aligned}$$

En este punto se puede notar que  $C_5(s, z)$  es la densidad bivalente de  $(Z(s) - Z(\beta U), Z(\beta U))$  en  $(0, z)$  y  $C_6(s)$  es la densidad univalente de  $Z(s) - Z(\beta U)$  en 0, donde  $U$  denota una variable aleatoria con densidad  $g_0$ . Luego,  $\frac{C_5(s, z)}{C_6(s)}$  es igual a la densidad de  $Z(\beta U)$  en  $z$ , condicionada a que  $Z(s) - Z(\beta U) = 0$ . Es decir,  $\frac{C_5(s, z)}{C_6(s)}$  es la densidad de  $Z(s)$  en  $z$ , lo que permite concluir que

$$\mathbb{E}[\tilde{\mu}(s)] \approx \int z \frac{C_5(s, z)}{C_6(s)} dz = \int z f_s(z) dz = \mu(s)$$

Entonces se tiene que Sesgo  $[\tilde{\mu}(s)]$  tiende a 0.

En cuanto a la varianza de  $\tilde{\mu}(s)$ , se puede descomponer

$$\mathbb{E}[\tilde{\mu}(s)^2] = \mathbb{E}[\mathbb{E}[\tilde{\mu}(s)^2 / s_k, \forall k]] \approx D_1(s) + D_2(s),$$

con

$$\begin{aligned}
D_1(s) &= \sum_i \sum_j \mathbb{E} \left[ \frac{K_d\left(\frac{s-s_i}{h}\right)^2}{\left(\sum_i K_d\left(\frac{s-s_i}{h}\right)\right)^2} \mathbb{E} \left[ \frac{L_1\left(\frac{Z(s_i)-Z(s_j)}{h_i}\right)^2 Z(s_j)^2}{\left(\sum_j L_1\left(\frac{Z(s_i)-Z(s_j)}{h_i}\right)\right)^2} \middle/ s_k, \forall k \right] \right] \\
D_2(s) &= \sum_{i, j, i', j'} \mathbb{E} \left[ \frac{K_d\left(\frac{s-s_i}{h}\right) K_d\left(\frac{s-s_{i'}}{h}\right)}{\left(\sum_i K_d\left(\frac{s-s_i}{h}\right)\right)^2} \mathbb{E} \left[ \frac{L_1\left(\frac{Z(s_i)-Z(s_j)}{h_i}\right) L_1\left(\frac{Z(s_{i'})-Z(s_{j'})}{h_i}\right) Z(s_j) Z(s_{j'})}{\sum_j L_1\left(\frac{Z(s_i)-Z(s_j)}{h_i}\right) \sum_{j'} L_1\left(\frac{Z(s_{i'})-Z(s_{j'})}{h_i}\right)} \middle/ s_k, \forall k \right] \right],
\end{aligned}$$

donde  $\sum_{i, j, i', j'}$  denota  $\sum_i \sum_j \sum_{i'} \sum_{j'}$ .

Siguiendo un procedimiento similar al de las pruebas anteriores

$$D_1(s) \approx \sum_i \sum_j \mathbb{E} \left[ \frac{K_d\left(\frac{s-s_i}{h}\right)^2}{\left(\sum_i K_d\left(\frac{s-s_i}{h}\right)\right)^2} \frac{D_3(s_i, s_j)}{n^2 C_2(s_i, s_j)^2} \right] \approx \frac{D_4(s)}{n^4 C_4(s)^2}$$

con

$$\begin{aligned} D_3(s_i, s_j) &= \int \int L_1 \left( \frac{z_1 - z_2}{h_i} \right)^2 z_2^2 f_{s_i, s_j}(z_1, z_2) dz_1 dz_2 \\ &\approx h_i \int L_1(t)^2 dt \int z^2 f_{s_i, s_j}(z, z) dz \\ D_4(s) &= n \left( \sum_i h_i \right) \int \int \int K_d \left( \frac{s - \beta u_1}{h} \right)^2 z^2 f_{\beta u_1, \beta u_2}(z, z) g_0(u_1) g_0(u_2) dz du_1 du_2 \\ &\approx n \left( \sum_i h_i \right) h^d \beta^{-d} g_0(0) \int L_1(t)^2 dt \int K_d(v)^2 dv \int \int z^2 f_{s, \beta u}(z, z) g_0(u) dz du \end{aligned}$$

En consecuencia

$$\begin{aligned} D_1(s) &\approx \frac{D_4(s)}{n^4 C_4(s)^2} = \frac{n \left( \sum_i h_i \right) h^d \beta^{-d} g_0(0) \int L_1(t)^2 dt \int K_d(v)^2 dv \int z^2 C_5(s, z) dz}{n^4 \left( \sum_i h_i \right)^2 h^{2d} g_0(0)^2 \beta^{-2d} C_6(s)} \\ &= \frac{\int L_1(t)^2 dt \int K_d(v)^2 dv \int z^2 f_s(z) dz}{n^3 \left( \sum_i h_i \right) h^d \beta^{-d} g_0(0)} = \frac{\mathbb{E}[Z(s)^2] \int L_1(t)^2 dt \int K_d(v)^2 dv}{n^3 \left( \sum_i h_i \right) h^d \beta^{-d} g_0(0)} = O\left(n^{-3} \left( \sum_i h_i \right)^{-1} h^{-d} \beta^d\right) \end{aligned}$$

En cuanto a  $D_2(s)$ , procediendo de manera análoga

$$\begin{aligned} D_2(s) &\approx \sum_{i, j, i', j'} \mathbb{E} \left[ \frac{K_d \left( \frac{s - s_i}{h} \right) K_d \left( \frac{s - s_{i'}}{h} \right)}{\left( \sum_i K_d \left( \frac{s - s_i}{h} \right) \right)^2} \frac{D_5(s_i, s_j, s_{i'}, s_{j'})}{n^2 C_2(s_i, s_j) C_2(s_{i'}, s_{j'})} \right] \\ &\approx \frac{D_6(s)}{C_4(s)^2} \approx \frac{\left( \sum_i h_i \right)^2 h^{2d} \beta^{-2d} g_0(0)^2 (D_7(s) + D_8(s))}{C_4(s)^2} \end{aligned}$$

con

$$\begin{aligned} D_5(s_i, s_j, s_{i'}, s_{j'}) &= \int \int \int \int L_1 \left( \frac{z_1 - z_2}{h_i} \right) L_1 \left( \frac{z_3 - z_4}{h_{i'}} \right) z_2 z_4 \cdot f_{s_i, s_j, s_{i'}, s_{j'}}(z_1, z_2, z_3, z_4) dz_1 dz_2 dz_3 dz_4 \\ &\approx \left( \sum_{i, i'} h_i h_{i'} \right) \int \int z z' f_{s_i, s_j, s_{i'}, s_{j'}}(z, z, z', z') dz dz' \\ D_6(s) &= \left( \sum_i h_i \right)^2 \int \int \int \int \int K_d \left( \frac{s - \beta u_1}{h} \right) K_d \left( \frac{s - \beta u_2}{h} \right) z z' \cdot \\ &\quad \cdot f_{\beta u_1, \beta u_3, \beta u_2, \beta u_4}(z, z, z', z') g_0(u_1) g_0(u_2) dz dz' du_1 du_2 g_0(u_3) g_0(u_4) du_3 du_4 \\ D_7(s) &= \left( \sum_i h_i \right)^2 \int_S \left( \int \int z z' f_{s, \beta u, s, \beta u'}(z, z, z', z') dz dz' \right) g_0(u) g_0(u') du du' \\ D_8(s) &= \left( \sum_i h_i \right)^2 \int_{S^c} \left( \int \int z z' f_{s, \beta u, s, \beta u'}(z, z, z', z') dz dz' \right) g_0(u) g_0(u') du du' \end{aligned}$$

donde  $S = \{(u, u') \in \mathbb{R}^{2d} : \|u - u'\| \leq \beta^{-1/2}\}$ .

De la definición de  $S$ , es fácil ver que  $D_7(s) = O(\beta^{-d/2})$ . Por otro lado

$$f_{s, \beta u, s, \beta u'}(z, z, z', z') = D_9(s, u, u', z, z') D_{10}(s, u) D_{10}(s, u') \quad (2.3)$$

donde  $D_9(s, u, u', z, z')$  es la densidad de  $(Z(\beta u), Z(\beta u'))$  condicionada a  $Z(s) - Z(\beta u) = 0$  y  $Z(s) - Z(\beta u') = 0$ , en  $(z, z')$  y  $D_{10}(s, u)$  representa la densidad de  $Z(s) - Z(\beta u)$  en 0.

Para  $(u, u') \in S^c$ , se tiene que  $\beta \|u - u'\| > \beta^{1/2}$ . Por tanto, por la hipótesis (iii), las variables aleatorias  $Z(\beta u)$  y  $Z(\beta u')$  están incorreladas asintóticamente y entonces:

$$D_9(s, u, u', z, z') \approx D_{11}(s, u, z) D_{11}(s, u', z') \quad (2.4)$$

donde  $D_{11}(s, u, z)$  es igual a la densidad de  $Z(\beta u)$  condicionada a  $Z(s) - Z(\beta u) = 0$  en  $z$ . Luego, por las relaciones (2.3) y (2.4),  $f_{s, \beta u, s, \beta u'}(z, z, z', z') \approx f_{s, \beta u}(z, z) f_{s, \beta u'}(z', z')$ , para  $(u, u') \in S^C$ .

Teniendo todo esto en cuenta se sigue que

$$\begin{aligned} D_2(s) &\approx \frac{(\sum_i h_i)^2 h^{2d} \beta^{-2d} g_0(0)^2 (D_7(s) + D_8(s))}{C_4(s)^2} = \frac{(\sum_i h_i)^2 h^{2d} \beta^{-2d} g_0(0)^2 D_8(s)}{C_4(s)^2} + O(\beta^{-d/2}) \\ &\approx \frac{(\sum_i h_i)^2 h^{2d} \beta^{-2d} g_0(0)^2 \int_{S^C} (\int \int z z' f_{s, \beta u}(z, z) f_{s, \beta u'}(z', z') dz dz') g_0(u) g_0(u') dud u'}{(\sum_i h_i)^2 h^{2d} \beta^{-2d} g_0(0)^2 C_6(s)^2} \\ &\approx \int \int z z' \frac{C_5(s, z)}{C_6(s)} \frac{C_5(s, z')}{C_6(s)} dz dz' = \left( \int z f_s(z) dz \right)^2 = \mu(s)^2. \end{aligned}$$

Entonces,  $\mathbb{E}[\tilde{\mu}(s)^2] \approx D_1(s) + D_2(s) \approx \mu(s)^2$ , lo que permite concluir que

$$\text{Var}[\tilde{\mu}(s)] = \mathbb{E}[(\tilde{\mu}(s))^2] - (\mathbb{E}[\tilde{\mu}(s)])^2 \approx 0.$$

## 2.4. Selección de la ventana

Se puede considerar una versión más general de los estimadores (2.1) y (2.2) donde, en lugar de un parámetro ventana  $h$ , aparezca una matriz de ventanas  $H$ :

$$\hat{\mu}(s) = \frac{\sum_{i=1}^n K_d(H^{-1}(s - s_i)) Z(s_i)}{\sum_{i=1}^n K_d(H^{-1}(s - s_i))} \quad \text{y} \quad \tilde{\mu}(s) = \frac{\sum_{i=1}^n K_d(H^{-1}(s - s_i)) \tilde{\mu}_1(s_i)}{\sum_{i=1}^n K_d(H^{-1}(s - s_i))}.$$

Para el estimador  $\tilde{\mu}(s)$  así definido también sería posible probar la consistencia bajo ciertas condiciones.

Es frecuente pensar que la matriz  $H$  es igual a  $hS^{-1/2}$ , donde  $S$  es la matriz de varianzas y covarianzas y  $h$  representa un parámetro de suavización. Así, una vez aplicada la función núcleo aparecen curvas de nivel que se orientan en la dirección de las posiciones, dependiendo de si la correlación entre posiciones es positiva o negativa. El grosor de dichas curvas de nivel varía en función de la varianza.

Además, se podría simplificar la selección de la matriz de ventanas  $H$  reduciéndola a una matriz diagonal  $H = \text{diag}(h_1, h_2)$ , donde cada elemento de la diagonal establece el nivel de suavizado en cada dirección. De este modo, para la selección de la matriz de ventanas  $H$  se proponen dos métodos.

El método más simple consiste en tomar  $H = \text{diag}(h, h)$ , donde  $h$  es un parámetro ventana local  $h = h(s)$ , obtenido como el percentil de orden  $P = P(s)$  de las menores distancias  $\|s - s_i\|$ , para cada  $i$  y algún  $0 < P < 1$ . Este mecanismo dará lugar a una ventana local.

La otra posibilidad es elegir la matriz resultante de aplicar el método de validación cruzada o bien el método de validación cruzada generalizada que elige la ventana minimizando, respectivamente, las expresiones

$$CV(H) = \frac{1}{n} \sum_{i=1}^n \left( Z(s_i) - \mu_{(-i)}^*(s_i) \right)^2$$

y

$$GCV(H) = \frac{1}{n} \sum_{i=1}^n \left( \frac{Z(s_i) - \mu^*(s_i)}{1 - \frac{1}{n} \text{tr}(S)} \right)^2,$$

donde  $\mu_{(-i)}^*$  representa alguno de los estimadores de la tendencia que se obtendría en  $s_i$  sin considerar el dato  $i$ -ésimo,  $Z(s_i)$ .

Sin embargo, estos últimos son criterios de selección de la ventana diseñados para datos independientes. En el caso de datos dependientes tienden a infrasuavizar la estimación como se puede ver, por ejemplo, en Opstomer et al. (2001). Por ello se han propuesto distintos criterios alternativos para el caso de datos dependientes, como el criterio de validación cruzada modificada (Chu y Marron, 1991) o el de validación cruzada generalizada con corrección de sesgo para dependencia (Francisco-Fernández y Opsomer, 2005). Estos métodos obtienen una matriz de ventanas minimizando, respectivamente, las expresiones

$$MCV(H) = \frac{1}{n} \sum_{i=1}^n \left( Z(s_i) - \mu_{-N(i)}^*(s_i) \right)^2 \quad (2.5)$$

y

$$CGCV(H) = \frac{1}{n} \sum_{i=1}^n \left( \frac{Z(s_i) - \mu^*(s_i)}{1 - \frac{1}{n\sigma^2} \text{tr}(S\hat{R})} \right),$$

siendo  $\mu_{-N(i)}^*(s_i)$  la estimación de la tendencia obtenida eliminando los datos en un vecindario  $N(i)$  de  $s_i$  y  $\hat{R}$  una estimación de la matriz de correlaciones.

En cuanto a la elección de las ventanas necesarias en la primera estimación de la tendencia,  $\tilde{\mu}_1(s_i)$ ,  $h_i$ , se podría considerar también un parámetro ventana local  $h_i = h_i(s_i)$  obtenido como el percentil de orden  $P_i = P_i(s_i)$  de los valores positivos  $|Z(s_i) - Z(s_j)|$ , para cada  $i$  y algún  $0 < P_i < 1$ . En este caso también sería posible seleccionar dichos parámetros  $h_i(s_i)$  mediante un mecanismo de validación cruzada clásico.



## Capítulo 3

# Estudio de simulación

En esta sección se describen los estudios de simulación llevados a cabo para analizar el comportamiento del estimador propuesto de la tendencia espacial en diferentes escenarios. Para ello se generarán datos de distintos procesos donde se hará variar tanto la estructura de dependencia como la tendencia.

Los objetivos son, por un lado estudiar si el estimador (2.2),  $\tilde{\mu}$ , es consistente en media cuadrática, para lo cual debemos analizar, en primer lugar cómo varían los resultados en función de la selección de los parámetros ventana. Por otro lado queremos comparar el funcionamiento del estimador propuesto con los estimadores clásicos en estadística espacial a la hora de realizar predicciones.

Para la programación de los estimadores tipo núcleo de la tendencia (2.1) y (2.2) se han empleado las librerías de R *npsp* de Fernández-Casal (2016) y *locpol* de Ojeda Cabrera (2018). Por otro lado, la simulación de datos espaciales y la aplicación de las técnicas kriging se han llevado a cabo usando la librería *geoR* de Ribeiro y Diggle (2018). Por último, para la aplicación a datos reales se recurrió al conjunto de datos *parana*, que se encuentra en la librería *geoR*.

### 3.1. Datos simulados

Se han simulado valores de distintos procesos estocásticos para varios tamaños muestrales  $n = 100$ ,  $n = 256$  y  $n = 400$  en la región de observación dada por el cuadrado unidad  $D = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ . Se considerarán el caso de diseño fijo empleando una rejilla regular para las observaciones y el caso de diseño aleatorio generando las localizaciones muestrales a partir de la distribución uniforme en  $D$ , como se puede ver en la parte izquierda de las Figuras 3.1 y 3.2 para el caso particular de tamaño muestral  $n = 256$ .

Para medir el ajuste de los estimadores de la tendencia a los valores reales se calculará el error cuadrático medio:

$$ECM(s) = \frac{1}{M} \sum_{j=1}^M \left( \tilde{\mu}^{(j)}(s) - \mu(s) \right)^2,$$

donde  $M$  es el número de muestras generadas y  $\tilde{\mu}^{(j)}(s)$  la estimación de la tendencia del proceso obtenida en la localización  $s$  para la muestra  $j$ -ésima. Por tratarse de una medida de error local, se han seleccionado cuatro puntos dentro de la región de observación,  $s^0 = (0.5, 0.5)$ ,  $s^1 = (0.25, 0.25)$ ,  $s^2 = (0.1, 0.1)$  y  $s^3 = (0.01, 0.01)$  y un punto fuera de  $D$ ,  $s^4 = (-0.01, -0.01)$ , para comprobar si el estimador funciona bien incluso fuera de la región de observación (ver Figuras 3.1 y 3.2).

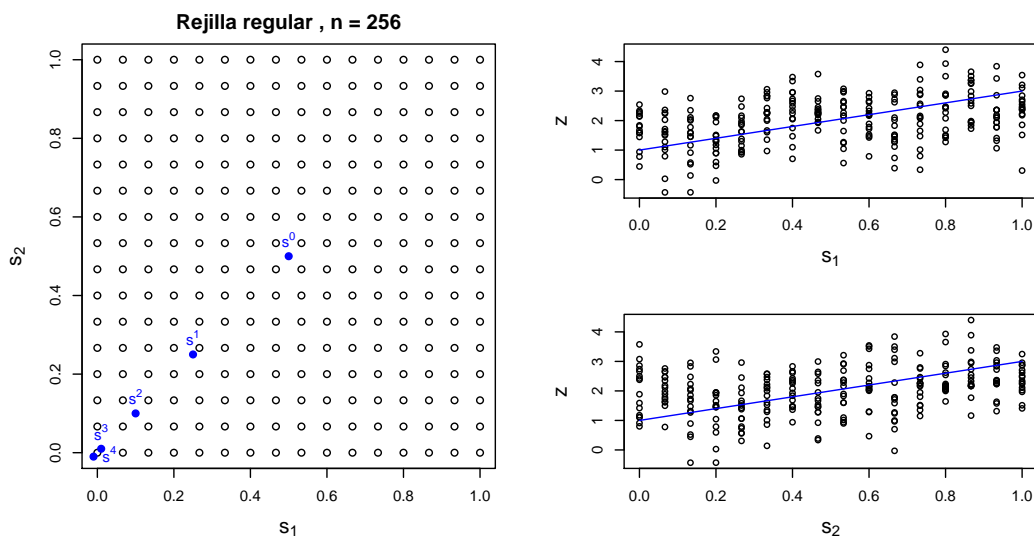


Figura 3.1: Localizaciones bajo diseño fijo (izquierda) y datos simulados frente a cada una de las coordenadas (derecha) para una muestra de tamaño  $n = 256$  con tendencia  $\mu(s) = 1 + s_1 + s_2$ .

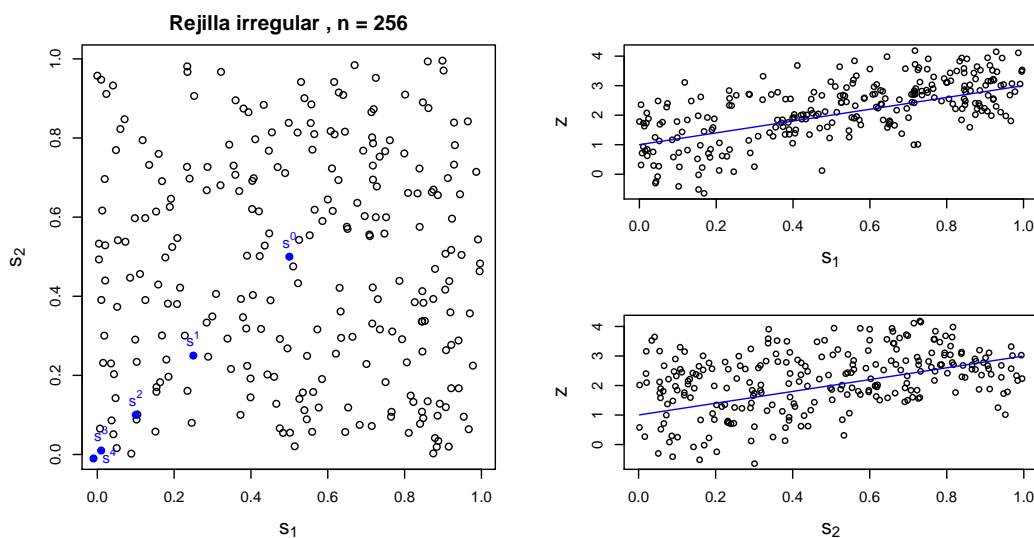


Figura 3.2: Localizaciones bajo diseño aleatorio (izquierda) y datos simulados frente a cada una de las coordenadas (derecha) para una muestra de tamaño  $n = 256$  con tendencia  $\mu(s) = 1 + s_1 + s_2$ .

### 3.2. Selección de las ventanas

La primera parte de este estudio se dedica a decidir qué método es más apropiado para seleccionar las ventanas  $h_i$  empleadas en la primera aproximación de la tendencia. Para esto se ha calculado el menor  $ECM$  cometido en una rejilla de ventanas constantes,  $h_{i,cte}$ , el  $ECM$  usando la ventana seleccionada como el 20% de las distancias entre observaciones,  $h_{i,20\%}$ , y el  $ECM$  para la ventana seleccionada por validación cruzada,  $h_{i,cv}$ . Nótese que la primera es una ventana global, igual para todas las localizaciones  $s_i$ , mientras que las otras dos son locales, diferentes para cada posición muestreada.

Para este primer objetivo, se han generado  $n$  datos del proceso espacial  $\{Z(s)/s \in D\}$ ,  $Z(s) = \mu(s) + Y(s)$  con  $\mu(s) = 1 + s_1 + s_2$  en las localizaciones  $s_i$  consideradas a partir de un proceso gaussiano con media dependiente de las localizaciones de manera lineal y cuya estructura de dependencia sigue un modelo de semivariograma isotrópico exponencial con parámetros  $c_0 = 0$ ,  $c_2 = 0.2$  y  $\sigma^2 = 0.5$ , es decir

$$\gamma_{c_0, c_1, c_2}(d) = \begin{cases} 0, & \text{si } d = 0 \\ 0.5 \left( 1 - \exp\left(\frac{-3d}{0.2}\right) \right), & \text{en otro caso.} \end{cases} \quad (3.1)$$

De este modo se han obtenido  $M = 1000$  muestras, para las cuales se ha estimado el valor de la tendencia en las localizaciones objetivo  $s^k$   $k = 0, \dots, 4$ , y su correspondiente error cuadrático. Los resultados se recogen en la Tabla 3.1 para datos simulados sobre una rejilla regular e irregular.

Es importante señalar que para obtener la estimación de la tendencia,  $\tilde{\mu}$ , es necesaria la selección de la ventana,  $h$ , o matriz de ventanas,  $H$ , para la segunda etapa del estimador. Por ahora se seleccionará por validación cruzada. En particular, se utiliza el método de validación cruzada modificada programado en el paquete *npsp*, que da lugar a la matriz ventana  $H_{cv} = \text{diag}(h_1, h_2)$ ,  $h_1 \neq h_2$ , seleccionada dentro de un conjunto prefijado de valores de tal manera que minimice la expresión (2.5).

Además, para reducir el elevado coste computacional que supone el método de validación cruzada para muestras grandes, se realiza previamente un binning de los datos, usando la técnica de binning lineal, de la que se pueden obtener más detalles en Wand y Jones (1995).

	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$		$s^0$	$s^1$	$s^2$	$s^3$	$s^4$		
	$h_{i,cv}$	1.30	12.19	31.19	47.21	51.24	$h_{i,cv}$	1.75	14.83	36.46	54.49	59.02	
100	$h_{i,cte}$	1.33	6.36	16.33	25.15	27.41	100	$h_{i,cte}$	1.76	9.19	22.18	33.31	36.13
	$h_{i,20\%}$	1.40	7.69	19.93	30.62	33.35		$h_{i,20\%}$	1.81	10.55	25.75	38.67	41.93
	$h_{i,cv}$	1.75	8.16	21.99	35.71	39.28		$h_{i,cv}$	2.13	9.86	25.91	41.44	45.46
256	$h_{i,cte}$	1.95	3.81	8.46	14.66	16.45	256	$h_{i,cte}$	2.31	5.05	11.81	19.93	22.23
	$h_{i,20\%}$	2.33	4.33	11.13	19.34	21.61		$h_{i,20\%}$	2.64	5.77	14.70	24.77	27.51
	$h_{i,cv}$	2.04	7.85	20.22	33.37	36.86		$h_{i,cv}$	2.18	8.14	21.63	35.69	39.42
400	$h_{i,cte}$	2.36	4.34	8.28	13.86	15.66	400	$h_{i,cte}$	2.47	4.87	9.94	16.71	18.82
	$h_{i,20\%}$	2.99	4.44	9.92	17.33	19.51		$h_{i,20\%}$	3.07	4.94	11.47	19.90	22.34

Tabla 3.1:  $100 \times ECM(s^k)$ ,  $k = 0, \dots, 4$ , en la estimación de  $\mu(s^k)$  mediante  $\tilde{\mu}(s^k)$  obtenido a partir de  $M = 1000$  muestras bajo diseño fijo (izquierda) y aleatorio (derecha).

Se puede observar que el  $ECM$  es mayor bajo diseño aleatorio en todos los casos y que los valores para una rejilla de  $h_i$  constantes y los obtenidos usando el percentil 20% son bastante similares. Además, se puede ver que en la localización  $s^0$  siempre se obtiene un  $ECM$  más pequeño con la ventana de validación cruzada que con  $h_{i,20\%}$ , mientras que en las localizaciones alejadas del centro de la región de observación el error cometido es mucho mayor para la ventana de validación cruzada que para la ventana  $h_{i,20\%}$ . Como en la práctica no se podrá conocer la mejor de las  $h_{i,cte}$  para seleccionar la ventana entre una rejilla de valores constantes, en adelante consideraremos la ventana local obtenida para cada localización como el 20% de las distancias entre observaciones. Esta ventana, además de presentar un buen comportamiento tiene un menor coste computacional que la ventana seleccionada mediante validación cruzada.

Para justificar la elección del 20% en la selección de la ventana calculada como el percentil de las diferencias entre observaciones se ha repetido el estudio anterior para el 15%, 20% y 25%. En la Tabla 3.2 se puede ver que tanto bajo diseño fijo como bajo diseño aleatorio el 15% produce menores errores en puntos alejados del centro y mayores errores en puntos cerca del centro, mientras que el efecto es el contrario con el 25%, siendo el 20% un buen punto de equilibrio entre ambos efectos.

		$s^0$	$s^1$	$s^2$	$s^3$	$s^4$			$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	15%	1.41	7.28	18.84	28.97	31.56	100	15%	1.81	10.16	24.70	37.09	40.23
	20%	1.40	7.69	19.93	30.62	33.35		20%	1.81	10.55	25.75	38.67	41.93
	25%	1.40	8.11	21.04	32.29	35.16		25%	1.81	10.97	26.83	40.29	43.69
256	15%	2.38	4.13	10.30	17.93	20.06	256	15%	2.69	5.54	13.84	23.34	25.95
	20%	2.33	4.33	11.13	19.34	21.61		20%	2.64	5.77	14.70	24.77	27.51
	25%	2.27	4.54	11.97	20.76	23.16		25%	2.59	6.03	15.59	26.20	29.08
400	15%	3.10	4.35	9.30	16.14	18.19	400	15%	3.18	4.86	10.88	18.78	21.11
	20%	2.99	4.44	9.92	17.33	19.51		20%	3.07	4.94	11.47	19.90	22.34
	25%	2.89	4.58	10.65	18.70	21.02		25%	2.98	5.06	12.17	21.15	23.72

Tabla 3.2:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$ ,  $M = 1000$  bajo diseño fijo (izquierda) y aleatorio (derecha).

Una vez elegida la ventana local  $h_i$ , el siguiente objetivo es el de decidir cómo construir la matriz de ventanas  $H$  global, necesaria para la segunda etapa en el cálculo del estimador (2.2). Para ello se ha tomado una secuencia de 21 valores fijos en el intervalo  $[0.1, 1]$  para los cuales se calcula el  $ECM$ , construyendo para cada valor  $h$  la matriz suavizadora  $H = diag(h, h)$ . Es decir, se está considerando el selector que aporta el mismo grado de suavizado para las dos coordenadas, puesto que el efecto sobre la tendencia es similar para ambas.

Los errores cuadráticos medios se muestran en las Tablas 3.3 y 3.5 para diseño fijo y en las Tablas 3.4 y 3.6 para diseño aleatorio. En negrita se han resaltado los valores de  $h$  que hacen mínimo el  $ECM$ . Además, en las Figuras 3.3 y 3.4 se recogen estos mismos resultados de un modo gráfico para tamaño muestral  $n = 256$ . Estos resultados no serán comparables con los obtenidos en la Tabla 3.1, ya que en este caso la matriz suavizadora es  $H = hI$ , mientras que en la Tabla 3.1 la matriz de ventanas es  $H_{cv} = diag(h_1, h_2)$ ,  $h_1 \neq h_2$ .

$n$	$h :$	0.100	0.145	0.190	0.235	0.280	0.325	0.370	0.415	0.460	0.505	
100		7.09	7.09	7.09	7.09	6.82	5.35	4.03	3.27	2.83	2.58	
256		6.88	6.88	5.87	4.77	4.26	3.91	3.44	3.03	2.72	2.45	
400		9.18	8.06	6.34	5.66	4.90	4.17	3.63	3.15	2.76	2.45	
$h :$		0.550	0.595	0.640	0.685	0.730	0.775	0.820	0.865	0.910	0.955	1.000
100		2.39	2.20	2.02	1.86	1.72	1.62	1.54	1.47	1.42	1.38	<b>1.35</b>
256		2.21	2.02	1.86	1.74	1.63	1.55	1.48	1.43	1.39	1.35	<b>1.32</b>
400		2.20	2.00	1.84	1.71	1.61	1.54	1.48	1.43	1.39	1.36	<b>1.34</b>

Tabla 3.3:  $100 \times ECM(s^0)$  en la estimación de la tendencia empleando el estimador  $\tilde{\mu}$  con ventanas  $H = diag(h, h)$  y  $h_i = h_{i,20\%}$  bajo diseño fijo para distintos tamaños de muestra,  $n$ .

$n$	$h :$	0.100	0.145	0.190	0.235	0.280	0.325	0.370	0.415	0.460	0.505	
100		8.02	8.02	8.02	8.02	7.52	5.68	4.25	3.46	3.02	2.77	
256		7.80	7.80	6.56	5.38	4.84	4.45	3.92	3.47	3.14	2.86	
400		8.78	7.63	6.07	5.46	4.77	4.11	3.63	3.19	2.83	2.53	
$h :$		0.550	0.595	0.640	0.685	0.730	0.775	0.820	0.865	0.910	0.955	1.000
100		2.59	2.42	2.28	2.16	2.06	1.99	1.93	1.89	1.85	1.82	<b>1.80</b>
256		2.62	2.43	2.27	2.15	2.05	1.97	1.91	1.85	1.81	1.78	<b>1.75</b>
400		2.29	2.11	1.96	1.84	1.75	1.68	1.63	1.59	1.55	1.53	<b>1.50</b>

Tabla 3.4:  $100 \times ECM(s^0)$  en la estimación de la tendencia empleando el estimador  $\tilde{\mu}$  con ventanas  $H = diag(h, h)$  y  $h_i = h_{i,20}\%$  bajo diseño aleatorio para distintos tamaños de muestra,  $n$ .

$n$	$h :$	0.100	0.145	0.190	0.235	0.280	0.325	0.370	0.415	0.460	0.505	0.550	0.595
100	$s^1$	7.45	7.45	7.45	7.45	7.25	6.16	5.20	4.68	4.42	4.28	<b>4.24</b>	4.33
	$s^2$	7.95	7.95	7.95	<b>7.95</b>	7.98	8.24	8.74	9.31	9.83	10.26	10.71	11.37
	$s^3$	15.47	15.47	15.47	15.47	15.37	<b>15.01</b>	15.14	15.73	16.45	17.13	17.85	18.82
	$s^4$	18.42	18.42	18.42	18.42	18.24	17.40	<b>17.18</b>	17.64	18.34	19.05	19.81	20.85
256	$s^1$	7.78	7.78	6.59	5.38	4.86	4.55	4.25	4.03	3.90	<b>3.88</b>	3.96	4.12
	$s^2$	9.48	9.48	8.94	8.46	8.32	<b>8.31</b>	8.46	8.75	9.13	9.64	10.29	11.03
	$s^3$	15.93	15.93	14.84	<b>14.21</b>	14.37	14.66	15.07	15.64	16.29	17.11	18.09	19.16
	$s^4$	20.55	20.55	18.23	16.60	<b>16.51</b>	16.76	17.13	17.69	18.36	19.22	20.26	21.39
400	$s^1$	10.07	8.78	6.88	6.16	5.50	4.88	4.47	4.20	4.04	<b>4.00</b>	4.06	4.22
	$s^2$	12.53	11.25	9.50	8.94	8.68	<b>8.62</b>	8.73	8.99	9.37	9.87	10.51	11.24
	$s^3$	16.79	15.70	14.52	<b>14.48</b>	14.54	14.78	15.22	15.82	16.55	17.41	18.42	19.53
	$s^4$	22.70	20.18	17.28	16.87	16.74	<b>16.88</b>	17.28	17.89	18.66	19.57	20.63	21.81

Tabla 3.5:  $100 \times ECM(s^k)$  en la estimación de la tendencia empleando el estimador  $\tilde{\mu}$  con ventanas  $H = diag(h, h)$  y  $h_i = h_{i,20}\%$  bajo diseño fijo para distintos tamaños de muestra,  $n$ .

$n$	$h :$	0.100	0.145	0.190	0.235	0.280	0.325	0.370	0.415	0.460	0.505	0.550	0.595
100	$s^1$	7.97	7.97	7.97	7.97	7.75	6.83	6.15	5.85	5.74	<b>5.73</b>	5.81	6.04
	$s^2$	11.49	11.49	11.49	11.49	<b>11.45</b>	11.69	12.27	12.91	13.48	13.94	14.50	15.32
	$s^3$	23.98	23.38	23.38	23.38	22.53	21.18	<b>21.14</b>	21.69	22.35	22.95	23.68	24.72
	$s^4$	27.76	27.76	27.76	27.76	26.50	24.26	<b>23.79</b>	24.21	24.83	25.44	26.18	27.25
256	$s^1$	8.83	8.83	7.54	6.36	5.84	5.54	5.27	5.11	<b>5.05</b>	5.11	5.27	5.53
	$s^2$	11.18	11.29	10.90	<b>10.69</b>	10.71	10.79	11.00	11.36	11.80	12.41	13.16	14.00
	$s^3$	22.78	23.00	20.10	18.90	<b>18.90</b>	19.06	19.32	19.82	20.46	21.29	22.31	23.42
	$s^4$	28.90	29.22	24.20	21.93	<b>21.69</b>	21.75	21.89	22.32	22.95	23.79	24.83	25.98
400	$s^1$	10.33	8.84	6.95	6.26	5.66	5.13	4.78	4.55	4.43	<b>4.42</b>	4.52	4.71
	$s^2$	13.74	12.70	11.15	10.60	10.22	<b>10.05</b>	10.08	10.28	10.63	11.12	11.76	12.52
	$s^3$	23.00	20.05	18.16	18.02	17.76	<b>17.65</b>	17.81	18.18	18.73	19.48	20.42	21.49
	$s^4$	30.87	25.08	21.36	20.94	20.44	<b>20.16</b>	20.23	20.55	21.08	21.84	22.82	23.94

Tabla 3.6:  $100 \times ECM(s^k)$  en la estimación de la tendencia empleando el estimador  $\tilde{\mu}$  con ventanas  $H = diag(h, h)$  y  $h_i = h_{i,20}\%$  bajo diseño aleatorio para distintos tamaños de muestra,  $n$ .

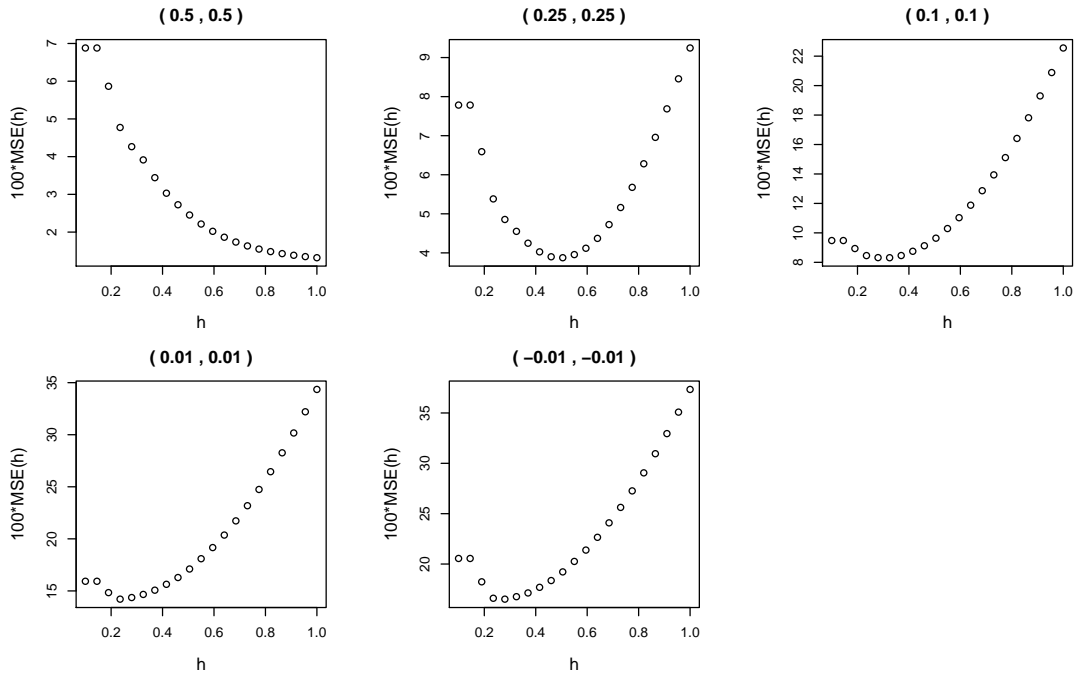


Figura 3.3:  $100 \times ECM(s^k)$  en la estimación de la tendencia empleando el estimador  $\tilde{\mu}$  con ventanas  $H = \text{diag}(h, h)$  y  $h_i = h_{i,20\%}$  bajo diseño fijo para tamaños de muestra  $n = 256$ .

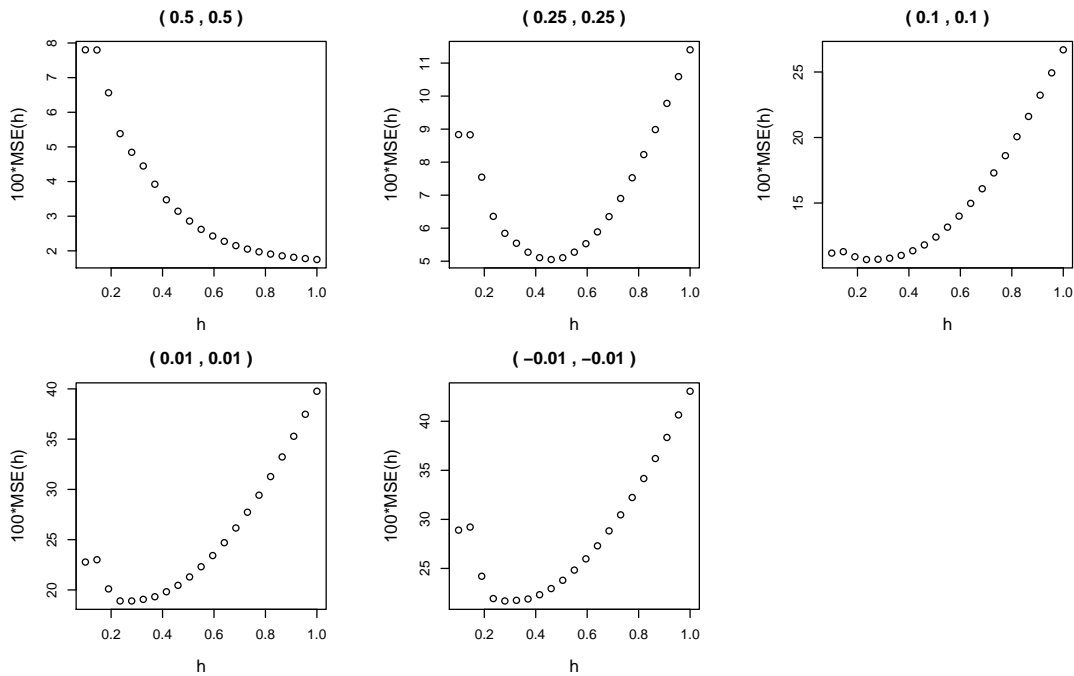


Figura 3.4:  $100 \times ECM(s^k)$  en la estimación de la tendencia empleando el estimador  $\tilde{\mu}$  con ventanas  $H = \text{diag}(h, h)$  y  $h_i = h_{i,20\%}$  bajo diseño aleatorio para tamaños de muestra  $n = 256$ .

En las Tablas 3.3 y 3.4 se ha estudiado el caso del centro de la región de observación por separado, que corresponde con el gráfico de la esquina superior izquierda de las Figuras 3.3 y 3.4, ya que el menor error se obtiene siempre en el extremo superior del intervalo de constantes. Este fenómeno ya ocurría en la selección de  $h_i$  y es debido a que las observaciones constituyen un cuadrado dentro de un hiperplano en  $\mathbb{R}^3$ . Al considerar una rejilla en una región cuadrada, la media muestral sobre esa rejilla coincide con el centro.

Por otro lado, en la Tabla 3.5 se aprecia que, en general, el valor de  $h$  que minimiza el error va decreciendo al aumentar  $n$ . Además, los errores aumentan a medida que la localización objetivo se aleja del centro. Cabe señalar que en bajo diseño fijo y tamaño  $n = 100$  la distancia vertical y horizontal entre localizaciones es 0.1, por lo que aparecen problemas para valores pequeños de  $h$  (de hecho el error cuadrático medio prácticamente se mantiene constante hasta  $h = 0.280$ , como se puede ver también en la Figura 3.3).

Como los valores resaltados de las Tablas 3.3, 3.4, 3.5 y 3.6 son bastante menores que los valores obtenidos en las filas sombreadas de la Tabla 3.1, se estudia si es necesario permitir la libertad a la matriz de suavizado de tener los dos elementos de la diagonal distintos,  $h_1 \neq h_2$ . Los resultados del menor error cuadrático medio variando  $h_1$  y  $h_2$  para cada tamaño muestral y cada localización objetivo se recogen en la Tabla 3.7 y de manera gráfica para las localizaciones objetivo  $s^1$  y  $s^2$  en la Figura 3.5 con tamaño muestral  $n = 256$  bajo diseño fijo.

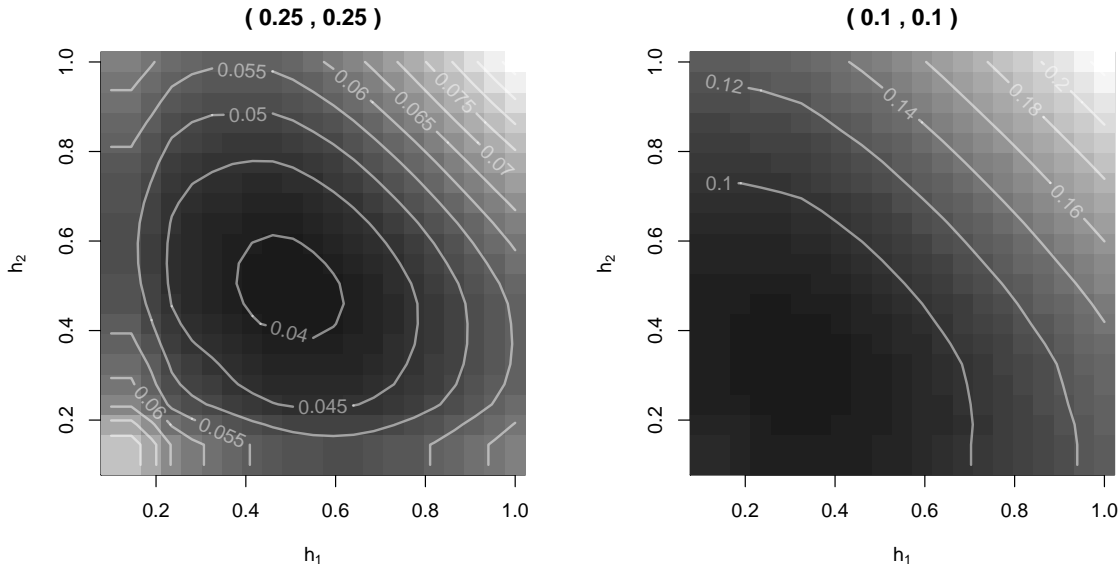


Figura 3.5:  $ECM(s^1)$  y  $ECM(s^2)$  con  $H = \text{diag}(h_1, h_2)$  para rejillas de  $h_1$  y  $h_2$  calculado a partir de  $M = 1000$  muestras de tamaño  $n = 256$  bajo diseño fijo.

	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$		$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	1.35	4.24	7.95	14.98	17.17	100	1.80	5.73	11.45	21.06	23.79
256	1.32	3.88	8.31	14.15	16.34	256	1.75	5.05	10.62	18.82	21.53
400	1.34	4.00	8.61	14.40	16.68	400	1.50	4.42	10.02	17.60	20.12

Tabla 3.7:  $100 \times ECM(s^k)$ ,  $k = 0, \dots, 4$ ,  $M = 1000$  bajo diseño fijo (izquierda) y aleatorio (derecha).

También se obtuvieron tablas similares con el selector de ventana  $H = \text{diag}(h, h)$ , siendo  $h$  el percentil de orden  $P$ , con  $0 < P < 1$ , de las distancias entre las localizaciones objetivo y las localizaciones muestrales. En la Tabla 3.8 se recogen los errores tomando  $P = 0.2$ .

	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$		$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	6.74	7.32	8.60	16.41	18.75	100	7.80	7.80	12.43	22.79	25.68
256	4.51	5.00	8.52	16.50	19.07	256	5.13	6.08	11.13	21.11	24.15
400	5.26	5.71	8.76	16.97	19.69	400	5.23	6.03	10.15	19.26	22.11

Tabla 3.8:  $100 \times ECM(s^k)$ ,  $k = 0, \dots, 4$ ,  $M = 1000$  bajo diseño fijo (izquierda) y aleatorio (derecha).

Por último se estudia si la elección de  $P$  como el percentil de orden 0.2 de las distancias entre la localización objetivo y las localizaciones muestrales es apropiado para la selección de  $H$ . Para ello se ha calculado el error cuadrático medio de estimación variando entre el 15 % y el 25 %. En la Tabla 3.9 parece que, de nuevo, el 20 % equilibra los efectos inversos entre puntos próximos al centro y puntos alejados del centro de los otros dos porcentajes. Así, en el caso de decidir usar esta ventana, el 20 % de las distancias entre localizaciones sigue siendo una buena opción.

	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$		$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
15 %	7.04	7.53	7.98	15.17	17.35	15 %	8.17	8.08	11.80	21.56	24.31
100 20 %	6.74	7.32	8.60	16.41	18.75	100 20 %	7.80	7.80	12.43	22.79	25.68
25 %	6.45	7.18	9.32	17.75	20.25	25 %	7.45	7.60	13.20	24.18	27.22
15 %	4.71	5.07	7.96	15.22	17.60	15 %	5.34	6.15	10.53	19.79	22.65
256 20 %	4.51	5.00	8.52	16.50	19.07	256 20 %	5.13	6.08	11.13	21.11	24.15
25 %	4.33	4.99	9.14	17.81	20.54	25 %	4.95	6.06	11.79	22.47	25.67
15 %	5.50	5.83	8.26	15.69	18.22	15 %	5.47	6.22	9.73	18.06	20.73
400 20 %	5.26	5.71	8.76	16.97	19.69	400 20 %	5.23	6.03	10.15	19.26	22.11
25 %	5.02	5.64	9.38	18.38	21.29	25 %	4.99	5.89	10.71	20.61	23.64

Tabla 3.9:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$ ,  $M = 1000$  bajo diseño fijo (izquierda) y aleatorio (derecha).

## Conclusión

A partir de los resultados analizados en esta sección se puede concluir que la ventana más apropiada para la selección de  $h_i$ , necesaria en la primera aproximación para la estimación de la tendencia a partir del método en dos etapas, es la que para cada localización muestral toma el percentil 20 % de las distancias entre las observaciones. Será esta la que se use en lo que resta de trabajo.

En cuanto a la matriz de ventanas necesaria para el segundo paso del método en dos etapas, tanto la matriz diagonal seleccionada por validación cruzada modificada como la matriz diagonal calculada como el 20 % de las distancias entre la localización objetivo y las localizaciones muestrales parecen una buena opción. En adelante se empleará la matriz de ventanas de validación cruzada.



### 3.3. Comparación con el estimador de NW

En esta sección se compara el error cometido usando el estimador de tipo NW de la ecuación (2.1),  $\hat{\mu}(s)$ , con el estimador en dos etapas de la ecuación (2.2),  $\tilde{\mu}(s)$ . Tal y como se demostró en el capítulo 2 el estimador de NW no es asintóticamente consistente. Se ha empleado para ambos estimadores la matriz de ventanas seleccionada por el método de validación cruzada modificada y además para  $\tilde{\mu}$  la ventana  $h_{i,20}\%$ . Esta comparación se llevará a cabo simulando datos de un proceso espacial gaussiano con la estructura de dependencia dada por el modelo de semivariograma (3.1) usado en la sección 3.2 pero modificando la forma de la tendencia. Además en este estudio se ha añadido un nuevo punto, más próximo al centro de la región de observación,  $s^* = (0.4, 0.4)$ .

En primer lugar se analiza el caso más simple de tendencia constante,  $\mu(s) = \mu = 1$ . Las Tablas 3.10 y 3.11 recogen los errores cuadráticos medios para ambos estimadores bajo diseño fijo y aleatorio. Se puede ver que en este caso es siempre mejor el estimador en dos etapas que el estimador de NW.

	$\hat{\mu}$						$\tilde{\mu}$					
	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$
100	1.28	1.33	1.66	2.11	2.53	2.63	1.26	1.30	1.56	1.91	2.23	2.31
256	2.19	2.28	2.46	3.32	4.18	4.44	1.95	2.00	2.13	2.78	3.43	3.63
400	2.87	2.90	3.26	4.84	6.29	6.80	2.43	2.47	2.77	3.91	4.95	5.31

Tabla 3.10:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para  $\hat{\mu}$  y  $\tilde{\mu}$  con tendencia  $\mu(s) = 1$ .

	$\hat{\mu}$						$\tilde{\mu}$					
	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$
100	1.68	1.78	2.24	2.89	3.47	3.61	1.67	1.74	2.11	2.62	3.06	3.17
256	2.55	2.60	3.04	4.21	5.20	5.51	2.35	2.37	2.69	3.55	4.29	4.53
400	3.02	3.10	3.58	5.42	7.11	7.68	2.64	2.71	3.07	4.46	5.75	6.19

Tabla 3.11:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para  $\hat{\mu}$  y  $\tilde{\mu}$  con tendencia  $\mu(s) = 1$ .

En las Tablas 3.12 y 3.13 se ha calculado el  $ECM$  al estimar tendencia lineal,  $\mu(s) = 1 + s_1 + s_2$ . Si se comparan los resultados para  $\hat{\mu}$  con los resultados para  $\tilde{\mu}$  (parte de los cuales ya se mostraban en las filas sombreadas de la Tabla 3.1) se puede comprobar que excepto en el centro de la región de observación,  $s_0$ , y en puntos cercanos, como  $(0.40, 0.40)$ , el error cometido al emplear el estimador de NW resulta siempre menor que el error usando  $\tilde{\mu}$ .

	$\hat{\mu}$						$\tilde{\mu}$					
	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$
100	1.44	2.22	6.36	16.33	25.15	27.41	1.40	2.42	7.69	19.93	30.62	33.35
256	2.58	2.79	3.81	8.46	14.66	16.45	2.33	2.57	4.33	11.13	19.34	21.61
400	3.39	3.41	4.35	8.28	13.86	15.66	2.99	3.10	4.44	9.92	17.33	19.51

Tabla 3.12:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para  $\hat{\mu}$  y  $\tilde{\mu}$  con tendencia  $\mu(s) = 1 + s_1 + s_2$ .

	$\hat{\mu}$						$\tilde{\mu}$					
	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$
100	1.81	2.98	9.19	22.18	33.31	36.13	1.81	3.20	10.55	25.75	38.67	41.93
256	2.86	3.16	5.05	11.81	19.93	22.23	2.64	3.05	5.77	14.70	24.77	27.51
400	3.43	3.73	4.88	9.94	16.71	18.82	3.07	3.35	4.94	11.47	19.90	22.34

Tabla 3.13:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para  $\hat{\mu}$  y  $\tilde{\mu}$  con tendencia  $\mu(s) = 1 + s_1 + s_2$ .

Por último se repite el estudio simulando los datos con la misma estructura de dependencia pero cambiando la tendencia lineal por una tendencia cuadrática,  $\mu(s) = 1 + 0.5 \cdot (s_1^2 + s_2^2)$ . Se obtienen los resultados de las Tablas 3.14 y 3.15. En este caso el estimador de la tendencia  $\tilde{\mu}$  presenta para los tamaños muestrales  $n = 256$  y  $n = 400$  y todas las localizaciones objetivo un error cuadrático medio inferior al del estimador de NW.

	$\hat{\mu}$						$\tilde{\mu}$					
	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$
100	2.06	2.91	3.90	4.46	4.43	4.41	2.06	3.08	4.30	4.95	4.82	4.76
256	2.58	2.86	3.49	4.68	5.59	5.87	2.37	2.74	3.51	4.59	5.21	5.37
400	3.22	3.38	4.00	5.84	7.37	7.91	2.84	3.12	3.81	5.29	6.36	6.72

Tabla 3.14:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para  $\hat{\mu}$  y  $\tilde{\mu}$  con tendencia  $\mu(s) = 1 + 0.5 \cdot (s_1^2 + s_2^2)$ .

	$\hat{\mu}$						$\tilde{\mu}$					
	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^*$	$s^1$	$s^2$	$s^3$	$s^4$
100	2.16	3.10	4.44	5.21	5.30	5.31	2.16	3.21	4.73	5.56	5.54	5.50
256	2.93	3.28	4.28	5.85	6.97	7.29	2.77	3.23	4.32	5.72	6.52	6.73
400	3.30	3.58	4.41	6.83	8.88	9.56	2.92	3.24	4.10	5.98	7.42	7.88

Tabla 3.15:  $100 \times ECM(s^k)$   $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para  $\hat{\mu}$  y  $\tilde{\mu}$  con tendencia  $\mu(s) = 1 + 0.5 \cdot (s_1^2 + s_2^2)$ .

En cada columna de las Figuras 3.6 y 3.7 se puede ver la representación de los datos frente a cada una de las coordenadas para las tres tendencias simuladas en este estudio.

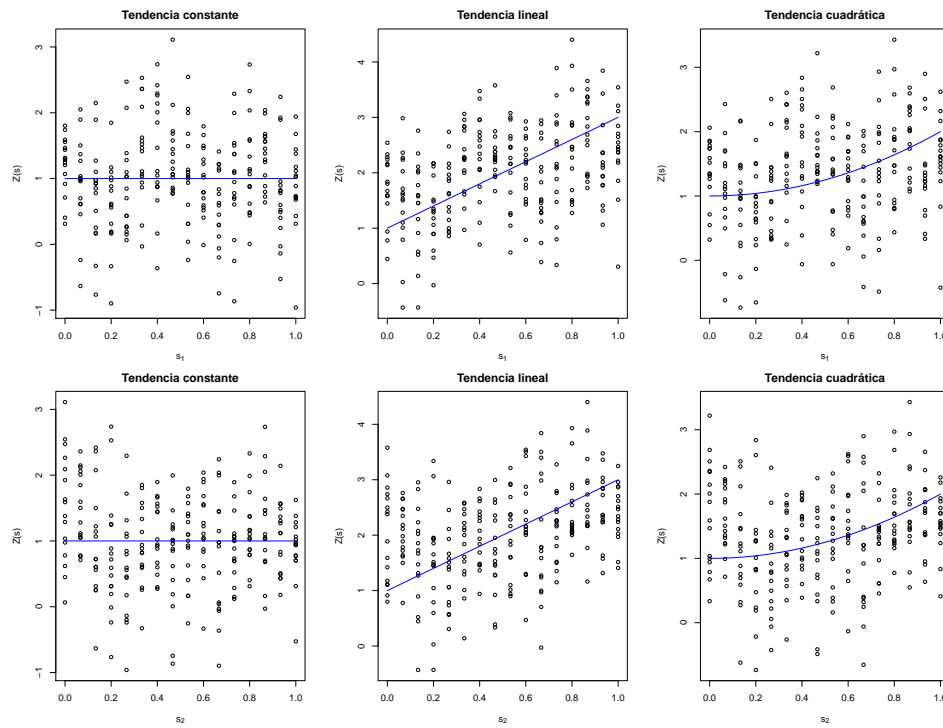


Figura 3.6: Datos para una muestra de tamaño  $n = 256$  bajo diseño fijo frente a la coordenada  $s_1$  (arriba) y  $s_2$  (abajo) y tendencia real (azul).

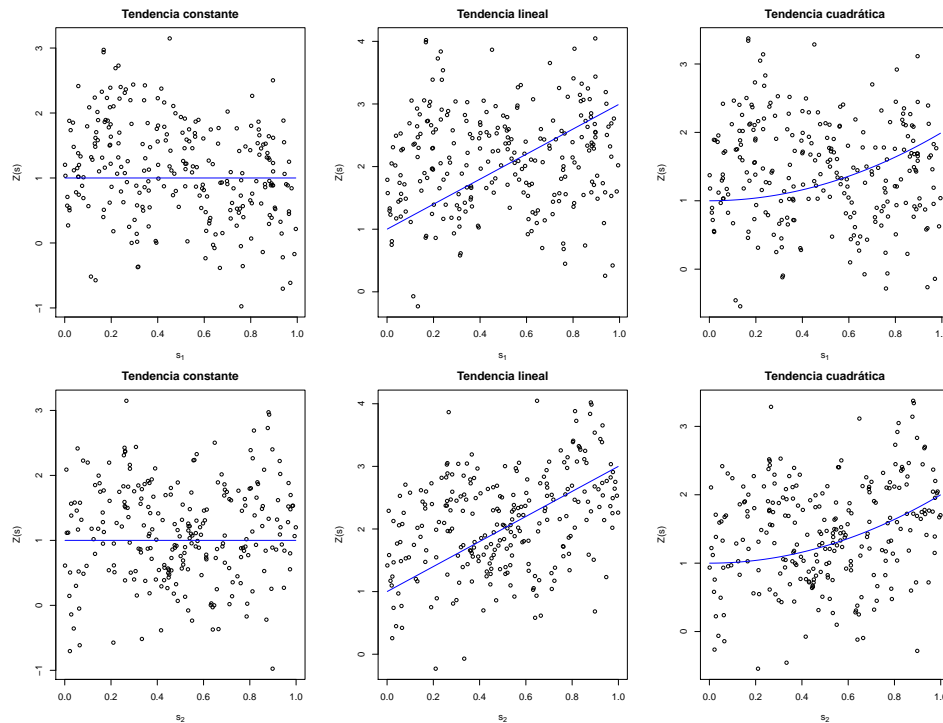


Figura 3.7: Datos para una muestra de tamaño  $n = 256$  bajo diseño aleatorio frente a la coordenada  $s_1$  (arriba) y  $s_2$  (abajo) y tendencia real (azul).

### 3.4. Predicción

El siguiente objetivo es el de comparar las predicciones que se obtienen a partir de la estimación no paramétrica de la tendencia con las predicciones que se obtienen al emplear la técnica clásica de kriging universal para la predicción a partir de la suposición de que la tendencia se puede modelizar de una determinada forma paramétrica.

Para poder llevar a cabo la predicción del proceso a partir de una estimación no paramétrica de la tendencia en las localizaciones muestreadas se recurrirá a aplicar la técnica del kriging ordinario al proceso residual, es decir, al proceso resultante de eliminar la tendencia al proceso original, ya que se puede suponer que es un proceso con media constante y desconocida.

Dado el proceso espacial  $Z(s) = \mu(s) + Y(s)$  y dada una estimación no paramétrica de la tendencia,  $\tilde{\mu}(s)$ , se define el proceso residual como  $R(s) = Z(s) - \tilde{\mu}(s)$ . De esta manera, la predicción del proceso en una nueva localización,  $s^0$ , será  $\tilde{Z}(s^0) = \tilde{\mu}(s^0) + \tilde{R}(s^0)$  y el error cuadrático medio de predicción en una nueva localización  $s^0$  vendrá dado por

$$ECMP(s^0) = \frac{1}{n} \sum_{j=1}^M \left( Z^{(j)}(s^0) - \tilde{Z}^{(j)}(s^0) \right)^2,$$

donde  $\tilde{Z}^{(j)}(s^0)$  es la predicción del proceso obtenida en la localización  $s^0$  para la muestra  $j$ -ésima.

Se supondrá que el proceso espacial  $\{Z(s)/s \in D \subset \mathbb{R}^d\}$  presenta una tendencia de la forma

$$\mu(s) = \mu(s_1, s_2) = 1 + a(s_1 + s_2) + b(s_1^2 + s_2^2),$$

de manera que al variar las constantes  $a$  y  $b$  se tendrán los distintos modelos de tendencia. Además, con el fin de analizar la variación de los resultados en función de la estructura de dependencia, se han considerado dos modelos isotrópicos exponenciales distintos: el modelo de variograma isotrópico exponencial de parámetros  $c_0 = 0$ ,  $\sigma^2 = 0.5$  y  $c_2 = 0.2$  de la ecuación (3.1) y otro modelo con efecto pepita  $c_0 = 0.2$ , mayor umbral  $\sigma^2 = 1$ , y un rango más grande,  $c_2 = 0.4$ , es decir

$$\gamma_{c_0, c_1, c_2}(d) = \begin{cases} 0, & \text{si } d = 0 \\ 0.2 + 0.8 \left( 1 - \exp\left(\frac{-3d}{0.4}\right) \right), & \text{en otro caso.} \end{cases} \quad (3.2)$$

Así, se simularán  $M = 1000$  muestras de tamaño  $n$  de datos espaciales  $Z(s_i)$  en las localizaciones simuladas mediante diseño fijo o aleatorio a partir de un proceso gaussiano de media  $\mu(s)$  y con modelo de variograma (3.1) o (3.2). Para cada una de las 1000 muestras se ha calculado el valor de la predicción del proceso en las 5 localizaciones objetivo  $s^k$ ,  $k = 0, \dots, 4$  y su correspondiente error cuadrático de predicción.

Los modelos de tendencia simulados son los siguientes:

Modelo (1): Si  $a = 0$  y  $b = 0$  se trata de un proceso estacionario con tendencia constante igual a 1. La predicción se ha llevado a cabo aplicando kriging ordinario al proceso residual y al proceso original. Los resultados correspondientes se muestran en las Tablas 3.16, 3.19, 3.22 y 3.25.

Modelo (2): Si  $a \neq 0$  y  $b = 0$  la tendencia del proceso es lineal. En este caso la predicción se ha realizado usando kriging universal con tendencia lineal en las localizaciones sobre el proceso original y kriging ordinario sobre el proceso residual. En particular se ha tomado  $a = 1$  y los resultados están en las Tablas 3.17, 3.20, 3.23 y 3.26.

Modelo (3): Si  $a = 0$  y  $b \neq 0$  la tendencia es cuadrática y para la predicción se ha aplicado kriging ordinario sobre el proceso residual y kriging universal con la suposición de que la tendencia es lineal sobre el proceso original. Es decir, es un ejemplo de qué ocurre cuando hay una mala especificación del modelo de tendencia. Los resultados para distintos valores de  $b$  y  $a = 0$  se muestran en las Tablas 3.18, 3.21, 3.24 y 3.27.

Además, para el caso de tamaño muestral  $n = 256$  se han representado los boxplot de los errores cuadráticos obtenidos en las  $M = 1000$  muestras simuladas. Dichas representaciones se muestran en las Figuras 3.8 y 3.9 para el modelo con semivariograma (3.1) bajo diseño fijo y aleatorio, respectivamente. Por otra parte, las Figuras 3.10 y 3.11 muestran resultados análogos para el modelo de semivariograma (3.2). Para una mejor interpretación de los boxplot se han eliminado los datos atípicos.

Tanto para los modelos (1) como (2) de tendencia, el mejor comportamiento se obtiene para el KU cuando las localizaciones están alejadas del centro del soporte. Por otra parte, la propuesta tiene un menor error en el centro y cuando aumenta el tamaño muestral consigue mejores resultados en puntos alejados del centro (ver Tablas 3.16, 3.17, 3.22 y 3.23). Este mismo efecto se aprecia mejor en las Tablas 3.19, 3.20, 3.25 y 3.26 correspondientes al diseño aleatorio.

Se puede ver que para los casos donde la especificación de la tendencia en el kriging ordinario o universal es correcta, ambos métodos de predicción presentan resultados muy similares. Mientras que para el modelo (3) de tendencia, cuando se hace una mala especificación de la tendencia en el kriging universal del proceso original, los errores cometidos son más grandes que los del kriging ordinario del proceso residual. Esto se hace mucho más evidente cuando la constante  $b$  que multiplica a la parte cuadrática de la tendencia se hace más grande.

Como es lógico, al comparar los resultados correspondientes para el semivariograma de la ecuación (3.1) con el de la ecuación (3.2), los errores crecen considerablemente. Esto es debido a que aumenta el efecto pepita de 0 a 0.2 y la meseta parcial de 0.5 a 0.8, obteniéndose una varianza del proceso de 1.

En general, el error en la predicción disminuye al aumentar el tamaño muestral y aumenta a medida que la localización objetivo se aleja del centro de la región de observación.

Entre las simulaciones bajo diseño fijo y diseño aleatorio no se aprecian grandes diferencias en cuanto a conclusiones. El error de predicción cometido sobre un rejilla regular es siempre inferior al cometido sobre una rejilla irregular con el mismo número de puntos.

## Conclusiones

En definitiva, el método propuesto para estimar la tendencia presenta una clara ventaja frente al estimador de NW al ser consistente. Además, es competitivo cuando la tendencia del proceso se conoce frente al modelo clásico de kriging universal y no está afectado por una mala especificación del modelo de tendencia, como sí le ocurre al kriging universal.

(1)	KO proceso residual					KO proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	37.59	35.41	<b>19.00</b>	<b>17.12</b>	<b>16.04</b>	<b>37.29</b>	<b>35.37</b>	19.35	17.54	16.67
256	<b>26.95</b>	<b>21.07</b>	<b>26.95</b>	<b>15.06</b>	<b>18.32</b>	27.23	21.71	27.29	15.98	18.67
400	<b>21.62</b>	16.83	<b>9.08</b>	<b>13.48</b>	<b>17.11</b>	21.86	<b>16.73</b>	9.81	14.05	17.69

Tabla 3.16:  $100 \times ECMP(s^k)$ ,  $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para el modelo de tendencia (1) y de semivariograma (3.1).

(2)	KO proceso residual					KU proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	37.34	35.47	20.29	18.93	18.90	<b>37.25</b>	<b>35.18</b>	<b>18.57</b>	<b>16.95</b>	<b>16.21</b>
256	<b>26.82</b>	<b>21.18</b>	<b>26.54</b>	<b>15.19</b>	18.87	26.93	21.51	27.03	15.45	<b>18.49</b>
400	<b>21.45</b>	<b>16.67</b>	<b>9.04</b>	<b>13.59</b>	18.48	21.81	16.78	9.45	13.65	<b>17.44</b>

Tabla 3.17:  $100 \times ECMP(s^k)$ ,  $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para el modelo de tendencia (2) y de semivariograma (3.1).

(3)	KO proceso residual					KU proceso original					
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	
$b = 1$	100	37.34	35.47	20.29	18.93	18.90	<b>37.25</b>	<b>35.18</b>	<b>18.57</b>	<b>16.95</b>	<b>16.21</b>
	256	<b>26.82</b>	<b>21.18</b>	<b>26.54</b>	<b>15.19</b>	18.87	26.93	21.51	27.03	15.45	<b>18.49</b>
	400	<b>21.45</b>	<b>16.67</b>	<b>9.04</b>	<b>13.59</b>	18.48	21.81	16.78	9.45	13.65	<b>17.44</b>
$b = 4$	100	<b>37.44</b>	35.69	22.53	19.81	18.21	38.11	<b>34.69</b>	<b>16.43</b>	<b>16.00</b>	<b>16.90</b>
	256	28.65	25.01	28.80	19.71	23.48	<b>26.73</b>	<b>20.88</b>	<b>27.03</b>	<b>15.12</b>	<b>18.67</b>
	400	23.70	18.63	13.64	16.69	22.04	<b>21.77</b>	<b>16.80</b>	<b>9.11</b>	<b>13.49</b>	<b>17.92</b>
$b = 10$	100	<b>38.31</b>	<b>35.10</b>	<b>16.55</b>	<b>15.94</b>	<b>17.20</b>	461.96	3414.72	4735.28	5000.31	5031.05
	256	<b>26.79</b>	<b>21.76</b>	<b>26.80</b>	<b>16.05</b>	<b>18.81</b>	409.28	3248.04	4471.41	4785.01	4783.16
	400	<b>22.52</b>	<b>18.38</b>	<b>12.77</b>	<b>15.11</b>	<b>20.29</b>	396.86	3170.13	4432.32	4756.93	4763.30

Tabla 3.18:  $100 \times ECMP(s^k)$ ,  $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para el modelo de tendencia (3) y de semivariograma (3.1).

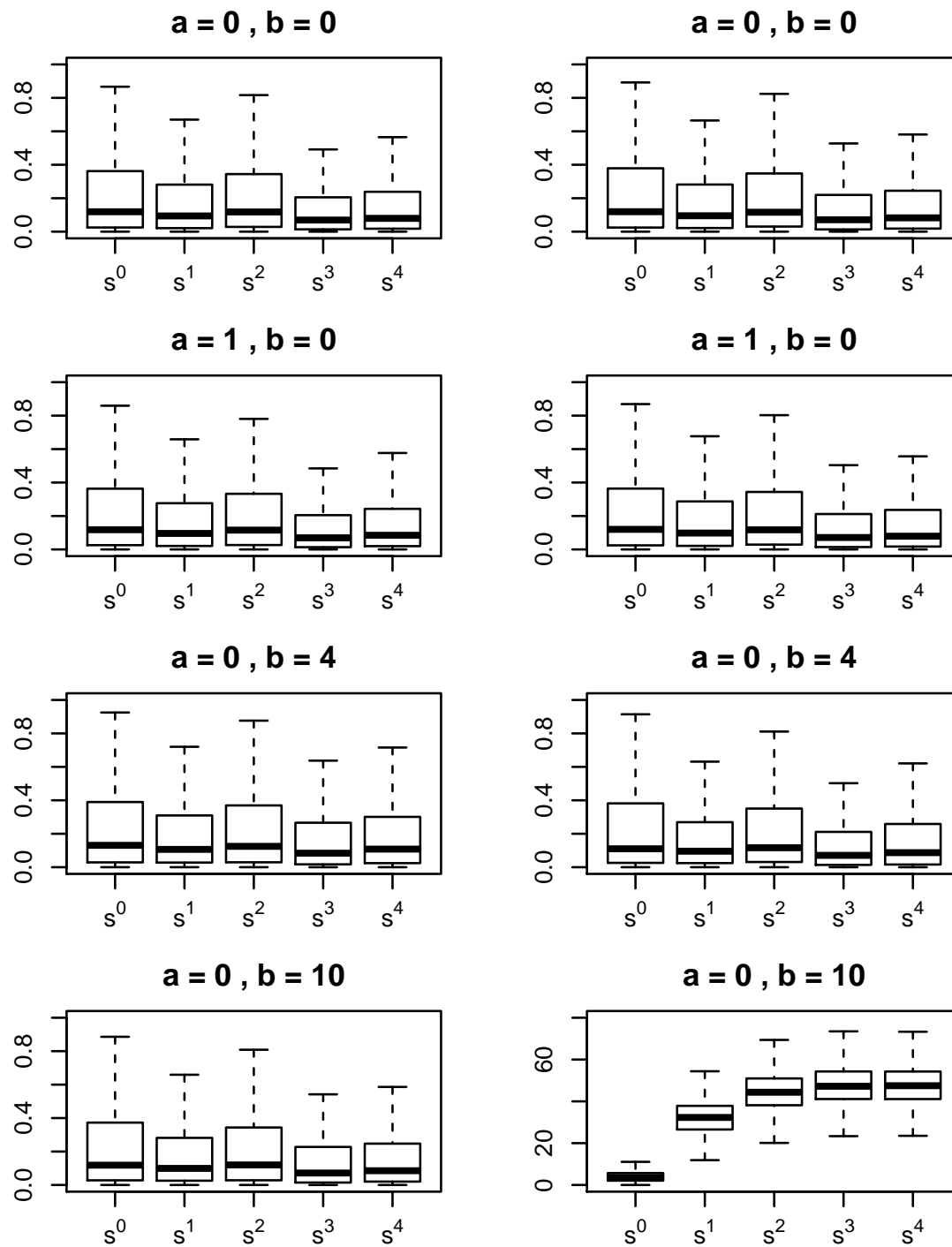


Figura 3.8: Boxplot de los 1000 errores cuadráticos de predicción para muestras de tamaño  $n = 256$  bajo diseño fijo y modelo de semivariograma (3.1).

(1)	KO proceso residual					KO proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	<b>33.36</b>	<b>33.98</b>	33.81	43.18	50.05	33.42	34.06	<b>33.76</b>	<b>42.39</b>	<b>49.18</b>
256	<b>23.10</b>	<b>25.36</b>	<b>23.43</b>	33.55	42.31	23.16	25.54	23.84	<b>32.99</b>	<b>42.09</b>
400	21.80	<b>18.58</b>	<b>20.23</b>	31.21	40.33	<b>21.69</b>	18.70	20.50	<b>30.26</b>	<b>38.80</b>

Tabla 3.19:  $100 \times ECMP(s^k), k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para el modelo de tendencia (1) y de semivariograma (3.1).

(2)	KO proceso residual					KU proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	33.56	34.67	36.36	54.58	64.80	<b>33.46</b>	<b>33.80</b>	<b>34.98</b>	<b>44.45</b>	<b>49.00</b>
256	<b>22.85</b>	<b>22.30</b>	23.55	41.59	56.22	23.05	22.54	<b>22.92</b>	<b>36.40</b>	<b>46.09</b>
400	<b>18.30</b>	20.10	20.39	31.43	48.00	18.31	<b>19.91</b>	<b>20.37</b>	<b>27.59</b>	<b>40.01</b>

Tabla 3.20:  $100 \times ECMP(s^k), k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para el modelo de tendencia (2) y de semivariograma (3.1).

(3)	KO proceso residual					KU proceso original					
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	
$b = 1$	100	<b>33.84</b>	<b>34.85</b>	<b>35.34</b>	<b>44.94</b>	<b>49.68</b>	34.28	35.52	35.87	50.59	56.79
	256	<b>22.91</b>	<b>22.46</b>	<b>23.17</b>	<b>36.99</b>	<b>47.14</b>	25.60	24.39	25.29	40.90	53.36
	400	<b>18.37</b>	<b>20.06</b>	<b>20.29</b>	<b>28.11</b>	<b>41.23</b>	20.34	22.20	23.21	31.47	45.65
$b = 4$	100	<b>34.34</b>	35.22	<b>35.13</b>	<b>52.50</b>	<b>60.01</b>	33.51	<b>34.41</b>	35.41	55.32	63.13
	256	25.39	24.53	25.17	<b>41.54</b>	<b>54.15</b>	<b>22.58</b>	<b>22.85</b>	<b>23.02</b>	42.10	55.77
	400	19.42	21.05	21.67	31.58	<b>47.01</b>	<b>18.08</b>	<b>19.86</b>	<b>19.90</b>	<b>31.15</b>	47.81
$b = 10$	100	<b>49.38</b>	<b>72.61</b>	<b>66.18</b>	<b>89.26</b>	<b>101.07</b>	264.65	2002.65	2742.58	2928.39	2945.06
	256	<b>23.44</b>	<b>23.10</b>	<b>23.67</b>	<b>43.22</b>	<b>57.05</b>	247.44	2019.63	2829.24	2946.99	2960.37
	400	<b>19.03</b>	<b>21.16</b>	<b>21.66</b>	<b>32.38</b>	<b>49.69</b>	240.42	2086.37	2885.34	3071.61	3060.49

Tabla 3.21:  $100 \times ECMP(s^k), k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para el modelo de tendencia (3) y de semivariograma (3.1).



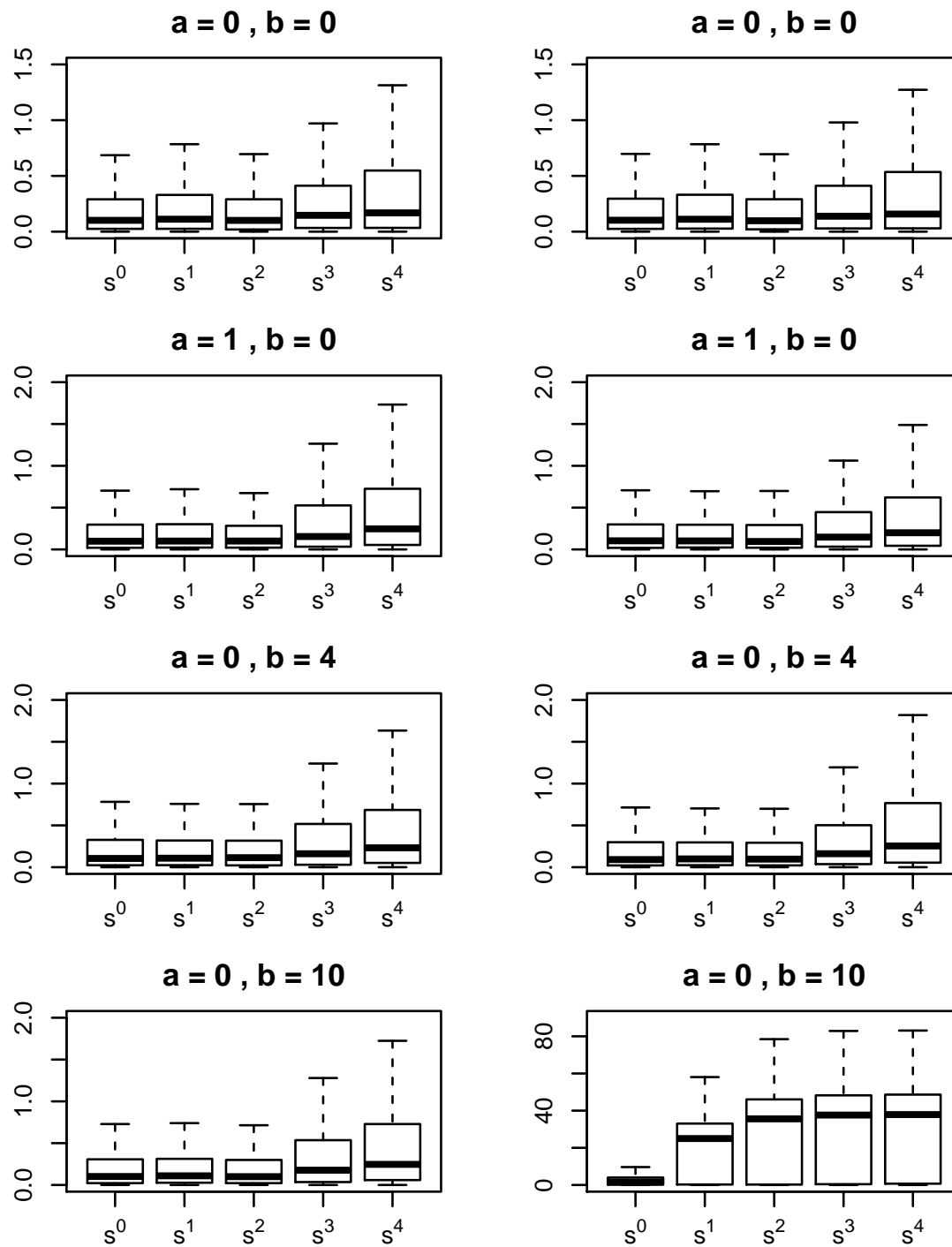


Figura 3.9: Boxplot de los 1000 errores cuadráticos de predicción para muestras de tamaño  $n = 256$  bajo diseño aleatorio y modelo de semivariograma (3.1).

(1)	KO proceso residual					KO proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	<b>67.42</b>	<b>67.68</b>	<b>48.12</b>	47.74	48.63	68.33	68.04	48.35	<b>47.64</b>	<b>48.16</b>
256	55.73	<b>48.88</b>	<b>55.34</b>	44.60	53.75	<b>55.37</b>	49.49	55.40	<b>44.47</b>	<b>53.44</b>
400	<b>49.44</b>	44.82	43.22	<b>41.62</b>	<b>49.13</b>	50.08	<b>43.47</b>	<b>40.94</b>	42.36	50.13

Tabla 3.22:  $100 \times ECMP(s^k)$ ,  $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para el modelo de tendencia (1) y de semivariograma (3.2).

(2)	KO proceso residual					KU proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	67.55	68.58	48.57	48.64	<b>49.45</b>	<b>67.22</b>	<b>67.59</b>	<b>47.14</b>	<b>47.38</b>	49.58
256	55.78	<b>48.83</b>	<b>54.43</b>	44.86	54.39	<b>55.16</b>	49.14	55.57	<b>44.86</b>	<b>53.98</b>
400	<b>49.35</b>	44.82	43.08	41.68	<b>50.43</b>	49.53	<b>44.26</b>	<b>42.25</b>	<b>41.64</b>	50.77

Tabla 3.23:  $100 \times ECMP(s^k)$ ,  $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para el modelo de tendencia (2) y de semivariograma (3.2).

	(3)	KO proceso residual					KU proceso original				
		$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
$b = 1$	100	<b>67.66</b>	68.46	<b>48.55</b>	<b>48.26</b>	<b>48.67</b>	67.81	<b>68.11</b>	50.16	49.75	49.85
	256	55.72	<b>48.83</b>	<b>54.96</b>	<b>44.65</b>	<b>53.95</b>	<b>55.20</b>	48.96	55.69	45.67	54.72
	400	49.43	44.83	42.88	<b>41.75</b>	<b>49.80</b>	<b>49.42</b>	<b>44.02</b>	<b>39.67</b>	42.35	51.52
$b = 4$	100	68.00	<b>68.43</b>	49.25	48.79	<b>49.45</b>	<b>67.40</b>	68.92	<b>46.81</b>	<b>48.63</b>	52.06
	256	55.55	<b>49.08</b>	<b>55.44</b>	<b>45.21</b>	<b>54.50</b>	<b>55.19</b>	50.49	57.42	46.20	54.60
	400	49.51	44.78	42.00	43.34	52.59	<b>49.34</b>	<b>43.37</b>	<b>40.46</b>	<b>40.16</b>	<b>50.60</b>
$b = 4$	100	<b>129.00</b>	<b>187.45</b>	<b>152.97</b>	<b>128.11</b>	<b>127.38</b>	518.74	3446.93	4781.11	5018.42	5062.61
	256	<b>54.86</b>	<b>48.22</b>	<b>54.71</b>	<b>44.66</b>	<b>53.34</b>	458.65	3288.10	4468.57	4784.51	4779.30
	400	<b>49.22</b>	<b>44.00</b>	<b>39.47</b>	<b>42.12</b>	<b>52.48</b>	455.99	3161.19	4383.79	4741.78	4762.16

Tabla 3.24:  $100 \times ECMP(s^k)$ ,  $k = 0, \dots, 4$  con  $M = 1000$  bajo diseño fijo para el modelo de tendencia (3) y de semivariograma (3.2).

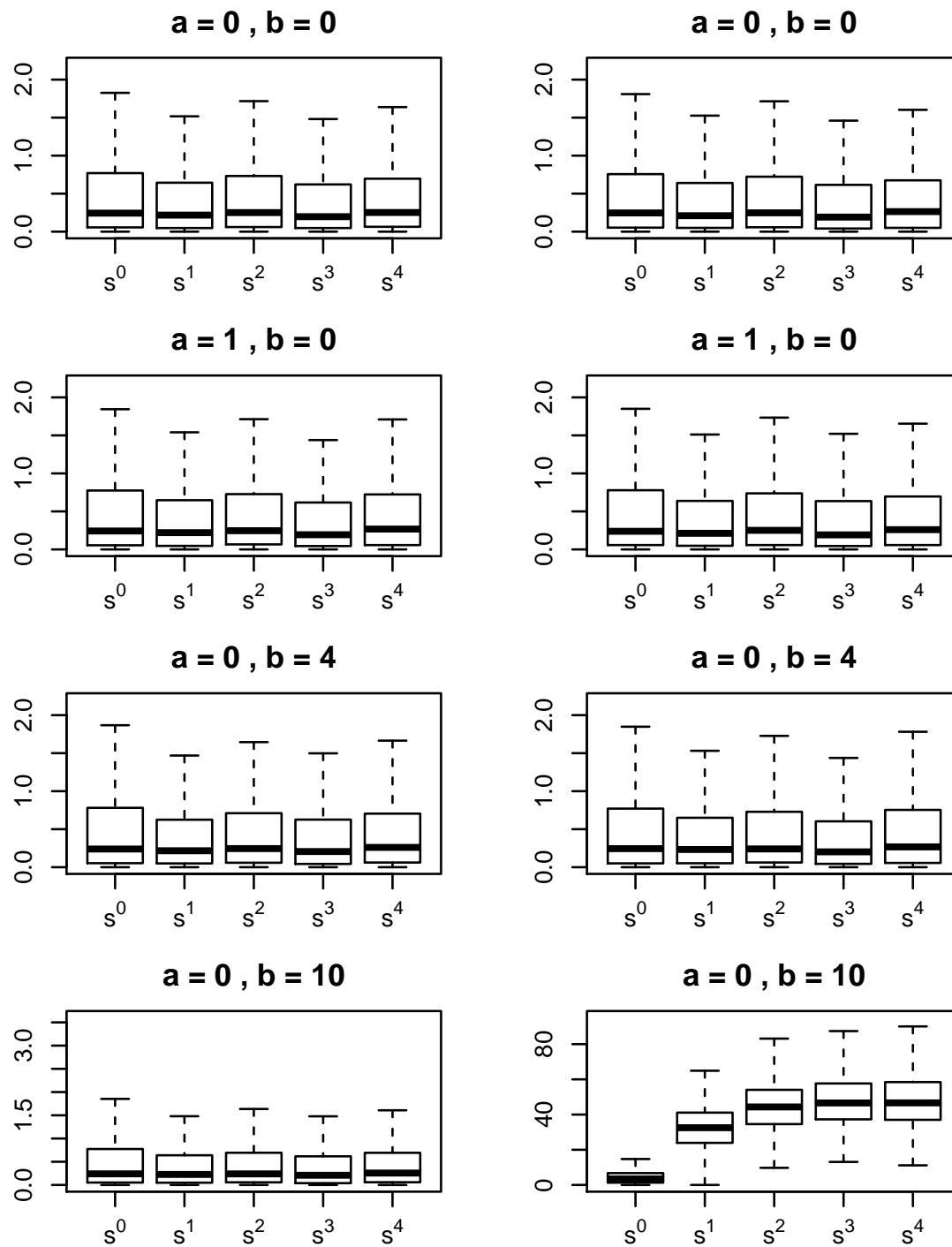


Figura 3.10: Boxplot de los 1000 errores cuadráticos de predicción para muestras de tamaño  $n = 256$  bajo diseño fijo y modelo de semivariograma (3.2).

(1)	KO proceso residual					KO proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	66.15	<b>66.12</b>	68.20	<b>90.26</b>	98.32	<b>65.98</b>	66.51	<b>67.74</b>	91.36	<b>97.25</b>
256	<b>51.25</b>	<b>50.31</b>	<b>51.16</b>	77.18	95.21	51.47	50.46	51.24	<b>76.39</b>	<b>93.26</b>
400	43.73	<b>47.75</b>	47.69	61.82	83.55	<b>43.54</b>	47.90	<b>47.48</b>	<b>61.13</b>	<b>80.33</b>

Tabla 3.25:  $100 \times ECMP(s^k), k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para el modelo de tendencia (1) y de semivariograma (3.2).

(2)	KO proceso residual					KU proceso original				
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$
100	<b>65.91</b>	66.66	<b>69.53</b>	97.85	111.43	66.40	<b>65.79</b>	69.67	<b>96.20</b>	<b>108.29</b>
256	<b>51.23</b>	<b>50.13</b>	<b>50.82</b>	<b>78.85</b>	<b>97.54</b>	51.33	50.27	51.51	81.97	103.86
400	43.80	47.80	47.62	65.31	89.83	<b>43.49</b>	<b>47.29</b>	<b>47.01</b>	<b>61.31</b>	<b>83.07</b>

Tabla 3.26:  $100 \times ECMP(s^k), k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para el modelo de tendencia (2) y de semivariograma (3.2).

(3)	KO proceso residual					KU proceso original					
	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	$s^0$	$s^1$	$s^2$	$s^3$	$s^4$	
$b = 1$	100	66.20	66.97	68.63	<b>91.21</b>	<b>100.62</b>	<b>65.83</b>	<b>66.85</b>	<b>68.22</b>	95.40	105.05
	256	<b>51.24</b>	<b>50.32</b>	<b>51.50</b>	<b>78.39</b>	<b>96.51</b>	52.18	50.39	51.77	79.50	99.25
	400	43.73	<b>47.85</b>	<b>47.61</b>	<b>62.38</b>	<b>83.59</b>	<b>43.72</b>	48.26	48.32	62.90	84.45
$b = 4$	100	<b>66.06</b>	<b>67.34</b>	<b>68.54</b>	<b>102.70</b>	<b>117.46</b>	66.73	68.18	69.52	105.96	120.87
	256	52.03	50.62	52.26	81.90	103.80	<b>49.88</b>	<b>49.79</b>	<b>51.13</b>	<b>81.53</b>	<b>102.32</b>
	400	43.45	47.71	47.81	65.54	89.57	<b>42.88</b>	<b>46.96</b>	<b>46.41</b>	<b>64.25</b>	<b>88.69</b>
$b = 10$	100	<b>141.77</b>	<b>255.13</b>	<b>221.36</b>	<b>211.19</b>	<b>219.19</b>	304.99	2089.80	2865.76	3047.56	3080.55
	256	<b>51.31</b>	<b>49.86</b>	<b>51.49</b>	<b>84.89</b>	<b>108.38</b>	301.93	2169.61	2975.30	3070.06	3110.82
	400	<b>43.24</b>	<b>48.12</b>	<b>48.43</b>	<b>67.31</b>	<b>94.72</b>	272.14	2121.59	2942.32	3170.16	3153.93

Tabla 3.27:  $100 \times ECMP(s^k), k = 0, \dots, 4$  con  $M = 1000$  bajo diseño aleatorio para el modelo de tendencia (3) y de semivariograma (3.2).

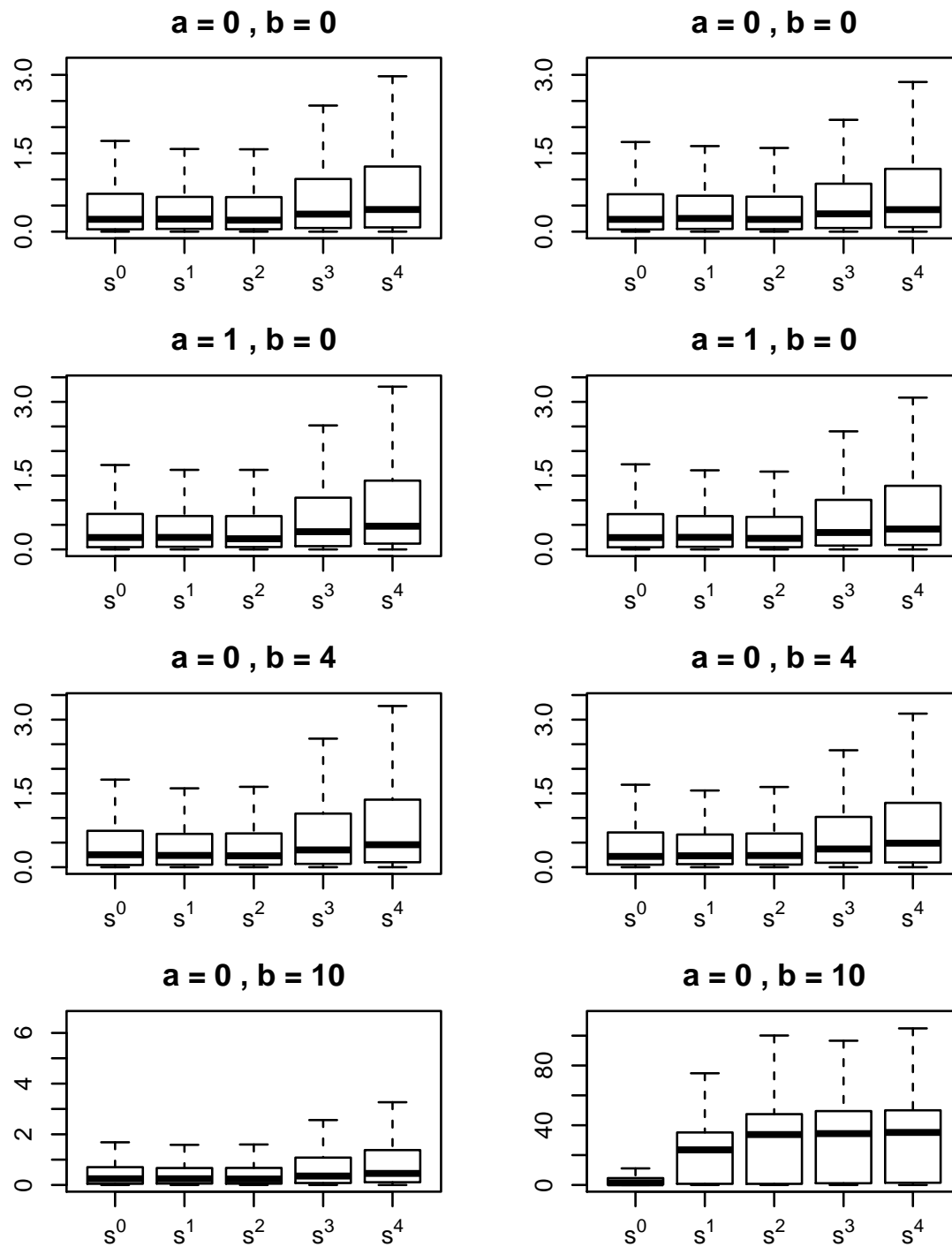


Figura 3.11: Boxplot de los 1000 errores cuadráticos de predicción para muestras de tamaño  $n = 256$  bajo diseño aleatorio y modelo de semivariograma (3.2).

### 3.5. Aplicación a datos reales

Para ilustrar la estimación de la tendencia de manera no paramétrica y su uso en la predicción para un conjunto de datos reales se tomará como ejemplo el conjunto `parana` de la librería `geoR`. Este conjunto de datos fue utilizado por Diggle y Ribeiro (2001) y recoge la precipitación promedio en el período de Mayo a Junio para diferentes años. Dichos datos fueron recogidos en 143 estaciones de medición en todo el estado de Paraná, Brasil. En la Figura 3.12 se muestra la región de observación y las localizaciones donde se han medido los datos. Las localizaciones donde se pretende obtener la predicción del proceso se muestran en color rojo en la Figura 3.12.

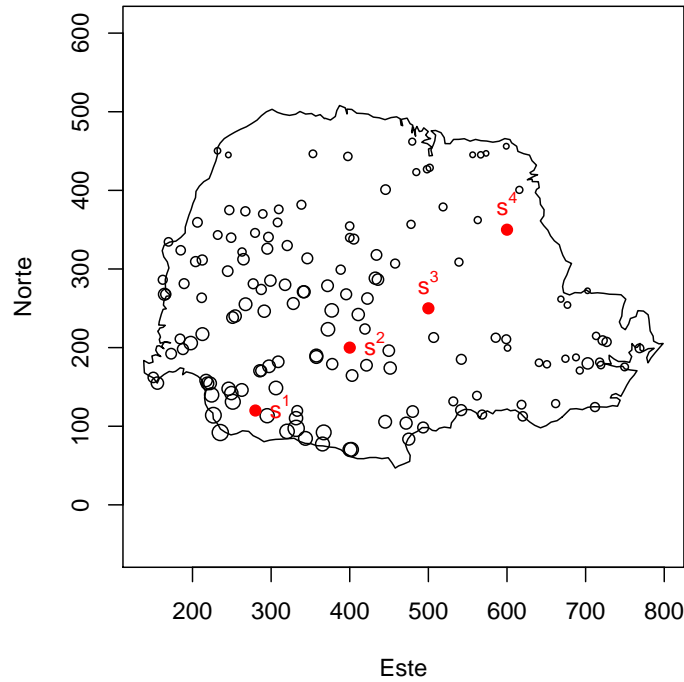


Figura 3.12: Región de observación, localizaciones muestrales con el tamaño del punto en función del valor observado y nuevas localizaciones (en rojo).

Si se examina la nube de puntos que constituyen las coordenadas y las observaciones (Figura 3.13) se puede ver que el proceso no es estacionario y que parece apropiado modelizar la tendencia como una función lineal en las coordenadas,  $\mu(s) = b_0 + b_1 s_1 + b_2 s_2$ . Esto también se puede comprobar en la Figura 3.14, donde se representan los datos frente a cada coordenada. Así, para comparar las predicciones obtenidas mediante el método propuesto con las obtenidas usando la técnica clásica de kriging universal se podrá suponer que la tendencia es lineal.

En primer lugar se calcula la estimación no paramétrica de la tendencia de acuerdo con el estimador  $\tilde{\mu}(s)$ . Para ello se utilizan la ventana percentil de orden 0.2 para obtener la primera estimación de la tendencia en las localizaciones muestrales y la ventana seleccionada por validación cruzada modificada para la segunda etapa. En este caso la matriz de ventanas ha resultado ser  $H = \text{diag}(541.1, 352.4)$ . La comparación entre la estimación no paramétrica de la tendencia y la estimación de la tendencia lineal en el kriging universal se puede ver en la Tabla 3.28.

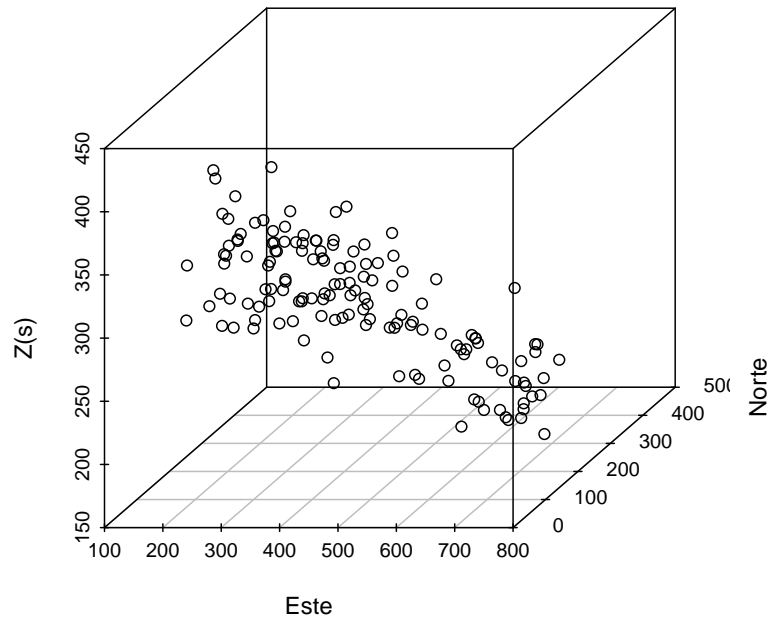


Figura 3.13: Datos de precipitación en 143 estaciones de medición en Paraná.

A continuación se calcula el proceso residual restando a los datos originales la estimación no paramétrica de la tendencia y se le aplica kriging ordinario. Para ello se ha ajustado un variograma gaussiano por mínimos cuadrados. Los parámetros del modelo resultantes son un efecto pepita de  $c_0 = 528.27$ ,  $\sigma^2 = 2479.07$  y rango  $c_2 = 269.52$ . Una vez obtenidas las predicciones de dicho proceso residual en las localizaciones objetivo se le suma la estimación de la tendencia en cada localización, obteniéndose los resultados de la primera fila de la Tabla 3.29.

Por otro lado se ajustó un variograma esférico a los datos originales por mínimos cuadrados, obteniéndose como parámetros ajustados  $c_0 = 363.44$ ,  $\sigma^2 = 803.12$  y  $c_2 = 266.37$ . A partir de este variograma se ha aplicado kriging universal con suposición de tendencia lineal a los datos originales. Con este método se obtienen las predicciones de la segunda fila de la Tabla 3.29 y los parámetros estimados para  $b_0$ ,  $b_1$  y  $b_2$  de tal modo que la tendencia lineal estimada resulta  $\hat{\mu}(s) = 424.5928 - 0.1465s_1 - 0.4006s_2$ . Así, se ha obtenido la estimación de la tendencia por kriging universal en las nuevas localizaciones (segunda fila de la Tabla 3.28).

	$s^1$	$s^2$	$s^3$	$s^4$
$\tilde{\mu}(s)$	306.03	287.75	272.92	250.34
Lineal	336.92	286.69	251.51	196.30

Tabla 3.28: Estimación de la tendencia en las nuevas localizaciones.

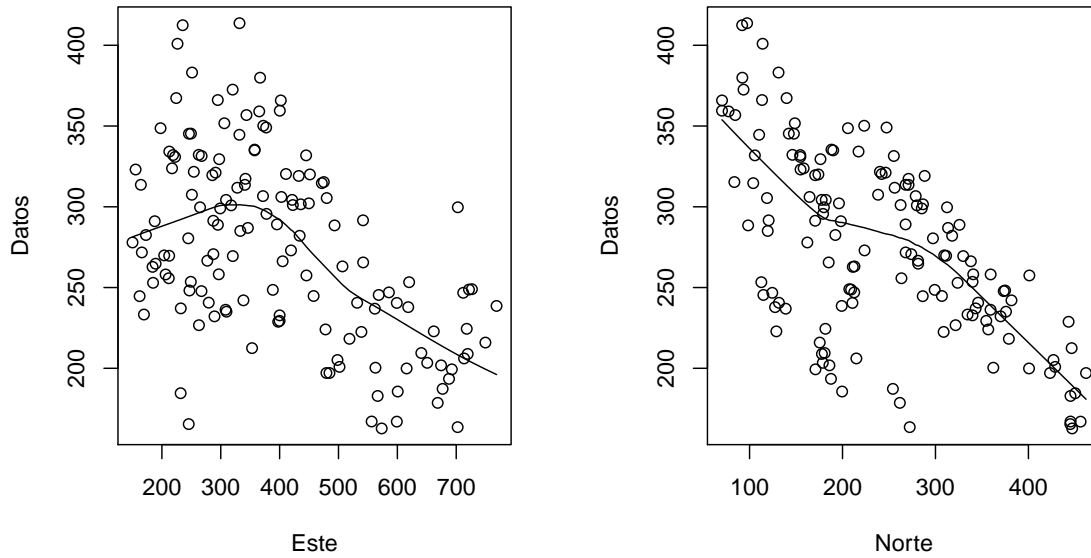


Figura 3.14: Datos de precipitación frente a cada una de las coordenadas junto con el ajuste de la curva de regresión polinómica local ponderada.

	$s^1$	$s^2$	$s^3$	$s^4$
KO residual	361.14	316.90	259.56	190.50
KU original	363.44	312.65	261.57	195.20

Tabla 3.29: Predicción en las nuevas localizaciones.

Las localizaciones están ordenadas de la esquina inferior izquierda hasta la esquina superior derecha de la región de observación, que es el mismo sentido en el que decrecen los valores de la precipitación promedio. Cabe esperar entonces que los valores de las predicciones vayan disminuyendo, como se ve en la Tabla 3.29, donde ambos métodos de predicción aportan resultados similares. El kriging del proceso residual da lugar a predicciones más pequeñas salvo en la localización  $s^2$ , mientras que en la estimación de la tendencia aportaba valores más grandes para todas las localizaciones salvo para  $s^1$ .



# Apéndice A

## Código de R

```
## Simulación estudio ventanas (sección 3.2) y comparación estimadores (sección 3.3)

# Librerías
library(geoR); library(locpol); library(npsp);

# Semilla
set.seed(123456)

n <- 10
ndata <- n*n # Tamaño muestral
nsim <- 1000 # Número de simulaciones

loc <- "irreg" # Diseño: reg (fijo) o irreg (aleatorio)
tend <- 1 # Tendencia: 0 cte, 1 lineal, 2 no lineal
a <- 0.5 # Coeficiente tendencia cuadrática

# parámetros modelo de semivariograma
tau <- 0 # pepita
sig2 <- 0.5 # sigma^2
phi <- 0.2/3 # rango o alcance

# nbin para el binning de los datos
nmbin <- ifelse(n==10,5,ifelse(n==16,8,10))

# Metodo de seleccion de ventanas
hicv <- FALSE ; hip <- TRUE ; hictes <- FALSE # ventana h_i
hcv <- FALSE ; hctes <- TRUE # matriz de ventanas H

# Cálculo del estimador de NW
hat <- FALSE

# Nuevas localizaciones
xnew <- matrix(c(0.5,0.5,0.4,0.4,0.25,0.25,0.1,0.1,0.01,0.01,-0.01,-0.01), ncol=2, byrow=T)

# Simulación en rejilla regular
if (loc=="reg") {
if (mod==1) sim1 <- grf(ndata, grid="reg", cov.model="exp", cov.pars=c(sig2, phi),
  nugget=tau, nsim=nsim, messages=FALSE)
if (tend==0){ media <- 1
```

```

medianew <- rep(1,nrow(xnew))}
if (tend==1){ media <- as.matrix(sim1$coords) %*% c(1,1) + 1
medianew <- xnew %*% c(1,1) + 1}
if (tend==2){ media <- as.matrix(a*(sim1$coords*sim1$coords)) %*% c(1,1) + 1
medianew <- a*(xnew*xnew) %*% c(1,1) + 1}
} # loc=="reg"

# Reserva de memoria para las distintas opciones de ventana
if (hcv==TRUE){
if (hicv==TRUE) { tilde.hicv.hcv <- matrix(0, nrow=nsim, ncol=nrow(xnew)) }
if (hip==TRUE) { tilde.p.hcv <- matrix(0, nrow=nsim, ncol=nrow(xnew))
p=0.2 }
if (hictes==TRUE) { hivec <- seq(0.005,1.2,by=0.02) # Rejilla de valores constantes para hi
tilde.ctes.hcv <- matrix(0, nrow=nsim, ncol=length(hivec)*nrow(xnew)) }
if (hat==TRUE) hat.hcv <- matrix(0, nrow=nsim, ncol=nrow(xnew))}

if (hctes==TRUE){
hvec <- hvec <- seq(0.1,1,len=21)
if (hicv==TRUE) { tilde.hicv.hctes <- matrix(0, nrow=nsim, ncol=nrow(xnew)*length(hvec))}
if (hip==TRUE) { tilde.p.hctes <- matrix(0, nrow=nsim, ncol=nrow(xnew)*length(hvec))
p=0.2 }
if (hictes==TRUE) { hivec <- seq(0.005,0.8,len=31) # Rejilla de valores constantes para hi
tilde.ctes.hcv <- array(0, dim=c(nsim,length(hivec)*nrow(xnew), length(hvec))) }
if (hat==TRUE) hat.hctes <- matrix(0, nrow=nsim, ncol=nrow(xnew)*length(hvec))
}

if (hp==TRUE){
pp <- 0.2
if (hicv==TRUE) { tilde.hicv.hp <- matrix(0, nrow=nsim, ncol=nrow(xnew)) }
if (hip==TRUE) { tilde.p.hp <- matrix(0, nrow=nsim, ncol=nrow(xnew)) }

if (hat==TRUE) hat.hp <- matrix(0, nrow=nsim, ncol=nrow(xnew))
}

pb <- txtProgressBar(min = 0, max = nsim , style = 3)

for (is in 1:nsim){

setTxtProgressBar(pb, is)

# Simulacion rejilla irregular
if (loc=="irreg"){
if (mod==1) sim1 <- grf(ndata, grid="irreg", cov.model="exp", cov.pars=c(sig2, phi),
nugget=tau, nsim=1, messages=FALSE)
if (mod==2) sim1 <- grf(ndata, grid="irreg", cov.model="gaussian", cov.pars=c(sig2, phi),
nugget=tau, nsim=1, messages=FALSE)

if (tend==0){ media <- 1
medianew <- rep(1,nrow(xnew))}
if (tend==1){ media <- sim1$coords %*% c(1,1) + 1
medianew <- xnew %*% c(1,1) + 1}
if (tend==2){ media <- a*(sim1$coords*sim1$coords)%*% c(1,1) + 1
medianew <- a*(xnew*xnew) %*% c(1,1) + 1}
datos <- data.frame(s=sim1$coords, z=(sim1$data + media))
}
if (loc=="reg") { datos <- data.frame(sim1$coords, z=(sim1$data[ ,is] + media)) }

```

```
#####

# H global por validación cruzada
if (hcv==TRUE){
if (hicv==TRUE){
# tilde con hi validación cruzada local
hh <- max(dist(datos$z))/6
mu.1 <- double(ndata)
for (i in 1:ndata){
pesos <- EpaK((datos$z[i]-datos$z)/hh)
hi <- regCVBwSelC(x=datos$z, y=datos$z, deg=0, kernel=EpaK, weig=pesos)
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi, EpaK)$beta0
}
newdata <- data.frame(sim1$coords, zz=mu.1)

binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin, nnbin))
h.opt <- h.cv(binned.data,degree=0)
mu.opt <- npsp::locpol(binned.data, h=h.opt$h, degree=0)
tilde.hicv.hcv[is,] <- predict(mu.opt, newx=xnew)$y
} # hicv

if (hictes==TRUE){
# tilde con rejilla de hi constantes
for (ihi in 1:length(hivec)){
hi <- hivec[ihi]
mu.1 <- numeric(ndata)
for (i in 1:ndata){
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi, EpaK)$beta0
} # i
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin, nnbin))
h.opt <- h.cv(binned.data,degree=0)
mu.opt <- npsp::locpol(binned.data, h=h.opt$h, degree=0)
tilde.ctes.hcv[is, (nrow(xnew)*(ihi-1)+1):(nrow(xnew)*ihi)] <- predict(mu.opt, newx=xnew)$y
} # ihi
} # hictes

if(hip==TRUE){
# tilde con hi local p%
hi <- apply(abs(outer(datos$z, datos$z, "-")), 1, sort)[ceiling(p*ndata), ]
mu.1 <- double(ndata)
for (i in 1:ndata){
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi[i], EpaK)$beta0
}
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin,nnbin))
h.opt <- h.cv(binned.data,degree=0)
mu.opt <- npsp::locpol(binned.data, h=h.opt$h, degree=0)
tilde.p.hcv[is,] <- predict(mu.opt, newx=xnew)$y
} # hip

# hat
if (hat==TRUE){
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data2 <- binning(x=newdata[,1:2], y=datos$z, nbin=c(nnbin,nnbin))
```

```

h.opt2 <- h.cv(binned.data2,degree=0)
mu.2.opt <- npsp::locpol(binned.data2, h=h.opt2$h, degree=0)
hat.hcv[is,] <- predict(mu.2.opt, newx=xnew)$y}
} #hcv

#####

# H global en una rejilla de constantes
if (hctes==TRUE){
if (hicv==TRUE){
# tilde con hi validación cruzada local (los pesos dependen de la localizacion)
hh <- max(dist(datos$z))/6
mu.1 <- double(ndata)
for (i in 1:ndata){
pesos <- EpaK((datos$z[i]-datos$z)/hh)
hi <- regCVBwSelC(x=datos$z, y=datos$z, deg=0, kernel=EpaK, weig=pesos)
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi, EpaK)$beta0
}
newdata <- data.frame(sim1$coords, zz=mu.1)

binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin,nnbin))
for (ih in 1:length(hvec)){
mu.opt <- npsp::locpol(binned.data, h=diag(hvec[ih],2), degree=0)
if (anyNA(mu.opt$est)==TRUE) { tilde.hicv.hctes[is, ((ih-1)*nrow(xnew)+1):(ih*nrow(xnew))] <- NA}
if (anyNA(mu.opt$est)==FALSE){ tilde.hicv.hctes[is, ((ih-1)*nrow(xnew)+1):(ih*nrow(xnew))] <-
  predict(mu.opt, newx=xnew)$y}
} #ih
} # hicv

if(hip==TRUE){
# tilde con hi local p%
hi <- apply(abs(outer(datos$z, datos$z, "-")), 1, sort)[ceiling(p*ndata), ]
mu.1 <- double(ndata)
for (i in 1:ndata){
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi[i], EpaK)$beta0
}
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin,nnbin))
for (ih in 1:length(hvec)){
mu.opt <- npsp::locpol(binned.data, h=diag(hvec[ih],2), degree=0)
if (anyNA(mu.opt$est)==TRUE) { tilde.p.hctes[is, ((ih-1)*nrow(xnew)+1):(ih*nrow(xnew))] <- NA}
if (anyNA(mu.opt$est)==FALSE){ tilde.p.hctes[is, ((ih-1)*nrow(xnew)+1):(ih*nrow(xnew))] <-
  predict(mu.opt, newx=xnew)$y}
} #ih
} #hip

if (hictes==TRUE){
# tilde con rejilla de hi constantes
for (ihi in 1:length(hivec)){
hi <- hivec[ihi]
mu.1 <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z, bw=hi, EpaK)$beta0
binned.data <- binning(x=sim1$coords, y=mu.1, nbin=c(nnbin,nnbin))
for (ii in 1:length(hvec)){
hii <- hvec[ii]
mu.opt <- npsp::locpol(binned.data, h=diag(hii,ncol=2, nrow=2), degree=0)
if (anyNA(mu.opt$est)==TRUE) {tilde.ctes.hcv[is, (nrow(xnew)*(ihi-1)+1):(nrow(xnew)*ihi), ii] <- NA}

```

```

if (anyNA(mu.opt$est)==FALSE){tilde.ctes.hcv[is, (nrow(xnew)*(ihi-1)+1):(nrow(xnew)*ihi), ii] <-
  predict(mu.opt, newx=xnew)$y}
} # ihi
} # hictes

# hat
if (hat==TRUE){
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data2 <- binning(x=newdata[,1:2], y=datos$z, nbin=c(nnbin,nnbin))
for (ih in 1:length(hvec)){
mu.2.opt <- npsp::locpol(binned.data2, h=diag(hvec[ih],2), degree=0)
if (anyNA(mu.2.opt$est)==TRUE) {hat.hctes[is, ((ih-1)*nrow(xnew)+1):(ih*nrow(xnew))] <- NA}
if (anyNA(mu.2.opt$est)==FALSE) {hat.hctes[is, ((ih-1)*nrow(xnew)+1):(ih*nrow(xnew))] <-
  predict(mu.2.opt, newx=xnew)$y}
} #ih
}
} # hctes

#####

# H local percentil de orden p
if (hp==TRUE){
if (hicv==TRUE){
# tilde con hi validación cruzada local (los pesos dependen de la localizacion)
hh <- max(dist(datos$z))/6
mu.1 <- double(ndata)
for (i in 1:ndata){
pesos <- EpaK((datos$z[i]-datos$z)/hh)
hi <- regCVBwSelC(x=datos$z, y=datos$z, deg=0, kernel=EpaK, weig=pesos)
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi, EpaK)$beta0
}
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin,nnbin))
for (iis in 1:nrow(xnew)){
h.p <- sort(sqrt(apply((xnew[iis,]-sim1$coords)^2,1,sum)))[ceiling(pp*ndata)]
mu.opt <- npsp::locpol(binned.data, h=diag(h.p,2), degree=0)
tilde.hicv.hp[is,iis] <- predict(mu.opt, newx=xnew)$y[iis]
}
} # hicv

if (hictes==TRUE){
# tilde con rejilla de hi constantes
for (ihi in 1:length(hivec)){
hi <- hivec[ihi]
mu.1 <- numeric(ndata)
for (i in 1:ndata){
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi, EpaK)$beta0
} # i
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data <- binning(x=newdata[, 1:2], y=newdata$zz, nbin=c(nnbin, nnbin))
h.opt <- h.cv(binned.data,degree=0)
mu.opt <- npsp::locpol(binned.data, h=h.opt$h, degree=0)
tilde.ctes.hp[is, (nrow(xnew)*(ihi-1)+1):(nrow(xnew)*ihi)] <- predict(mu.opt, newx=xnew)$y
} # ihi
} # hictes

```

```

if(hip==TRUE){
# tilde con hi local p%
hi <- apply(abs(outer(datos$z, datos$z, "-")), 1, sort)[ceiling(pp*ndata), ]
mu.1 <- double(ndata)
for (i in 1:ndata){
mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi[i], EpaK)$beta0
}
newdata <- data.frame(sim1$coords, zz=mu.1)
binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(nnbin,nnbin))
for (iis in 1:nrow(xnew)){
h.p <- sort(sqrt(apply((xnew[iis,]-sim1$coords)^2,1,sum)))[ceiling(p*ndata)]
mu.opt <- npsp::locpol(binned.data, h=diag(h.p,2), degree=0)
if (anyNA(mu.opt$est)==TRUE) tilde.p.hp[is,iis] <- NA
if (anyNA(mu.opt$est)==FALSE) tilde.p.hp[is,iis] <- predict(mu.opt, newx=xnew)$y[iis]
}
} # hip
# hat
if (hat==TRUE){
binned.data2 <- binning(x=newdata[,1:2], y=datos$z, nbin=c(nnbin,nnbin))
for (iis in 1:nrow(xnew)){
h.p2 <- sort(sqrt(apply((xnew[iis,]-sim1$coords)^2,1,sum)))[ceiling(p*ndata)]
mu.2.opt <- npsp::locpol(binned.data2, h=diag(h.p2,2), degree=0)
if (anyNA(mu.2.opt$est)==TRUE) hat.hp[is,iis] <- NA
if (anyNA(mu.2.opt$est)==FALSE) hat.hp[is,iis] <- predict(mu.2.opt, newx=xnew)$y[iis]
}
} # hat
} #hp
} # is

#####
#####

# Calculo del ECM para h global por validacion cruzada
if (hcv == TRUE){
mnew.mat <- matrix(as.vector(medianew), ncol=nrow(xnew), nrow=nsim, byrow=TRUE)
hat_ecms.hcv <- apply((hat.hcv-mnew.mat)^2, 2, mean)
if (hip==TRUE){
tilde.p_ecms.hcv <- apply((tilde.p.hcv-mnew.mat)^2, 2, mean)}
if (hicv==TRUE){
tilde.hicv_ecms.hcv <- apply((tilde.hicv.hcv-mnew.mat)^2, 2, mean)}
if (hictes==TRUE){
mnew.mat.ctes <- matrix(as.vector(medianew), ncol=nrow(xnew)*length(hivec), nrow=nsim, byrow=TRUE)
tilde.ctes_ecms.hcv <- apply((tilde.ctes.hcv-mnew.mat.ctes)^2, 2, mean)
tilde.ctes_ecms.hcv <- matrix(tilde.ctes_ecms.hcv,ncol=nrow(xnew),byrow=T)
}
} # hcv

round(hat_ecms.hcv*100, 2)
round(tilde.p_ecms.hcv*100, 2)
round(tilde.hicv_ecms.hcv*100, 2)
round(apply(tilde.ctes_ecms.hcv*100,2,min), 2)
#####

```

```

# Calculo del ECM para h global en rejilla de constantes con p y cv para hi
if (hctes == TRUE){
indices <- matrix(1:(nrow(xnew)*length(hvec)),nrow=nrow(xnew),ncol=length(hvec),byrow=F)
if (hat==TRUE) hat_ecms.hctes <- matrix(0,ncol=length(hvec),nrow=nrow(xnew))
if (hip==TRUE) tilde.p_ecms.hctes <- matrix(0,ncol=length(hvec),nrow=nrow(xnew))
if (hicc==TRUE) tilde.hicc_ecms.hctes <- matrix(0,ncol=length(hvec),nrow=nrow(xnew))
for (ii in 1:nrow(xnew)){
if (hat==TRUE) hat_ecms.hctes[ii,] <- apply((hat.hctes[,indices[ii,]]-medianew[ii])^2,2,mean,
na.rm=TRUE)
if (hip==TRUE) tilde.p_ecms.hctes[ii,] <- apply((tilde.p.hctes[,indices[ii,]]-medianew[ii])^2,2,
mean,na.rm=TRUE)
if (hicc==TRUE) {tilde.hicc_ecms.hctes[ii,] <- apply((tilde.hicc.hctes[,indices[ii,]]-medianew[ii])^2,2,
mean,na.rm=TRUE)}
}
} # hctes

round(hat_ecms.hctes*100, 2)
round(tilde.p_ecms.hctes*100, 2)
round(tilde.hicc_ecms.hctes*100, 2)
#####

#####
# Calculo del ECM para h local percentil de orden pp
if (hp == TRUE){
mnew.mat <- matrix(as.vector(medianew), ncol=nrow(xnew), nrow=nsim, byrow=TRUE)
hat_ecms.hp <- apply((hat.hp-mnew.mat)^2, 2, mean, na.rm=TRUE)
if (hip==TRUE) tilde.p_ecms.hp <- apply((tilde.p.hp-mnew.mat)^2, 2, mean, na.rm=TRUE)
if (hicc==TRUE) tilde.hicc_ecms.hp <- apply((tilde.hicc.hp-mnew.mat)^2, 2, mean, na.rm=TRUE)
} # hp

round(hat_ecms.hp*100, 2)
round(tilde.p_ecms.hp*100, 2)
round(tilde.hicc_ecms.hp*100, 2)
#####

## Simulación predicción (sección 3.4)
library(geoR); library(locpol); library(npsp);
set.seed(123456)

# nuevas localizaciones
xnew <- matrix(c(0.5, 0.5, 0.25, 0.25, 0.1, 0.1, 0.01, 0.01, -0.01, -0.01), ncol=2, byrow=T)
colnames(xnew) <- c("s1","s2")
a <- 0; b <- 10

# Simulación en rejilla regular
if (loc=="reg") {
s <- rbind(expand.grid(s1=seq(0,1,len=n),s2=seq(0,1,len=n)),data.frame(xnew))
if (mod==1) sim1 <- grf(grid=s, cov.model="exp", cov.pars=c(sig2, phi), nugget=tau, nsim=nsim,
messages=FALSE)
if (mod==2) sim1 <- grf(grid=s, cov.model="gaussian", cov.pars=c(sig2, phi), nugget=tau, nsim=nsim,
messages=FALSE)
media <- 1 + a*(as.matrix(sim1$coords[1:ndata,]) %*% c(1,1)) +
b*((sim1$coords[1:ndata,]*sim1$coords[1:ndata,]) %*% c(1,1))
medianew <- 1 + a*(sim1$coords[-(1:ndata),] %*% c(1,1)) +
b*((sim1$coords[-(1:ndata),]*sim1$coords[-(1:ndata),]) %*% c(1,1))
} # loc=="reg"

```

```

# Reserva de memoria para los errores
error.ku <- matrix(0,nrow=nsim,ncol=nrow(xnew))
error.res <- matrix(0,nrow=nsim,ncol=nrow(xnew))

pb <- txtProgressBar(min = 0, max = nsim , style = 3)
for (is in 1:nsim){
  setTxtProgressBar(pb, is)
  # Simulacion rejilla irregular
  if (loc=="irreg"){
    s <- rbind(data.frame(s1=runif(ndata),s2=runif(ndata)),data.frame(xnew))
    if (mod==1) sim1 <- grf(grid=s, cov.model="exp", cov.pars=c(sig2, phi), nugget=tau, nsim=1,
      messages=FALSE)
    if (mod==2) sim1 <- grf(grid=s, cov.model="gaussian", cov.pars=c(sig2, phi), nugget=tau, nsim=1,
      messages=FALSE)
    media <- 1+a*(as.matrix(sim1$coords[1:ndata,]) %*% c(1,1)) +
      b*((sim1$coords[1:ndata,]*sim1$coords[1:ndata,]) %*% c(1,1))
    medianew <- 1+a*(sim1$coords[-(1:ndata),] %*% c(1,1)) +
      b*((sim1$coords[-(1:ndata),]*sim1$coords[-(1:ndata),]) %*% c(1,1))
    datos <- data.frame(s=sim1$coords[1:ndata,], z=(sim1$data[1:ndata] + media[1:ndata]))
  }
  if (loc=="reg") { datos <- data.frame(sim1$coords[1:ndata,], z=(sim1$data[1:ndata ,is] + media)) }

  # ventana local percentil 0.2
  hi <- apply(abs(outer(datos$z, datos$z, "-")), 1, sort)[ceiling(0.2*ndata), ]
  mu.1 <- double(ndata)
  for (i in 1:ndata){
    mu.1[i] <- locpol::locCteSmootherC(x=datos$z, y=datos$z, xeval=datos$z[i], bw=hi[i], EpaK)$beta0
  }
  newdata <- data.frame(sim1$coords[1:ndata,], zz=mu.1)
  binned.data <- binning(x=newdata[, 1:2], y=newdata$zz, nbin=c(nnbin,nnbin))

  h.cv <- h.cv(binned.data,degree=0)$h
  mu.opt <- npsp::locpol(binned.data, h=h.cv, degree=0)
  muu <- predict(mu.opt, newx=sim1$coords)$y
  muu.pred <- muu[-(1:ndata)]
  muu <- muu[1:ndata]

  # kriging ordinario del proceso residual
  datos$res <- datos$z-muu # ya no hay tendencia
  dat.geoR <- as.geodata(datos, coords.col=1:2, data.col=4)
  vario <- variog(dat.geoR, max.dist=0.6, trend="cte")
  ini.vals <- expand.grid(seq(0.1,1.5,l=11), seq(0.05,0.6,l=11))
  v.geoR <- variofit(vario=vario, ini.cov.pars=ini.vals, cov.model="exponential", nugget=0,
    weights="cressie", messages=FALSE)#c(sig2,phi)
  if (v.geoR$nugget<0){ error.res[is] <- NA }
  if (v.geoR$nugget>=0){
    ko.wls <- krige.conv(dat.geoR, loc=xnew, krige=krige.control(type.krige="ok", trend.d="cte",
      trend.l="cte", obj.m=v.geoR))
    if (loc=="reg") error.res[is,] <- sim1$data[-(1:ndata) ,is] + medianew - (muu.pred + ko.wls$predict)
    if (loc=="irreg") error.res[is,] <- sim1$data[-(1:ndata)] + medianew - (muu.pred + ko.wls$predict)
  }
}

```



```

# kriging universal con tendencia lineal sobre los datos originales
dat.geoR <- as.geodata(datos, coords.col=1:2, data.col=3)
if (a==0) vario <- variog(dat.geoR, max.dist=0.6, trend="cte", messages=FALSE)
if (a!=0) vario <- variog(dat.geoR, max.dist=0.6, trend="1st", messages=FALSE)
ini.vals <- expand.grid(seq(0.1,1.5,l=11), seq(0.05,0.6,l=11))
v.geoR <- variofit(vario=vario, ini.cov.pars=ini.vals, cov.model="exponential", nugget=0,
  weights="cressie", messages=FALSE)#c(sig2,phi)
if (v.geoR$nugget<0) { error.ku[is] <- NA }
if (v.geoR$nugget>=0){
if (a==0) ku.wls <- krige.conv(dat.geoR, loc=xnew, krige=krige.control(type.krige="ok",
  trend.d="cte", trend.l="cte", obj.m=v.geoR))
if (a!=0) ku.wls <- krige.conv(dat.geoR, loc=xnew, krige=krige.control(type.krige="ok",
  trend.d="1st", trend.l="1st", obj.m=v.geoR))

if (loc=="reg") error.ku[is,] <- (sim1$data[-(1:ndata),is] + medianew) - ku.wls$predict
if (loc=="irreg") error.ku[is,] <- (sim1$data[-(1:ndata)]) + medianew - ku.wls$predict
}

} # is

round(apply(error.ku^2,2,mean,na.rm=TRUE)*100,3)
round(apply(error.res^2,2,mean,na.rm=TRUE)*100,3)

#####
## Datos reales

library(geor); library(npsp); library(locpol); library("scatterplot3d")

data(parana)
summary(parana)

# Figura 3.13
scatterplot3d(parana$coords[,1],parana$coords[,2],parana$data,xlab="Este",ylab="Norte",zlab="Z(s)")

par(mfrow=c(1,2),mar=c(4,4,3,3)) # Figura 3.14
# Gráficos parciales
# Coordenada x
plot(parana$coords[,1], parana$data, xlab="Este", ylab="Datos")
foo <- lowess(parana$data ~ parana$coords[, 1])
lines(foo[[1]], foo[[2]])
# Coordenada y
plot(parana$coords[,2], parana$data, xlab="Norte", ylab="Datos")
foo <- lowess(parana$data ~ parana$coords[, 2])
lines(foo[[1]], foo[[2]])

# nuevas localizaciones
loc <- cbind(east=c(280, 400, 500, 600), north=c(120, 200, 250, 350))
points(parana,xlab="Este",ylab="Norte")
points(loc,col=2,pch=19)
text(loc,col=2,c(expression(s^1),expression(s^2),expression(s^3),expression(s^4)),pos=c(4,4,3,3))

## K0 del proceso residual
# 1. Estimación de la tendencia

ndata <- length(parana$data)

```

```

hi <- apply(abs(outer(parana$data, parana$data, "-")), 1, sort)[ceiling(0.2*ndata), ]
mu.1 <- double(ndata)
for (i in 1:ndata){
mu.1[i] <- locpol::locCteSmootherC(x=parana$data, y=parana$data, xeval=parana$data[i], bw=hi[i],
  EpaK)$beta0
}
newdata <- data.frame(parana$coords[,1:2], zz=mu.1)
head(newdata)

binned.data <- binning(x=newdata[,1:2], y=newdata$zz, nbin=c(5,5))
h.cv <- h.cv(binned.data,degree=0)$h
mu.opt <- npsp::locpol(binned.data, h=h.cv, degree=0)
muu <- predict(mu.opt, newx=rbind(newdata[,1:2],loc))$y

# Estimación de la tendencia en las nuevas localizaciones
muu.pred <- muu[-(1:ndata)]
round(muu.pred,2)

# Tendencia estimada
muu <- muu[1:ndata]

# 2. KO
parana$res <- parana$data-muu
vario <- vario(parana, data=parana$res, max.dist=400, trend="cte")
plot(vario)
v.geoR <- variofit(vario=vario, ini.cov.pars=c(2000,200), cov.model="gaus", nugget=500)
lines(v.geoR)
ko.wls <- krige.conv(parana, data=parana$res, loc=loc, krige=krige.control(type.krige="ok",
  trend.d="cte", trend.l="cte", obj.m=v.geoR))

# Predicción en las nuevas localizaciones
round(ko.wls$predict + muu.pred, 2)

## KU del proceso original
vario.ku <- variog(parana, data=parana$data, max.dist=400, trend="1st")
v.geoR.ku <- variofit(vario=vario.ku, ini.cov.pars=c(800,250), cov.model="sph", nugget=300)
plot(vario.ku); lines(v.geoR.ku)
ku.wls <- krige.conv(parana, data=parana$data, loc=loc, krige=krige.control(type.krige="ok",
  trend.d="1st", trend.l="1st", obj.m=v.geoR.ku))

# Predicción en las nuevas localizaciones
round(ku.wls$predict,2)

# Tendencia del KU
round(loc %*% ku.wls$beta[-1] + ku.wls$beta[1], 2)

```

# Bibliografía

- [1] Bivand R, Pebesma E, Gomez-Rubio V (2013). Applied spatial data analysis with R, Second edition. Springer, NY. <http://www.asdar-book.org/>.
- [2] Chu CK, Marron JS (1991) Comparison of two bandwidth selectors with dependent errors. *Annals of Statistics* 19: 1906-1918.
- [3] Cressie N (1986) Kriging nonstationary data. *Journal of the American Statistical Association* 81: 625-634.
- [4] Cressie N (1993) *Statistics for Spatial Data*. John Wiley and Sons, New York.
- [5] Crujeiras RM, Van Keilegom I (2010) Least squares estimation of nonlinear spatial trends. *Computational Statistics & Data Analysis* 54: 452-465.
- [6] Fernandez-Casal R (2016) npsp: Nonparametric Spatial Statistics. R package version 0.5-3. <https://CRAN.R-project.org/package=npsp>.
- [7] Francisco-Fernández M, Opsomer JD (2005) Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics* 33: 539-558.
- [8] Francisco-Fernández M, Vilar-Fernández JM (2001) Local polynomial regression estimation with correlated errors. *Communications in Statistics - Theory and Methods* 30: 1271-1293.
- [9] Francisco-Fernández M, Quintela-del-Río A, Fernández-Casal R (2012) Nonparametric methods for spatial regression. An application to seismic events. *Environmetrics* 23: 85-93.
- [10] García-Soidán P, González-Manteiga W, Febrero-Bande M (2003) Local linear regression estimation of the variogram. *Statistics and Probability Letters* 64: 169-179.
- [11] García-Soidán P, González-Manteiga W, Febrero-Bande M (2004) Nonparametric kernel estimation of an isotropic variogram. *Journal of Statistical Planning and Inference* 121: 65-92.
- [12] García-Soidán P (2007) Asymptotic normality of the Nadaraya-Watson semivariogram estimators. *TEST* 16: 479-503.
- [13] García-Soidán P, Menezes R (2012) Estimation of the spatial distribution through the kernel indicator variogram. *Environmetrics* 23: 535-548.
- [14] García-Soidán P, Menezes R (2017) Nonparametric construction of probability maps under local stationarity: Nonparametric probability maps. *Environmetrics* 28: 1-14.
- [15] Hall P, Fisher I, Hoffmann B (1994) On the nonparametric estimation of covariance functions. *Annals of Statistics* 22: 2115-2134.
- [16] Journel AG y Huijbregts CJ (2003) *Mining geostatistics*. Academic Press, New York.
- [17] Matheron G (1963) Principles of geostatistics. *Economic Geology* 58: 1246-1266.

- [18] Menezes R, García-Soidán P, Ferreira C (2010) Nonparametric spatial prediction under stochastic sampling design. *Journal of Nonparametric Statistics* 22: 363-377.
- [19] Neuman SP, Jacobson EA (1984) Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. *Mathematical Geology* 16: 499-521.
- [20] Opsomer JD, Wang Y, Yang Y (2001) Nonparametric regression with correlated errors. *Statistical Science* 16: 134-153.
- [21] Ojeda Cabrera JL (2018). locpol: Kernel Local Polynomial Regression. R package version 0.7-0. <https://CRAN.R-project.org/package=locpol>.
- [22] Ribeiro P and Diggle P (2018) geoR: Analysis of Geostatistical Data. R package version 1.7-5.2.1. <https://CRAN.R-project.org/package=geoR>.
- [23] Ribeiro, P.J. (2018-12-30) Practicals - geoR. Acceso desde <http://www.leg.ufpr.br/geoR/PASI/>.
- [24] Ribeiro P and Diggle P (2001) geoR: a package for geostatistical analysis. *R news*, 1(2), 14-18.
- [25] Wand MP and Jones MC (1995) *Kernel Smoothing*. Chapman and Hall, London.
- [26] Yu K, Mateu J, Porcu E (2006) A kernel-based method for nonparametric estimation of variograms. *Statistica Neerlandica* 60: 1-25.