



UNIVERSIDAD NACIONAL DE ROSARIO

Facultad de Ciencias Económicas y Estadística

Carrera de posgrado

MAESTRÍA EN ESTADÍSTICA APLICADA

# Predicción no paramétrica para datos espaciales univariados

Autora: Mariel Lovatto

---

8 de junio de 2022

Directora: Dra. Pamela Llopp

Co - Director: Dr. Rodrigo García Arancibia

# Agradecimientos

texto de agradecimientos



# Resumen

Predicir una variable de interés en determinado punto geográfico a partir de mediciones de dicha variable en otras localizaciones, es comúnmente llamado *predicción espacial* y es, en geoestadística, una de las aplicaciones más recurrentes. Para esta tarea son necesarios datos indexados en el espacio, que siguen un modelo determinado del cual se asume que tienen cierta estructura de covarianza. Dicha estructura, debe poder captada por los modelos para predicción espacial para que la predicción sea eficiente.

En esta dirección, uno de los métodos más utilizados para tal fin es el clásico método *kriging* que consiste en un promedio ponderado del valor de la variable de interés cuyos pesos son estimados a partir de una función paramétrica, la cual surge con el objetivo de modelar la variabilidad de los datos mediante el concepto de *variograma*.

En esta tesis presentamos una propuesta que flexibiliza el cálculo paramétrico de dichos pesos, siguiendo el espíritu de *kriging* pero donde los pesos son estimados de forma no paramétrica, es decir, sin imponer restricciones sobre la estructura de covarianza.

Bajo diferentes hipótesis de existencia del fenómeno de interés, propusimos cinco predictores que logran captar no solo la hipótesis de que valores cercanos presentan mayor correlación, sino que también tienen en cuenta que pueden existir estructuras de covarianza diferentes y vecinas, provocando valores cercanos pero disímiles. Cuatro de esas cinco versiones propuestas revelaron buen desempeño en términos de error de predicción, bajo diferentes escenarios. Los predictores propuestos también fueron aplicados sobre datos reales donde revelan ventajas predictivas frente a los predictores clásicos, siendo en algunos casos más significativas que en otros donde los métodos paramétricos muestran ser una buena estrategia para predecir a pesar de su menor flexibilidad.

Palabras claves: Predicción espacial, kriging, estadística no paramétrica



# Índice de figuras

1.1. (a) Conjunto de datos regionalizados; (b) Regionalización. . . . .	3
1.2. Ejemplo del variograma empírico para los datos Meuse. (Rikken y Van Rijn, 1993) . . . . .	14
1.3. Semivariograma empírico y semivariograma estimado, a partir de un modelo exponencial (Ecuación 1.2.13), para los datos Meuse. . . . .	17
1.4. Semivariograma empírico y semivariograma estimado, a partir de una combinación entre el modelo pepita puro (1.2.11) y un modelo exponencial (1.2.13), para los datos Meuse. . . . .	18
1.5. Comportamiento del semivariograma cerca al origen: (a) intersección positiva (pepita); (b) comportamiento lineal continuo, no diferenciable; (c) comportamiento parabólico, continuo y diferenciable. . . . .	20
1.6. Mapas de caja. . . . .	25
1.7. Distribución de la variable concentración de zinc de los Datos Meuse en función de las coordenadas (Cressie, 1993, pág. 37). . . . .	26
1.8. Gráfico de dispersión $h$ para los datos Meuse. . . . .	27
3.1. Diferentes funciones núcleo. . . . .	47
3.2. Ejemplo del predictor introducido por Biau y Cadre (2004) para $d = 2$ . Los puntos que están contenidos en la línea de puntos son los que pertenecen a la vecindad y serán utilizados para la predicción (Biau y Cadre, 2004, pág. 329). . . . .	49
3.3. (a) Conjunto de valores regionalizados simulados a partir de un proceso gaussiano $Z(\mathbf{s})$ , no estacionario con función de covarianza exponencial para $n = 99$ y $d = 2$ . (b) Comportamiento de los valores regionalizados según su ubicación . . . . .	52
3.4. Comportamiento de los pesos del predictor K1S sobre el conjunto de valores regionalizados para: (a) el parámetro de suavizado obtenido a través de validación cruzada, (b) diferentes valores del parámetro de suavizado. . . .	53
3.5. (a) Comportamiento de los pesos del predictor K2IM sobre el conjunto de valores regionalizados. (b) Comportamiento de la función núcleo que contiene al Índice $I_i$ . . . . .	54

3.6.	Comportamiento de los pesos del predictor K2ME sobre el conjunto de: (a) valores regionalizados, (b) medianas de los valores vecinos para cada sitio.	56
3.7.	Comportamiento de los pesos del predictor K2ME sobre: (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra. . . . .	57
3.8.	Panel superior: Comportamiento de los pesos del predictor K1M para el sitio $s_0 = (4, 4)$ sobre (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra. Panel inferior: Comportamiento de los pesos del predictor K1M para el sitio $s_0 = (8, 5)$ sobre (c) el conjunto de valores regionalizados, (d) el conjunto de medianas de los vecindarios para cada sitio de la muestra. . . . .	59
3.9.	Panel superior: Comportamiento de los pesos del predictor K1M para el sitio $s_0 = (4, 4)$ sobre (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra. Panel inferior: Comportamiento de los pesos del predictor K1M para el sitio $s_0 = (8, 5)$ sobre (c) el conjunto de valores regionalizados, (d) el conjunto de medianas de los vecindarios para cada sitio de la muestra. . . . .	60
3.10.	(a) Error absoluto de predicción para el sitio $s_0 = (4, 4)$ , (b) Error absoluto de predicción para el sitio $s_0 = (8, 5)$ . . . . .	61
4.1.	Función de correlación para el rango: (a) $\theta_2 = 2$ ; (b) $\theta_2 = 100$ . . . . .	65
4.2.	Caso heterocedástico: distribución de las varianzas en función de los sitios. En el caso homocedástico $\sigma^2 = 10$ para todo $s_i \in D$ y para todos los modelos propuestos. . . . .	66
4.3.	Realización de un proceso con modelo de covarianza exponencial y rango $\theta_2 = 100$ : (a) Caso homocedástico; (b) Caso heterocedástico. . . . .	67
4.4.	Combinación de parámetros considerados en cada escenario partiendo del modelo de covarianza exponencial. . . . .	68
4.5.	Combinación de parámetros considerados en cada escenario partiendo del modelo de covarianza combinado. . . . .	68
4.6.	Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza exponencial y rango $\theta_2 = 2$ . Desde arriba hacia abajo, los tres modelos considerados para el error $\epsilon$ y de izquierda a derecha, los casos homocedásticos y heterocedásticos. . . . .	71
4.7.	Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza exponencial y rango $\theta_2 = 100$ . Desde arriba hacia abajo, los tres modelos considerados para el error $\epsilon$ y de izquierda a derecha, los casos homocedásticos y heterocedásticos. . . . .	72

4.8. Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza combinada y rangos $\theta_{21} = 0.2$ y $\theta_{22} = 10$ . Desde arriba hacia abajo, los tres modelos considerados para el error $\epsilon$ y de izquierda a derecha, los casos homocedásticos y heterocedásticos. . . . .	73
5.1. Distribución de la variable de interés sin considerar el espacio. Saltillo y Valle de Bravo en miles de pesos mexicanos y Singapur en Dólar Singapur. . . . .	78
5.2. Mapa de caja: precio de inmuebles en Saltillo. . . . .	79
5.3. Datos Saltillo: distribución de la variable de interés en función las coordenadas geográficas. . . . .	80
5.4. Gráfico de dispersión h del precio de inmuebles en Saltillo, México. . . . .	81
5.5. Distribución del MAE, sobre 50 repeticiones, para los datos Saltillo. . . . .	82
5.6. Datos Saltillo: Mapas predictivos para cada método. . . . .	83
5.7. Mapa de caja: precio de inmuebles en Valle de Bravo, México. . . . .	84
5.8. Datos Valle Bravo: Distribución de la variable de interés en función las coordenadas. . . . .	85
5.9. Gráfico de dispersión h del precio de inmuebles en Valle de Bravo, México. . . . .	85
5.10. Distribución del MAE, sobre 50 repeticiones, para los datos Valle de Bravo. . . . .	86
5.11. Datos Valle Bravo: Mapas predictivos para cada método. . . . .	87
5.12. Mapa de caja: precio de hospedajes en Singapur. . . . .	88
5.13. Datos Singapur: Distribución de la variable de interés en función de las coordenadas. . . . .	89
5.14. Gráfico de dispersión h del precio de hospedaje en Singapur. . . . .	90
5.15. Distribución del MAE, sobre 50 repeticiones, para los datos de Singapur. . . . .	91
5.16. Datos de Singapur: Mapas predictivos para cada método. . . . .	91



# Índice de tablas

5.1. Resumen numérico para cada conjunto de datos. Saltillo y Valle de Bravo en miles de pesos mexicanos y Singapur en Dólar Singaporense. . . . .	78
5.2. Promedio del MAE, del coeficiente de correlación de Pearson y del RSME (en miles), sobre 50 repeticiones, para los datos Saltillo. Mejores resultados en negrita. . . . .	82
5.3. Promedio del MAE, del coeficiente de correlación de Pearson y del RSME (en miles), para los datos Valle Bravo. Mejores resultados en negrita. . . . .	86
5.4. Promedio del MAE, del coeficiente de correlación de Pearson y del RSME (en miles) sobre 50 repeticiones, para los datos de Singapur. Mejores resultados en negrita. . . . .	90



# Índice general

Agradecimientos	I
Resumen	II
Índice de figuras	V
Índice de tablas	IX
Introducción	XIV
<b>1. Datos espaciales</b>	<b>1</b>
1.1. Análisis estructural . . . . .	2
1.1.1. Proceso estocástico espacial . . . . .	2
1.1.2. Estacionariedad . . . . .	5
1.2. Modelo espacial . . . . .	9
1.2.1. Semivariograma y covariograma empíricos . . . . .	11
1.2.2. Semivariograma teórico . . . . .	14
1.3. Análisis exploratorio . . . . .	24
1.3.1. Tipos de gráficos exploratorios . . . . .	24
1.3.2. Tendencia . . . . .	25
1.3.3. Variabilidad . . . . .	27
1.3.4. Índices de asociación espacial . . . . .	28
1.4. Conclusiones . . . . .	29
<b>2. Predicción espacial paramétrica: kriging</b>	<b>31</b>
2.1. Kriging . . . . .	31
2.2. Kriging simple . . . . .	32
2.2.1. Insegadez del predictor . . . . .	33
2.2.2. Mínima varianza del error de predicción . . . . .	33
2.2.3. Forma final del predictor . . . . .	35
2.3. Kriging ordinario . . . . .	35
2.3.1. Insegadez del predictor . . . . .	36
2.3.2. Mínima varianza del error de predicción . . . . .	36
2.3.3. Forma final del predictor . . . . .	39

2.4. Kriging universal . . . . .	39
2.4.1. Inesgidez del predictor . . . . .	41
2.4.2. Mínima varianza del error de predicción . . . . .	41
2.4.3. Forma final del predictor . . . . .	43
2.5. Conclusiones . . . . .	44
<b>3. Predicción espacial no paramétrica</b>	<b>45</b>
3.1. Estimador de núcleo de la función de regresión . . . . .	45
3.2. Estado actual de la literatura . . . . .	48
3.3. La nueva propuesta: kriging no paramétrico . . . . .	51
3.3.1. Kriging de un núcleo simple (K1S) . . . . .	52
3.3.2. Kriging de dos núcleos con Índice de Moran (K2IM) . . . . .	53
3.3.3. Kriging de dos núcleos con medianas (K2ME) . . . . .	55
3.3.4. Kriging de un núcleo Multiplicativo (K1M) . . . . .	57
3.3.5. Kriging de un núcleo Aditivo (K1A) . . . . .	58
3.3.6. Comparación de errores . . . . .	60
3.4. Conclusiones . . . . .	61
<b>4. Estudios de simulación</b>	<b>63</b>
4.1. Método de simulación . . . . .	63
4.2. Escenarios . . . . .	64
4.2.1. Generalidades . . . . .	64
4.2.2. Con respecto a la función de covarianza . . . . .	64
4.2.3. Con respecto al vector aleatorio . . . . .	67
4.3. Estimación de parámetros y validación cruzada . . . . .	68
4.4. Resultados . . . . .	70
4.4.1. Con respecto a la función de covarianza . . . . .	73
4.4.2. Con respecto a la heterogeneidad de la varianza . . . . .	74
4.4.3. Con respecto al vector aleatorio . . . . .	75
4.5. Conclusiones . . . . .	75
<b>5. Aplicación a datos reales</b>	<b>77</b>
5.1. Presentación de los datos . . . . .	77
5.2. Precios de Inmuebles: Saltillo . . . . .	79
5.2.1. Análisis exploratorio . . . . .	79
5.2.2. Resultados . . . . .	81
5.3. Precios de Inmuebles: Valle de Bravo . . . . .	83
5.3.1. Análisis exploratorio . . . . .	83
5.3.2. Resultados . . . . .	84
5.4. Precios de hospedajes: Singapur . . . . .	86
5.4.1. Análisis exploratorio . . . . .	86
5.4.2. Resultados . . . . .	90

---

5.5. Conclusiones . . . . .	92
<b>Conclusiones generales</b>	<b>93</b>
<b>Bibliografía</b>	<b>95</b>



# Introducción

Una de las herramientas fundamentales en el área de estadística es la predicción de fenómenos de interés a partir de ciertas variables observadas. Esto implica construir modelos que se adecúen a los datos disponibles y a los objetivos buscados. Por ejemplo, una característica muy importante a tener en cuenta es la independencia o no de los mismos. Es bien sabido que modelos desarrollados para datos independientes no tienen un buen desempeño en datos que presentan cierta dependencia, y de hecho, no es correcto utilizarlos. Ejemplos de datos dependientes abarcan, entre otras, series de tiempo, donde las mediciones están recolectadas en momentos de tiempo sucesivos; modelos de medidas repetidas, donde las mediciones se toman repetidamente sobre el mismo sujeto; o datos espaciales, donde las mediciones están recolectadas en diferentes locaciones espaciales. Hacer predicciones considerando este último tipo de datos es de nuestro interés en esta tesis y es lo que comúnmente se llama *predicción espacial*.

La predicción espacial fue estudiada en sus comienzos en el contexto de geoestadística a partir de los predictores clásicos kriging (no utiliza covariables) y cokriging (permite covariables) (Anselin, 2004; Cressie, 1993; Montero et al., 2015; Legendre y Legendre, 2012; Wackernagel, 1995; Webster y Oliver, 2007). Luego fue estudiada en el contexto de regresión espacial con errores correlacionados (Anselin, 2004), utilizando modelos lineales generalizados espaciales (Diggle, 2014; Zhang, 2002), bajo modelos de procesos puntuales de Poisson (Diggle, 2014), entre otras. También existen muchas técnicas de predicción espacial no paramétrica (Biau y Cadre, 2004; Dabo-Niang y Yao, 2007; Dabo-Niang et al., 2016; Menezes et al., 2010), de machine learning (Pham et al., 2018; Liu et al., 2022) y de ensemble (Yang, 2018).

En los últimos años, junto a la expansión de la tecnología y las facilidades en la recolección de datos espaciales, el uso de geoestadística en las ciencias aplicadas ha sido creciente. Por ejemplo, en ciencias hídricas (Rizo-Decelis et al., 2017), en agricultura de precisión (Oliver, 2010), en procesamiento de imágenes, entre otras. Ejemplos de estos datos son: mediciones de contaminantes registradas en estaciones de monitoreo ambiental, mediciones atmosféricas; efectos en el suelo de los explosivos nucleares; precios de venta o alquiler de un lote o una propiedad en los mercados de bienes raíces comerciales en una ubicación determinada. En cada uno de estos casos, anticipamos la dependencia entre las mediciones en una

ubicación particular y la asociación de las mediciones entre ubicaciones. En particular en el área de economía y ciencias sociales una de las líneas desarrolladas recientemente es el uso de modelos predictivos para la reconstrucción de mapas que reflejen algún fenómeno socio-económico de interés. Así por ejemplo, Chica-Olmo (2007), Kuntz y Helbich (2014) y Bajat et al. (2017) utilizan y comparan diferentes métodos de estadística espacial para la predicción de precios de las viviendas en diferentes localidades. También los métodos clásicos de geoestadística han sido aplicados para la predicción de niveles de pobreza, y con ello, en la generación de mapas para estudiar las privaciones económicas y sociales de hogares y/o individuos en determinados territorios geográficos (Mathiassen, 2009; Pokhriyal y Jacques, 2017; Mohamed y Mohamed, 2009; Chandra et al., 2018; Norman, 2009; Vasan y Alcantara, 2016). Esta aplicación de metodologías de predicción espacial constituye una poderosa herramienta para el diseño de políticas públicas ya que posibilita contar con un mapa más completo de las privaciones sociales o de otras características socioeconómicas de una región, como así también, evaluar *ex-post* determinadas políticas focalizadas en locaciones geográficas que habitualmente no están abarcadas por los instrumentos de medición de las estadísticas oficiales.

Un problema fundamental que surge a la hora de aplicar las técnicas clásicas de predicción espacial es la violación de los supuestos de los modelos utilizados, los cuales muchas veces son fuertes y hasta difíciles de probar en la práctica. Dichos supuestos pueden incluir la distribución subyacente del error, la cual se ve reflejada en la distribución de la variable de interés y/o el modelo sugerido para la estructura de covarianza presente en los datos. En esta dirección, en esta tesis proponemos y desarrollamos nuevos métodos no paramétricos de predicción espacial, sin imponer restricciones sobre dicha estructura de covarianza.

La única restricción que mantendremos a lo largo de la tesis es que para predecir una variable de interés en una localización determinada solo contamos con información de la misma variable pero medida en otras localizaciones. Es decir, no tenemos otras características de dicha observación que puedan ser usadas como variables predictoras de nuestra variable respuesta de interés, sino que simplemente podemos explotar la información que nos brinda la cercanía/lejanía espacial, como así también la similaridad de la variable en distintas vecindades de la región explorada. Por lo tanto, el problema que enfrentamos en la presente tesis es de predicción espacial univariada. Tomando esta limitación que puede encontrarse en varios datos espaciales, propondremos metodologías más flexibles respecto a los supuestos inherentes de los métodos clásicos de predicción espacial univariada.

La tesis está organizada y desarrollada de la siguiente manera. En el **Capítulo 1** definimos y describimos los datos espaciales, hacemos explícita la estructura formal subyacente sobre la cual, se supone, se generan los mismos. También explicamos de qué manera se considera la correlación de los datos y cómo se modela bajo los diferentes supuestos de estacionariedad. Particularmente definimos el variograma y presentamos las herramientas para su estimación. Finalizamos con una pequeña descripción del análisis exploratorio adecuado

para este tipo de datos.

En el **Capítulo 2** presentamos los métodos de predicción espacial clásicos, comúnmente denominados métodos kriging. Más precisamente, explicitamos el modelo general y las diferentes metodologías de acuerdo a los supuestos asumidos en la especificación del mismo. En base a esto presentamos los predictores para cada caso y describimos el razonamiento y procedimiento para su obtención.

El **Capítulo 3** está dedicado a introducir los nuevos predictores junto con las consideraciones que motivaron este trabajo, haciendo un repaso de los métodos desarrollados hasta el momento, comparándolos y presentado la bibliografía relacionada. Dado que estas clases de predictores están basadas en la estimación no paramétrica de la función de regresión, destinamos una sección del capítulo para su repaso haciendo especial hincapié en el estimador de núcleo de la función de regresión.

En el **Capítulo 4** presentamos estudios de simulación que muestran el desempeño de todos los métodos, nuevos y existentes, detallados en los capítulos previos. En particular, detallamos el método de simulación utilizado, los escenarios considerados, las técnicas utilizadas para estimar parámetros y computar errores y finalmente los resultados obtenidos.

El **Capítulo 5** está destinado a la aplicación de los métodos a conjuntos de datos reales. Comenzamos presentando los datos, haciendo un análisis exploratorio de los mismos para luego presentar los resultados obtenidos. Incluimos en cada caso, además, un mapa predictivo generado con cada método.

# Capítulo 1

## Datos espaciales

Los datos espaciales son aquellos cuya particularidad es que cada observación esta asociada a una unidad espacial, como ser coordenadas o áreas geográficas. Por ello es imprescindible contar no solo con el dato en sí mismo (i.e. el valor observado de la/s variable/s de interés) sino también con información sobre la posición o referencia geográfica indexada a cada observación. Con esta caracterización de los datos espaciales surge de forma natural el concepto de lo que llamaremos **valor regionalizado** o **dato regionalizado**, definido como el valor de la variable de interés en el lugar o punto geográfico donde fue medido.

Dentro del contexto estadístico, la estadística espacial (o geoestadística) es, en parte, la aplicación de la teoría de procesos estocásticos a datos espaciales o, dicho de otra forma, a datos correlacionados espacialmente. Según Wackernagel (2006) la necesidad de tener en cuenta al espacio como fuente de variabilidad se originó en el campo de la minería y la geología a mediados del siglo XX, al tratar de analizar la distribución de las reservas de oro en un depósito mineral a partir de muestras tomadas en algunas ubicaciones espaciales, la cual resultó ser sesgada, revelando la omisión del espacio como fuente de variabilidad.

En este capítulo definimos algunos conceptos y definiciones claves en el estudio de datos espaciales, como así también el marco teórico sobre el cual suelen modelarse. Con este capítulo buscamos introducir las herramientas teóricas y metodológicas básicas para el tratamiento de datos espaciales, especificando los supuestos teóricos sobre los que descansan mayormente las tareas de inferencia y predicción estadística. Así mismo, sobre la base de este marco teórico, se derivan herramientas de visualización y descripción para el estudio del fenómeno espacial de interés.

## 1.1. Análisis estructural

En este apartado presentamos los elementos esenciales para comprender el desarrollo y funcionamiento de los modelos espaciales y el análisis clásico de los mismos. La información con la que contamos es una muestra de datos regionalizados, los cuales, desde la perspectiva probabilística, pueden considerarse como el resultado de un mecanismo aleatorio, al cual llamaremos **proceso estocástico espacial**. En la bibliografía también podemos hallarlo con el nombre de función aleatoria o campo aleatorio (Montero et al. (2015), Wackernagel (1995)). Para comprender mejor las características de este proceso, los supuestos subyacentes y sus implicancias sobre el tratamiento de datos espaciales, en esta sección focalizamos en la formalización del mismo y de sus propiedades.

### 1.1.1. Proceso estocástico espacial

Sea  $d \in \mathbb{N}$ ,  $D \subset \mathbb{R}^d$  y  $\Omega$  un espacio de probabilidad, se llama **proceso estocástico espacial** al conjunto

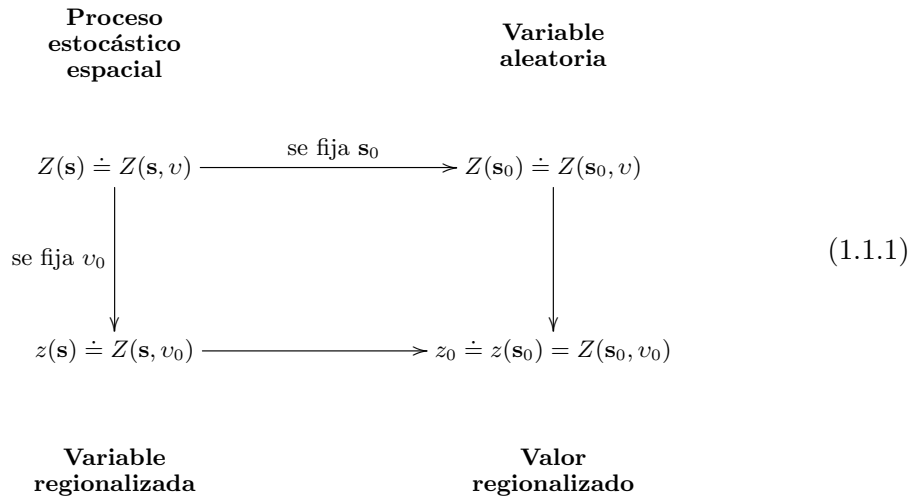
$$\{Z(\mathbf{s}, v) : \mathbf{s} \in D, v \in \Omega\},$$

con  $Z$  una función a valores reales tal que  $Z : D \times \Omega \rightarrow \mathbb{R}$ .

Si tomamos al dato regionalizado como resultado de un proceso estocástico, obtenemos dos caracterizaciones del mismo. En particular, si fijamos una localización  $\mathbf{s}_0 \in D$ , tenemos una variable aleatoria  $Z(\mathbf{s}_0, v)$  con su variabilidad asociada. Si en su lugar fijamos  $v_0 \in \Omega$ , entonces tenemos la realización  $z : D \subset \mathbb{R}^d \rightarrow \mathbb{R}$ , denominada comúnmente como **variable regionalizada o regionalización** (e.g. Wackernagel (2006)).

Esta doble caracterización nos permite tener en cuenta las diferentes fuentes de variabilidad. Por un lado la variabilidad del fenómeno en cada sitio, entender que el dato existe con cierta probabilidad y que existen otros valores posibles. Por otro lado nos permite, de alguna manera, formalizar la variabilidad entre sitios, considerando las distancias entre ellos y la forma de la dependencia con la variable de interés.

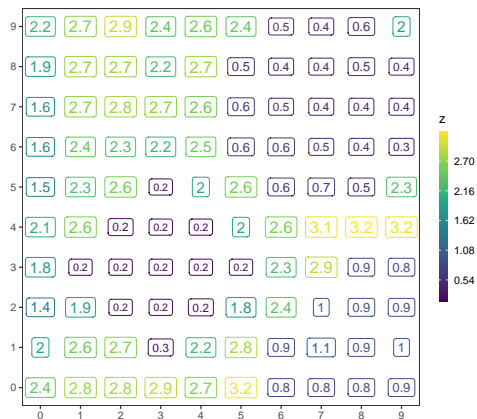
En el Diagrama 1.1.1 se ilustran estas definiciones junto con la notación a ser utilizada.



Fuente: Wackernagel (2006), pág. 28.

Luego, si fijamos un conjunto de localizaciones  $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$  y un  $v_0 \in \Omega$ , tenemos  $(z_1, \dots, z_n)^T$  un vector de valores regionalizados, los cuales representan los **datos empíricos** a utilizar.

Para ilustrar la generación de datos regionalizados, en la Figura 1.1 (a) presentamos un ejemplo de datos empíricos simulados sobre una grilla a partir de un determinado proceso estocástico. En la Figura 1.1 (b) observamos una regionalización, esto es, una realización de dicho proceso, para un  $v_0 \in \Omega$  dado. Teniendo ya una definición formal de proceso



(a)

(b)

Figura 1.1: (a) Conjunto de datos regionalizados; (b) Regionalización.

estocástico espacial, para su estudio estadístico resulta crucial una definición de su distri-

bución. Específicamente, dado un conjunto finito de sitios  $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ , podemos definir la **distribución del proceso estocástico** como la colección de distribuciones conjuntas de dimensión finita

$$\begin{aligned} F_{\mathbf{s}_1, \dots, \mathbf{s}_n}(z_1, \dots, z_n) &\doteq \mathbb{P}(\{v : Z(\mathbf{s}_1, v) \leq z_1, \dots, Z(\mathbf{s}_n, v) \leq z_n\}) \\ &= \mathbb{P}(Z(\mathbf{s}_1) \leq z_1, \dots, Z(\mathbf{s}_n) \leq z_n), \end{aligned} \quad (1.1.2)$$

del vector aleatorio  $(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ .

De esta manera, con la especificación de (1.1.2) para todos los órdenes  $n$  y todos los posibles conjuntos de puntos de muestreo  $\mathbf{s}_j$  (con  $j = 1, \dots, n$ ) tenemos una caracterización completa de un proceso estocástico. Un caso particular de especial importancia es el denominado **proceso gaussiano** donde las distribuciones de dimensión finita definidas en la Ecuación (1.1.2) son normales multivariadas. Para denotar la densidad conjunta de las variables aleatorias obtenidas de una muestra específica de localizaciones (asumiendo que el proceso es de valores continuos), usaremos  $p_{\mathbf{s}_1, \dots, \mathbf{s}_n}(z_1, \dots, z_n)$ .

Dada la distribución del proceso  $Z(\mathbf{s})$ , la información contenida en la misma puede resumirse a partir de los momentos del proceso para un conjunto fijo de localizaciones. El primer momento o **valor esperado**, denotado por  $\mu(\mathbf{s})$ , es una función de la localización, definido por

$$\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})] = \int_{-\infty}^{\infty} z p_{\mathbf{s}}(z) dz,$$

para un proceso de valores continuos, donde  $p_{\mathbf{s}}(z)$  es la densidad de la variable aleatoria  $Z(\mathbf{s})$ . Para un proceso de valores discretos, la integral se reemplaza por la suma y en este caso  $p_{\mathbf{s}}(z)$  es la función de probabilidad puntual conjunta. A su vez, los tres momentos de segundo orden más utilizados en geoestadística son la **varianza**, la **covarianza** y el **variograma**. Para ello, podemos definir en primer lugar, la **función de autocorrelación** de la forma

$$R_Z(i, j) \doteq \mathbb{E}[Z(\mathbf{s}_i)Z(\mathbf{s}_j)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z_i z_j p_{\mathbf{s}_i, \mathbf{s}_j}(z_i, z_j) dz_i dz_j,$$

para el caso de un proceso a valores continuos. Para el caso discreto, las integrales se reemplazan por una doble sumatoria. Luego la función de covarianza o autocovarianza queda definida, para todo  $\mathbf{s}, \mathbf{s}_i, \mathbf{s}_j \in D$ , por

$$\begin{aligned} C(\mathbf{s}_i, \mathbf{s}_j) &\doteq Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \\ &= \mathbb{E}[(Z(\mathbf{s}_i) - \mu(\mathbf{s}_i))(Z(\mathbf{s}_j) - \mu(\mathbf{s}_j))] \\ &= R_Z(i, j) - \mu(\mathbf{s}_i)\mu(\mathbf{s}_j), \end{aligned} \quad (1.1.3)$$

y la varianza de  $Z(\mathbf{s}_i)$  para algún  $\mathbf{s}_i \in D$  será

$$\begin{aligned} Var(Z(\mathbf{s}_i)) &\doteq \mathbb{E}[(Z(\mathbf{s}_i) - \mu(\mathbf{s}_i))^2] \\ &= R_Z(i, i) - \mu^2(\mathbf{s}_i). \end{aligned}$$

Por último, el variograma del proceso se define como

$$\begin{aligned} 2\gamma(\mathbf{s}_i, \mathbf{s}_j) &\doteq \text{Var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) \\ &= \text{Var}(Z(\mathbf{s}_i)) + \text{Var}(Z(\mathbf{s}_j)) - 2\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \\ &= R_Z(i, i) - \mu^2(\mathbf{s}_i) + R_Z(j, j) - \mu^2(\mathbf{s}_j) - 2(R_Z(i, j) - \mu(\mathbf{s}_i)\mu(\mathbf{s}_j)). \end{aligned}$$

La función  $\gamma(\mathbf{s}_i, \mathbf{s}_j)$  es conocida como el **semivariograma** del proceso  $Z(\mathbf{s})$  y es una herramienta clásica muy importante y muy utilizada que sirve para cuantificar la variabilidad espacial de muchos procesos y es esencial para obtener predicciones espaciales precisas. La estudiaremos más en detalle a lo largo del capítulo.

### 1.1.2. Estacionariedad

El tipo de estacionariedad más fuerte que podemos considerar para un proceso  $Z(\mathbf{s})$  es la **estacionariedad en sentido estricto** o simplemente **estacionariedad estricta**. La misma implica que dado un proceso  $Z(\mathbf{s})$  con esperanza y varianza finita y dos conjuntos de puntos  $(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k)$  y  $(\mathbf{s}_1 + \mathbf{h}, \mathbf{s}_2 + \mathbf{h}, \dots, \mathbf{s}_k + \mathbf{h})$ , para cualquier vector de traslación  $\mathbf{h} \in \mathbb{R}^d$ ,  $(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_k))$  y  $(Z(\mathbf{s}_1 + \mathbf{h}), \dots, Z(\mathbf{s}_k + \mathbf{h}))$  tienen la misma función de distribución conjunta para todo  $k$ .

Un tipo de estacionariedad más débil que la estricta es la conocida como **estacionariedad débil o de segundo orden**. Más precisamente, diremos que un proceso  $Z(\mathbf{s})$  es **estacionario de segundo orden o débilmente estacionario**, si tiene momento de segundo orden finito (es decir, existe covarianza) y verifica las siguientes dos condiciones:

- (i) La esperanza de  $Z(\mathbf{s})$  existe y es constante,

$$\mathbb{E}[Z(\mathbf{s})] = \mu, \quad \forall \mathbf{s} \in D. \quad (1.1.4)$$

- (ii) La covarianza existe para cada par de variables  $Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})$  y solo depende del vector distancia  $\mathbf{h}$  entre los sitios,

$$\text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})) = C(\mathbf{s} + \mathbf{h}, \mathbf{s}) \doteq C(\mathbf{h}). \quad (1.1.5)$$

En este caso la función  $C(\cdot)$  es llamada **covariograma** o **función de covarianza estacionaria**.

La varianza de un proceso débilmente estacionario existe, es finita y constante, esto es,

$$\text{Var}(Z(\mathbf{s})) = \text{Cov}(Z(\mathbf{s}), Z(\mathbf{s})) = C(\mathbf{0}) \doteq \sigma^2 < \infty, \quad \forall \mathbf{s} \in D. \quad (1.1.6)$$

En este caso el variograma está dado por

$$2\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))$$

$$\begin{aligned}
&= \text{Var}(Z(\mathbf{s} + \mathbf{h})) + \text{Var}(Z(\mathbf{s})) - 2\text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})) \\
&= C(\mathbf{0}) + C(\mathbf{0}) - 2C(\mathbf{h}) \quad (\text{por (1.1.5)}) \\
&= 2(C(\mathbf{0}) - C(\mathbf{h})),
\end{aligned}$$

por lo que el semivariograma de un proceso débilmente estacionario solo depende de la distancia entre los puntos  $\mathbf{h}$  y queda expresado como

$$\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \doteq \gamma(\mathbf{h}). \quad (1.1.7)$$

Otra propiedad importante de la función de covarianza  $C$  es que está acotada por su valor en el origen, esto es,

$$|C(\mathbf{h})| \leq C(\mathbf{0}). \quad (1.1.8)$$

En efecto, de  $0 \leq \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 2(C(\mathbf{0}) - C(\mathbf{h}))$  se tiene que  $C(\mathbf{h}) \leq C(\mathbf{0})$  y, análogamente, de  $0 \leq \text{Var}(Z(\mathbf{s} + \mathbf{h}) + Z(\mathbf{s})) = 2(C(\mathbf{0}) + C(\mathbf{h}))$  se tiene que  $-C(\mathbf{0}) \leq C(\mathbf{h})$  de donde se sigue (1.1.8).

Observemos que si el proceso  $Z(\mathbf{s})$  es estacionario de segundo orden,  $\gamma(\mathbf{h})$  también está acotada pues, de (1.1.7) y (1.1.8) se tiene que,

$$|\gamma(\mathbf{h})| = |C(\mathbf{0}) - C(\mathbf{h})| \leq |C(\mathbf{0})| + |C(\mathbf{h})| \leq 2C(\mathbf{0}).$$

Esto nos dice que si  $\gamma(\mathbf{h})$  no está acotada, el proceso  $Z(\mathbf{s})$  no puede ser estacionario de segundo orden.

De la Ecuación (1.1.7) se revela que un semivariograma queda determinado a partir de una función de covarianza. Sin embargo, el recíproco no es cierto, ya que el semivariograma puede crecer indefinidamente. Por ejemplo, la función de semivariograma dada por<sup>1</sup>

$$\gamma(\mathbf{h}) = b\|\mathbf{h}\|^p, \quad \text{con } 0 < p < 2 \text{ y } b > 0, \quad (1.1.9)$$

no puede dar lugar a una función de covarianza ya que crece sin límites. Esta función de variograma supera el marco de los procesos débilmente estacionarios; esto es,  $\gamma(\mathbf{h})$  existe pero no representa la variabilidad de un proceso estacionario de segundo orden.

Dado que la hipótesis de media constante es bastante restrictiva y muchos procesos reales no la verifican, en la práctica suele considerarse un tercer (y más débil) tipo de estacionariedad que es llamada **estacionariedad intrínseca**. La misma se basa en la idea de que si bien la media puede no ser constante para todo  $\mathbf{s}$  en el dominio, que por lo menos lo sea para valores pequeños de  $\mathbf{h}$  (Matheron, 1965). Esto significa que para ubicaciones cercanas, las diferencias esperadas sean lo más pequeñas posibles. Además, para la estacionariedad débil, suponemos que la varianza del proceso existe y es finita. Esta condición también es restrictiva ya que en este contexto no podríamos modelar un fenómeno con

<sup>1</sup>Ejemplo 5.1, (Wackernagel, 2006, pág. 37).

capacidad de variación infinita. Sin embargo, existen procesos  $Z(\mathbf{s})$  cuyas diferencias (o incrementos)  $Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})$  tienen varianza finita y, por lo tanto, (dichos incrementos) son estacionarios de segundo orden. Precisamente estos procesos, son los que se conocen como procesos intrínsecamente estacionarios.

Formalmente, diremos que un proceso  $Z(\mathbf{s})$  es **intrínsecamente estacionario** o simplemente **intrínseco**, si el proceso diferencia  $Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})$  es estacionario de segundo orden para cualquier vector de traslación  $\mathbf{h}$ . Esto es, la esperanza de la diferencia es constante en  $\mathbf{s}$ ,

$$\mathbb{E}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})] = m(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D, \quad (1.1.10)$$

para alguna función  $m(\mathbf{h})$  y, para cada  $\mathbf{h}$  fijo,

$$\text{Cov}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{s} + \mathbf{h}')] \doteq \tilde{C}_{\mathbf{h}}(\mathbf{h}'), \quad (1.1.11)$$

para todo  $\mathbf{s}, \mathbf{s} + \mathbf{h}, \mathbf{s} + \mathbf{h}', \mathbf{s} + \mathbf{h} + \mathbf{h}' \in D$ , esto es, la covarianza del proceso diferencia solo depende de  $\mathbf{h}'$ .

Una consecuencia de (1.1.10) es que  $m(\mathbf{h})$  es necesariamente lineal, pues,

$$\begin{aligned} m(\mathbf{h} + \mathbf{h}') &= \mathbb{E}[Z(\mathbf{s} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{s})] \\ &= \mathbb{E}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}) + Z(\mathbf{s} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{s} + \mathbf{h})] \\ &= \mathbb{E}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})] + \mathbb{E}[Z(\mathbf{s} + \mathbf{h} + \mathbf{h}') - Z(\mathbf{s} + \mathbf{h})] \\ &= m(\mathbf{h}) + m(\mathbf{h}'). \end{aligned}$$

Además, tomando  $\mathbf{h}' = \mathbf{0}$  en (1.1.11) resulta que,

$$2\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = \tilde{C}_{\mathbf{h}}(\mathbf{0}),$$

por lo que la varianza tiene valor finito y no depende de  $\mathbf{s}$ .

Observemos que podemos considerar  $m(\mathbf{h}) = 0$  pues,  $m(\mathbf{h}) \neq 0$  implicaría, de (1.1.10), que la  $\mathbb{E}[Z(\mathbf{s})]$  no es constante con respecto al espacio  $\mathbf{s}$ , y por lo tanto  $Z(\mathbf{s})$  no sería estacionario.

Por ende, a partir de este momento y a menos que se indique explícitamente lo contrario, consideraremos  $m(\mathbf{h}) = 0$  y como consecuencia, podemos decir que un proceso  $Z(\mathbf{s})$  es **intrínsecamente estacionario** si verifica que

$$\mathbb{E}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})] = 0, \quad (1.1.12)$$

$$\text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = 2\gamma(\mathbf{h}). \quad (1.1.13)$$

Estas condiciones son consideradas, generalmente, como las condiciones de estacionariedad intrínseca (véase por ejemplo Cressie, 1993; Webster y Oliver, 2007).

Luego, si el proceso  $Z(\mathbf{s})$  es intrínseco, de (1.1.12) tenemos que

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = \frac{1}{2} \mathbb{E}[(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2]. \quad (1.1.14)$$

Cabe destacar que la ecuación (1.1.14) vale también bajo un proceso estacionario de segundo orden.

Resulta sencillo probar que **todo proceso estacionario de segundo orden es intrínsecamente estacionario**. En efecto, si la esperanza del proceso es constante, la esperanza de las diferencias es cero. Esto es,

$$\mathbb{E}[Z(\mathbf{s})] = \mu \quad \Rightarrow \quad \mathbb{E}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})] = 0. \quad (1.1.15)$$

Luego resta probar que si la covarianza del proceso existe y depende solo de la distancia entre los sitios  $\mathbf{h}$ , es decir,

$$\text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})) = C(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D, \quad (1.1.16)$$

entonces la función de semivariograma  $\gamma$  también depende solo de  $\mathbf{h}$ . Para ello tenemos que

$$\begin{aligned} 2\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) &= \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) \\ &= \text{Var}(Z(\mathbf{s} + \mathbf{h})) + \text{Var}(Z(\mathbf{s})) - 2\text{Cov}(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})) \\ &= C(\mathbf{0}) + C(\mathbf{0}) - 2C(\mathbf{h}) \quad (\text{por (1.1.16)}) \\ &= 2(C(\mathbf{0}) - C(\mathbf{h})), \end{aligned}$$

Luego, dado que la varianza  $C(\mathbf{0})$  es constante para procesos estacionarios de segundo orden, tenemos que

$$\gamma(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \frac{1}{2} \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})) = \gamma(\mathbf{h}), \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D,$$

lo cual asegura, junto con (1.1.15), que  $Z(\mathbf{s})$  es intrínsecamente estacionario.

El recíproco de este resultado no es cierto. Es decir, existen procesos intrínsecamente estacionarios que no son estacionarios de segundo orden. Por ejemplo, hemos visto que la función de semivariograma dada en (1.1.9) no es acotada y como consecuencia no cuenta con una función de covarianza asociada, lo que hace que supere el marco de los procesos débilmente estacionarios. Sin embargo, esta función de semivariograma podría estar asociada a un proceso intrínseco. Otro ejemplo es el proceso discreto Wiener-Levy (Montero et al., 2015, pág. 17) dado por  $Z(s_{k+1}) = Z(s_k) + \epsilon_k$ , con  $\epsilon_k \sim N(0, 1)$  variables aleatorias independientes,  $s_i \in \mathbb{R}$  con  $i = 1, 2, \dots, k, k+1, \dots$ . Puede probarse que este proceso cumple con las condiciones intrínsecas (1.1.12) y (1.1.13), en efecto,

$$\mathbb{E}[Z(s_{k+h}) - Z(s_k)] = \mathbb{E}\left[\sum_{i=k}^{k+h-1} \epsilon_i\right] = 0,$$

$$\text{Var}(Z(s_{k+h}) - Z(s_k)) = \text{Var}\left(\sum_{i=k}^{k+h-1} \epsilon_i\right) = h.$$

Sin embargo, para este proceso, la covarianza no depende solo de  $h$  pues

$$\begin{aligned} \text{Var}(Z(s_{k+h})) &= \text{Var}(Z(s_{k+h-1})) + \epsilon_{k+h-1} \\ &= \text{Var}(Z(s_{k+h-2})) + \epsilon_{k+h-2} + \epsilon_{k+h-1} \\ &\vdots \\ &= \text{Var}(Z(s_k)) + \epsilon_k + \epsilon_{k+1} + \cdots + \epsilon_{k+h-1} \\ &= \text{Var}(Z(s_k)) + h, \end{aligned}$$

y despejando la covarianza de la siguiente ecuación,

$$\text{Var}(Z(s_{k+h}) - Z(s_k)) = \text{Var}(Z(s_{k+h})) + \text{Var}(Z(s_k)) - 2\text{Cov}(Z(s_{k+h}), Z(s_k)),$$

se deduce que,

$$\begin{aligned} \text{Cov}(Z(s_{k+h}), Z(s_k)) &= \frac{1}{2} \left( \text{Var}(Z(s_{k+h})) + \text{Var}(Z(s_k)) - \text{Var}(Z(s_{k+h}) - Z(s_k)) \right) \\ &= \frac{1}{2} \left( (\text{Var}(Z(s_k)) + h) + \text{Var}(Z(s_k)) - h \right) \\ &= \text{Var}(Z(s_k)), \end{aligned}$$

de donde se sigue que la covarianza no depende solo de  $h$ . De aquí podemos concluir que el solo hecho de que un proceso  $Z(s_k)$  sea intrínseco no garantiza que el mismo sea débilmente estacionario. Ello se debe a que por más que sea intrínseco, dicha propiedad no nos brinda información acerca de su varianza de forma tal que podamos asegurar que  $\text{Var}(Z(s_k))$  depende solo de  $h$ .

## 1.2. Modelo espacial

El modelo geoestadístico clásico utilizado para dar estructura a un proceso espacial  $Z(\mathbf{s})$  esta dado por

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + Y(\mathbf{s}), \quad (1.2.1)$$

donde  $\mu(\mathbf{s}) = \mathbb{E}[Z(\mathbf{s})]$  es la función media, la cual consideramos continua y determinística y el proceso de error  $Y(\mathbf{s})$  es un proceso estocástico con media cero y función de covarianza  $\text{Cov}(Y(\mathbf{s}_i), Y(\mathbf{s}_j))$ , para todo  $\mathbf{s}_i, \mathbf{s}_j \in D$ .

Si el proceso  $Y(\mathbf{s})$  es débilmente estacionario, para todo  $\mathbf{s}_i, \mathbf{s}_j \in D$  tenemos que,

$$\text{Cov}(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = C(\mathbf{s}_i - \mathbf{s}_j).$$

Si  $Y(\mathbf{s})$  es intrínsecamente estacionario, para todo  $\mathbf{s}_i, \mathbf{s}_j \in D$  tenemos que,

$$\text{Var}(Y(\mathbf{s}_i) - Y(\mathbf{s}_j)) = 2\gamma_Y(\mathbf{s}_i - \mathbf{s}_j), \quad (1.2.2)$$

donde  $2\gamma_Y(\mathbf{s}_i, \mathbf{s}_j)$  es el **variograma del residuo** el cual es llamado así ya que, en la práctica, en lugar de usar los datos originales para hallarlo, se utilizan los residuos obtenidos a partir de la estimación de la función media (Wackernagel, 2006, pag. 181).

El modelo (1.2.1) tiene en cuenta la variación espacial a gran escala (tendencia) a través de la función media  $\mu(\mathbf{s})$ , y la variación espacial a pequeña escala (dependencia espacial) a través del proceso  $Y(\mathbf{s})$ . Además de capturar la variación espacial a pequeña escala, el proceso de error  $Y(\mathbf{s})$  en (1.2.1) explica, en parte, el error de medición que puede ocurrir en el proceso de recopilación de datos. Este componente generalmente no tiene estructura espacial, por lo tanto, para algunos propósitos puede ser deseable separarlo explícitamente del componente espacialmente dependiente. Es decir, podemos escribir

$$Y(\mathbf{s}) = \eta(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (1.2.3)$$

donde  $\eta(\cdot)$  y  $\epsilon(\cdot)$  son independientes,  $\eta(\cdot)$  es el componente que mide la dependencia espacial y  $\epsilon(\cdot)$  el proceso con media cero y no correlacionado espacialmente, que modela el error de medición. Más precisamente, su función de covarianza está dada por

$$Cov(\epsilon(\mathbf{s}), \epsilon(\mathbf{s} + \mathbf{h})) = \begin{cases} \sigma_\epsilon^2 \geq 0, & \text{si } \mathbf{h} = \mathbf{0}, \\ 0, & \text{si } \mathbf{h} \neq \mathbf{0}. \end{cases} \quad (1.2.4)$$

$\epsilon(\cdot)$  es conocido como **efecto pepita** (*nugget effect*) o proceso de ruido blanco.

Conociendo los supuestos sobre el proceso de error  $Y(\mathbf{s})$  y la función media  $\mu(\mathbf{s})$  deducimos el comportamiento del proceso  $Z(\mathbf{s})$  a partir de los siguientes lemas.

**Lema 1.2.1.** *Si el proceso  $Y(\mathbf{s})$  es estacionario de segundo orden y  $\mu(\mathbf{s}) \doteq \mu$  es constante entonces el proceso  $Z(\mathbf{s})$  también es estacionario de segundo orden.*

*Demostración.* Para probarlo consideramos el modelo definido por la Ecuación (1.2.1) a partir del cual tenemos que

$$\mathbb{E}(Z(\mathbf{s})) = \mu + \mathbb{E}(Y(\mathbf{s})) = \mu,$$

y

$$\begin{aligned} Cov(Z(\mathbf{s} + \mathbf{h}), Z(\mathbf{s})) &= Cov(\mu(\mathbf{s} + \mathbf{h}) + Y(\mathbf{s} + \mathbf{h}), \mu(\mathbf{s}) + Y(\mathbf{s})) \\ &= Cov(Y(\mathbf{s} + \mathbf{h}), Y(\mathbf{s})) \\ &= C(\mathbf{h}), \end{aligned}$$

luego  $Z(\mathbf{s})$  es estacionario de segundo orden.

□

**Lema 1.2.2.** *Si  $Y(\mathbf{s})$  es intrínsecamente estacionario y  $\mu(\mathbf{s}) = \mu$  constante entonces el proceso  $Z(\mathbf{s})$  también es intrínsecamente estacionario.*

*Demostración.* A partir del mismo razonamiento que en el Lema anterior, tenemos que

$$\mathbb{E}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})] = \mathbb{E}[\mu(\mathbf{s} + \mathbf{h}) + Y(\mathbf{s} + \mathbf{h}) - \mu(\mathbf{s}) - Y(\mathbf{s})] = 0,$$

y

$$\begin{aligned} \text{Var}[Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})] &= \text{Var}([\mu(\mathbf{s} + \mathbf{h}) + Y(\mathbf{s} + \mathbf{h}) - \mu(\mathbf{s}) - Y(\mathbf{s})]) \\ &= \text{Var}([Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})]) = 2\gamma_Y(\mathbf{h}), \end{aligned}$$

con lo cual se prueba que  $Z(\mathbf{s})$  es intrínsecamente estacionario.  $\square$

En la práctica estas cantidades, en particular el semivariograma y el covariograma, se estiman empíricamente a partir de los datos y luego, a partir de los mismos se deducen sus modelos paramétricos como veremos en las próximas secciones.

### 1.2.1. Semivariograma y covariograma empíricos

En muchos casos, incluso en aquellos donde se cuenta con información para suponer que el proceso es débilmente estacionario, se prefiere utilizar el semivariograma antes que covariograma pues, éste último requiere conocer la media del proceso  $Z(\mathbf{s})$ . En la práctica, ésta es desconocida y debe estimarse a partir de los datos, lo que introduce un sesgo en la estimación del covariograma. Para demostrar este hecho, definimos las versiones empíricas del covariograma y semivariograma de un proceso  $Z(\mathbf{s})$ .

Si se cumplen las condiciones de estacionariedad intrínseca (1.1.12) y (1.1.13) o las condiciones de estacionariedad de segundo orden (1.1.4) y (1.1.5), los estimadores empíricos del semivariograma y el covariograma construidos a partir del método de los momentos son, respectivamente,

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i))^2 \quad (1.2.5)$$

y

$$\hat{C}(\mathbf{h}) = \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - \bar{Z})(Z(\mathbf{s}_i) - \bar{Z}), \quad (1.2.6)$$

donde  $m(\mathbf{h})$  es la cantidad de pares de puntos muestrales  $(\mathbf{s}_i, \mathbf{s}_j)$  que están a distancia  $\mathbf{h}$ , esto es,  $\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}$  para un vector  $\mathbf{h}$  dado y  $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z(\mathbf{s}_i)$  un estimador de la media constante  $\mu$ .

Observemos entonces que el covariograma empírico (1.2.6) utiliza la media muestral,  $\bar{Z}$ , como un estimador de la media  $\mu(\mathbf{s})$  del proceso. Sin embargo, utilizar dicho promedio como estimador de la esperanza del proceso hace que  $\hat{C}(\mathbf{h})$  sea sesgado. En efecto, consideremos

la siguiente versión simplificada de la esperanza de (1.2.6)

$$\mathbb{E}(\hat{C}(\mathbf{h})) = \mathbb{E} \left[ \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - \bar{Z}_{\mathbf{h}})(Z(\mathbf{s}_i) - \bar{Z}_{(\mathbf{h})}) \right], \quad (1.2.7)$$

donde las medias muestrales se calculan considerando solo los  $m(\mathbf{h})$  pares de puntos que están a distancia  $\mathbf{h}$  y no sobre toda la muestra; esto es

$$\bar{Z}_{\mathbf{h}} \doteq \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} Z(\mathbf{s}_i + \mathbf{h}) \quad \text{y} \quad \bar{Z}_{(\mathbf{h})} \doteq \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} Z(\mathbf{s}_i). \quad (1.2.8)$$

Operando en (1.2.7) obtenemos que,

$$\begin{aligned} \mathbb{E}(\hat{C}(\mathbf{h})) &= \mathbb{E} \left[ \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - \bar{Z}_{\mathbf{h}}) (Z(\mathbf{s}_i) - \bar{Z}_{(\mathbf{h})}) \right] \\ &= \mathbb{E} \left[ \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h})Z(\mathbf{s}_i) - Z(\mathbf{s}_i + \mathbf{h})\bar{Z}_{(\mathbf{h})} - Z(\mathbf{s}_i)\bar{Z}_{\mathbf{h}} + \bar{Z}_{\mathbf{h}}\bar{Z}_{(\mathbf{h})}) \right] \\ &= \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \mathbb{E} [Z(\mathbf{s}_i + \mathbf{h})Z(\mathbf{s}_i)] - \mathbb{E} \left[ \bar{Z}_{(\mathbf{h})} \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} Z(\mathbf{s}_i + \mathbf{h}) \right] \\ &\quad - \mathbb{E} \left[ \bar{Z}_{\mathbf{h}} \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} Z(\mathbf{s}_i) \right] + \mathbb{E} [\bar{Z}_{\mathbf{h}}\bar{Z}_{(\mathbf{h})}] \quad (1.2.9) \\ &= \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \mathbb{E} [Z(\mathbf{s}_i + \mathbf{h})Z(\mathbf{s}_i)] - \mathbb{E} [\bar{Z}_{\mathbf{h}}\bar{Z}_{(\mathbf{h})}]. \quad (\text{por (1.2.8)}) \end{aligned}$$

Calculamos la esperanza  $\mathbb{E} [\bar{Z}_{\mathbf{h}}\bar{Z}_{(\mathbf{h})}]$  considerando el proceso  $Z(\mathbf{s})$  estacionario de segundo orden, esto es,  $\mathbb{E}[Z(\mathbf{s})] = \mu$  y  $Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_i + \mathbf{h})) = C(\mathbf{h})$ ,

$$\begin{aligned} \mathbb{E}(\bar{Z}_{\mathbf{h}}\bar{Z}_{(\mathbf{h})}) &= Cov(\bar{Z}_{\mathbf{h}}, \bar{Z}_{(\mathbf{h})}) + \mathbb{E}(\bar{Z}_{\mathbf{h}})\mathbb{E}(\bar{Z}_{(\mathbf{h})}) \\ &= \frac{1}{m(\mathbf{h})^2} \sum_{i=1}^{m(\mathbf{h})} \sum_{j=1}^{m(\mathbf{h})} Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j + \mathbf{h})) + \mu^2 \\ &= \frac{1}{m(\mathbf{h})^2} \sum_{i=1}^{m(\mathbf{h})} Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_i + \mathbf{h})) \\ &\quad + \frac{1}{m(\mathbf{h})^2} \sum_{i=1}^{m(\mathbf{h})} \sum_{j \neq i}^{m(\mathbf{h})} Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j + \mathbf{h})) + \mu^2 \\ &= \frac{1}{m(\mathbf{h})} C(\mathbf{h}) + \frac{1}{m(\mathbf{h})^2} \sum_{i=1}^{m(\mathbf{h})} \sum_{j \neq i}^{m(\mathbf{h})} Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j + \mathbf{h})) + \mu^2. \quad (1.2.10) \end{aligned}$$

Luego reemplazamos (1.2.10) en (1.2.9) y, teniendo en cuenta que

$$\mathbb{E} [Z(\mathbf{s}_i + \mathbf{h})Z(\mathbf{s}_i)] = C(\mathbf{h}) + \mu^2,$$

obtenemos

$$\begin{aligned}
\mathbb{E}(\hat{C}(\mathbf{h})) &= \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \mathbb{E}[Z(\mathbf{s}_i + \mathbf{h})Z(\mathbf{s}_i)] - \mathbb{E}[\bar{Z}_{\mathbf{h}}\bar{Z}(\mathbf{h})] \\
&= C(\mathbf{h}) + \mu^2 - \left( \frac{1}{m(\mathbf{h})}C(\mathbf{h}) + \frac{1}{m(\mathbf{h})^2} \sum_{i=1}^{m(\mathbf{h})} \sum_{j \neq i}^{m(\mathbf{h})} Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j + \mathbf{h})) + \mu^2 \right) \\
&= \left(1 - \frac{1}{m(\mathbf{h})}\right) C(\mathbf{h}) - \frac{1}{m(\mathbf{h})^2} \sum_{i=1}^{m(\mathbf{h})} \sum_{j \neq i}^{m(\mathbf{h})} Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j + \mathbf{h})).
\end{aligned}$$

Con lo cual podemos observar que el estimador  $\hat{C}(\mathbf{h})$  es sesgado y el sesgo disminuye a medida que aumenta la cantidad de puntos  $m(\mathbf{h})$  considerada.

Por otra parte podemos probar que  $\hat{\gamma}(\mathbf{h})$  es insesgado, esto es,

$$\begin{aligned}
\mathbb{E}[\hat{\gamma}(\mathbf{h})] &= \mathbb{E} \left[ \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} (Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i))^2 \right] \\
&= \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \frac{1}{2} \mathbb{E} [(Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i))^2] \\
&= \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \frac{1}{2} Var [Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)] \\
&= \frac{1}{m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} \gamma(\mathbf{h}) \\
&= \gamma(\mathbf{h}).
\end{aligned}$$

Entonces, a partir de la muestra disponible podemos calcular el semivariograma empírico (1.2.5), el cual resume la relación espacial presente en los datos. Para cada dirección  $\mathbf{h}$  podemos hallar también un gráfico  $\hat{\gamma}(h)$  con  $h \doteq \|\mathbf{h}\|$  como se muestra en la Figura 1.2, el cual juega un papel muy importante en el análisis exploratorio de datos espaciales ya que brinda información acerca de la variabilidad del proceso. Si para cada dirección  $\mathbf{h}$  encontramos un gráfico diferente quiere decir que el variograma depende de  $\mathbf{h}$  tanto en magnitud y dirección. A este comportamiento se le llama **anisotrópico**. Cuando depende solo de la magnitud (distancia)  $h$ , se llama **isotrópico**. En este caso solo trabajamos con variogramas isotrópicos y utilizaremos la notación  $\gamma^*$  para diferenciarlo del anisotrópico.

A partir del gráfico  $\hat{\gamma}(h)$  que observamos en la Figura 1.2 podemos inferir cuál podría ser el modelo para el semivariograma teórico (Sección 1.2.2) para luego poder estimar sus parámetros. Hallamos el semivariograma empírico a partir de una dirección  $\mathbf{h}$  y una distancia máxima  $h_{max}$ . Luego dividimos el intervalo  $[0, h_{max}]$  de forma tal que en cada subintervalo exista una cantidad mínima dada de pares de sitios  $m(h)$  a considerar en cada uno. Luego, para cada intervalo  $[h_i, h_j]$  calculamos  $\hat{\gamma}(h)$  tal que  $h \in [h_i, h_j]$ .

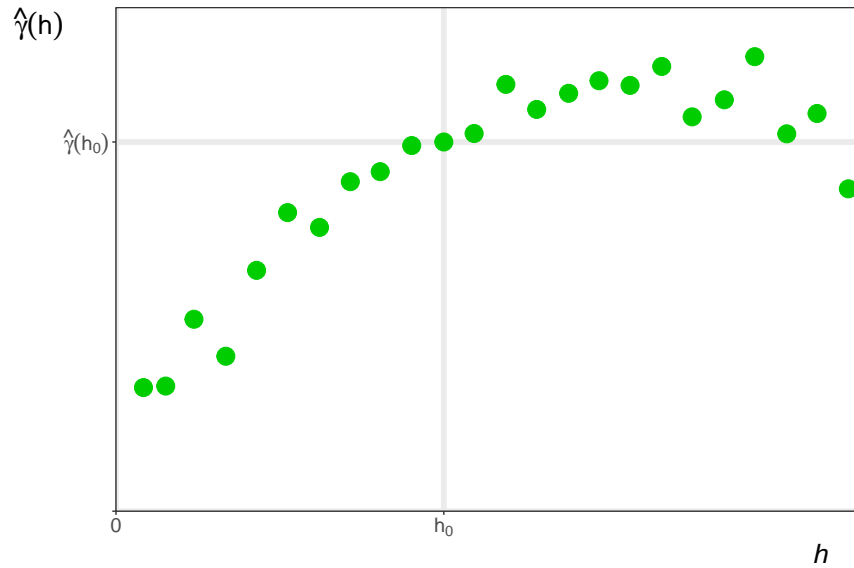


Figura 1.2: Ejemplo del variograma empírico para los datos Meuse. (Rikken y Van Rijn, 1993)

### 1.2.2. Semivariograma teórico

Como puede observarse en la Figura 1.2, la versión empírica del semivariograma no nos brinda información sobre la dispersión de variables definidas en sitios cuyas distancias no hayan sido observadas en la muestra (i.e. no es continuo en  $\mathbf{h}$ ). Además no cumple con ciertas condiciones que son necesarias al momento de la predicción, las cuales explicitaremos más adelante. Por lo tanto, utilizaremos el semivariograma empírico (1.2.5) para modelar y estimar los parámetros correspondientes del semivariograma teórico, a partir de modelos que son funciones válidas de semivariograma.

Primero revisaremos las principales propiedades matemáticas de los semivariogramas, caracterizando así la clase de funciones que pueden representar a los mismos. Como hemos mencionado en (1.1.13), el **semivariograma teórico** para un proceso intrínseco está dado por

$$\gamma_{\boldsymbol{\theta}}(\mathbf{h}) = \gamma(\mathbf{h}) = \frac{1}{2} \text{Var}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s})).$$

En lo que sigue, utilizaremos la notación  $\gamma(\mathbf{h})$  o  $\gamma_{\boldsymbol{\theta}}(\mathbf{h})$  indistintamente. La notación  $\gamma_{\boldsymbol{\theta}}(\mathbf{h})$  hace hincapié en que el semivariograma no es más que una función paramétrica con una expresión analítica simple que depende de la distancia entre ubicaciones espaciales  $h$  y de un vector de parámetros,  $\boldsymbol{\theta}$ . Para ser considerada una función de semivariograma válida, la función  $\gamma_{\boldsymbol{\theta}}(\mathbf{h})$ , debe satisfacer las siguientes propiedades, para cada  $\boldsymbol{\theta} \in \Theta$ , donde  $\Theta$  es el espacio de parámetros:

- I)  $\gamma_{\boldsymbol{\theta}}(\mathbf{0}) = 0$ ;
- II)  $\gamma_{\boldsymbol{\theta}}(\mathbf{h}) \geq 0$ ;

$$\text{III) } \gamma_{\boldsymbol{\theta}}(-\mathbf{h}) = \gamma_{\boldsymbol{\theta}}(\mathbf{h});$$

IV) Si el proceso  $Z(\mathbf{s})$  es intrínsecamente estacionario,  $\gamma$  es condicionalmente definido negativo. Esto es, para cualquier vector  $\mathbf{a}^T = (a_1, \dots, a_n)$ ,  $a_i \in \mathbb{R}$  que verifique  $\mathbf{a}^T \mathbf{1} = 0$ , se tiene que  $\mathbf{a}^T \boldsymbol{\Gamma} \mathbf{a} \leq 0$ , donde  $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$  es la matriz semivariograma entre los puntos de la muestra tal que  $\Gamma_{i,j} \doteq \gamma_{\boldsymbol{\theta}}(\mathbf{s}_i - \mathbf{s}_j)$  para  $i, j = 1, \dots, n$ .

Las tres primeras propiedades se prueban fácilmente mediante la definición de  $\gamma$ . Para probar la Propiedad IV), utilizaremos la condición de estacionariedad intrínseca (1.1.14) y que la sumatoria de los  $a_i$  es cero. Luego,

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathbb{E}[(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2] && \text{(de (1.1.14))} \\ &= \frac{1}{2} \mathbb{E} \sum_{i=1}^n \left( \sum_{j=1}^n a_j \right) a_i Z(\mathbf{s}_i)^2 \\ &\quad - \mathbb{E} \sum_{i=1}^n \sum_{j=1}^n a_i a_j Z(\mathbf{s}_i) Z(\mathbf{s}_j) \\ &\quad + \frac{1}{2} \mathbb{E} \sum_{j=1}^n \left( \sum_{i=1}^n a_i \right) a_j Z(\mathbf{s}_j)^2 \\ &= -\mathbb{E} \sum_{i=1}^n \sum_{j=1}^n a_i a_j Z(\mathbf{s}_i) Z(\mathbf{s}_j) && \left( \text{pues } \sum_{i=1}^n a_i = \sum_{j=1}^n a_j = 0 \right) \\ &= -\mathbb{E} \left[ \sum_{i=1}^n a_i Z(\mathbf{s}_i) \right]^2 \leq 0. \end{aligned}$$

La forma clásica de modelar el semivariograma teórico es utilizando modelos paramétricos y luego eligiendo un método para estimar los parámetros del mismo. La elección del modelo, entre la colección de modelos de semivariograma válidos, se basa en un examen del semivariograma empírico, y también, en el conocimiento previo que se tenga del problema, la simplicidad computacional, entre otras cosas.

Una característica que tienen los datos espaciales es que, generalmente, la dependencia espacial decrece con la distancia, por lo que el covariograma es una función decreciente en  $h$ . Dado que el semivariograma es simétrico, con respecto a la constante  $C(\mathbf{0})/2$ , al covariograma (ver (1.1.7)), generalmente  $\gamma_{\boldsymbol{\theta}}(h)$  es una función monótona creciente en función de la distancia entre sitios.

Si bien existe una gran variedad de modelos que satisfacen los requisitos de validez antes mencionados, y que además son monótonos e isotrópicos, los más utilizados son los que se detallan a continuación.

**Modelo pepita puro:** Para  $\theta_1 > 0$ ,

$$\gamma^*(h, \theta_1) = \begin{cases} \theta_1, & \text{si } h \neq 0, \\ 0, & \text{si } h = 0. \end{cases} \quad (1.2.11)$$

**Modelo esférico:** Para  $\boldsymbol{\theta} = (\theta_1, \theta_2)$  con  $\theta_1 > 0$ ,  $\theta_2 > 0$ ,

$$\gamma^*(h, \boldsymbol{\theta}) = \begin{cases} \theta_1 \left[ \frac{3h}{2\theta_2} - \frac{1}{2} \left( \frac{h}{\theta_2} \right)^3 \right], & \text{si } 0 < h \leq \theta_2, \\ \theta_1, & \text{si } h > \theta_2. \end{cases} \quad (1.2.12)$$

**Modelo potencia:** Para  $\boldsymbol{\theta} = (\theta_1, \theta_2)$  con  $\theta_1 > 0$ ,  $0 \leq \theta_2 < 2$ ,

$$\gamma^*(h, \boldsymbol{\theta}) = \theta_1 h^{\theta_2}.$$

**Modelo exponencial:** Para  $\boldsymbol{\theta} = (\theta_1, \theta_2)$  con  $\theta_1 > 0$ ,  $\theta_2 > 0$ ,

$$\gamma^*(h, \boldsymbol{\theta}) = \theta_1 \left[ 1 - \exp \left( -\frac{h}{\theta_2} \right) \right]. \quad (1.2.13)$$

**Modelo gaussiano:** Para  $\boldsymbol{\theta} = (\theta_1, \theta_2)$  con  $\theta_1 > 0$ ,  $\theta_2 > 0$ ,

$$\gamma^*(h, \boldsymbol{\theta}) = \theta_1 \left[ 1 - \exp \left( -\frac{h^2}{\theta_2^2} \right) \right].$$

**Modelo de Matérn:** Para  $\boldsymbol{\theta} = (\theta_1, \theta_2, \nu)$  con  $\nu > -1$ ,  $\theta_1 > 0$  y  $\theta_2 > 0$ ,

$$\gamma^*(h, \boldsymbol{\theta}) = \theta_1 \left[ 1 - \left( -\frac{\left( \frac{h}{\theta_2} \right)^\nu \kappa_\nu \left( \frac{h}{\theta_2} \right)}{2^{\nu-1} \Gamma(\nu)} \right) \right], \quad (1.2.14)$$

donde  $\kappa_\nu(\cdot)$  es la función Bessel modificada de segunda especie y orden  $\nu$ . En este modelo,  $\nu > -1$  es un parámetro de suavizado y se puede observar que si  $\nu = 0.5$  y  $\nu \rightarrow \infty$  coincide con el exponencial y el gaussiano, respectivamente.

Una cuestión esencial a la hora de ajustar un modelo paramétrico no es solo la estimación de sus parámetros sino también la elección del mismo. En este caso en particular, además, es importante comprender que estas decisiones implican una interpretación del comportamiento del semivariograma tanto en el origen como a grandes distancias. El comportamiento mencionado lo estudiamos a través del semivariograma empírico que observamos en la Figura 1.2, el cual nos brinda la única información disponible sobre el proceso subyacente para la estimación de los parámetros y la interpretación de los mismos. Los parámetros asociados al comportamiento del semivariograma a grandes distancias son la meseta y el rango, los cuales están graficados en la Figura 1.3. La tasa de crecimiento del semivariograma refleja el grado de disimilaridad entre valores cada vez más distantes, la misma puede

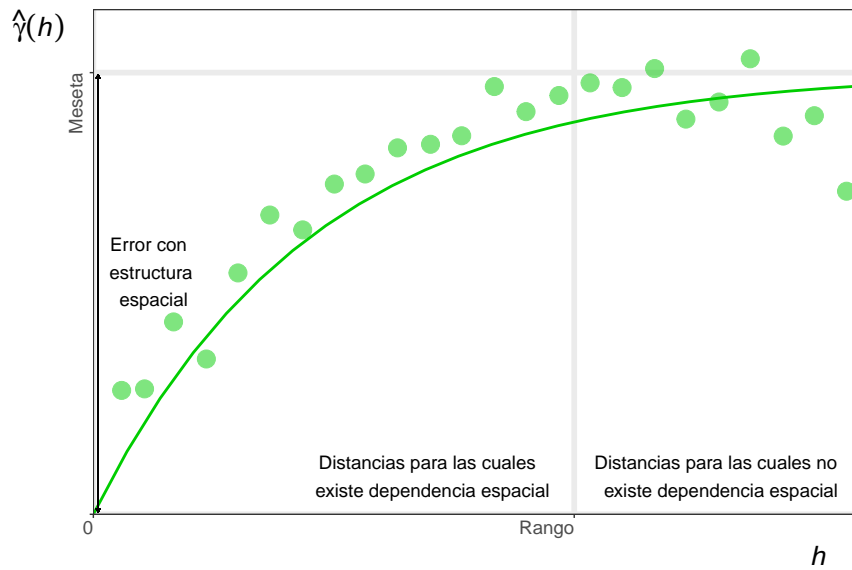


Figura 1.3: Semivariograma empírico y semivariograma estimado, a partir de un modelo exponencial (Ecuación 1.2.13), para los datos Meuse.

mantenerse si la variabilidad del fenómeno no tiene límite a grandes distancias o puede tender a desaparecer a partir de cierto  $h$  en cuyo caso el semivariograma se estabiliza en lo que llamamos meseta. El valor a partir del cual el semivariograma se estabiliza es el rango y es la distancia a partir de la cual el semivariograma no tiene cambio, esto es,  $Z(\mathbf{s})$  y  $Z(\mathbf{s} + \mathbf{h})$  no están correlacionadas espacialmente. Formalmente se llama **meseta** de  $\gamma^*(h, \boldsymbol{\theta})$  al

$$\lim_{h \rightarrow \infty} \gamma^*(h, \boldsymbol{\theta}),$$

siempre que este límite exista. Los variogramas de procesos estacionarios de segundo orden alcanzan este límite en los que permanecen a partir de un cierto  $h$  si la covarianza tiende a cero cuando  $h$  tiende a infinito. Esto es, si

$$\lim_{h \rightarrow \infty} C(h) = 0.$$

Por lo tanto, y a partir de la ecuación (1.1.7), tenemos que

$$\lim_{h \rightarrow \infty} \gamma^*(h, \boldsymbol{\theta}) = C(\mathbf{0}), \quad (1.2.15)$$

por lo que a priori y en muchos casos la meseta está determinada por la varianza del proceso. El caso en el que el semivariograma empírico sea no acotado puede ser un síntoma de la invalidez del supuesto de estacionariedad débil como ya lo hemos mencionado anteriormente. Los modelos esféricos, exponenciales, gaussianos y de Matérn tienen mesetas y en la parametrización utilizada esta dada por  $\theta_1$ . Si la meseta existe, entonces el **rango** de  $\gamma^*(h, \boldsymbol{\theta})$  es el valor más pequeño de  $h$  para el cual  $\gamma^*(h, \boldsymbol{\theta})$  es igual a su meseta. Formalmente,

$$\min \{h : \gamma^*(h, \boldsymbol{\theta}) = \lim_{h \rightarrow \infty} \gamma^*(h, \boldsymbol{\theta})\}.$$

Si el rango se alcanza de manera asintótica, se dice que no existe (estrictamente hablando), para este caso existe una noción relacionada llamada **rango efectivo** que se define como el mínimo valor de  $h$  para el cual  $\gamma^*(h, \boldsymbol{\theta})$  es igual al 95 % de la meseta. En este caso, el rango efectivo es a menudo una función de un parámetro llamado **parámetro de rango**. De los modelos enumerados anteriormente que tienen meseta, por definición solo el esférico tiene rango (igual a  $\theta_2$ ). De los otros modelos, puede probarse que el modelo exponencial tiene rango efectivo de aproximadamente 3 veces el rango  $\theta_2$ , esto es,  $3\theta_2$ . En efecto, si igualamos  $\gamma^*(h, \boldsymbol{\theta})$  al 95 % de la meseta tenemos que,

$$\theta_1 \left[ 1 - \exp\left(-\frac{h^2}{\theta_2^2}\right) \right] = 0.95\theta_1,$$

de donde se sigue que,

$$h = -\ln(0.5)\theta_2 \approx 3\theta_2.$$

De manera análoga puede probarse que el modelo gaussiano tiene rango efectivo de aproximadamente  $\sqrt{3}\theta_2$ , donde  $\theta_2$  es parámetro de rango.

Los parámetros de rango pueden ser difíciles de estimar incluso con conjuntos de datos bastante grandes, en particular, cuando el rango no es mucho más pequeño que las dimensiones de la región en observación. Esta dificultad es quizás un argumento para usar el modelo potencia, que es esencialmente la clase Matérn para  $\nu < 1$  con el rango establecido en infinito, evitando así la necesidad de estimar el mismo.

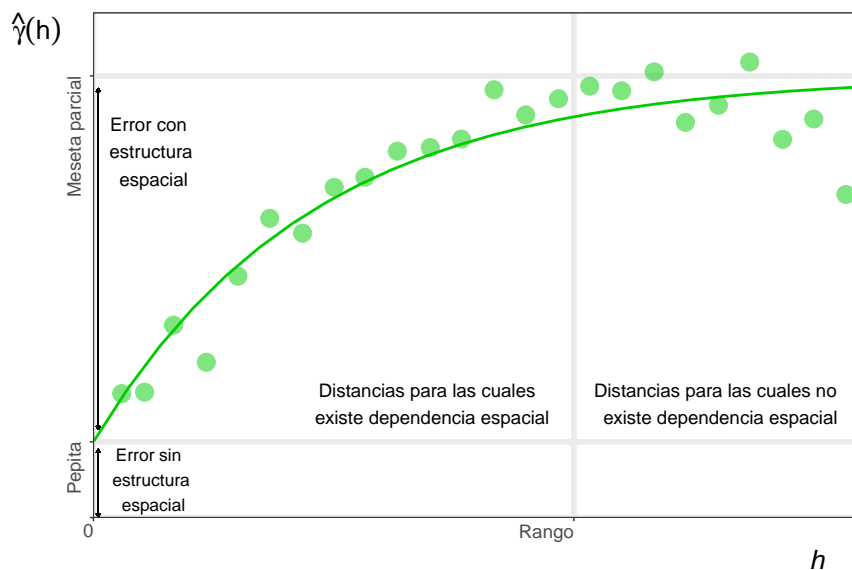


Figura 1.4: Semivariograma empírico y semivariograma estimado, a partir de una combinación entre el modelo pepita puro (1.2.11) y un modelo exponencial (1.2.13), para los datos Meuse.

Como mencionamos en la Propiedad 1), una condición para que  $\gamma^*(h, \boldsymbol{\theta})$  sea una función de semivariograma válida es que

$$\gamma^*(0, \boldsymbol{\theta}) = 0.$$

En la práctica esto muchas veces no sucede como por ejemplo en el semivariograma dado en la Figura 1.4, donde los valores del mismo cuando  $h$  se acerca a 0 parecen acercarse a algún valor estrictamente positivo. Este fenómeno puede explicarse a partir de la descomposición del error dada en (1.2.3), lo cual implica un cambio en las funciones presentadas anteriormente (1.2.11)-(1.2.14). Esto significa que, utilizando dicha descomposición tenemos que

$$\begin{aligned}\gamma_Y^*(h, \boldsymbol{\theta}) &= \text{Var} [Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})] \\ &= \text{Var} [\eta(\mathbf{s} + \mathbf{h}) + \epsilon(\mathbf{s} + \mathbf{h}) - \eta(\mathbf{s}) - \epsilon(\mathbf{s})] \\ &= \text{Var} [\eta(\mathbf{s} + \mathbf{h}) - \eta(\mathbf{s})] + \text{Var} [\epsilon(\mathbf{s} + \mathbf{h}) - \epsilon(\mathbf{s})] \quad (\eta \text{ y } \epsilon \text{ independientes}) \\ &= \gamma_\eta^*(h, \boldsymbol{\theta}) + \gamma_\epsilon^*(h, \boldsymbol{\theta}),\end{aligned}\tag{1.2.16}$$

donde  $\gamma_{\eta(\mathbf{s})}^*(h, \boldsymbol{\theta})$  es el semivariograma del proceso que captura la variabilidad espacial y  $\gamma_{\epsilon(\mathbf{s})}^*(h, \boldsymbol{\theta})$  es el semivariograma del proceso asociado al error de medición, el cual es el que hemos llamado *Modelo Pepita Puro* en (1.2.11). Si en el semivariograma empírico observamos que los valores del mismo no se acercan a cero cuando  $h$  sí lo hace y además el comportamiento cercano a cero es aproximadamente lineal, una buena opción para el ajuste puede ser el dado por

$$\gamma_{Y(\mathbf{s})}^*(h, \boldsymbol{\theta}) = \begin{cases} 0, & \text{si } h = 0, \\ \theta_1^\epsilon + \theta_1^\eta \left[ 1 - \exp\left(-\frac{h}{\theta_2}\right) \right], & \text{si } h > 0, \end{cases}$$

el cual es la combinación de los modelos (1.2.11) y (1.2.13), en términos de (1.2.16) donde  $\theta_1^\epsilon$  es la parametrización para el efecto pepita, que no es más que el parámetro meseta en el modelo (1.2.11), y  $\theta_1^\eta$  la meseta del modelo (1.2.13), llamada **meseta parcial**, con lo cual la meseta del modelo  $\gamma_{Y(\mathbf{s})}^*(h, \boldsymbol{\theta})$  es  $\theta_1^\epsilon + \theta_1^\eta$ , esto es, la meseta se divide en meseta parcial y efecto pepita, que es la meseta del modelo pepita puro, lo cual podemos observar en la Figura 1.4. Por ello y si elegimos ajustar un semivariograma teórico que considere un efecto pepita, tendremos que estimar tres parámetros la meseta parcial y el rango dados por  $\theta_1$  y  $\theta_2$ , respectivamente, y el efecto pepita que parametrizamos como  $\theta_3$ .

Para elegir el mejor modelo que se ajuste a los datos es importante analizar el comportamiento del semivariograma cerca del origen, y particularmente en el origen, ya que esto nos brinda información sobre intervalos de distancias donde la dependencia espacial es grande. Dicho comportamiento es analizado a través de qué tan suave es la superficie generada por el proceso subyacente, la cual clásicamente se caracteriza a partir de la continuidad y diferenciabilidad del mismo.

Por ejemplo, si el comportamiento de  $\gamma^*(h, \boldsymbol{\theta})$  cerca de cero es lineal, como por ejemplo en el modelo exponencial, estamos asumiendo una variable regionalizada menos suave que en el caso de elegir un modelo gaussiano, el cual posee un comportamiento parabólico para distancias cercanas a 0. Estos comportamientos se observan en las Figuras 1.5 (b) y (c). Si consideramos que el semivariograma tiene un efecto pepita, tendremos una discontinuidad

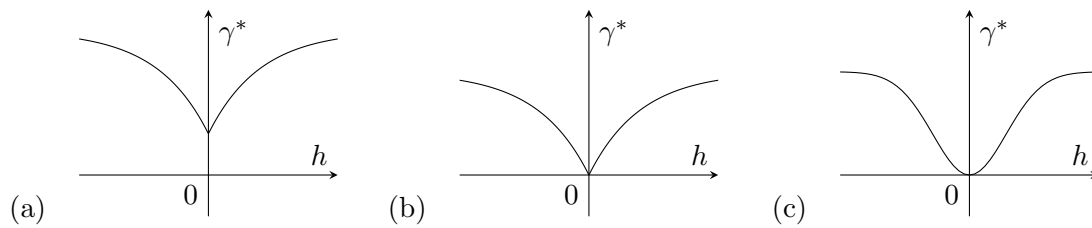


Figura 1.5: Comportamiento del semivariograma cerca al origen: (a) intersección positiva (pepita); (b) comportamiento lineal continuo, no diferenciable; (c) comportamiento parabólico, continuo y diferenciable.

en el origen como la observada en la Figura 1.5 (a). En este caso asumimos que la variable regionalizada presenta irregularidades para distancias pequeñas.

Un tipo de continuidad muy utilizada es la **continuidad en media cuadrática** que se define como

$$\mathbb{E} [(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2] \rightarrow 0, \quad \text{cuando } \|\mathbf{h}\| \rightarrow 0.$$

en cuyo caso diremos que el proceso  $Z(\mathbf{s})$  es continuo en media cuadrática. A partir de esta definición podemos afirmar que si el variograma  $2\gamma(\cdot)$  es continuo en el origen entonces el proceso  $Z(\mathbf{s})$  es continuo en media cuadrática, ya que  $\mathbb{E} [(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2] \rightarrow 0$  si y sólo si  $2\gamma(\mathbf{h}) \rightarrow 0$  cuando  $\|\mathbf{h}\| \rightarrow 0$ . Los modelos (1.2.11)-(1.2.14) presentados anteriormente están definidos de manera tal que  $\gamma^*(h, \boldsymbol{\theta}) \rightarrow 0$  cuando  $h \rightarrow 0$  además, por la Propiedad 1)  $\gamma^*(0, \boldsymbol{\theta}) = 0$ , por lo que son funciones continuas. Esto significa que el proceso subyacente asociado a dichas funciones de semivariograma genera superficies más o menos suaves.

Según Wackernagel (1995) no es tan relevante qué tan bien la función semivariograma se ajusta a la secuencia de puntos hallados a partir del semivariograma empírico, lo importante es el tipo de continuidad asumido para la variable regionalizada y la hipótesis de estacionariedad asociada al proceso. Estos supuestos guiarán la selección de una función de semivariograma apropiada y esto tiene muchas más implicaciones que la forma en que la función teórica se ajusta al semivariograma empírico.

### 1.2.2.1. Estimación de los parámetros del semivariograma teórico

Los modelos descritos en (1.2.11)-(1.2.14) son los que comúnmente se usan para modelar semivariogramas. Resta ver cómo se estiman sus parámetros (pepita, meseta y rango) para lograr el mejor ajuste al semivariograma empírico. Como lo afirman los autores Webster y Oliver (2007, pág. 101), elegir modelos y ajustarlos a los valores del semivariograma empírico se encuentra entre los temas más controvertidos en geoestadística. Este trabajo se puede tornar difícil ya que la precisión de cada estimación varía debido a las diferencias de tamaños de muestra; la variabilidad puede no ser la misma en todas las direcciones; la gráfica del semivariograma empírico puede mostrar mucha fluctuación punto a punto; la

mayoría de los modelos no son lineales en uno o más parámetros y se deben usar métodos iterativos para estimar los mismos. Además existe la posibilidad de que los valores de los semivariogramas empíricos estén correlacionados.

A partir todos estos problemas es que se sugiere (Wackernagel, 2006; Webster y Oliver, 2007; Montero et al., 2015) dos formas de ajustar el modelo: la inspección y ajuste manual del mismo y los correspondientes métodos estadísticos. La forma manual consiste en utilizar métodos gráficos, los valores del semivariograma empírico, el conocimiento del fenómeno ya que la forma analítica específica del semivariograma no importa tanto como el hecho de que respete las características principales del fenómeno. El valor del efecto de pepita se puede obtener a partir de los primeros valores del semivariograma empírico extrapolando hasta que corten el eje de las ordenadas. Este valor, el cual representa, el comportamiento cerca del origen (para distancias inferiores a la primera distancia de retraso) es otra decisión central en el ajuste manual porque tiene una gran influencia en los resultados de la predicción. La pendiente del semivariograma en el origen puede derivarse también de los primeros valores del semivariograma empírico. En cuanto al comportamiento para distancias grandes, representado por el rango (o rango práctico en caso de tener una meseta asintótica) y la meseta, generalmente se detectan fácilmente, especialmente en el caso estacionario de segundo orden: la distancia más allá de la cual el semivariograma se estabiliza y el valor del semivariograma cuando se estabiliza, respectivamente. Estas cuestiones traen consigo el problema de la subjetividad por parte del investigador. La forma no manual tiene que ver con la utilización de métodos estadísticos conocidos:

- **Métodos de mínimos cuadrados:** ordinarios (OLS, Ordinary Least Square), generalizados (GLS, Generalized Least Square) y pesados (WLS, Weighted Least Square).
- **Métodos basados en la máxima verosimilitud:** incluyen los métodos de máxima verosimilitud (ML, Maximum Likelihood) y los de máxima verosimilitud restringida (REML, Restricted Maximum Likelihood).

### Métodos de mínimos cuadrados

Dada una dirección del vector de traslación  $\mathbf{h}$ , deseamos ajustar un modelo de semivariograma dentro de la familia de semivariogramas teóricos  $\gamma^*(h, \boldsymbol{\theta})$  dados en (1.2.11)-(1.2.14). El vector  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$  contiene la meseta, el parámetro de rango y el efecto pepita, respectivamente. Obtenemos el estimador del semivariograma empírico  $\hat{\gamma}(\mathbf{h})$  en  $K$  rezagos  $\hat{\gamma} \doteq (\hat{\gamma}(h_1), \dots, \hat{\gamma}(h_K))^T$  donde  $K$  es fijo y la cantidad de pares de sitios  $m(h)$  que contribuyen al estimador en cada rezago es grande (al menos 30 pares según Journel y Huijbregts (1978), pág. 194). Sea  $\gamma^*(h, \boldsymbol{\theta})$  un modelo de variograma cuya forma exacta se conoce excepto por los parámetros desconocidos  $\boldsymbol{\theta}$  con el que hallamos  $\boldsymbol{\gamma}_{\boldsymbol{\theta}}^* \doteq (\gamma^*(h_1, \boldsymbol{\theta}), \dots, \gamma^*(h_K, \boldsymbol{\theta}))^T$ . Supongamos que  $\hat{\gamma}$  es el vector  $K \times 1$  variables aleatorias con matriz de covarianzas  $Var(\hat{\gamma}) = \mathbf{V}$  que depende de  $\boldsymbol{\theta}$ . Los métodos de mínimos cuadrados, que no requieren

el conocimiento de la distribución de probabilidad de los valores de semivariogramas experimentales, se basan en el modelo

$$\hat{\gamma} = \gamma_{\theta}^* + \mathbf{e},$$

donde  $\mathbf{e}$  tiene media  $\mathbf{0}$  y matriz de covarianza  $\mathbf{V}$ .

Con el **método OLS** estimamos  $\theta$  minimizando

$$Q(\theta) = \sum_{i=1}^K (\hat{\gamma}(h_i) - \gamma^*(h_i, \theta))^2,$$

donde  $\gamma^*(h_i, \theta)$  es el modelo teórico valuado en la distancia  $h_i$ . Este método no tiene en cuenta el comportamiento del semivariograma cerca del origen ni la posibilidad de que los valores  $\hat{\gamma}$  estén correlacionados. Para conservar la geometría, pero también para introducir el concepto de covarianza en el procedimiento, consideramos el **método GLS** y estimamos  $\theta$  minimizando

$$Q(\theta) = (\hat{\gamma} - \gamma_{\theta}^*)^T \mathbf{V}^{-1} (\hat{\gamma} - \gamma_{\theta}^*).$$

Es este caso el problema es determinar la matriz  $\mathbf{V}$ , por lo que generalmente se supone el proceso  $Z(\mathbf{s})$  gaussiano y luego se procede iterativamente a partir de una solución inicial.

En el caso del **método WLS** se minimiza la siguiente expresión (Cressie, 1985)

$$Q(\theta) = \sum_{i=1}^K \frac{N(h_i)}{[\hat{\gamma}(h_i, \theta)]^2} (\gamma(h_i) - \gamma(h_i, \theta))^2, \quad (1.2.17)$$

donde  $N(h_i)$  es la cantidad de unidades espaciales que se encuentran a una distancia  $h_i$ . WLS puede verse como un compromiso entre la eficiencia del método GLS (donde las entradas fuera de la diagonal de  $\mathbf{V}$  no son cero) y la simplicidad del método OLS que produce buenos resultados en la práctica. Podemos observar que  $\frac{N(h_i)}{[\hat{\gamma}(h_i, \theta)]^2}$  es pequeño si  $N(h_i)$  es pequeño o  $\hat{\gamma}(h_i, \theta)$  grande, lo cual implica asignarle menos peso a los valores del semivariograma para datos que estén más alejados. Dado que los pesos dependen de  $\theta$ , el estimador WLS debe obtenerse de forma iterativa, actualizando los pesos en cada iteración hasta que se considere que converge.

### Métodos basados en la máxima verosimilitud

Estos métodos tratan con los datos observados suponiendo que siguen una distribución gaussiana multivariante. Por lo tanto, no se requiere la construcción del semivariograma empírico. Teniendo en cuenta el modelo dado en (1.2.1), supongamos que

$$Z(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + Y(\mathbf{s}), \quad (1.2.18)$$

donde  $\mathbf{X}(\mathbf{s})$  es una matriz de  $n \times q$  con  $q < n$ , cuyos elementos son covariables asociadas a las ubicaciones  $(\mathbf{s}_1, \dots, \mathbf{s}_n)$ ,  $\boldsymbol{\beta}$  es un vector de parámetros desconocidos e  $Y(\mathbf{s})$  un proceso

estocástico gaussiano cuya media es cero y su función de covarianza es

$$\text{Cov}(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}),$$

donde  $\boldsymbol{\theta}$  es un vector de parámetros desconocidos que pertenecen a un espacio de parámetros dado,  $\Theta \in \mathbb{R}^m$ , dentro del cual la función de covarianza es definida positiva. El espacio de parámetros conjunto para  $\boldsymbol{\beta}$  y  $\boldsymbol{\theta}$  se considera simplemente el producto cartesiano de  $\mathbb{R}^p$  y  $\Theta$ .

Sea  $\mathbf{z} \doteq (z_1, \dots, z_n)^T$ , el vector de valores observados en la muestra y  $\mathbf{X}$  una matriz de covariables asociadas al conjunto de ubicaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ . Como el proceso  $Z(\mathbf{s})$  es gaussiano, suponemos que los datos  $\mathbf{z}$  provienen de una distribución normal multivariada, con vector de medias  $\mathbf{X}(\mathbf{s})\boldsymbol{\beta}$  y matriz de covarianza  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  donde el elemento  $ij$ -ésimo es  $C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ . Por lo tanto, la función de verosimilitud queda determinada por

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp\{-(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})/2\}.$$

Los estimadores de máxima verosimilitud  $\hat{\boldsymbol{\beta}}$  y  $\hat{\boldsymbol{\theta}}$  de  $\boldsymbol{\beta}$  y  $\boldsymbol{\theta}$ , respectivamente, se definen como los valores (en  $\mathbb{R}^p$  y  $\Theta$ , respectivamente) que maximizan  $l(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z})$  o su logaritmo,

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

En general, se supone  $\boldsymbol{\theta}_0$  fijo y se halla  $\hat{\boldsymbol{\beta}}$ . Luego se utilizan métodos iterativos para encontrar  $\hat{\boldsymbol{\theta}}$ . Aunque los estimadores de máxima verosimilitud de los parámetros de covarianza espacial que componen  $\boldsymbol{\theta}$  tienen varias propiedades deseables, tienen un defecto, están sesgados como consecuencia de estimar  $\boldsymbol{\beta}$ . Este sesgo puede ser sustancial incluso para muestras de gran tamaño si la correlación espacial es fuerte o  $p$  (la dimensión de  $\boldsymbol{\beta}$ ) es grande. Con lo cual surge (Patterson y Thompson, 1971) el método de **máxima verosimilitud restringida**. El método consiste en eliminar la media de la función de verosimilitud de modo que quede definida solo en términos de la matriz de varianzas y covarianzas. Para esto, una posibilidad es transformar los datos a un conjunto de combinaciones lineales de las observaciones que tenga una distribución que no dependa de la media. Con lo cual, partiendo de (1.2.18),  $\mathbf{W} = \mathbf{K}^T \mathbf{z}$  un vector de  $n - q$  contrastes linealmente independientes, lo cual implica que  $\mathbf{K}$  posee  $n - q$  columnas linealmente independientes y luego  $\mathbf{K}^T \mathbf{X} = \mathbf{0}$ , luego la esperanza  $\mathbb{E}(\mathbf{K}^T \mathbf{z}) = \mathbf{0}$  y

$$\mathbf{W} \sim N(\mathbf{0}, \mathbf{K}^T \boldsymbol{\Sigma} \mathbf{K}).$$

Según Patterson y Thompson (1971) es posible hallar  $\mathbf{K}$  de forma tal que  $\mathbf{K}\mathbf{K}^T = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}$  y  $\mathbf{K}^T \mathbf{K} = \mathbf{I}$  aunque cualquier conjunto de  $n - p$  elementos linealmente independientes de  $\mathbf{W}$  puede servir como contrastes de error necesarios para este método, pero en realidad no importa qué conjunto de contrastes de error se use siempre que sean  $n - p$  e independientes, porque la función de máxima verosimilitud asociada con cualquiera de estos conjuntos difiere como máximo de una constante aditiva (que no

depende de  $\beta$  o  $\theta$ ) de la función

$$\log l(\theta; \mathbf{Z}) = -\frac{1}{2} \log |\Sigma(\theta)| - \frac{1}{2} (\mathbf{X})^T \Sigma^{-1}(\theta) (\mathbf{X}) - \frac{1}{2} \mathbf{Z} \mathbf{P}(\theta) \mathbf{Z},$$

donde  $\mathbf{P}(\theta) = \Sigma^{-1}(\theta) - \Sigma^{-1}(\theta) \mathbf{X} (\mathbf{X}^T \Sigma^{-1}(\theta) \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1}(\theta)$ . Luego el estimador REML de  $\theta$  es el valor  $\hat{\theta} \in \Theta$  que maximiza  $\log l(\theta; \mathbf{Z})$ . Una vez obtenido  $\hat{\theta}$  se halla el estimador  $\hat{\beta}$  a partir de mínimos cuadrados generalizados utilizando  $\hat{\theta}$ .

### 1.3. Análisis exploratorio

Como es bien sabido, el análisis exploratorio de datos es la primera herramienta que nos permite conocer y resumir los datos de la muestra y con ello por ejemplo, descubrir errores en la codificación, determinar casos atípicos, comprobar supuestos (estacionariedad, por ejemplo) e incluso encontrar pistas para la modelización. Nuestro mayor objetivo en el análisis exploratorio de datos será identificar, si existe, una estructura de correlación en los datos. Más precisamente, trataremos de observar el hecho de que valores cercanos sean parecidos y que esa similitud desaparezca a medida que los datos se alejan. Como vimos en (1.1.2), las funciones de semivariograma o covariograma del proceso son la forma de representar esta asociación estimando relaciones espaciales en lo que es llamado análisis estructural (Sección 1.1). Para esto, es necesario ver qué nos dicen los datos sobre la existencia o no de una estructura de correlación, de algún patrón estacionario o de tendencias. En esta sección presentamos algunos gráficos y medidas que podrían ser útiles en esta parte del análisis.

#### 1.3.1. Tipos de gráficos exploratorios

En los últimos años con el desarrollo de software y la posibilidad de manejar grandes cantidades de datos es posible realizar una primera visualización graficando fácilmente los datos sobre mapas o solamente las unidades espaciales ya sean puntos o polígonos. Los **gráficos de puntos** están definidos por sus coordenadas geográficas (latitud y longitud) y pueden ser estaciones meteorológicas, inmuebles, un punto en la tierra para medir diferentes componentes de la misma (Figura 1.6 (a)). Los **polígonos** son áreas delimitadas por líneas, como pueden ser los países, departamentos, estados, radios censales, entre otras (Figura 1.6 (b)).

Para el caso que tengamos datos por áreas o polígonos, el análisis de la distribución espacial se categoriza la variable en función a diferentes intervalos que se calculan a partir de percentiles y luego se pinta cada polígono en función de la categoría a la que pertenece. Si dichos intervalos están particionados a partir de los cuartiles el gráfico se llama **mapa de caja** y es el análogo al gráfico de caja. Este gráfico también se puede utilizar si contamos

con coordenadas como unidades espaciales (gráfico de puntos) y podemos observar cada caso en la Figura 1.6.

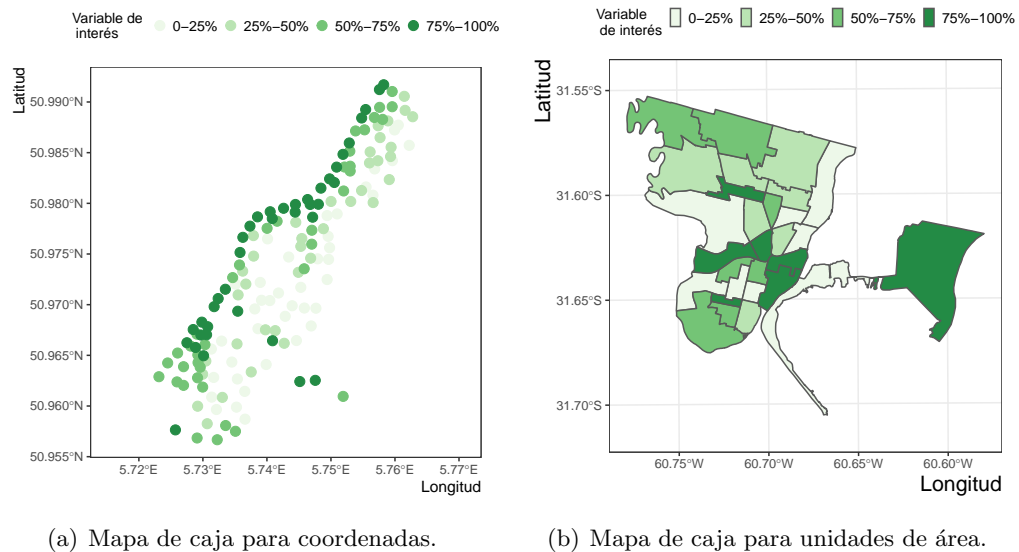
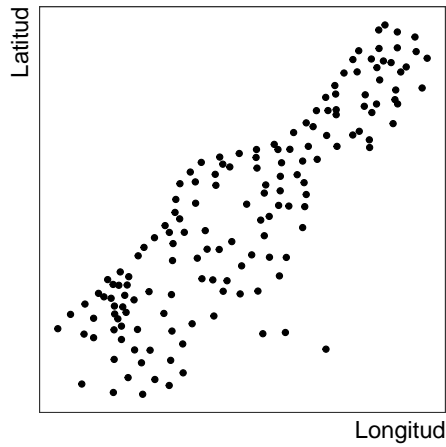


Figura 1.6: Mapas de caja.

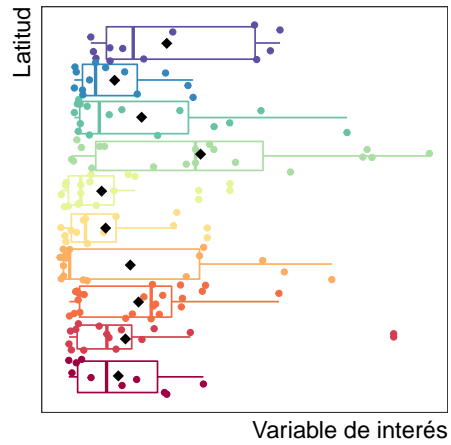
Podemos observar cómo los colores se concentran en zonas, lo cual nos da una pista de la posible correlación espacial existente.

### 1.3.2. Tendencia

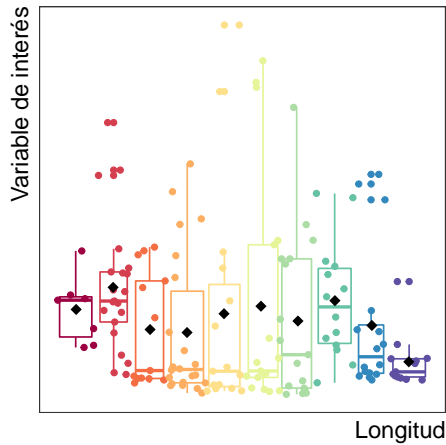
Para estudiar el comportamiento de la variable con respecto a su ubicación, es decir, si existe tendencia, y en caso de contar con los datos dados por coordenadas geográficas podemos graficar la variable de interés con respecto a la longitud y otro gráfico con respecto a la latitud. En particular, en términos de tendencia, nos interesa el comportamiento del promedio de la variable de interés en función de las coordenadas pero solo contamos con un dato por ubicación. Por lo tanto lo que podemos hacer para hallar esa información, en el caso de contar con una grilla regular, es calcular medias y medianas por filas y columnas (oeste-este, norte-sur, respectivamente) y graficarlas en función de las coordenadas. En el caso de no contar con una grilla regular se pueden agrupar los datos para obtener el gráfico. En la Figura 1.7 (b) y (c) se observan los mismos donde los puntos negros representan la media según el agrupamiento de los datos, ya sea por latitud o longitud. En dicha figura, contamos con el conjunto de datos meuse, el cual proporciona concentraciones de zinc en la capa superior del suelo en una llanura aluvial del río Meuse, cerca del pueblo de Stein (Países Bajos). En la Figura 1.7 no observamos tendencia en ninguna dirección, esto nos da la pauta de que la condición que la esperanza sea constante, se cumplirá.



(a) Ubicaciones donde fue medida la variable de interés.



(b) Variación de la variable de interés en dirección norte-sur.



(c) Variación de la variable de interés en dirección este-oeste.

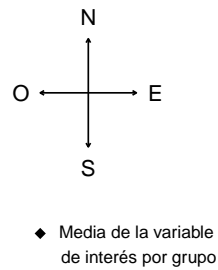


Figura 1.7: Distribución de la variable concentración de zinc de los Datos Meuse en función de las coordenadas (Cressie, 1993, pág. 37).

### 1.3.3. Variabilidad

A partir de los aportes de Waldo Tobler con la primera ley de la geografía que establece que todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas que las lejanas (Siabato y Guzmán-Manrique, 2019), podemos inferir que dos puntos cercanos asumen valores similares porque estos se generaron en condiciones similares. Por el contrario, a grandes distancias las condiciones son diferentes y se esperan mayores variaciones. Esta variabilidad puede visualizarse con **diagramas de dispersión  $h$** . En lugar de describir la relación entre dos variables, un diagrama de dispersión  $h$  describe la relación entre la variable de interés en alguna ubicación  $s$  y la misma variable en una ubicación  $s + hi$  para alguna dirección  $i$ .

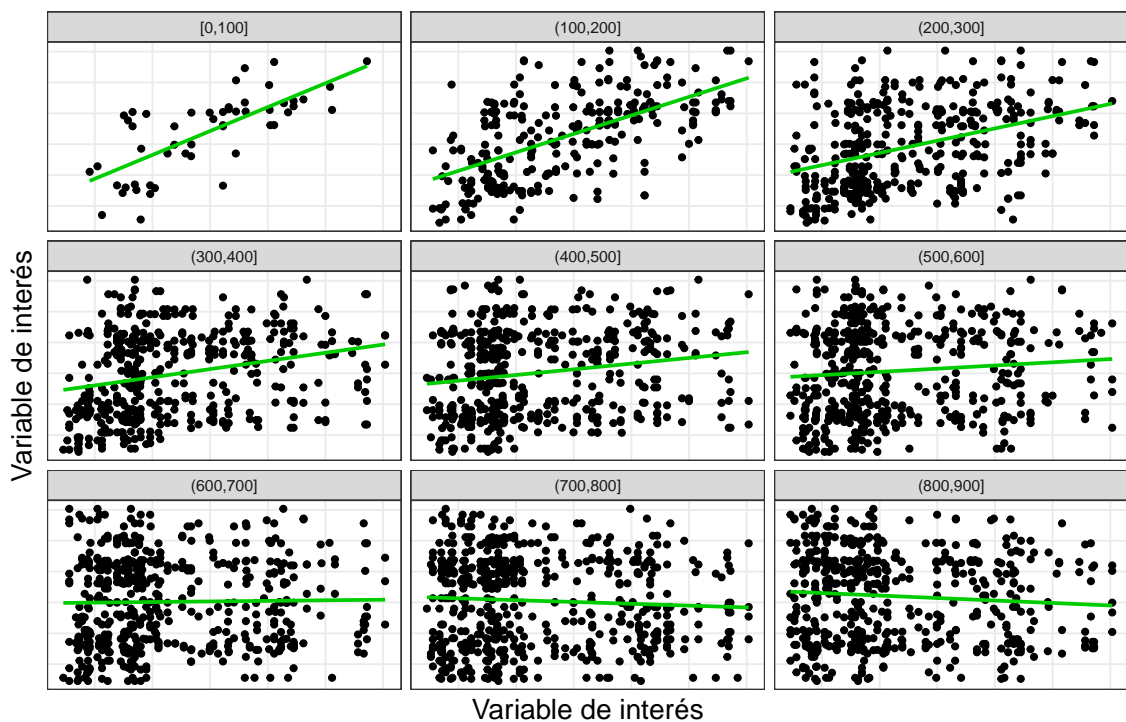


Figura 1.8: Gráfico de dispersión  $h$  para los datos Meuse.

La Figura 1.8 muestra varios paneles donde hemos graficado la variable de interés y su asociación para diferentes rangos de la distancia  $h$ . Como puede observarse, para distancias chicas la asociación es mayor y la misma desaparece a medida que  $h$  aumenta. La recta observada es la recta de regresión cuya pendiente tiende a cero cuando la distancia entre sitios aumenta.

### 1.3.4. Índices de asociación espacial

Además del análisis gráfico que podemos realizar para describir la autocorrelación espacial de la variable de interés, podemos cuantificar la misma a partir de ciertos índices los cuales generalmente se utilizan para medir fuerza de la asociación espacial o describir la forma en la que se distribuyen los fenómenos en el espacio analizado. Los más utilizados son el **Índice  $I$  de Moran** (Moran, 1950),

$$I = \frac{n \sum_i^n \sum_j^n p_{ij} (Z(\mathbf{s}_i) - \bar{Z})(Z(\mathbf{s}_j) - \bar{Z})}{(\sum_{i \neq j} p_{ij}) \sum_i^n (Z(\mathbf{s}_i) - \bar{Z})^2}, \quad (1.3.1)$$

y el **Índice  $C$  de Geary** (Geary, 1954),

$$C = \frac{(n-1) \sum_i^n \sum_j^n p_{ij} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2}{(\sum_{i \neq j} p_{ij}) \sum_i^n (Z(\mathbf{s}_i) - \bar{Z})^2},$$

donde, para todo  $i, j = 1, \dots, n$ ,  $p_{ij}$  denota los **pesos** que se le asignan a la conexiones entre  $i$  y  $j$ . Existen muchas posibilidades para los pesos  $p_{ij}$  (Siabato y Guszmán-Manrique, 2019) las cuales dependen de la unidad espacial. Por ejemplo, si consideramos  $i$  y  $j$  como unidades espaciales dadas por polígonos,  $p_{ij}$  podría ser un peso binario cuyo valor es  $p_{ij} = 1$  si  $i$  y  $j$  comparten algún límite común y  $p_{ij} = 0$  en otro caso. Alternativamente,  $p_{ij}$  podría reflejar distancias entre unidades (por ejemplo entre centroides de un polígono o entre coordenadas) o también podríamos establecer  $p_{ij} = 1$  para todo  $i$  y  $j$  dentro de una distancia especificada  $d$  y  $p_{ij} = 0$  en otro caso. Los pesos en general se acomodan en una matriz  $\mathbf{P}$  de elementos  $\mathbf{P}_{ij} = p_{ij}$  conocida como **matriz de vecindario**. En la misma se resume entonces la información que definen los vecindarios como el peso de la asociación, es decir, se asociará un peso mayor para el dato medido en la unidad espacial  $j$  más cercana (en algún sentido) a la unidad espacial  $i$  que a aquellos medidos en unidades alejadas.

Tanto el Índice de Moran como el Índice de Gearys se pueden decomponer en indicadores locales de asociación espacial (Anselin, 2010). La versión local del índice de Moran (1.3.1) es el **índice de Moran local** dado por

$$I_i = \frac{(Z(\mathbf{s}_i) - \bar{Z}) \sum_{j=1}^k p_{ij} (Z(\mathbf{s}_j) - \bar{Z})}{n^{-1} \sum_{l=1}^n (Z(\mathbf{s}_l) - \bar{Z})^2}. \quad (1.3.2)$$

En el caso de que los pesos estén estandarizados por fila, la suma de los  $I_i$  es igual al índice global  $I$  dado en (1.3.1). Un valor positivo para  $I_i$  indica agrupación espacial de valores similares (ya sea altos o bajos), y valores negativos una agrupación de valores diferentes (por ejemplo, una ubicación con valores altos rodeada de vecinos con valores bajos).

No cualquier expresión puede considerarse como un indicador local de asociación espacial (LISA), Anselin (2010) los define de forma que cumplan los siguientes requisitos:

1. Para cada observación, el LISA proporciona una medida del grado de agrupamiento espacial significativo con relación a los valores que se localizan alrededor de ella;
2. La suma de las medidas LISA de todas las observaciones es proporcional al indicador global de asociación espacial,  $\sum_i L_i = \gamma\Lambda$  donde  $\Lambda$  es el indicador global.

Considerando a

$$L_i = f(Z(\mathbf{s}_i), Z_{J_i}),$$

como la expresión formal del índice LISA para una variable  $Z(\mathbf{s}_i)$  ubicada en un sitio  $\mathbf{s}_i$  donde  $f$  es una función, que posiblemente requiera de la estimación de algún parámetro, y  $z_{J_i}$  son los valores observados en el vecindario  $J_i$  de  $i$ . Los valores de la variable  $Z$  pueden ser las observaciones originales o alguna estandarización para evitar la dependencia de escala, como se observa en el índice  $I_i$  dado en (1.3.2) donde las observaciones se toman como desviaciones de su media. Los sitios considerados en  $J_i$  se definen en términos de los pesos  $p_{ij}$  los cuales cuantifican, de alguna manera, la distancia entre las ubicaciones de la muestra.

Los gráficos y medidas nombrados en esta sección son algunos de los que se pueden utilizar en esta parte del análisis, en Cressie (1993); Banerjee et al. (2004); Chiles y Delfiner (2012) se puede encontrar más información. A partir del análisis estructural y del análisis exploratorio contamos con la información necesaria para continuar con el modelado de los datos.

## 1.4. Conclusiones

En este capítulo se describieron las ideas que conforman un marco teórico sobre el cual nos apoyaremos al analizar datos espaciales, desde el enfoque probabilístico para su construcción hasta su análisis exploratorio, pasando principalmente por la definición y el ajuste de las funciones que modelan la correlación de los datos y los supuestos para ello.



## Capítulo 2

# Predicción espacial paramétrica: kriging

El análisis de predicción espacial fue la herramienta que inicialmente dió lugar al desarrollo de la geoestadística como tal (Cressie, 1990) y se basa en predecir la variable de interés  $Z(\mathbf{s}_0)$  en una ubicación no muestreada  $\mathbf{s}_0 \in D \subseteq \mathbb{R}^d$ . El método clásico utilizado para este propósito es el conocido **kriging**, cuyo nombre se debe al ingeniero de minería sudafricano Daniel Gerhardus Krige, quien fue el primero en desarrollarlo y aplicarlo. El método original de Krige, ahora llamado kriging ordinario, se basó en el caso especial del modelo (1.2.1) en el que se supone que la media es constante. El término kriging fue acuñado originalmente como krigeage por Pierre Carlier, pero Matheron (Matheron, 1963) lo introdujo al idioma inglés en reconocimiento a la contribución de Krige para mejorar la precisión en la predicción de las concentraciones de oro y otros metales en los cuerpos minerales.

En este capítulo desarrollaremos la base teórica del kriging considerando el modelo subyacente de los datos, los supuestos necesarios y la forma del predictor para cada variante del método (Wackernagel, 2006; Montero et al., 2015; Webster y Oliver, 2007; Cressie, 1993; Gelfand et al., 2010).

### 2.1. Kriging

Kriging es un método de predicción espacial cuyo predictor se caracteriza por ser un promedio pesado de observaciones muestreadas sobre diferentes unidades espaciales. Formalmente, para un sitio  $\mathbf{s}_0 \in D \subseteq \mathbb{R}^d$  donde la variable de interés no fue observada, el

**predicador de kirging** está dado por

$$Z_{\omega}(\mathbf{s}_0) = \sum_{i=1}^n \omega(\mathbf{s}_i, \mathbf{s}_0) Z(\mathbf{s}_i), \quad (2.1.1)$$

donde  $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_i), \dots, Z(\mathbf{s}_n)\}$  son las variables de interés en el conjunto de sitios muestreados  $\{\mathbf{s}_1, \dots, \mathbf{s}_i, \dots, \mathbf{s}_n\}$  y obtenemos los pesos  $\omega(\mathbf{s}_i, \mathbf{s}_0)$  de manera tal que el predictor sea insesgado y de mínima varianza. Por simplicidad llamaremos  $\omega_i \doteq \omega(\mathbf{s}_i, \mathbf{s}_0)$ . Definiendo el vector de pesos  $\boldsymbol{\omega} \doteq (\omega_1, \dots, \omega_n)^T \in \mathbb{R}^n$  y el vector de variables  $\mathbf{Z} \doteq (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_i), \dots, Z(\mathbf{s}_n)) \in \mathbb{R}^n$ , podemos escribir a la ecuación (2.1.1) de manera matricial como

$$Z_{\omega}(\mathbf{s}_0) = \boldsymbol{\omega}^T \mathbf{Z}. \quad (2.1.2)$$

Una de las principales ventajas (ver Montero et al. (2015)) con las que cuenta kriging radica en que considera la estructura de la correlación espacial de los datos. Esto es, para poder aplicar este método se necesita modelar la estructura de correlación del fenómeno a estudiar, para lo cual es necesaria la estimación del semivariograma. Si bien esto significa que no se tiene en cuenta solo las características geométricas, el número y organización de los sitios, no quiere decir que no sean importantes. Como en cualquier modelo estadístico, la calidad de las predicciones depende de la calidad de la información disponible, en este caso la distribución uniforme de los sitios muestrados, su relación con el fenómeno estudiado y el conocimiento que tengamos del mismo.

Como mencionamos anteriormente, para hallar los pesos  $\omega_i$  le imponemos al predictor que sea insesgado y de mínima varianza, para lo cual es necesario contar con un marco que defina el proceso subyacente de los datos para poder explicitar los supuestos sobre el mismo, necesarios para la predicción. Para ello utilizamos el modelo definido en la ecuación (1.2.1).

En los siguientes apartados revisamos y desarrollamos diferentes conceptos relacionados con los tipos de kriging más utilizados y nombrados en la literatura (Wackernagel, 2006; Montero et al., 2015; Webster y Oliver, 2007): kriging simple, ordinario y universal. En cada caso se listan los supuestos sobre el modelo, los cuales determinan el tipo de kriging; se define el predictor y se hallan las ecuaciones de kriging alcanzadas a partir de las condiciones de optimalidad pedidas.

## 2.2. Kriging simple

Es la forma más simple de kriging en el cual se asume que el proceso  $Z(\mathbf{s})$  se modela a partir de la Ecuación (1.2.1) considerando

- (I)  $\mu(\mathbf{s}) = \mu, \forall \mathbf{s} \in D, \mu$  conocida.

(II)  $Y(\mathbf{s})$  es un proceso estocástico estacionario de segundo orden.

A partir de estos supuestos, y utilizando el Lema (1.2.1), podemos asegurar que el proceso  $Z(\mathbf{s})$  es también estacionario de segundo orden con media  $\mu$  y función de covarianza dada por

$$C(\mathbf{h}) = \text{Cov}(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) = \mathbb{E}[Z(\mathbf{s})Z(\mathbf{s} + \mathbf{h})] - \mu^2, \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D.$$

Como en este caso la media del proceso es conocida, consideramos el proceso  $X(\mathbf{s}) \doteq Z(\mathbf{s}) - \mu$  para  $\mathbf{s} \in D$ , con  $\mathbb{E}(X(\mathbf{s})) = 0$  el cual da lugar al predictor

$$\begin{aligned} X_\omega(\mathbf{s}_0) &= \sum_{i=1}^n \omega_i X(\mathbf{s}_i) \\ Z_\omega(\mathbf{s}_0) - \mu &= \sum_{i=1}^n \omega_i (Z(\mathbf{s}_i) - \mu), \end{aligned}$$

de donde, despejando  $Z_\omega(\mathbf{s}_0)$  tenemos que,

$$\begin{aligned} Z_\omega(\mathbf{s}_0) &= \mu + \sum_{i=1}^n \omega_i (Z(\mathbf{s}_i) - \mu) \\ &= \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) + \mu \left( 1 - \sum_{i=1}^n \omega_i \right), \end{aligned} \tag{2.2.1}$$

donde el factor  $k \doteq 1 - \sum_{i=1}^n \omega_i$  es conocido como el **peso de la media** en kriging simple. Con el objetivo de hallar los pesos  $\omega_i$ , a partir de las condiciones de insesgadez y mínima varianza, escribimos la ecuación (2.2.1) de forma matricial como,

$$Z_\omega(\mathbf{s}_0) = \mu + \boldsymbol{\omega}^T (\mathbf{Z} - \mu \mathbf{1}), \tag{2.2.2}$$

donde  $\mathbf{1}$  es un vector de unos en  $\mathbb{R}^n$ .

### 2.2.1. Insesgadez del predictor

De (2.2.2) se tiene que,

$$\begin{aligned} \mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] &= \mathbb{E}[\mu + \boldsymbol{\omega}^T (\mathbf{Z} - \mu \mathbf{1}) - Z(\mathbf{s}_0)] \\ &= \mu + \boldsymbol{\omega}^T (\mathbb{E}[\mathbf{Z}] - \mu \mathbf{1}) - \mathbb{E}[Z(\mathbf{s}_0)] \\ &= \mu + 0 - \mu && \text{de (I)} \\ &= 0, \end{aligned} \tag{2.2.3}$$

por lo cual no necesitamos condiciones sobre los pesos  $\boldsymbol{\omega}$  para que el estimador de kriging simple sea insesgado.

### 2.2.2. Mínima varianza del error de predicción

En esta sección encontramos la forma de la varianza del error de predicción y mostramos las condiciones sobre los pesos  $\boldsymbol{\omega}$  para que sea mínima. Luego,

$$\begin{aligned}
\sigma_E^2(\boldsymbol{\omega}) &\doteq \text{Var}(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)) \\
&= \mathbb{E}[(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] - (\mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)])^2 \\
&= \mathbb{E}[(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] && \text{(de (2.2.3))} \\
&= \mathbb{E}\left[\left(\sum_{i=1}^n \omega_i (Z(\mathbf{s}_i) - \mu) + (\mu - Z(\mathbf{s}_0))\right)^2\right] \\
&= \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \mathbb{E}[(Z(\mathbf{s}_i) - \mu)(Z(\mathbf{s}_j) - \mu)] \\
&\quad - 2 \sum_{i=1}^n \omega_i \mathbb{E}[(Z(\mathbf{s}_i) - \mu)(Z(\mathbf{s}_0) - \mu)] + \mathbb{E}[(Z(\mathbf{s}_0) - \mu)^2] \\
&= \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \\
&\quad - 2 \sum_{i=1}^n \omega_i \text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_0)) + \text{Cov}(Z(\mathbf{s}_0), Z(\mathbf{s}_0)) && \text{(de (1.1.3))} \\
&= \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j C(\mathbf{s}_i - \mathbf{s}_j) - 2 \sum_{i=1}^n \omega_i C(\mathbf{s}_i - \mathbf{s}_0) + C(\mathbf{0}) && \text{(de (1.1.5))} \\
&\doteq \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} - 2\boldsymbol{\omega}^T \mathbf{c}_0 + C(\mathbf{0}), && \text{(2.2.4)}
\end{aligned}$$

donde  $\mathbf{c}_0^T = (C(\mathbf{s}_1 - \mathbf{s}_0), \dots, C(\mathbf{s}_n - \mathbf{s}_0))$  y  $\boldsymbol{\Sigma}$  es la matriz de elementos  $\Sigma_{i,j} = C(\mathbf{s}_i - \mathbf{s}_j)$ . Observemos que, si definimos  $\mathbf{x}^T \doteq (-1, \boldsymbol{\omega}^T)$  la matriz de covarianza del vector  $(Z(\mathbf{s}_0), Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$  está dada por

$$\boldsymbol{\Sigma}_0 \doteq \left( \begin{array}{c|c} C(\mathbf{0}) & \mathbf{c}_0^T \\ \hline \mathbf{c}_0 & \boldsymbol{\Sigma} \end{array} \right),$$

la cual es simétrica y semidefinida positiva. Luego resulta que

$$\sigma_E^2(\boldsymbol{\omega}) = \mathbf{x}^T \boldsymbol{\Sigma}_0 \mathbf{x} \geq 0,$$

ya que

$$\begin{aligned}
\mathbf{x}^T \boldsymbol{\Sigma}_0 \mathbf{x} &= (-C(\mathbf{0}) + \boldsymbol{\omega}^T \mathbf{c}_0, -\mathbf{c}_0^T + \boldsymbol{\omega}^T \boldsymbol{\Sigma}) (-1, \boldsymbol{\omega}^T)^T \\
&= C(\mathbf{0}) - \boldsymbol{\omega}^T \mathbf{c}_0 + (-\mathbf{c}_0^T + \boldsymbol{\omega}^T \boldsymbol{\Sigma}) \boldsymbol{\omega} \\
&= C(\mathbf{0}) - \boldsymbol{\omega}^T \mathbf{c}_0 - \mathbf{c}_0^T \boldsymbol{\omega} + \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} \\
&= -2\boldsymbol{\omega}^T \mathbf{c}_0 + \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega} + C(\mathbf{0}).
\end{aligned}$$

Ahora calculamos los pesos  $\boldsymbol{\omega} \in \mathbb{R}^n$  que minimizan  $\sigma_E^2(\boldsymbol{\omega})$ . Para ello hallamos la derivada primera de  $\sigma_E^2(\boldsymbol{\omega})$  con respecto a  $\boldsymbol{\omega}$ ,

$$\frac{\partial \sigma_E^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} = 2\boldsymbol{\Sigma}\boldsymbol{\omega} - 2\mathbf{c}_0,$$

e igualando a cero tenemos que

$$\boldsymbol{\Sigma}\boldsymbol{\omega} = \mathbf{c}_0. \quad (2.2.5)$$

La segunda derivada de  $\sigma_E^2(\boldsymbol{\omega})$  con respecto a  $\boldsymbol{\omega}$  es

$$\frac{\partial^2 \sigma_E^2(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}^2} = 2\boldsymbol{\Sigma},$$

y a partir de que  $\boldsymbol{\Sigma}$  es definida positiva, el peso hallado en (2.2.5) minimiza  $\sigma_E^2(\boldsymbol{\omega})$  y está dado por

$$\boldsymbol{\omega}_{SK} \doteq \boldsymbol{\Sigma}^{-1}\mathbf{c}_0. \quad (2.2.6)$$

Reemplazando (2.2.6) en (2.2.4) tenemos que la varianza del error de predicción está dada por

$$\begin{aligned} \sigma_{SK}^2 &= C(\mathbf{0}) + \boldsymbol{\omega}_{SK}^T \boldsymbol{\Sigma} \boldsymbol{\omega}_{SK} - 2\boldsymbol{\omega}_{SK}^T \mathbf{c}_0 \\ &= C(\mathbf{0}) + \mathbf{c}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \mathbf{c}_0 - 2\mathbf{c}_0^T \boldsymbol{\Sigma}^{-1} \mathbf{c}_0 \\ &= C(\mathbf{0}) - \mathbf{c}_0^T \boldsymbol{\Sigma}^{-1} \mathbf{c}_0. \end{aligned}$$

### 2.2.3. Forma final del predictor

Reemplazando los pesos  $\boldsymbol{\omega}_{SK}$  calculados en (2.2.6) en (2.2.2) tenemos que el predictor de kriging simple está dado por

$$\begin{aligned} Z_\omega(\mathbf{s}_0) &= \mu + \boldsymbol{\omega}_{SK}^T (\mathbf{Z} - \mu \mathbf{1}) \\ &= \mu + \mathbf{c}_0^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mu \mathbf{1}). \end{aligned}$$

Con esto obtenemos la predicción kriging que depende del sitio en el que queremos predecir  $\mathbf{s}_0$  a través del vector  $\mathbf{c}_0$ . Cabe aclarar que la predicción también depende de los sitios muestreados y el modelado de la covarianza.

## 2.3. Kriging ordinario

Sea  $Z(\mathbf{s})$  un proceso estocástico espacial cuya forma es la dada en (1.2.1). Supondremos que:

- (I)  $\mu(\mathbf{s}) = \mu, \forall \mathbf{s} \in D$ , esto es,  $\mu$  constante en  $D$  y desconocida.
- (II)  $Y(s)$  es un proceso estocástico intrínsecamente estacionario.

A partir de estos supuestos, y utilizando el Lema (1.2.2), podemos asegurar que el proceso  $Z(\mathbf{s})$  es también intrínsecamente estacionario, en este caso, con media  $\mu$  y semivariograma dado por

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E}[(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2], \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D.$$

De (2.1.2), tenemos que el predictor en este caso es

$$Z_\omega(\mathbf{s}_0) = \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) = \boldsymbol{\omega}^T \mathbf{Z}, \quad (2.3.1)$$

donde los pesos  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T \in \mathbb{R}^n$  serán elegidos de manera tal que el predictor sea insesgado y de mínima varianza.

### 2.3.1. Inssegadez del predictor

Para probar que el predictor de kriging ordinario es insesgado tenemos que

$$\begin{aligned} \mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] &= \mathbb{E}[\boldsymbol{\omega}^T \mathbf{Z} - Z(\mathbf{s}_0)] \\ &= \boldsymbol{\omega}^T \mathbb{E}[\mathbf{Z}] - \mu \\ &= \boldsymbol{\omega}^T \mu \mathbf{1} - \mu \\ &= \mu(\boldsymbol{\omega}^T \mathbf{1} - 1). \end{aligned}$$

Luego

$$\mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] = \mu(\boldsymbol{\omega}^T \mathbf{1} - 1) = 0 \Leftrightarrow \boldsymbol{\omega}^T \mathbf{1} = 1. \quad (2.3.2)$$

Por lo tanto, la condición de que la suma de los pesos  $\omega_i$  sea 1, garantiza que el predictor sea insesgado.

### 2.3.2. Mínima varianza del error de predicción

En esta sección encontramos la forma de la varianza del error de predicción y mostramos las condiciones sobre los pesos  $\boldsymbol{\omega}$  para que sea mínima. Esto es,

$$\begin{aligned} \sigma_E^2(\boldsymbol{\omega}) &\doteq \text{Var}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] \\ &= \mathbb{E}[(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] - (\mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)])^2 \\ &= \mathbb{E}[(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] \quad (\text{de (2.3.2)}) \\ &= \mathbb{E} \left[ \left( \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0) \right)^2 \right] \\ &= \mathbb{E} \left[ \left( \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) \right)^2 - 2Z(\mathbf{s}_0) \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) + Z(\mathbf{s}_0)^2 \right] \end{aligned}$$

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

Dado que,

$$-\frac{1}{2} \sum_{i=1}^n \omega_i Z(\mathbf{s}_i)^2 - \frac{1}{2} \sum_{j=1}^n \omega_j Z(\mathbf{s}_j)^2 + \sum_{i=1}^n \omega_i Z(\mathbf{s}_i)^2 = 0,$$

y

$$\sum_{i=1}^n \omega_i = \sum_{j=1}^n \omega_j = 1, \quad (\text{de (2.3.2)})$$

tenemos que,

$$\begin{aligned}
\sigma_E^2(\boldsymbol{\omega}) &= - \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \frac{1}{2} \mathbb{E}[Z(\mathbf{s}_i)^2] + \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \mathbb{E}[Z(\mathbf{s}_i)Z(\mathbf{s}_j)] - \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \frac{1}{2} \mathbb{E}[Z(\mathbf{s}_j)^2] \\
&\quad + \sum_{i=1}^n \omega_i \mathbb{E}[Z(\mathbf{s}_i)^2] - 2 \sum_{i=1}^n \omega_i \mathbb{E}[Z(\mathbf{s}_i)Z(\mathbf{s}_0)] + \sum_{i=1}^n \omega_i \mathbb{E}[Z(\mathbf{s}_0)^2] \\
&= - \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \frac{1}{2} \mathbb{E}[(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2] + 2 \sum_{i=1}^n \omega_i \frac{1}{2} \mathbb{E}[(Z(\mathbf{s}_i) - Z(\mathbf{s}_0))^2].
\end{aligned}$$

Luego, a partir de la definición de variograma para procesos intrínsecamente estacionarios (1.1.14) tenemos que

$$\begin{aligned}
\sigma_E^2(\boldsymbol{\omega}) &= - \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2 \sum_{i=1}^n \omega_i \gamma(\mathbf{s}_i - \mathbf{s}_0) \\
&\doteq -\boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega} + 2\boldsymbol{\omega}^T \boldsymbol{\gamma}_0,
\end{aligned} \tag{2.3.3}$$

donde  $\boldsymbol{\gamma}_0 = (\gamma(\mathbf{s}_1 - \mathbf{s}_0), \dots, \gamma(\mathbf{s}_n - \mathbf{s}_0))^T \in \mathbb{R}^n$  y  $\boldsymbol{\Gamma}$  es la matriz de elementos  $\boldsymbol{\Gamma}_{i,j} = \gamma(\mathbf{s}_i - \mathbf{s}_j)$ .

Para asegurar que la varianza del error de predicción sea positiva, definimos  $\mathbf{x}^T \doteq (-1, \boldsymbol{\omega}^T)$  de manera tal que  $\mathbf{x}^T \mathbf{1} = 0$  ya que  $\boldsymbol{\omega}^T \mathbf{1} = 1$  por (2.3.2). También definimos

$$\boldsymbol{\Gamma}_0 \doteq \left( \begin{array}{c|c} 0 & \boldsymbol{\gamma}_0^T \\ \hline \boldsymbol{\gamma}_0 & \boldsymbol{\Gamma} \end{array} \right),$$

la matriz de semivariogramas del vector  $(Z(\mathbf{s}_0), Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))$ . La Propiedad IV) del semivariograma, esto es, considerar un modelo de variograma que sea condicionalmente definido negativo, nos asegura que

$$\mathbf{x}^T \boldsymbol{\Gamma}_0 \mathbf{x} \leq 0,$$

de donde se sigue que,

$$\begin{aligned}
0 &\geq \mathbf{x}^T \mathbf{\Gamma}_0 \mathbf{x} \\
&= (\boldsymbol{\omega}^T \boldsymbol{\gamma}_0, -\boldsymbol{\gamma}_0^T + \boldsymbol{\omega}^T \mathbf{\Gamma}) (-1, \boldsymbol{\omega}^T)^T \\
&= -\boldsymbol{\omega}^T \boldsymbol{\gamma}_0 + (-\boldsymbol{\gamma}_0^T + \boldsymbol{\omega}^T \mathbf{\Gamma}) \boldsymbol{\omega} \\
&= -\boldsymbol{\omega}^T \boldsymbol{\gamma}_0 - \boldsymbol{\gamma}_0^T \boldsymbol{\omega} + \boldsymbol{\omega}^T \mathbf{\Gamma} \boldsymbol{\omega} \\
&= -2\boldsymbol{\omega}^T \boldsymbol{\gamma}_0 + \boldsymbol{\omega}^T \mathbf{\Gamma} \boldsymbol{\omega},
\end{aligned}$$

y como consecuencia de (2.3.3) tenemos que,

$$\sigma_E^2(\boldsymbol{\omega}) = -\boldsymbol{\omega}^T \mathbf{\Gamma} \boldsymbol{\omega} + 2\boldsymbol{\omega}^T \boldsymbol{\gamma}_0 = -\mathbf{x}^T \mathbf{\Gamma}_0 \mathbf{x} \geq 0.$$

Ahora calculamos los pesos  $\boldsymbol{\omega} \in \mathbb{R}^n$ , que minimizan  $\sigma_E^2(\boldsymbol{\omega})$ , utilizando para ello multiplicadores de Lagrange con la condición sobre los pesos  $\boldsymbol{\omega}^T \mathbf{1} = 1$ . Por lo tanto, definimos una función  $\varphi$  como sigue,

$$\begin{aligned}
\varphi : \mathbb{R}^n \times \mathbb{R} &\rightarrow \mathbb{R} \\
(\boldsymbol{\omega}, \lambda) &\mapsto \varphi(\boldsymbol{\omega}, \lambda) \doteq -\boldsymbol{\omega}^T \mathbf{\Gamma} \boldsymbol{\omega} + 2\boldsymbol{\omega}^T \boldsymbol{\gamma}_0 - 2\lambda(\boldsymbol{\omega}^T \mathbf{1} - 1),
\end{aligned}$$

donde  $\lambda \in \mathbb{R}$  representa el multiplicador de Lagrange. En primer lugar, hallamos la derivada de  $\varphi$  con respecto a  $\boldsymbol{\omega}$ ,

$$\frac{\partial \varphi(\boldsymbol{\omega}, \lambda)}{\partial \boldsymbol{\omega}} = -2\mathbf{\Gamma} \boldsymbol{\omega} + 2\boldsymbol{\gamma}_0 - 2\lambda \mathbf{1},$$

igualamos a cero y obtenemos la siguiente condición

$$\mathbf{\Gamma} \boldsymbol{\omega} + \lambda \mathbf{1} = \boldsymbol{\gamma}_0. \quad (2.3.4)$$

En segundo lugar derivamos  $\varphi$  con respecto a  $\lambda$ ,

$$\frac{\partial \varphi(\boldsymbol{\omega}, \lambda)}{\partial \lambda} = -2(\boldsymbol{\omega}^T \mathbf{1} - 1),$$

de la misma manera que en (2.3.4), igualamos a cero y obtenemos la condición

$$\boldsymbol{\omega}^T \mathbf{1} = 1. \quad (2.3.5)$$

Las ecuaciones (2.3.4) y (2.3.5) sobre los pesos  $\omega_i$  son las llamadas ecuaciones de kriging ordinario para un proceso intrínsecamente estacionario con esperanza desconocida. A partir de dichas ecuaciones calculamos los pesos y el valor del multiplicador del Lagrange (Cressie, 1993). Primero, de (2.3.4) despejamos los pesos  $\boldsymbol{\omega}$ ,

$$\boldsymbol{\omega} = \mathbf{\Gamma}^{-1}(\boldsymbol{\gamma}_0 - \lambda \mathbf{1}), \quad (2.3.6)$$

que luego reemplazamos en (2.3.5) y transponemos para obtener  $\lambda$ ,

$$1 = \mathbf{1}^T \boldsymbol{\omega} = \mathbf{1}^T \mathbf{\Gamma}^{-1}(\boldsymbol{\gamma}_0 - \lambda \mathbf{1}) = \mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \mathbf{1}^T \mathbf{\Gamma}^{-1} \lambda \mathbf{1},$$

o equivalentemente,

$$\lambda \mathbf{1}^T \mathbf{\Gamma}^{-1} \mathbf{1} = \mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1,$$

de donde se sigue que,

$$\lambda_{OK} \doteq \frac{\mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1}{\mathbf{1}^T \mathbf{\Gamma}^{-1} \mathbf{1}}. \quad (2.3.7)$$

Finalmente, reemplazando el multiplicador en (2.3.6) tenemos que

$$\boldsymbol{\omega}_{OK} \doteq \mathbf{\Gamma}^{-1} \left( \boldsymbol{\gamma}_0 - \frac{\mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1}{\mathbf{1}^T \mathbf{\Gamma}^{-1} \mathbf{1}} \mathbf{1} \right). \quad (2.3.8)$$

Para ver la forma de la varianza del error de predicción del kriging ordinario, reescribimos (2.3.3) para obtener

$$\begin{aligned} \sigma_E^2(\boldsymbol{\omega}) &= -\boldsymbol{\omega}^T \mathbf{\Gamma} \boldsymbol{\omega} + \boldsymbol{\omega}^T \boldsymbol{\gamma}_0 + \boldsymbol{\omega}^T \boldsymbol{\gamma}_0 & (2.3.9) \\ &= -\boldsymbol{\omega}^T (\mathbf{\Gamma} \boldsymbol{\omega} - \boldsymbol{\gamma}_0) + \boldsymbol{\omega}^T \boldsymbol{\gamma}_0 \\ &= \lambda \boldsymbol{\omega}^T \mathbf{1} + \boldsymbol{\omega}^T \boldsymbol{\gamma}_0 & (\text{de (2.3.4), } \mathbf{\Gamma} \boldsymbol{\omega} - \boldsymbol{\gamma}_0 = -\lambda \mathbf{1}) \\ &= \lambda + \boldsymbol{\omega}^T \boldsymbol{\gamma}_0. & (\text{de (2.3.5)}) \end{aligned}$$

Ahora, reemplazamos (2.3.6) en (2.3.9) y de esta manera la varianza queda escrita en función de  $\lambda$  como sigue,

$$\sigma_E^2(\lambda) = \lambda + (\boldsymbol{\gamma}_0 - \lambda \mathbf{1})^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0. \quad (2.3.10)$$

Entonces, reemplazando  $\lambda_{OK}$  de (2.3.7) en (2.3.10) tenemos que,

$$\begin{aligned} \sigma_{OK}^2 &= \lambda_{OK} + (\boldsymbol{\gamma}_0 - \lambda_{OK} \mathbf{1})^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 \\ &= \lambda_{OK} + \boldsymbol{\gamma}_0^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \lambda_{OK} \mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 \\ &= \boldsymbol{\gamma}_0^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \lambda_{OK} (\mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1) \\ &= \boldsymbol{\gamma}_0^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \frac{(\mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1)^2}{\mathbf{1}^T \mathbf{\Gamma}^{-1} \mathbf{1}}. \end{aligned}$$

### 2.3.3. Forma final del predictor

Reemplazando en (2.3.1) los pesos  $\boldsymbol{\omega}_{OK}$  calculados en (2.3.8) tenemos que el predictor de kriging ordinario está dado por

$$Z_\omega(\mathbf{s}_0) = \left( \boldsymbol{\gamma}_0 - \frac{\mathbf{1}^T \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1}{\mathbf{1}^T \mathbf{\Gamma}^{-1} \mathbf{1}} \mathbf{1} \right)^T \mathbf{\Gamma}^{-1} \mathbf{Z}. \quad (2.3.11)$$

En las dos ecuaciones anteriores podemos observar que tanto el predictor  $Z_\omega(\mathbf{s}_0)$  como la varianza del error de kriging ordinario  $\sigma_{OK}^2$  dependen de los valores del semivariograma ( $\mathbf{\Gamma}$  y  $\boldsymbol{\gamma}_0$ ). En el caso de considerar que el proceso  $Y(\mathbf{s})$  es también estacionario de segundo orden, los resultados son similares teniendo en cuenta que (de (1.1.7))  $C(\mathbf{h}) = \gamma(\mathbf{h}) - C(\mathbf{0})$ .

## 2.4. Kriging universal

Este caso surge de suponer que la media en (1.2.1) es **desconocida** y **dependiente** del sitio  $\mathbf{s}$ . Sea  $Z(\mathbf{s})$  un proceso estocástico espacial cuya forma es la dada en (1.2.1), suponemos que:

(I)  $\mathbb{E}[Z(\mathbf{s})] = \mu(\mathbf{s}), \forall \mathbf{s} \in D$ , desconocida.

(II)  $Y(\mathbf{s})$  es un proceso estocástico intrínsecamente estacionario (recordemos que  $\mathbb{E}(Y(\mathbf{s})) = 0, \forall \mathbf{s}$ ) con semivariograma dado por

$$\gamma_Y(\mathbf{h}) = \frac{1}{2} \text{Var}(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) = \frac{1}{2} \mathbb{E}[(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s}))^2], \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D.$$

(III) Sean  $f_0, f_1, \dots, f_P$  funciones determinísticas que dependen de las coordenadas geográficas  $\mathbf{s} \in D$  y sea  $P \in \mathbb{N}$  la cantidad de funciones  $f_p : D \mapsto \mathbb{R}, p = 0, \dots, P$ . Asumiremos que  $\mu(\mathbf{s})$ , comúnmente llamado **drift**, es una combinación lineal de dichas funciones, esto es,

$$\mu(\mathbf{s}) = \sum_{p=0}^P a_p f_p(\mathbf{s}), \quad (2.4.1)$$

donde los pesos  $a_p$  son números reales no nulos y  $f_0(\mathbf{s}) \doteq 1$ .

Observemos que kriging ordinario es un caso particular de kriging universal tomando  $P = 0$  y  $\mu(\mathbf{s}) = a_0$  constante y desconocida.

Para escribir el modelo (1.2.1) de forma matricial, definimos

$$\mathbf{a} \doteq (a_0, \dots, a_P)^T \in \mathbb{R}^{P+1}$$

y

$$\mathbf{F} \doteq \begin{pmatrix} 1 & f_1(\mathbf{s}_1) & \dots & f_P(\mathbf{s}_1) \\ 1 & f_1(\mathbf{s}_2) & \dots & f_P(\mathbf{s}_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_1(\mathbf{s}_n) & \dots & f_P(\mathbf{s}_n) \end{pmatrix} \in \mathbb{R}^{n \times (P+1)}.$$

A partir de esta notación y de los supuestos anteriormente establecidos, tenemos que

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + Y(\mathbf{s}_i) = \sum_{p=0}^P a_p f_p(\mathbf{s}_i) + Y(\mathbf{s}_i) = (\mathbf{F}\mathbf{a} + \mathbf{Y})_i, \quad i = 1, \dots, n, \quad (2.4.2)$$

donde  $\mathbf{Y} \doteq (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$ . Reemplazando (2.4.2) en (2.1.2) tenemos que el predictor queda expresado como

$$Z_\omega(\mathbf{s}_0) = \boldsymbol{\omega}^T \mathbf{Z} = \boldsymbol{\omega}^T (\mathbf{F}\mathbf{a} + \mathbf{Y}), \quad (2.4.3)$$

donde los pesos  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^T \in \mathbb{R}^n$  en este caso quedarán establecidos a partir de imponer al predictor que sea insesgado y de mínima varianza, lo cual realizamos en las secciones siguientes.

### 2.4.1. Inssegadez del predictor

En esta sección daremos las condiciones sobre los pesos para que el predictor de kriging universal sea insesgado. Para ello escribimos,

$$\begin{aligned}
\mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] &= \mathbb{E} \left[ \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0) \right] \\
&= \mathbb{E} \left[ \sum_{i=1}^n \omega_i (\mu(\mathbf{s}_i) + Y(\mathbf{s}_i)) - (\mu(\mathbf{s}_0) + Y(\mathbf{s}_0)) \right] \quad (\text{de (2.4.2)}) \\
&= \sum_{i=1}^n \omega_i (\mu(\mathbf{s}_i) + \mathbb{E}[Y(\mathbf{s}_i)]) - \mu(\mathbf{s}_0) - \mathbb{E}[Y(\mathbf{s}_0)] \\
&= \sum_{i=1}^n \omega_i \mu(\mathbf{s}_i) - \mu(\mathbf{s}_0) \quad (\text{de (II)}) \quad (2.4.4) \\
&= \sum_{i=1}^n \omega_i \left( \sum_{p=0}^P a_p f_p(\mathbf{s}_i) \right) - \left( \sum_{p=0}^P a_p f_p(\mathbf{s}_0) \right) \quad (\text{de (2.4.1)}) \\
&= \sum_{p=0}^P a_p \left( \sum_{i=1}^n \omega_i f_p(\mathbf{s}_i) - f_p(\mathbf{s}_0) \right) \\
&\doteq \mathbf{a}^T (\mathbf{F}^T \boldsymbol{\omega} - \mathbf{f}_0),
\end{aligned}$$

donde  $\mathbf{f}_0 \doteq (1, f_1(\mathbf{s}_0), \dots, f_P(\mathbf{s}_0))^T \in \mathbb{R}^{P+1}$ . Luego,

$$\mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] = \mathbf{a}^T (\mathbf{F}^T \boldsymbol{\omega} - \mathbf{f}_0) = 0 \Leftrightarrow \mathbf{F}^T \boldsymbol{\omega} = \mathbf{f}_0. \quad (2.4.5)$$

Observemos que de la ecuación (2.4.4) también tenemos que

$$\sum_{i=1}^n \omega_i \mu(\mathbf{s}_i) - \mu(\mathbf{s}_0) = 0, \quad (2.4.6)$$

lo cual será utilizado más adelante.

### 2.4.2. Mínima varianza del error de predicción

En esta sección mostramos la forma de la varianza del error de predicción y las condiciones para que sea mínima. A saber,

$$\begin{aligned}
\sigma_E^2(\boldsymbol{\omega}) &\doteq \text{Var}(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)) \\
&= \mathbb{E} [(Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] - (\mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)])^2
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[ (Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0))^2 \right] && \text{(de (2.4.5))} \\
&= \mathbb{E} \left[ \left( \sum_{i=1}^n \omega_i (\mu(\mathbf{s}_i) + Y(\mathbf{s}_i)) - (\mu(\mathbf{s}_0) + Y(\mathbf{s}_0)) \right)^2 \right] && \text{(de (2.4.2))} \\
&= \mathbb{E} \left[ \left( \sum_{i=1}^n \omega_i \mu(\mathbf{s}_i) - \mu(\mathbf{s}_0) + \sum_{i=1}^n \omega_i Y(\mathbf{s}_i) - Y(\mathbf{s}_0) \right)^2 \right] \\
&= \mathbb{E} \left[ \left( \sum_{i=1}^n \omega_i Y(\mathbf{s}_i) + Y(\mathbf{s}_0) \right)^2 \right]. && \text{(de (4.1.1))}
\end{aligned}$$

Dado que  $Y(\mathbf{s})$  es un proceso intrínsecamente estacionario con semivariograma dado por  $\gamma_Y$  (ver (II)), análogamente a lo realizado en la Sección (2.3.2) para obtener la varianza del error de predicción para el kriging ordinario dada en la Ecuación (2.3.3), tenemos que,

$$\begin{aligned}
\sigma_E^2(\boldsymbol{\omega}) &= - \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \gamma_Y(\mathbf{s}_i - \mathbf{s}_j) - 2 \sum_{i=1}^n \omega_i \gamma_Y(\mathbf{s}_i - \mathbf{s}_0) \\
&= -\boldsymbol{\omega}^T \boldsymbol{\Gamma}_Y \boldsymbol{\omega} + 2\boldsymbol{\omega}^T \boldsymbol{\gamma}_{Y,0},
\end{aligned} \tag{2.4.7}$$

donde  $\boldsymbol{\Gamma}_Y \in \mathbb{R}^{n \times n}$  es la matriz simétrica con elementos  $(\boldsymbol{\Gamma}_Y)_{i,j} \doteq \gamma_Y(\mathbf{s}_i - \mathbf{s}_j)$ ,  $i, j = 1, \dots, n$  y  $\boldsymbol{\gamma}_{Y,0} = (\gamma_Y(\mathbf{s}_1 - \mathbf{s}_0), \dots, \gamma_Y(\mathbf{s}_n - \mathbf{s}_0))^T \in \mathbb{R}^n$ .

Para encontrar los pesos  $\boldsymbol{\omega} \in \mathbb{R}^n$  que minimizan  $\sigma_E^2(\boldsymbol{\omega})$  también utilizamos los multiplicadores de Lagrange con la condición sobre los pesos  $\boldsymbol{\omega}^T \mathbf{F} = \mathbf{f}_0^T$ . Para ello, definimos la función  $\varphi$  a minimizar como sigue,

$$\begin{aligned}
\varphi : \mathbb{R}^n \times \mathbb{R}^{P+1} &\rightarrow \mathbb{R} \\
(\boldsymbol{\omega}, \boldsymbol{\lambda}) &\mapsto \varphi(\boldsymbol{\omega}, \boldsymbol{\lambda}) \doteq -\boldsymbol{\omega}^T \boldsymbol{\Gamma}_Y \boldsymbol{\omega} + 2\boldsymbol{\omega}^T \boldsymbol{\gamma}_{Y,0} - 2(\boldsymbol{\omega}^T \mathbf{F} - \mathbf{f}_0^T) \boldsymbol{\lambda},
\end{aligned}$$

donde el vector  $\boldsymbol{\lambda}^T \doteq (\lambda_0, \dots, \lambda_P) \in \mathbb{R}^{P+1}$  representa los  $P+1$  multiplicadores de Lagrange. La derivada de  $\varphi$  con respecto a  $\boldsymbol{\omega}$  está dada por

$$\frac{\partial \varphi(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\omega}} = -2\boldsymbol{\Gamma}_Y \boldsymbol{\omega} + 2\boldsymbol{\gamma}_{Y,0} - 2\mathbf{F} \boldsymbol{\lambda},$$

igualando a cero obtenemos,

$$\boldsymbol{\Gamma}_Y \boldsymbol{\omega} + \mathbf{F} \boldsymbol{\lambda} = \boldsymbol{\gamma}_{Y,0}. \tag{2.4.8}$$

Por otro lado, la derivada de  $\varphi$  con respecto a  $\boldsymbol{\lambda}$  es

$$\frac{\partial \varphi(\boldsymbol{\omega}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = -2(\boldsymbol{\omega}^T \mathbf{F} - \mathbf{f}_0^T),$$

igualando a cero deducimos la segunda condición para el kriging universal

$$\boldsymbol{\omega}^T \mathbf{F} = \mathbf{f}_0^T. \tag{2.4.9}$$

Luego, de (2.4.8) y (2.4.9) tenemos el siguiente sistema de ecuaciones

$$\begin{cases} \boldsymbol{\Gamma}_Y \boldsymbol{\omega} + \mathbf{F} \boldsymbol{\lambda} &= \boldsymbol{\gamma}_{Y,0} \\ \boldsymbol{\omega}^T \mathbf{F} &= \mathbf{f}_0^T, \end{cases}$$

o equivalentemente,

$$\begin{pmatrix} \mathbf{\Gamma}_Y & \mathbf{F} \\ \mathbf{F}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \gamma_{Y,0} \\ \mathbf{f}_0^T \end{pmatrix},$$

el cual tiene solución única si  $\mathbf{F}$  es de rango completo, es decir, si las funciones  $f_p$  son linealmente independientes.

A partir de las ecuaciones (2.4.8) y (2.4.9) podemos calcular los pesos y los multiplicadores de Lagrange de la misma manera a lo realizado en (2.3.6) y (2.3.7), esto es, despejando  $\boldsymbol{\omega}$  de (2.4.8) para obtener

$$\boldsymbol{\omega} = \mathbf{\Gamma}_Y^{-1}(\gamma_{Y,0} - \mathbf{F}\boldsymbol{\lambda}). \quad (2.4.10)$$

y reemplazándolo en  $\mathbf{F}^T\boldsymbol{\omega} = \mathbf{f}_0$ , ecuación que obtenemos de transponer ambos miembros en (2.4.9) para así obtener,

$$\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}(\gamma_{Y,0} - \mathbf{F}\boldsymbol{\lambda}) = \mathbf{f}_0,$$

la cual es equivalente a,

$$\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F}\boldsymbol{\lambda} = \mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0.$$

Ahora, despejando  $\boldsymbol{\lambda}$  de la ecuación anterior tenemos que,

$$\boldsymbol{\lambda}_{UK} \doteq (\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0). \quad (2.4.11)$$

Finalmente, reemplazando (2.4.11) en (2.4.10) obtenemos los pesos para kriging universal,

$$\boldsymbol{\omega}_{UK} \doteq \mathbf{\Gamma}_Y^{-1}(\gamma_{Y,0} - \mathbf{F}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0)). \quad (2.4.12)$$

Para obtener el valor del predictor y de la varianza del error de predicción, en (2.4.7) tenemos que

$$\begin{aligned} \sigma^2(\boldsymbol{\omega}) &= -\boldsymbol{\omega}^T\mathbf{\Gamma}_Y\boldsymbol{\omega} + 2\boldsymbol{\omega}^T\gamma_{Y,0} & (2.4.13) \\ &= -\boldsymbol{\omega}^T\mathbf{\Gamma}_Y\boldsymbol{\omega} + \boldsymbol{\omega}^T\gamma_{Y,0} + \boldsymbol{\omega}^T\gamma_{Y,0} \\ &= -\boldsymbol{\omega}^T(\mathbf{\Gamma}_Y\boldsymbol{\omega} - \gamma_{Y,0}) + \boldsymbol{\omega}^T\gamma_{Y,0} \\ &= \boldsymbol{\omega}^T\mathbf{F}\boldsymbol{\lambda} + \boldsymbol{\omega}^T\gamma_{Y,0} & (\text{de (2.4.8), } -\mathbf{F}\boldsymbol{\lambda} = \mathbf{\Gamma}_Y\boldsymbol{\omega} - \gamma_{Y,0}) \\ &= \mathbf{f}_0^T\boldsymbol{\lambda} + \gamma_{Y,0}^T\boldsymbol{\omega}, & (\text{de (2.4.5), } \boldsymbol{\omega}^T\mathbf{F} = \mathbf{f}_0^T \text{ y } \boldsymbol{\omega}^T\gamma_{Y,0} \in \mathbb{R}) \end{aligned}$$

Entonces, reemplazando  $\boldsymbol{\omega}_{UK}$  y  $\boldsymbol{\lambda}_{UK}$  calculado en (2.4.11) y (2.4.12), respectivamente, en (2.4.13) tenemos que la varianza del error de predicción para el kriging universal está dada por

$$\begin{aligned} \sigma_{UK}^2 &\doteq \mathbf{f}_0^T(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0) \\ &\quad + \gamma_{Y,0}^T\mathbf{\Gamma}_Y^{-1}(\gamma_{Y,0} - \mathbf{F}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0)) \\ &= \mathbf{f}_0^T(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0) \\ &\quad + \gamma_{Y,0}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \gamma_{Y,0}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0) \\ &= \gamma_{Y,0}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - (\gamma_{Y,0}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F} - \mathbf{f}_0^T)(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\mathbf{F})^{-1}(\mathbf{F}^T\mathbf{\Gamma}_Y^{-1}\gamma_{Y,0} - \mathbf{f}_0). \end{aligned} \quad (2.4.14)$$

### 2.4.3. Forma final del predictor

Reemplazando, los pesos  $\omega_{UK}$  calculados en (2.4.12), en (2.4.3), obtenemos el predictor de kriging universal

$$Z_\omega(\mathbf{s}_0) \doteq (\boldsymbol{\gamma}_{Y,0} - \mathbf{F}(\mathbf{F}^T \boldsymbol{\Gamma}_Y^{-1} \mathbf{F})^{-1} (\mathbf{F}^T \boldsymbol{\Gamma}_Y^{-1} \boldsymbol{\gamma}_{Y,0} - \mathbf{f}_0))^T \boldsymbol{\Gamma}_Y^{-1} \mathbf{Z}. \quad (2.4.15)$$

Es importante notar que, tanto el predictor  $Z_\omega(\mathbf{s}_0)$  hallado en (2.4.15) como la varianza del error de predicción  $\sigma_{UK}^2$  calculada en (2.4.14), dependen del semivariograma residual (ver (1.2.2))  $\boldsymbol{\Gamma}_Y$  y  $\boldsymbol{\gamma}_{Y,0}$ . En otras palabras, dado que el proceso  $Z(\mathbf{s})$  es el observado, para poder estimar el variograma residual es necesario un ajuste provisorio de la función media  $\mu(\mathbf{s})$ , o bien que no requiera del conocimiento de la estructura de covarianza, o bien considerando la independencia de los datos. Una vez estimada la estructura de covarianza,  $\boldsymbol{\Gamma}_Y$ , logramos hallar los pesos de kriging para el cálculo del predictor, a partir de la ecuación obtenida en (2.4.15).

## 2.5. Conclusiones

En este capítulo explicitamos el desarrollo teórico de algunas de las versiones más utilizadas del método de predicción Kriging, detallando los puntos más importantes para la construcción de los pesos del predictor.

## Capítulo 3

# Predicción espacial no paramétrica

En este capítulo presentamos y discutimos los predictores desarrollados en esta tesis. En general nuestra propuesta se basa en la forma del predictor de kriging, esto es, una combinación lineal pesada de los datos de la muestra donde los pesos son los protagonistas. A diferencia de los predictores kriging, estos pesos no necesitan de la estimación del semivariograma y, por ello, tampoco ningún supuesto de estacionariedad sobre el proceso. Esta idea está inspirada en los resultados de Dabo-Niang et al. (2016), en particular, considerando en los pesos, núcleos que controlen tanto la geometría de los datos como su autocorrelación. Por ello, y en primer lugar, abordamos algunos conceptos de estimación no paramétrica de la función de regresión, tanto el caso univariado como el multivariado. Presentamos también algunas contribuciones recientes sobre propuestas no paramétricas aplicadas a predicción de datos espaciales.

### 3.1. Estimador de núcleo de la función de regresión

Los estimadores de núcleo son los más utilizados en estimación no paramétrica de la función de regresión, los mismos son conocidos como estimadores de **Nadaraya-Watson** debido a sus creadores Nadaraya (1964, 1965) y Watson (1964). Para definir este tipo de estimadores, consideremos una muestra aleatoria  $(X_1, Y_1), \dots, (X_n, Y_n)$  proveniente de una población de variables aleatorias  $(X, Y)$  relacionadas linealmente de la forma:

$$Y = r(X) + e, \tag{3.1.1}$$

donde  $r : \mathbb{R} \rightarrow \mathbb{R}$  es la función de regresión desconocida y  $e$  es el error con  $\mathbb{E}(e|X) = 0$  por lo que,

$$r(x) = \mathbb{E}(Y|X = x).$$

Queremos estimar de manera no paramétrica la función de regresión  $r$ . Para ello, sean  $f_{X,Y}(x, y)$  la densidad conjunta de  $(X, Y)$  y  $f_{Y|X}(y|x)$  la densidad de  $Y|X$  (desconocidas).

Por definición de esperanza condicional resulta que

$$\mathbb{E}(Y|X = x) = \int y f_{Y|X}(y|x) dy = \int y \frac{f_{X,Y}(x, y)}{f_X(x)} dy,$$

por lo tanto, podríamos definir

$$\hat{r}(x) = \hat{\mathbb{E}}_Y(Y|X) = \int y \frac{\hat{f}_{X,Y}(x, y)}{\hat{f}_X(x)} dy. \quad (3.1.2)$$

Luego, estimando las funciones de densidad conjunta y marginal podremos estimar la función de regresión. Siguiendo a Bosq (1998), los estimadores no paramétricos de la densidad  $f_X(x)$  y de la densidad conjunta  $f_{X,Y}(x, y)$  están dados por

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (3.1.3)$$

y

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right), \quad (3.1.4)$$

donde  $K(x)$  una **función núcleo** que definimos a continuación. Reemplazando (3.1.3) y (3.1.4) en (3.1.2) obtenemos

$$\begin{aligned} \hat{r}_n(x) &= \int y \frac{\frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)} dy \\ &= \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \frac{1}{h} \int y K\left(\frac{y - Y_i}{h}\right) dy}{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)}. \end{aligned} \quad (3.1.5)$$

Ahora, definiendo  $z = \frac{y - Y_i}{h}$  y considerando  $K$  tal que satisfaga

$$\int K(z) dz = 1 \quad \text{y} \quad \int z K(z) dz = 0,$$

tenemos que

$$\begin{aligned} \frac{1}{h} \int y K\left(\frac{y - Y_i}{h}\right) dy &= \frac{1}{h} \int (hz + Y_i) K(z) h dz \\ &= h \int z K(z) dz + Y_i \int K(z) dz \\ &= Y_i, \end{aligned}$$

por lo tanto en (3.1.5) tenemos que

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{j=1}^n K\left(\frac{x - X_j}{h}\right)} \doteq \sum_{i=1}^n \omega_{ni}(x) Y_i,$$

donde  $\omega_{ni}(x) \doteq \frac{K\left(\frac{x-X_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h}\right)}$  son los denominados **pesos** del estimador. Si el denominador de esta expresión es cero definimos,

$$\hat{r}_n(x) = \begin{cases} \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)Y_i}{\sum_{j=1}^n K\left(\frac{x-X_j}{h}\right)}, & \text{si } \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i, & \text{si } \sum_{j=1}^n K\left(\frac{x-X_j}{h}\right) = 0. \end{cases} \quad (3.1.6)$$

Algunas de las funciones núcleo más utilizadas son:

- *Núcleo uniforme*:  $K(x) = \frac{1}{2}\mathbb{I}_{[-1,1]}(x)$ .
- *Núcleo triangular*:  $K(x) = (x+1)\mathbb{I}_{[-1,0]}(x) + (1-x)\mathbb{I}_{[0,1]}(x)$ .
- *Núcleo cuadrático* (de Epanechnikov):  $K(x) = \frac{3}{4}(1-x^2)\mathbb{I}_{[-1,1]}(x)$ .
- *Núcleo gaussiano*:  $K(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\}$ .

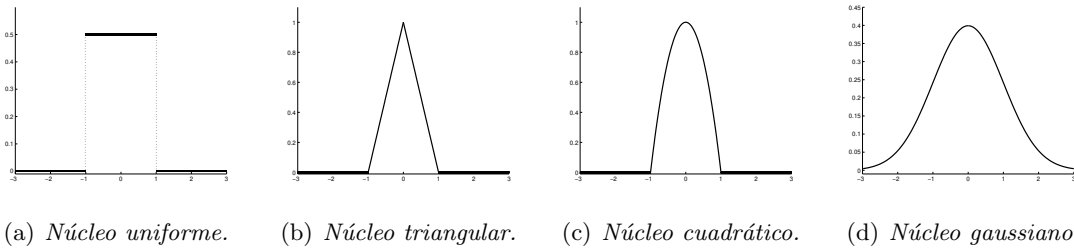


Figura 3.1: Diferentes funciones núcleo.

El estimador dado en la Ecuación (3.1.6) puede extenderse al caso multivariado (Härdle y Müller, 1997), considerando  $\mathbf{X} = (X_1, \dots, X_p)^T \in \mathbb{R}^p$ , estimando

$$r(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x}) = \int y \frac{f_{\mathbf{X},Y}(\mathbf{x}, y)}{f_{\mathbf{X}}(\mathbf{x})} dy.$$

En base al mismo razonamiento que en el caso univariado, para estimar  $r(\mathbf{x})$  necesitamos estimadores de la densidad, pero en este caso estimadores multivariados, esto es, dada una muestra  $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$  tenemos

$$\hat{f}_{\mathbf{X}}(\mathbf{x}; \mathbf{H}) = \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i),$$

y

$$\hat{f}_{\mathbf{X},Y}(\mathbf{x}, y; \mathbf{H}, h) = \frac{1}{n} \sum_{i=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) K\left(\frac{y - Y_i}{h}\right),$$

donde

$$\mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) = \frac{1}{\det(\mathbf{H})} \mathcal{K}(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)),$$

siendo  $\mathbf{H}$  una matriz no singular y diagonal cuyos elementos son los parámetros de suavizado y  $\mathcal{K} : \mathbb{R}^p \rightarrow \mathbb{R}$  una función núcleo multivariada. Considerar parámetros de suavizado iguales en todas las dimensiones implica que  $\mathbf{H} = h\mathbf{I}$ , donde  $\mathbf{I}$  es la matriz identidad.

Luego el estimador Nadaraya-Watson multivariado queda dado por

$$\hat{r}_n(\mathbf{x}) = \begin{cases} \frac{\sum_{i=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) Y_i}{\sum_{j=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_j)}, & \text{si } \sum_{j=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_j) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i, & \text{si } \sum_{j=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_j) = 0. \end{cases} \quad (3.1.7)$$

El núcleo multivariado  $\mathcal{K}$  puede ser definido como  $\tilde{\mathcal{K}} : \mathbb{R}^p \rightarrow \mathbb{R}$ , combinando una función núcleo  $K$  en  $\mathbb{R}$  y una norma  $\|\cdot\|$  en  $\mathbb{R}^p$  de la siguiente manera

$$\tilde{\mathcal{K}}(\mathbf{x}) = K(\|\mathbf{x}\|), \quad \text{para todo } \mathbf{x} \in \mathbb{R}^p.$$

Otra manera supone considerar un núcleo multivariado  $\mathcal{K}^* : \mathbb{R}^p \rightarrow \mathbb{R}$  como un producto de  $P$  funciones núcleo univariadas  $K_1, \dots, K_p$

$$\mathcal{K}^*(\mathbf{x}) = K_1(x_1) \cdot K_2(x_2) \cdots K_p(x_p), \quad \text{para todo } \mathbf{x} = (x_1, \dots, x_p)^T \in \mathbb{R}^p.$$

El suavizado obtenido en (3.1.6) o (3.1.7), es simplemente una combinación lineal pesada de los  $n$  valores observados de la variable respuesta  $Y$ , donde los pesos o ponderadores dependen de los valores observados de las variables predictoras. Además de la capacidad descriptiva de dicho estimador, es de interés su uso para obtener predicciones de la variable respuesta a partir de un nuevo valor de  $\mathbf{X}$ ; esto es, para una nueva observación  $\mathbf{x}_0$ ,  $\hat{r}_n(\mathbf{x}_0)$  constituye una predicción del valor  $Y_0$ , que *a priori* es desconocido. Precisamente, este uso predictivo es el que nos interesa en la presente tesis. Sin embargo, la singularidad de nuestro problema respecto al planteamiento general de estimación no paramétrica de una función de regresión es que, por un lado estamos interesados en predecir datos espaciales; y por otro, contamos con solo una variable, aunque medida en distintas locaciones. A partir de estas dos consideraciones, en el espíritu del *kriging*, tal variable de interés es usada como predictora y como variable respuesta, la cual se busca predecir en un determinado punto espacial. En las siguientes secciones se presentarán las contribuciones más recientes en la literatura respecto al enfoque no paramétrico de predicción espacial, para posteriormente presentar nuestra propuesta de diferentes estimadores alternativos en el marco de lo que denominaremos como **kriging no paramétrico**.

## 3.2. Estado actual de la literatura

Los métodos de predicción espacial no paramétrica son relativamente nuevos ya que, hasta el año 2004 aproximadamente (Biau y Cadre, 2004) la mayoría de los resultados existentes

para datos correlacionados correspondían a series de tiempo (Bosq, 1998; Györfi et al., 1989) y las pocas generalizaciones existentes para datos espaciales, estaban relacionadas con la estimación no paramétrica de la densidad (Carbon et al., 1996, 1997). En este apartado presentamos algunos de los estimadores no paramétricos utilizados para predicción espacial.

Una contribución temprana es la de Biau y Cadre (2004), quienes introdujeron un método para predecir la variable de interés en un sitio  $\mathbf{s}_0$  a partir de la función de regresión (3.1.7), considerando como predictora la misma variable medida en un vecindario del mismo.

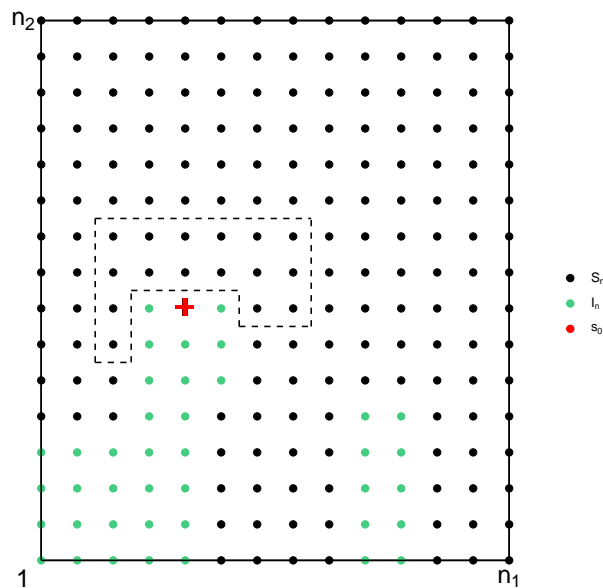


Figura 3.2: Ejemplo del predictor introducido por Biau y Cadre (2004) para  $d = 2$ . Los puntos que están contenidos en la línea de puntos son los que pertenecen a la vecindad y serán utilizados para la predicción (Biau y Cadre, 2004, pág. 329).

Formalmente, dado  $\mathbf{n} = (n_1, \dots, n_d) \in (\mathbb{N}^*)^d$ ,  $d \geq 1$ , donde  $\mathbb{N}^*$  es el conjunto de los naturales sin el cero, construimos una región rectangular  $I_{\mathbf{n}} = \{\mathbf{s} \in (\mathbb{N}^*)^d : 1 \leq s_k \leq n_k, k = 1, \dots, d\}$  con  $n = n_1 \dots n_d$  la cantidad total de puntos en la misma, y sea  $S_{\mathbf{n}} \subseteq I_{\mathbf{n}}$  el subconjunto de sitios donde la variable de interés fue observada,  $\mathbf{s}_0 \in I_{\mathbf{n}} - S_{\mathbf{n}}$  denota el sitio donde la misma no fue observada y, por consiguiente, donde se desea predecir. Para cada  $\mathbf{s}$ , tenemos el vector  $p$ -dimensional de vecinos,  $\tilde{Z}_{\mathbf{s}}$ , como aquel cuyos componentes son  $\{Z_{\mathbf{i}}, \mathbf{i} \in V_{\mathbf{s}}\}$  donde  $V_{\mathbf{s}}$  son los índices  $\mathbf{i}$  tal que  $\mathbf{s}_{\mathbf{i}}$  es un vecino de  $\mathbf{s}$ .

Con todas estos elementos, Biau y Cadre (2004) definen a el predictor de la variable de

interés en el sitio  $\mathbf{s}_0$  de la forma

$$\hat{Z}(\mathbf{s}_0) \doteq \hat{r}(\tilde{Z}_{\mathbf{s}_0}) = \frac{\sum_{\substack{\mathbf{s} \in S_{\mathbf{n}} \\ V_{\mathbf{s}} \subset S_{\mathbf{n}}}} \mathcal{K}_{\mathbf{H}}(\tilde{Z}_{\mathbf{s}_0} - \tilde{Z}_{\mathbf{s}}) Z_{\mathbf{s}}}{\sum_{\substack{\mathbf{s} \in S_{\mathbf{n}} \\ V_{\mathbf{s}} \subset S_{\mathbf{n}}}} \mathcal{K}_{\mathbf{H}}(\tilde{Z}_{\mathbf{s}_0} - \tilde{Z}_{\mathbf{s}})}, \quad (3.2.1)$$

donde  $\mathcal{K}_{\mathbf{H}} : \mathbb{R}^p \rightarrow \mathbb{R}^+$  es una función núcleo  $p$ -dimensional con sus respectivos parámetros de suavizado dados por  $\mathbf{H} = h\mathbf{I}$ , siendo  $\mathbf{I} \in \mathbb{R}^{p \times p}$  la matriz identidad. Es fácil ver que el mismo se obtiene a partir de la estimación por núcleos (3.1.7), considerando  $Y_{\mathbf{s}} = Z_{\mathbf{s}}$  y  $\mathbf{X}_{\mathbf{s}} = \tilde{Z}_{\mathbf{s}}$  y evaluando en  $\tilde{Z}_{\mathbf{s}_0}$ . Para este estimador, los autores demostraron la consistencia uniforme del mismo como así también la normalidad asintótica cuando  $K$  es el núcleo uniforme.

Más tarde, Dabo-Niang y Yao (2007) extendieron la idea de Biau y Cadre (2004) al dominio continuo, esto es, considerando los sitios  $\mathbf{s}_i \in \mathbb{R}_+^d$ . Más precisamente, dado  $\mathbf{T} \in \mathbb{R}_+^d$ ,  $d \geq 1$ , definieron la región rectangular  $I_{\mathbf{T}} = \{\mathbf{s} \in \mathbb{R}_+^d : 0 \leq s_k \leq T_k, T_k \geq 1, k = 1, \dots, d\}$ , siendo  $T = T_1 \cdots T_d$  la cantidad total de puntos en la misma. Sea  $S_{\mathbf{T}} \subseteq I_{\mathbf{T}}$  el subconjunto de sitios donde la variable de interés fue observada, y  $\mathbf{s}_0 \in I_{\mathbf{T}} - S_{\mathbf{T}}$  el sitio donde la misma no fue observada y se desea predecir, luego la estimación por núcleos de la regresión, a partir de observaciones del proceso  $Z(\mathbf{s}) = (\mathbf{X}(\mathbf{s}), Y(\mathbf{s})) \in \mathbb{R}^p \times \mathbb{R}$ , está dada por

$$\hat{r}_{\mathbf{T}}(\mathbf{x}) = \begin{cases} \frac{\int_{I_{\mathbf{T}}} Y(\mathbf{s}) \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}(\mathbf{s})) d\mathbf{s}}{\int_{I_{\mathbf{T}}} \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}(\mathbf{u})) d\mathbf{u}}, & \text{si } \int_{I_{\mathbf{T}}} \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}(\mathbf{u})) d\mathbf{u} \neq 0, \\ \frac{1}{\hat{\mathbf{T}}} \int_{I_{\mathbf{T}}} Y(\mathbf{s}) d\mathbf{s}, & \text{si } \int_{I_{\mathbf{T}}} \mathcal{K}_{\mathbf{H}}(\mathbf{x} - \mathbf{X}(\mathbf{u})) d\mathbf{u} = 0, \end{cases} \quad (3.2.2)$$

donde  $\mathbf{x} \in \mathbb{R}^p$  con lo cual el núcleo  $\mathcal{K}_{\mathbf{H}}$  está definido en  $\mathbb{R}^p$  al igual que el núcleo utilizado en (3.2.1). Las autoras demostraron la consistencia débil y fuerte del estimador dado en (3.2.2) y también las tasas de convergencia del mismo.

Por otro lado, en Menezes et al. (2010) utilizaron un diseño de muestreo estocástico para la selección de sitios en lugar de una grilla determinística. Esto es, dadas  $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$  observaciones del proceso  $Z(\mathbf{s})$ , y suponiendo que el mismo es estacionario de segundo orden, para  $\mathbf{s}_0 \in \mathbb{R}^d$ , el predictor propuesto está dado por

$$\hat{Z}(\mathbf{s}_0) = \frac{\sum_{i=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{s}_0 - \mathbf{s}_i) Z(\mathbf{s}_i)}{\sum_{j=1}^n \mathcal{K}_{\mathbf{H}}(\mathbf{s}_0 - \mathbf{s}_j)}, \quad (3.2.3)$$

donde  $\mathcal{K}_{\mathbf{H}}$  es una función de núcleo  $d$ -dimensional y  $\mathbf{H} \in \mathbb{R}^{d \times d}$  ya que  $\mathbf{s} \in \mathbb{R}^d$ . Los autores estudian el comportamiento asintótico del predictor asumiendo un diseño no determinístico

y demuestran que el error cuadrático medio de predicción converge a cero a medida que aumenta el tamaño de la muestra.

Una diferencia importante entre el predictor presentado por Menezes et al. (2010) y los dos anteriores es que, este último utiliza un núcleo que controla las distancias entre sitios, mientras que los otros dos utilizan un núcleo que controla las distancias entre predictoras. Una combinación de estas dos ideas fue considerada en Dabo-Niang et al. (2016), quienes utilizaron dos núcleos en la definición del predictor, uno que controla la distancia entre las predictoras y otro que controla la distancia entre sitios. Consideraron un proceso espacial  $Z(\mathbf{s})$  tal que para cada  $\mathbf{s}_i \in \mathbb{Z}^d$  (donde  $\mathbb{Z}$  representa a los números enteros)  $Z(\mathbf{s}_i) = (\mathbf{X}(\mathbf{s}_i), Y(\mathbf{s}_i)) \in \mathbb{R}^p \times \mathbb{R}$ , para el cual se verifica el modelo de regresión dado en (3.1.1) y es observable en una grilla rectangular  $\mathcal{O}_n = \{\mathbf{s} \in \mathbb{Z}^d : 1 \leq s_k \leq n_k, k = 1, \dots, d\}$ . El predictor propuesto, entonces, está dado por

$$\hat{Y}(\mathbf{s}_0) = \frac{\sum_{\mathbf{s} \in \mathcal{O}_n} Y(\mathbf{s}) \mathcal{K}_{\mathbf{H}_1}(\mathbf{X}(\mathbf{s}_i) - \mathbf{X}(\mathbf{s}_0)) \mathcal{K}_{\mathbf{H}_2}(\mathbf{s} - \mathbf{s}_0)}{\sum_{\mathbf{s} \in \mathcal{O}_n} \mathcal{K}_{\mathbf{H}_1}(\mathbf{X}(\mathbf{s}_i) - \mathbf{X}(\mathbf{s}_0)) \mathcal{K}_{\mathbf{H}_2}(\mathbf{s} - \mathbf{s}_0)}, \quad (3.2.4)$$

donde  $\mathbf{s}_0 \in \mathbb{Z}^d$  y  $\mathbf{s}_0 \notin \mathcal{O}_n$ ,  $\mathcal{K}_{\mathbf{H}_1} : \mathbb{R}^p \rightarrow \mathbb{R}$  y  $\mathcal{K}_{\mathbf{H}_2} : \mathbb{R}^d \rightarrow \mathbb{R}$ .

Podemos observar que este predictor es una generalización del dado en (3.2.3) cuando se consideran covariables y un diseño no aleatorio. Esta última idea también se ha presentado en Dabo-Niang et al. (2014) en el contexto de la estimación de la densidad y en Ternynck (2014) para tratar un problema de regresión de datos funcionales.

### 3.3. La nueva propuesta: kriging no paramétrico

En esta sección presentamos el predictor al que llamamos **kriging no paramétrico** el cual está basado en la idea de kriging (Ecuación (2.1.1)) pero con pesos estimados de manera no paramétrica. Formalmente, sean  $d \in \mathbb{N}$ , y  $D \subset \mathbb{R}^d$  el conjunto de sitios muestreados cuyo cardinal es  $n$ . Supongamos que la variable de interés  $Z(\mathbf{s}_i)$  es observada en los sitios  $\mathbf{s}_i \in D$  y que deseamos predecir el valor la variable  $Z(\mathbf{s}_0)$  para  $\mathbf{s}_0 \notin D$ . El predictor de kriging no paramétrico está dado por,

$$Z_\omega(\mathbf{s}_0) = \sum_{i=1}^n \omega_i Z(\mathbf{s}_i).$$

Considerando diferentes formas (no paramétricas) para los pesos  $\omega_i$ , obtendremos diferentes versiones de predictores que serán estudiadas en las siguientes secciones. Para describir el comportamiento de dichas versiones, mostramos en la Figura 3.3 una realización de un proceso  $Z(\mathbf{s})$  gaussiano, no estacionario con función de covarianza exponencial, con  $d = 2$ ,  $n = 99$ . En el panel (a) de dicha figura consideramos desconocido el valor regionalizado en el sitio  $\mathbf{s}_0 = (4, 4)$ , es decir, suponemos que el valor  $z \approx 0.2$  es desconocido y será predicho a partir del valor de la variable de interés en los sitios restantes. En el panel (b)

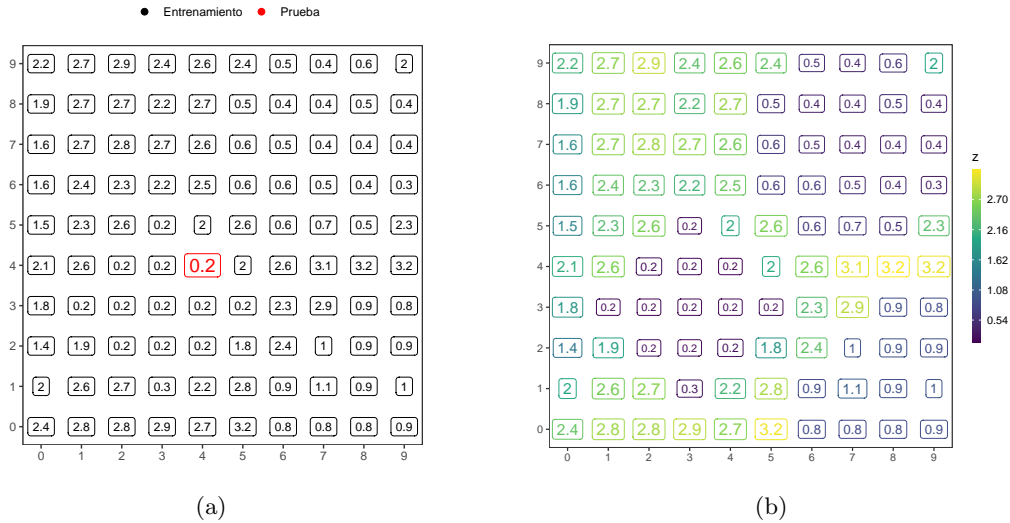


Figura 3.3: (a) Conjunto de valores regionalizados simulados a partir de un proceso gaussiano  $Z(\mathbf{s})$ , no estacionario con función de covarianza exponencial para  $n = 99$  y  $d = 2$ . (b) Comportamiento de los valores regionalizados según su ubicación

de la misma figura agregamos una escala de colores para visualizar la distribución de tales valores regionalizados.

### 3.3.1. Kriging de un núcleo simple (K1S)

En esta versión del predictor, tomando como base a Menezes et al. (2010) y Dabo-Niang et al. (2016), los pesos  $\omega_i$  consisten en un sólo núcleo que controla la distancia espacial entre los puntos muestreados y el sitio donde deseamos realizar la predicción, esto es

$$\omega_i^{K1S} \doteq \frac{K_1\left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h}\right)}{\sum_{j=1}^n K_1\left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h}\right)}, \quad (3.3.1)$$

donde  $K_1 : \mathbb{R} \rightarrow \mathbb{R}$  es una función núcleo univariada,  $h$  es el parámetro de suavizado correspondiente y  $d(\cdot)$  es la distancia euclidiana entre sitios.

En la Figura 3.4 (a) mostramos el comportamiento de los pesos (3.3.1), tomando como  $K_1$  una función de núcleo gaussiana y  $h$  el parámetro de suavizado elegido mediante validación cruzada (5-fold), utilizando como muestra de entrenamiento los 99 datos restantes. En dicha figura observamos que a medida que aumenta la distancia al sitio de interés, el valor de los pesos disminuye. Por otro lado, en la Figura 3.4 (b) mostramos el comportamiento de los pesos para diferentes valores del parámetro de suavizado, donde observamos como la magnitud de los pesos aumenta cuando el parámetro de suavizado disminuye, concentrándose sus valores alrededor del sitio  $\mathbf{s}_0 = (4, 4)$ .

La principal ventaja de este predictor es su simpleza, al involucrar un solo núcleo univariado y con ello, un único parámetro de suavizado. La desventaja es que solamente tiene en cuenta la distancia entre los sitios, ignorando la posible relación entre el valor de la variable en algún entorno del punto que se busca predecir.

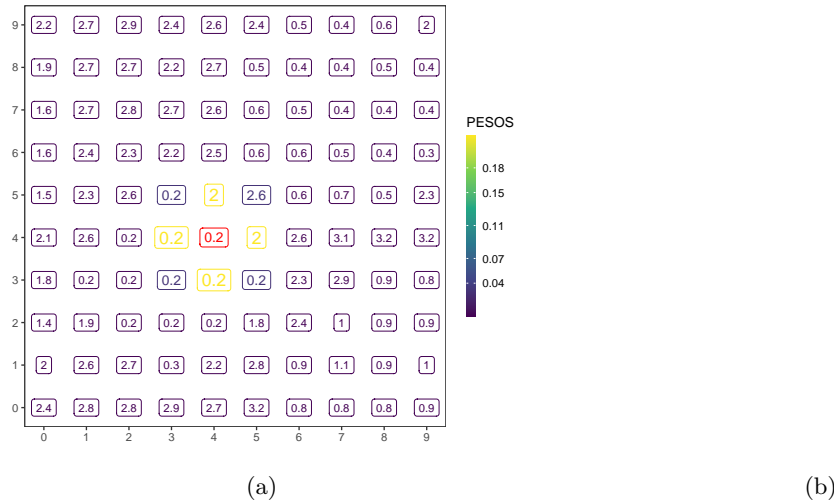


Figura 3.4: Comportamiento de los pesos del predictor K1S sobre el conjunto de valores regionalizados para: (a) el parámetro de suavizado obtenido a través de validación cruzada, (b) diferentes valores del parámetro de suavizado.

### 3.3.2. Kriging de dos núcleos con Índice de Moran (K2IM)

En una segunda versión del predictor incluimos, además del núcleo que controla las distancias, un segundo núcleo que controla la autocorrelación local medida a través del Índice de Moran Local  $I_i$  definido en la Ecuación (1.3.2). Los pesos en este caso quedan determinados por

$$\omega_i^{K2IM} \doteq \frac{K_1 \left( \frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h} \right) K_2 \left( \frac{1}{e^{I_i}} \right)}{\sum_{j=1}^n K_1 \left( \frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h} \right) K_2 \left( \frac{1}{e^{I_i}} \right)}. \quad (3.3.2)$$

El objetivo de incluir un segundo núcleo al predictor (idea basada en Dabo-Niang et al. (2016), Ecuación (3.2.4)) es que el mismo pueda identificar sitios donde la variable de interés se encuentre rodeada por valores similares, asignándole en dicho caso un mayor peso a la misma. Como fue mencionado en la Sección 1.3.4, un Índice de Morán positivo,  $I_i > 0$ , indica que la variable de interés en el sitio  $\mathbf{s}_i$  está rodeada por valores similares a ella. De esta manera, la transformación  $\frac{1}{e^{I_i}}$  toma un valor pequeño haciendo que en ese caso el peso tome un valor grande. Esto es, la variable de interés con mayor Índice de Moran, obtiene mayor peso en el predictor.

En la Figura 3.5 (a) observamos los valores regionalizados y el comportamiento de los pesos dados en la Ecuación (3.3.2), tomando como  $K_1$  y  $K_2$  funciones núcleos gaussianas y 8-vecinos más cercanos para la construcción de la matriz de vecindario  $P$  utilizada en la Ecuación (1.3.2), cuyos elementos  $p_{ij} = 1$  si los sitios  $s_i$  y  $s_j$  son vecinos y  $p_{ij} = 0$  si no lo son. En la Figura 3.5 (b) podemos observar el comportamiento del término  $K_2\left(\frac{1}{e^{I_i}}\right)$  incluido en la Ecuación (3.3.2). En dicha figura podemos ver que se asignan mayores pesos a observaciones que están rodeadas de valores similares a ellas, formando así grupos homogéneos, lo cual es uno de los objetivos del análisis de estos índices locales (Anselin, 2010).

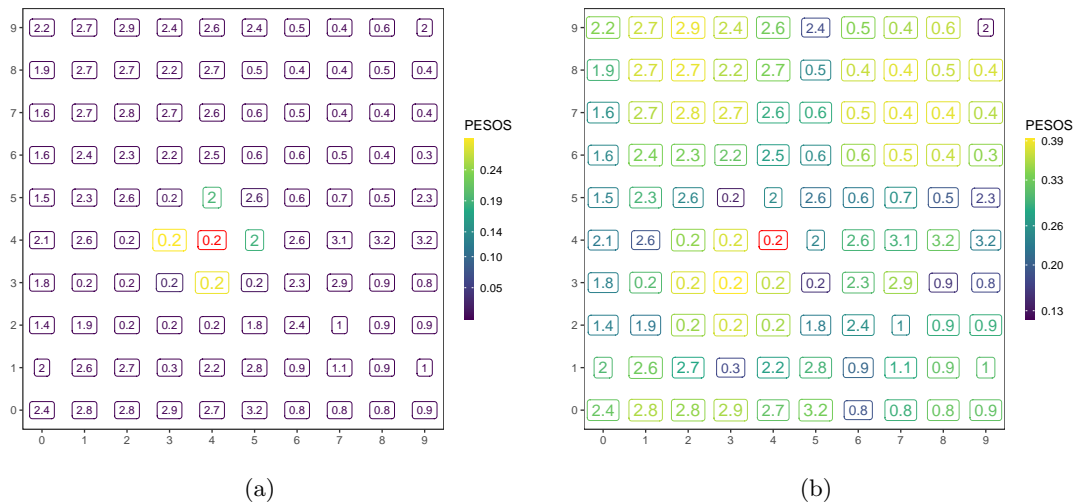


Figura 3.5: (a) Comportamiento de los pesos del predictor  $K2IM$  sobre el conjunto de valores regionalizados. (b) Comportamiento de la función núcleo que contiene al Índice  $I_i$ .

Específicamente, podemos observar tres grupos bien definidos (aquellos para los cuales el valor que toma  $K_2\left(\frac{1}{e^{I_i}}\right)$  es más alto). El valor de la variable de interés entre dichos grupos difiere, aunque el valor del término  $K_2\left(\frac{1}{e^{I_i}}\right)$  es parecido. En este caso en particular, esto no afecta la predicción ya que, al estar dichos grupos alejados espacialmente, el núcleo que controla la distancia espacial devuelve valores pequeños dejando como resultado pesos pequeños. Sin embargo, podría darse el caso en el que el valor de la variable de interés en dichos grupos sea similar, en cuyo caso estaríamos perdiendo información ya que, al igual que en el caso anterior, a estos valores espacialmente alejados se le asigna un peso pequeño. Un caso concreto en el que podría darse este fenómeno es en el precio de viviendas de diferentes barrios privados no vecinos. Observemos además que no puede ocurrir que tengamos grupos donde el término  $K_2\left(\frac{1}{e^{I_i}}\right)$  sea parecido, el valor de la variable de interés sea diferente y que sean cercanos espacialmente, ya que dos sitios no podrían tener un alto índice de Moran y no ser similares pues están rodeados por los mismos valores.

### 3.3.3. Kriging de dos núcleos con medianas (K2ME)

En esta versión del predictor incluimos, además del núcleo que controla las distancias, un segundo núcleo que incluye cierta medida de similaridad entre vecindarios definida a partir de la mediana. Más precisamente, para cada sitio consideramos la mediana de los valores de la variable de interés en sitios vecinos y luego definimos una distancia basada en la diferencia de dichas medianas. Seleccionamos la mediana por ser esta una medida robusta a valores atípicos o extremos los cuales pueden afectar innecesariamente la medida de similaridad entre vecindarios. Formalmente, sea  $\tilde{Z}_s$  el vector cuyos componentes están dados por  $\{Z(\mathbf{s}_i), \mathbf{i} \in V_s\}$  donde  $V_s$  es el conjunto de índices  $\mathbf{i}$  tal que  $\mathbf{s}_i$  es un vecino de  $\mathbf{s}$ . Supondremos que el cardinal de  $V_s$  es  $p$  por lo que  $\tilde{Z}_s \in \mathbb{R}^p$ . Definimos,

$$d_m(\tilde{Z}_s, \tilde{Z}_{s_0}) \doteq Med\tilde{Z}_s - Med\tilde{Z}_{s_0},$$

donde  $Med\tilde{Z}_s$  es la mediana del vector  $\tilde{Z}_s$ . De esta manera, los pesos quedan dados por

$$\omega_i^{K2ME} \doteq \frac{K_1\left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h_\beta}\right) K_2\left(\frac{d_m(\tilde{Z}_{\mathbf{s}_i}, \tilde{Z}_{\mathbf{s}_0})}{h_\alpha}\right)}{\sum_{j=1}^n K_1\left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h_\beta}\right) K_2\left(\frac{d_m(\tilde{Z}_{\mathbf{s}_j}, \tilde{Z}_{\mathbf{s}_0})}{h_\alpha}\right)}. \quad (3.3.3)$$

Observemos que este predictor está basado en el predictor introducido por Biau y Cadre (2004) (Ecuación (3.2.1)) el cual compara las distancias entre vectores de vecinos. La diferencia principal entre ambas propuestas es que en éste último, los autores consideran la distancia entre vectores, necesitando así considerar un núcleo multivariado. En esta nueva versión sin embargo, proponemos considerar la distancia entre las medianas de dichos valores necesitando de esta manera un núcleo univariado.

En la Figura 3.6 (a) observamos el comportamiento de los pesos dados en la Ecuación (3.3.3) sobre los valores regionalizados. Los mismos fueron calculados tomando como  $K_1$  y  $K_2$  funciones núcleo gaussianas y, para la construcción del vector  $\tilde{Z}$ , 8-vecinos más cercanos (lo cual se mantiene para los casos siguientes). En la Figura 3.6 (b) observamos en cada sitio  $\mathbf{s}_i$  la mediana del vector  $\tilde{Z}_{\mathbf{s}_i}$ .

Observemos que en este tipo de pesos, los cuales están contruidos por el producto de dos funciones núcleo, los pesos mayores están asociados a observaciones para las cuales ambas distancias son pequeñas. Sin embargo, un peso pequeño no necesariamente está asociado a una observación para la cual ambas distancias son grandes. De hecho, basta con que solo una de las distancias sea grande (o un parámetro de suavizado pequeño) para que una función núcleo se haga muy pequeña y anule el peso a la observación, sin considerar el valor del otro núcleo. Esto es, aun cuando exista similaridad entre los vecindarios, si los mismos están alejados geográficamente, el peso asignado a ese sitio será pequeño. Lo mismo ocurre con sitios que se ubican cerca del sitio de interés pero su mediana asociada difiere considerablemente de la mediana asociada al mismo. Este comportamiento puede

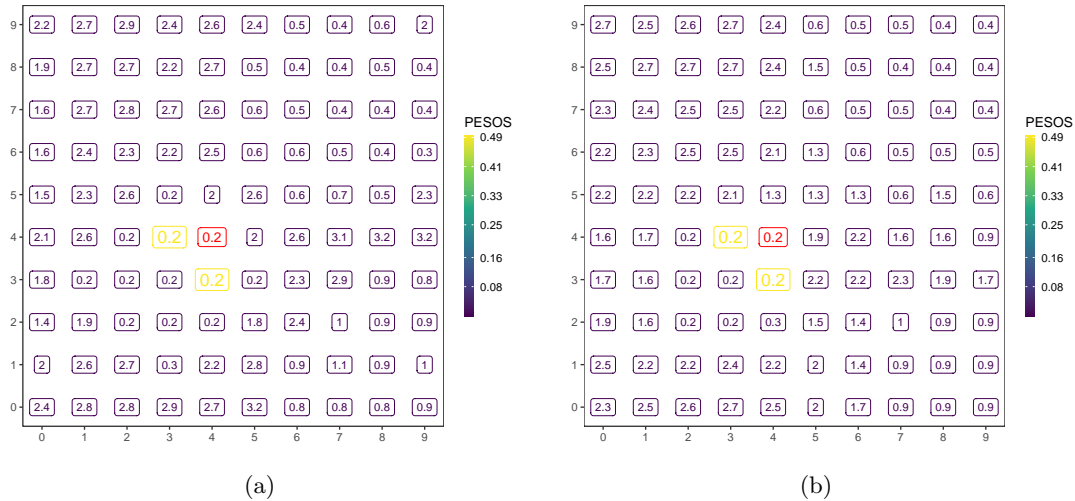


Figura 3.6: Comportamiento de los pesos del predictor  $K2ME$  sobre el conjunto de: (a) valores regionalizados, (b) medianas de los valores vecinos para cada sitio.

observarse en la Figura 3.6 (a) y (b), donde solamente un par de sitios cercanos a  $\mathbf{s}_0$  reciben un peso alto. Esto se debe a que son sitios que están espacialmente más cerca de  $\mathbf{s}_0$  y cuya mediana del vecindario es similar a la mediana del vecindario del mismo. También observamos que al sitio ubicado en la coordenada (4, 5), el cual se encuentra a la misma distancia de  $\mathbf{s}_0$  que el que está en (3, 4), se le asigna menor peso ya que la distancia entre medianas es grande.

Considerando la autocorrelación supuesta y la primera ley de la geografía (Sección 1.3.3) esperamos que la mediana sea representativa de la distribución de la variable  $Z(\mathbf{s}_0)$ , lo cual significa que el valor de la mediana asignado al sitio  $\mathbf{s}_0$  sea similar al valor regionalizado  $z_0$ , el cual es no observable. Bajo dicha hipótesis, esta versión del predictor puede captar el comportamiento de la variable de interés cuyos valores son muy disímiles incluso entre sitios cercanos. En la Figura 3.6 observamos que el valor regionalizado en  $\mathbf{s}_0$  (en el panel (a)) y la mediana  $Med\tilde{Z}_{\mathbf{s}_0}$  (en el panel (b)) son iguales, esto significa que en este caso la hipótesis se cumple. Si retomamos el ejemplo del fenómeno de precios de inmuebles, también es frecuente encontrar zonas fronteras donde podemos observar como vecinos barrios populares con barrios privados de alto poder adquisitivo, teniendo con esto una variabilidad de precios alta entre viviendas vecinas.<sup>1</sup>

A la elección del sitio  $\mathbf{s}_0 = (4, 4)$ , como sitio a predecir, no la realizamos de manera aleatoria sino de forma intencional. Si observamos la Figura 3.3 donde coloreamos en función de la variable de interés podemos ver que dicho sitio representa un límite en cuanto al comportamiento de la variable, pues, si observamos a su alrededor podemos ver que el vecindario es fuertemente heterogéneo. En la Figura 3.7 (a) observamos la misma

<sup>1</sup><https://unequalscenes.com/buenos-aires>

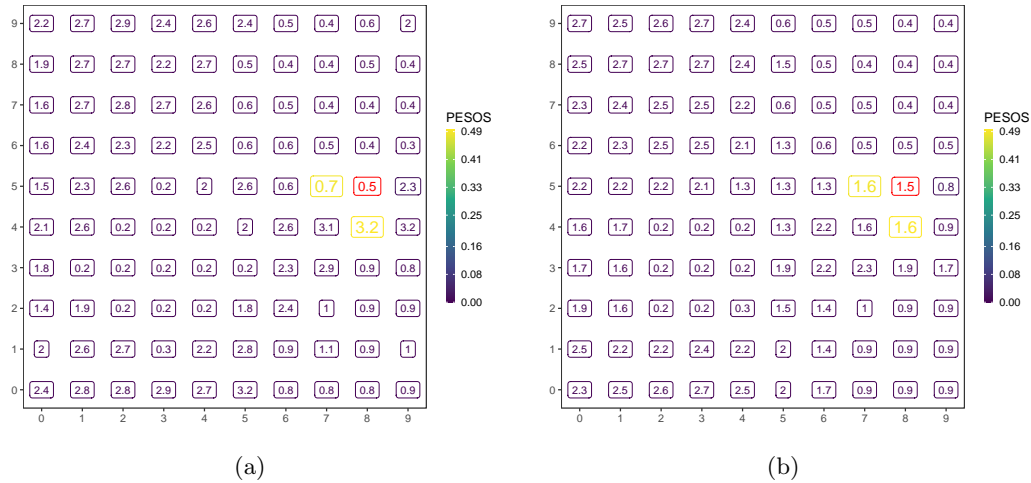


Figura 3.7: Comportamiento de los pesos del predictor  $K2ME$  sobre: (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra.

realización del proceso observado en la Figura 3.3 pero en este caso cambiamos el sitio donde queremos predecir con el objetivo de ejemplificar el caso donde la mediana *no sea representativa* de la distribución de la variable  $Z(\mathbf{s}_0)$ . La coordenada del nuevo sitio está dada por  $\mathbf{s}_0 = (8, 5)$  y tiene la particularidad de que la mediana de los valores de la variable en el vecindario de  $\mathbf{s}_0$ , esta dada por  $Med\tilde{Z}_{\mathbf{s}_0} = 1.5$ , mientras que el valor de la variable en  $\mathbf{s}_0$  es  $z_0 = 0.5$ . Esto se da porque que los valores de la variable en sitios vecinos a  $\mathbf{s}_0$  son muy diferentes al valor en este sitio, lo que se le llama autocorrelación espacial de datos disímiles (Anselin, 2010) o autocorrelación negativa (Griffith y Arbia, 2010). En consecuencia, en la Figura 3.7 (a), observamos que el comportamiento de los pesos no es el esperado. Aparecen mayores pesos asociados a valores que no son similares al que queremos predecir, como ser el valor ubicado en el sitio  $(8, 4)$ , el cual es muy diferente al valor que queremos predecir y se le asocia un peso alto. Por lo tanto, si bien esta versión no paramétrica es enriquecida con la introducción de la similaridad entre vecindarios de acuerdo a un valor representativo de la variable de interés, los pesos se comportarán de manera errática sobre realizaciones de procesos en los que tengamos evidencia de autocorrelación negativa; esto es, donde una observación se encuentre rodeada de valores muy diferentes a la misma.

### 3.3.4. Kriging de un núcleo Multiplicativo (K1M)

Teniendo en cuenta lo expuesto en el apartado anterior, en este consideramos un predictor cuyos pesos tienen en cuenta la interacción entre la distancia geográfica y la distancia entre medianas, pero considerando un solo núcleo. En esta dirección, introducimos los pesos de

núcleo multiplicativo dados por

$$\omega_i^{K1M} \doteq \frac{K_1 \left( \frac{d(\mathbf{s}_i, \mathbf{s}_0) * d_m(\tilde{Z}_{\mathbf{s}_i}, \tilde{Z}_{\mathbf{s}_0})}{h} \right)}{\sum_{j=1}^n K_1 \left( \frac{d(\mathbf{s}_j, \mathbf{s}_0) * d_m(\tilde{Z}_{\mathbf{s}_j}, \tilde{Z}_{\mathbf{s}_0})}{h} \right)},$$

los cuales, a diferencia de los pesos dados en la Ecuación (3.3.3), utilizan solo una función núcleo evaluada en el producto de las distancias entre sitios y las distancias entre las medianas de vecindarios, considerando en este caso un solo parámetro de suavizado común. Esta estrategia brinda una ventaja computacional al requerir la búsqueda un solo parámetro de suavizado en contraposición con el de dos núcleos.

En la Figura 3.8 observamos el comportamiento de estos pesos para los dos sitios de interés  $\mathbf{s}_0 = (4, 4)$  en el panel superior y  $\mathbf{s}_0 = (8, 5)$  en el panel inferior. A diferencia de las Figuras 3.6 y 3.7 correspondientes al predictor K2ME para los mismos sitios, en este caso observamos un mayor aporte no solo de sitios cercanos (esto es, cuya distancia espacial es pequeña) sino también en sitios alejados pero cuya mediana del vecindario es similar a la del sitio donde queremos realizar la predicción. Particularmente, en la Figura 3.8 (a) y (b) vemos que los pesos mayores se concentran cerca del sitio de interés y se observa un leve cambio en la magnitud del peso a medida que nos alejamos del mismo. Esto se debe a que la distancia entre medianas se mantiene constante sobre dichos sitios y por lo tanto el cambio en el comportamiento lo aporta su distancia espacial. Por otro lado, en la Figura 3.8 (c) y (d) vemos que, debido a que la mediana es más variable cerca del punto donde se desea realizar la predicción, los sitios a los que se le asigna mayor peso están relacionados con la distancia entre medianas, desestimando la distancia espacial a la que se encuentran dichos sitios.

Este predictor también tiene un mejor comportamiento en sitios donde la mediana es representativa de la distribución de  $Z(\mathbf{s}_0)$ . La ventaja del mismo radica en que permite detectar vecindarios similares aún estando espacialmente alejados y así poder usar toda la información disponible en la muestra.

### 3.3.5. Kriging de un núcleo Aditivo (K1A)

Si en el peso anterior consideramos la función núcleo evaluada en la suma en lugar del producto, obtenemos el peso de núcleo aditivo,

$$\omega_i^{K1A} \doteq \frac{K_1 \left( \frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h_\beta} + \frac{d_m(\tilde{Z}_{\mathbf{s}_i}, \tilde{Z}_{\mathbf{s}_0})}{h_\alpha} \right)}{\sum_{j=1}^n K_1 \left( \frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h_\beta} + \frac{d_m(\tilde{Z}_{\mathbf{s}_j}, \tilde{Z}_{\mathbf{s}_0})}{h_\alpha} \right)}. \quad (3.3.4)$$

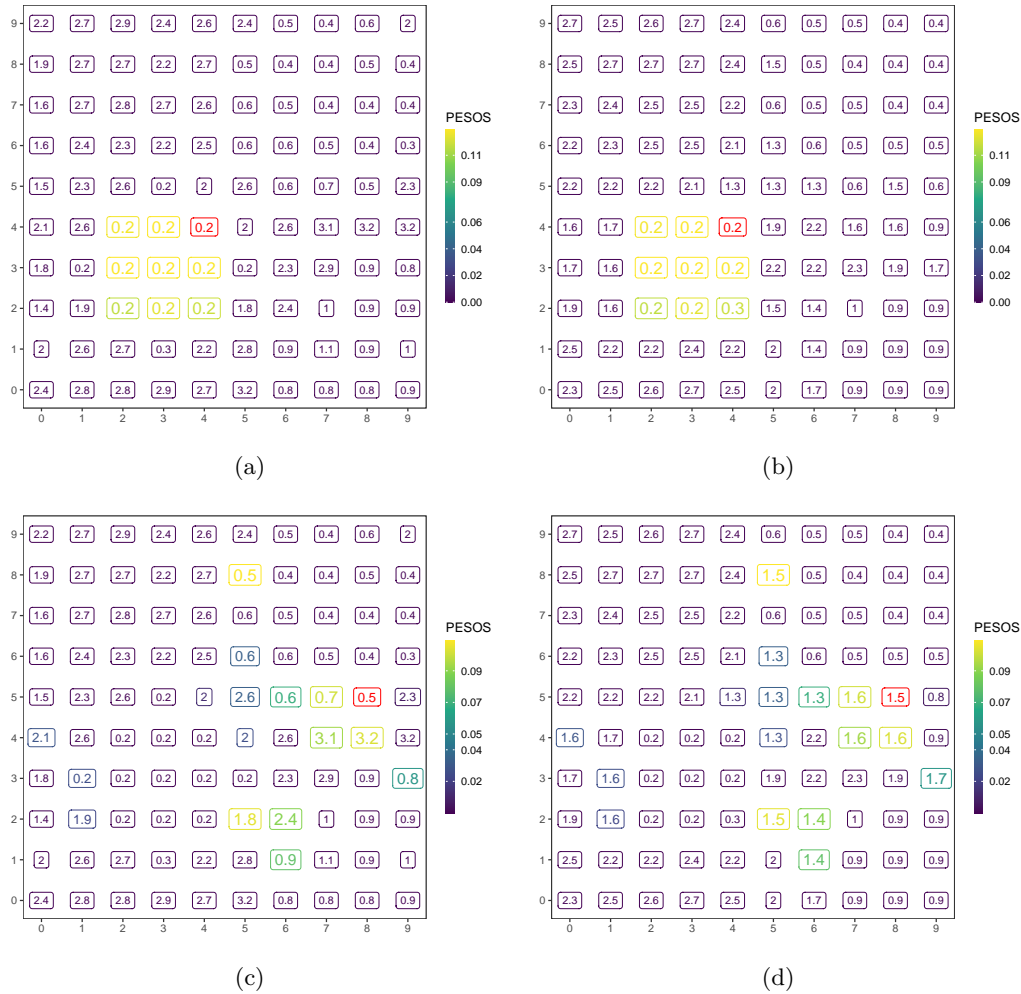


Figura 3.8: Panel superior: Comportamiento de los pesos del predictor K1M para el sitio  $s_0 = (4, 4)$  sobre (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra. Panel inferior: Comportamiento de los pesos del predictor K1M para el sitio  $s_0 = (8, 5)$  sobre (c) el conjunto de valores regionalizados, (d) el conjunto de medianas de los vecindarios para cada sitio de la muestra.

Nuevamente, en la Figura 3.9 observamos el comportamiento de los pesos para los dos sitios de interés  $s_0 = (4, 4)$  y  $s_0 = (8, 5)$ , respectivamente. En particular, comparando las Figuras 3.9 y 3.6 se puede observar que el comportamiento de los pesos dados en la Ecuación (3.3.4) es similar al de los pesos dados en (3.3.3), donde en el producto de los dos núcleos tiene una fuerte influencia el que está asociado a la mayor distancia, dejando en ese caso un valor del núcleo cercano a cero. En esta nueva versión es también la distancia más grande la que cuenta con la mayor contribución al peso, haciendo que la función núcleo devuelva valores pequeños y dejando sin influencia el comportamiento de la otra distancia.

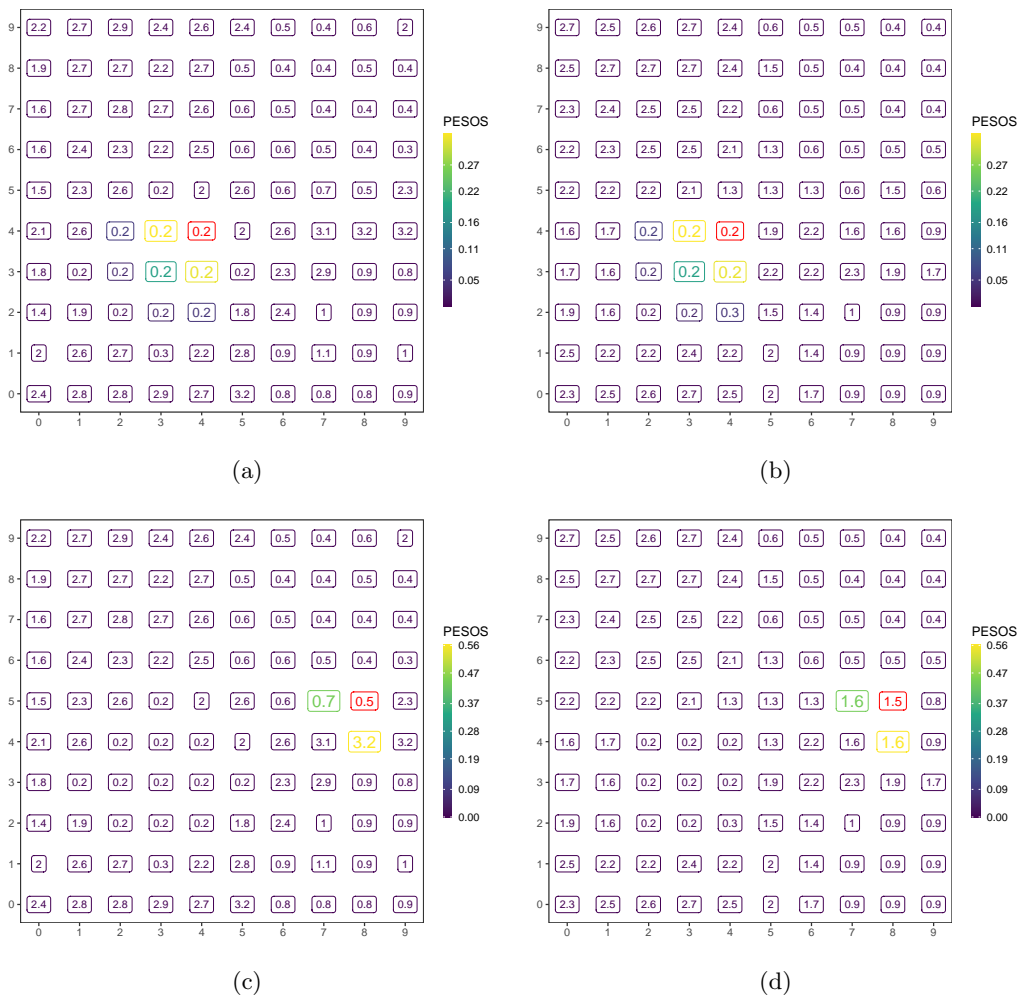


Figura 3.9: Panel superior: Comportamiento de los pesos del predictor  $K1M$  para el sitio  $s_0 = (4, 4)$  sobre (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra. Panel inferior: Comportamiento de los pesos del predictor  $K1M$  para el sitio  $s_0 = (8, 5)$  sobre (c) el conjunto de valores regionalizados, (d) el conjunto de medianas de los vecindarios para cada sitio de la muestra.

### 3.3.6. Comparación de errores

Para finalizar, en esta sección comparamos el error absoluto de predicción para todos los predictores presentados anteriormente e incluimos, además, los predictores de kriging ordinario en dos versiones, a partir del variograma esférico (**KOsp**) y del variograma exponencial (**KOexp**), y el kriging universal (**KUexp**) con variograma exponencial cuya función media está dada por un polinomio de primer orden en las coordenadas.

En la Figura 3.10 graficamos el error de predicción comparando entre los dos sitios de prueba considerados anteriormente en el capítulo, esto es,  $s_0 = (4, 4)$  y  $s_0 = (8, 5)$ . En dichas figuras podemos observar en el extremo inferior y en línea punteada el verdadero

valor (el de prueba) de la variable de interés, en el extremo superior el valor predicho y en rojo el error. Podemos observar el comportamiento de cada predictor en ambos contextos. Recordemos que la mediana  $Med\tilde{Z}_{s_0}$  asociada al sitio  $s_0 = (4, 4)$  es representativa de la distribución de la variable  $Z(s_0)$ , en cambio la mediana  $Med\tilde{Z}_{s_0}$  asociada a  $s_0 = (8, 5)$  no, aunque ambos son frontera en la distribución de la muestra, como vemos en la Figura 3.3 (b). De este simple ejercicio evaluativo podemos ver que cuando la mediana del entorno del sitio donde queremos predecir es representativa de dicho valor, los métodos no paramétricos que utilizan tal información mejoran sustancialmente en la predicción (i.e. los métodos K2ME, K1A y K1M). Sin embargo, cuando este no es el caso, tales métodos pueden arrojar errores mayores a los clásicos paramétricos o al no paramétrico simple (K1S), como ocurre con K1M y K1A. No obstante, vemos que aun cuando la mediana deja de ser representativa, aquel que incluye las distancias entre las medianas de los vecindarios con dos funciones de núcleos (K2ME) da menores errores predictivos que los paramétricos y el no paramétrico de un núcleo simple. También observamos en tal situación, el método que incluye el índice de Moran, termina brindando una predicción con el menor error.

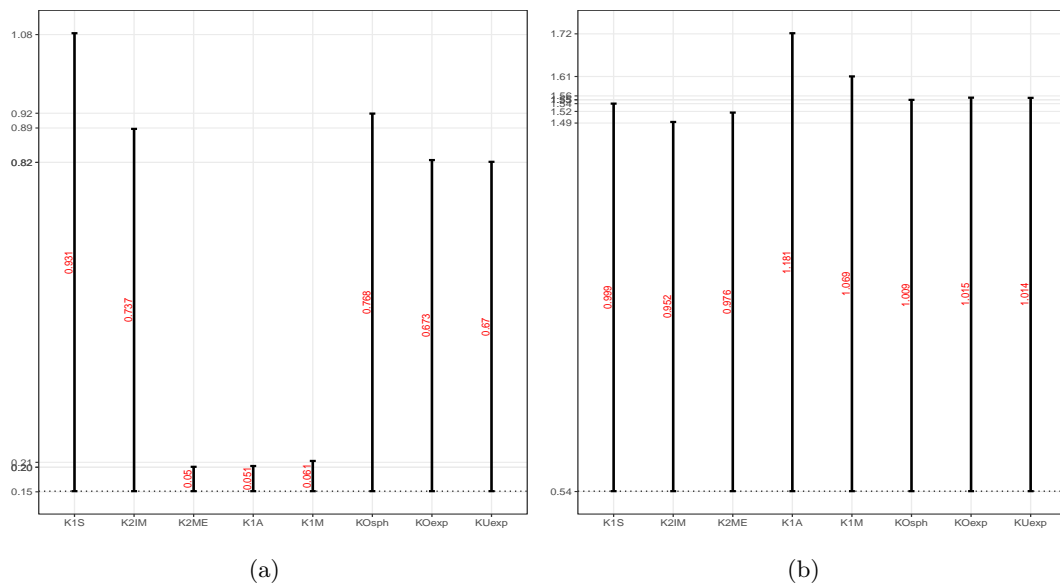


Figura 3.10: (a) Error absoluto de predicción para el sitio  $s_0 = (4, 4)$ , (b) Error absoluto de predicción para el sitio  $s_0 = (8, 5)$ .

### 3.4. Conclusiones

En este capítulo hemos desarrollado las diferentes propuestas metodológicas dentro de lo que consideramos versiones no paramétricas del kriging para la predicción espacial univariada. En esta sección vamos a explicitar algunas consideraciones sobre los predictores presentados en relación a otras propuestas no paramétricas encontradas en la literatura

como así también respecto a los kriging paramétricos. En primer lugar, todos los predictores propuestos en la tesis utilizan núcleos univariados evitando así problemas heredados de la *maldición de la dimensionalidad*. En otro orden, nuestra propuesta evita también la estimación de la función de regresión, y continuando en el espíritu del kriging paramétrico presentado en el Capítulo 2, proponemos pesos no paramétricos que contemplan en lo posible la mayor información contenida en los datos espaciales, aun en un contexto univariado. En este sentido, si bien los predictores presentados no fueron construídos teniendo en cuenta las propiedades de insesgadez y varianza mínima (aunque podrían verificarlas) como el estimador kriging, cuenta con varias ventajas:

- i. Evita el modelado del variograma y, como consecuencia, la elección del modelo a ajustar y la estimación de todos los parámetros involucrados en el mismo. En este punto se evitan, también, problemas relacionados con la forma de calcular los parámetros computacionalmente, esto es, problemas asociados a la convergencia o no de ciertos métodos numéricos.
- ii. Evita la inversión de la matriz  $\Gamma$  en (2.3.11). Esto se asocia a los llamados problemas mal condicionados (Tikhonov y Arsenin, 1977). En otras palabras, problemas que no tienen solución, la misma no es única o no depende continuamente de los datos. Esto, en muchos casos, se traduce en inestabilidad cuando se emplean métodos numéricos tradicionales para aproximar las soluciones.
- iii. No necesitamos hacer supuestos fuertes sobre el proceso subyacente a los datos para poder utilizarlo. Esto implica que, a diferencia del método kriging (Capítulo 2), los supuestos de estacionariedad no son necesarios para la predicción.
- iv. Como no hacemos supuestos *a priori* sobre el proceso que consideramos generador de los datos, nuestro predictor no depende del dominio sobre el cual fue medida la variable de interés. Esto significa que podemos aplicar el método a datos que provienen de procesos en dominio discreto como continuo.

## Capítulo 4

# Estudios de simulación

En este capítulo mostramos el desempeño de los métodos de predicción propuestos en el Capítulo 3 junto con los métodos Kriging descritos en el Capítulo 2. Más precisamente, a partir de estudios de simulación mostramos el comportamiento los métodos no paramétricos **K1S**, **K2IM**, **K2ME**, **K1M**, **K1A** y los paramétricos clásicos, Kriging Ordinario en dos versiones, a partir del variograma esférico, **KOsph**, y del variograma exponencial, **KOexp**, si los datos se simulan con el modelo de covarianza exponencial o variograma sinc, **KOsinc**, si los datos se simulan con el modelo de covarianza combinado. De igual manera para el Kriging Universal, **KUexp** y **KOsinc**, y en este caso debemos estimar la función media lo cual hacemos a partir de un polinomio de primer orden en las coordenadas.

### 4.1. Método de simulación

Dado  $n \in \mathbb{N}$ ,  $n > 1$  y un conjunto de localizaciones  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ , para simular una realización  $\mathbf{z} \doteq (z_1, \dots, z_n)$  de un vector aleatorio  $\mathbf{Z} \doteq (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$  con media  $\boldsymbol{\mu}$  y matriz de covarianza  $\boldsymbol{\Sigma}$ , utilizamos la descomposición de Cholesky de la matriz de covarianza (Cressie, 1993). Esto es, dada la esperanza

$$\mathbb{E}(\mathbf{Z}) = (\mathbb{E}(Z(\mathbf{s}_1)), \dots, \mathbb{E}(Z(\mathbf{s}_n)))^T \doteq \boldsymbol{\mu}, \quad (4.1.1)$$

y la varianza

$$\text{Var}(\mathbf{Z}) = C(\mathbf{s}_i, \mathbf{s}_j) = \boldsymbol{\Sigma},$$

donde  $\boldsymbol{\Sigma}$  es una matriz simétrica y definida positiva, utilizando la descomposición de Cholesky hallamos

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^T,$$

donde  $\mathbf{L}$  es una matriz de  $n \times n$ , triangular inferior. A partir de dicha matriz  $\mathbf{L}$  podemos escribir

$$\mathbf{Z} = \boldsymbol{\mu} + \mathbf{L}\boldsymbol{\epsilon}, \quad (4.1.2)$$

donde  $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \dots, \epsilon(\mathbf{s}_n))^T$  es un vector aleatorio de variables no correlacionadas e idénticamente distribuidas con esperanza  $\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}$  y  $Var(\boldsymbol{\epsilon}) = \mathbf{I}_n$  que nos asegura que  $\mathbb{E}(\mathbf{Z}) = \boldsymbol{\mu}$  y

$$Var(\mathbf{Z}) = Var(\mathbf{L}\boldsymbol{\epsilon}) = \mathbf{L}Var(\boldsymbol{\epsilon})\mathbf{L}^T = \boldsymbol{\Sigma}. \quad (4.1.3)$$

A partir de esta descomposición de  $\boldsymbol{\Sigma}$  y utilizando la Ecuación (4.1.2), consideraremos diferentes distribuciones para  $\boldsymbol{\epsilon}$  y diferentes modelos para la matriz de covarianza  $\boldsymbol{\Sigma}$ , de manera tal de generar diferentes escenarios para el proceso, los cuales serán detallados a continuación.

## 4.2. Escenarios

### 4.2.1. Generalidades

Para todos los casos generamos una grilla regular  $I_n$  de tamaño  $n = n_1 \cdot n_2 = 400$  sobre  $[0, 9]^2$ . Sobre la misma obtenemos la realización  $\mathbf{z}$  donde  $\boldsymbol{\Sigma}_{ij} = C(\mathbf{s}_i - \mathbf{s}_j)$ .

### 4.2.2. Con respecto a la función de covarianza

En la Sección 1.2.2 definimos diferentes funciones teóricas para el modelado del semivariograma y en el mismo capítulo probamos que si un proceso es estacionario de segundo orden podemos escribir la función de semivariograma a partir de la función de covarianza como  $\gamma(h) = C(0) - C(h)$  (Ecuación 1.1.7). Utilizando esta información y para construir la matriz de covarianza  $\boldsymbol{\Sigma}$ , consideramos diferentes modelos de covarianza isotrópicos  $C(h)$  (Sección 1.2.1), variando el valor de la meseta, el parámetro de rango y el efecto pepita,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ , respectivamente.

- Una de las funciones de covarianza elegida fue la que define el *modelo exponencial*  $C(h)$  dado en la Ecuación (1.2.13) a partir del cual, y utilizando la relación entre  $\gamma$  y  $C$  dada anteriormente, tenemos que,

$$C(h) = \theta_1 \exp\left(-\frac{h}{\theta_2}\right).$$

En este caso elegimos dos opciones de valores para el rango,  $\theta_2 = 2$  y  $\theta_2 = 100$ , en cuyo caso el rango efectivo está dado por 6 y 300, respectivamente. Recordemos que este último representa la distancia a partir de la cual la covarianza es nula. En la Figura 4.1 comparamos el comportamiento de la correlación (esto es  $\frac{C(h)}{C(0)}$ )

para cada caso y observamos que para el rango mayor, menor es la velocidad en la que disminuye la correlación en función de la distancia. Para distancias mayores a 6, cuando  $\theta_2 = 2$  la correlación es casi nula mientras que cuando  $\theta_2 = 100$  la correlación es poco variable y de magnitud cercana a 1.

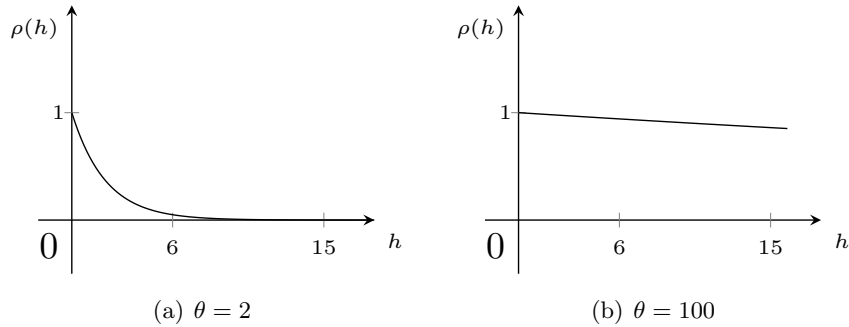


Figura 4.1: Función de correlación para el rango: (a)  $\theta_2 = 2$  ; (b)  $\theta_2 = 100$

- En segundo lugar elegimos un modelo combinado dado por

$$C(h) = \begin{cases} C_1(0) + C_2(0) + C_3(0), & \text{si } h = 0, \\ C_1(h) + C_2(h), & \text{si } 0 < h \leq \theta_{22}, \\ C_1(h), & \text{si } h > \theta_{22}, \end{cases}$$

donde el modelo  $C_1(h)$  es el *modelo seno cardinal* (sinc), un caso particular del modelo J-Bessel para el cual  $\alpha = \frac{1}{2}$ , el cual está dado por

$$C_1(h) = \begin{cases} \theta_{11} \left[ \left( \frac{2\theta_{21}}{h} \right)^{\frac{1}{2}} \Gamma\left(\frac{3}{2}\right) J_{\frac{1}{2}}\left(\frac{h}{\theta_{21}}\right) \right], & \text{si } h \neq 0, \\ 0, & \text{si } h = 0. \end{cases}$$

donde  $\theta_{ij}$  representa el parámetro  $\theta_i$  para el modelo  $j$ ,  $\Gamma$  es la función gamma y  $J_{1/2}$  es la función J-bessel de primer tipo de orden  $1/2$ ,

$$J_{1/2}(v) = \left(\frac{1}{2}\right)^2 \sum_{k=0}^{\infty} \frac{-1^k}{k! \Gamma(1/2 + k + 1)} \left(\frac{1}{2}\right)^{2k}.$$

El modelo  $C_2(h)$  es el modelo esférico (Ecuación 1.2.12) dado por

$$C_2(h) = \begin{cases} \theta_{12} \left[ 1 - \left( 1.5 \left(\frac{h}{\theta_{22}}\right) + 0.5 \left(\frac{h}{\theta_{22}}\right)^3 \right) \right], & \text{si } h \leq \theta_{22}, \\ \theta_{12}, & \text{si } h > \theta_{22}. \end{cases}$$

Y por último,  $C_3(h)$  es el modelo pepita definido en (1.2.4), el cual está dado por

$$C_3(h) = \begin{cases} \theta_{13}, & \text{si } h = 0, \\ 0, & \text{si } h \neq 0, \end{cases}$$

el cual modela un proceso no correlacionado espacialmente, ya que vale 0 si la distancia entre sitios es diferente de cero.

La idea de combinar modelos para la construcción de la matriz de covarianza (Montero et al., 2015, pág. 50) surge de suponer que el proceso subyacente de los datos es suma de procesos independientes que tienen diferentes niveles de asociación espacial. Para este modelo consideramos  $\theta_{21} = 0.2$  y  $\theta_{22} = 10$ .

Recordemos que en la Ecuación (1.1.6) definimos a la varianza del proceso como  $Var(Z(\mathbf{s})) = C(0)$  y, en la que la Ecuación 1.2.15, a la meseta como  $\theta_1 = C(0)$ , pues es el valor de la covarianza en  $h = 0$ . Para la elección de este parámetro consideramos dos casos: el *caso homocedástico* donde, para los dos modelos propuestos consideramos varianza constante para todo  $\mathbf{s}_i \in D$ , esto es,  $\theta_1 = \sigma^2 = 10$  y  $\theta_{11} = \theta_{12} = \theta_{13} = \sigma^2 = 10$ , respectivamente; y el *caso heterocedástico* donde la varianza varía en función del sitio donde fue medida la variable. En la Figura 4.2 graficamos, a partir de las coordenadas dadas, cuatro zonas donde cada color representa una varianza diferente con el fin de ilustrar cómo varía la misma en el espacio. Formalmente, sea  $\mathbf{s} = (s_1, s_2)$  la coordenada del sitio, luego

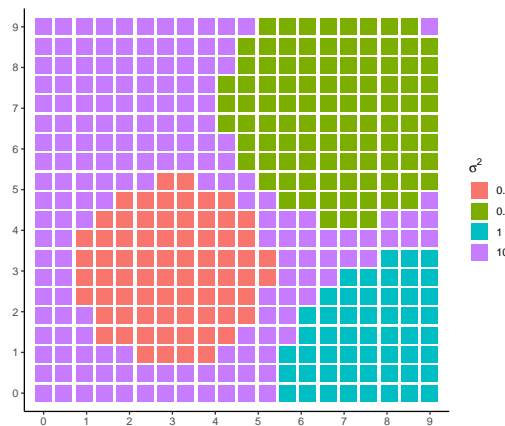


Figura 4.2: Caso heterocedástico: distribución de las varianzas en función de los sitios. En el caso homocedástico  $\sigma^2 = 10$  para todo  $\mathbf{s}_i \in D$  y para todos los modelos propuestos.

$$\sigma^2(\mathbf{s}) = \begin{cases} 0.1, & \text{si } s_1^2 + s_2^2 - 6s_1 - 6s_2 < -13, \\ 0.5, & \text{si } s_1^2 + s_2^2 - 14s_1 - 14s_2 < -90, \\ 1, & \text{si } s_1^2 + s_2^2 - 18s_1 < -68, \\ 10, & \text{en otro caso.} \end{cases}$$

En la Figura 4.3 graficamos dos realizaciones, una para el caso homocedástico (panel (a)) y otra para el caso heterocedástico (panel (b)). Comparando ambas figuras podemos observar la diferencia entre los valores que toma el proceso en cada caso y, además, cómo en el caso heterocedástico, las zonas de cambio de varianza están bien reflejadas por dichos valores de la correspondiente realización.

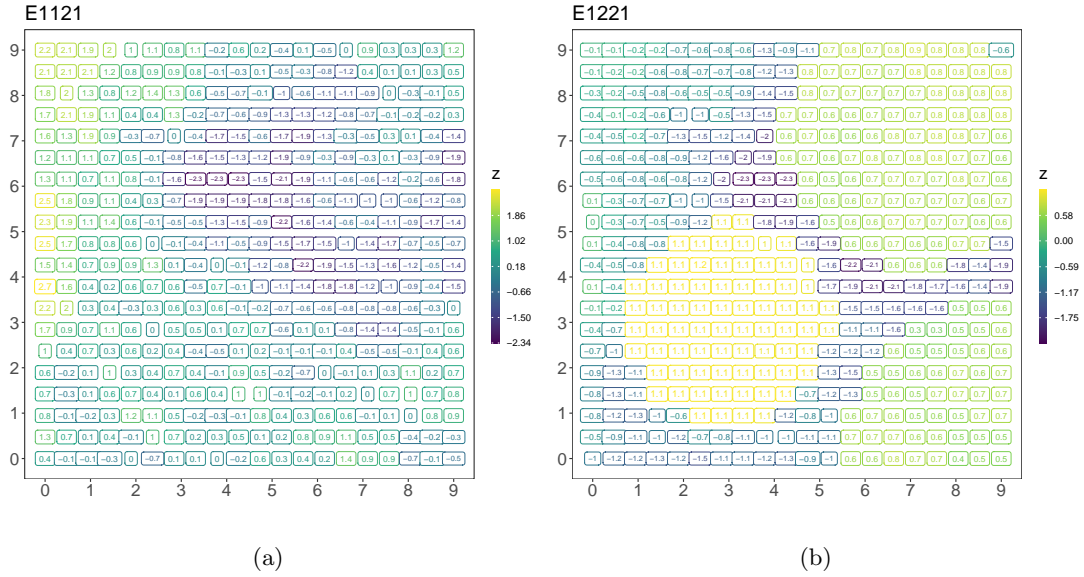


Figura 4.3: Realización de un proceso con modelo de covarianza exponencial y rango  $\theta_2 = 100$ : (a) Caso homocedástico; (b) Caso heterocedástico.

#### 4.2.3. Con respecto al vector aleatorio $\epsilon$

Para el vector  $\epsilon$  de variables no correlacionadas e idénticamente distribuidas consideramos diferentes distribuciones, las cuales se detallan a continuación:

- Distribución normal  $\epsilon \sim N(\mathbf{0}, \mathbf{I}_n)$ : en este caso, a partir de las Ecuaciones (4.1.2) y (4.1.3) tenemos que  $\mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .
- Distribución Bernoulli dada por:

$$\epsilon = \begin{cases} -10, & \text{con probabilidad } \frac{1}{2}, \\ 10, & \text{con probabilidad } \frac{1}{2}, \end{cases}$$

con  $\mathbb{E}[\epsilon] = \mathbf{0}$  y  $\text{Var}(\epsilon) = 100$ , lo cual implica que  $\text{Var}(\mathbf{Z}) = 100\boldsymbol{\Sigma}$  a partir de la cual se construye la realización  $\mathbf{Z}$ . Además sabemos que, a partir de la forma en al que construimos al vector  $\mathbf{Z}$  y usando el Teorema de Lindeberg-Feller, para cada elemento  $z_n$  de  $\mathbf{Z}$ , se cumple que  $z_n/\text{Var}(z_n) \xrightarrow{D} N(0, 1)$  cuando  $n \rightarrow \infty$ .

- Distribución  $t$ -student no centrada con  $\nu = 5$  grados de libertad y parámetro de no centralidad  $\delta = 2$ . Dado que  $\nu > 2$  tenemos que, para cada elemento  $\epsilon_n$  de  $\epsilon$ ,

$$\mathbb{E}[\epsilon_n] = \delta \sqrt{\frac{\nu}{2}} \Gamma\left(\frac{\nu-1}{2}\right) \Gamma^{-1}\left(\frac{\nu}{2}\right),$$

y

$$\sigma_{\epsilon_n}^2 = \frac{\nu(1+\delta^2)}{\nu-2} - \frac{\delta^2\nu}{2} \left( \Gamma\left(\frac{\nu-1}{2}\right) \Gamma^{-1}\left(\frac{\nu}{2}\right) \right)^2,$$

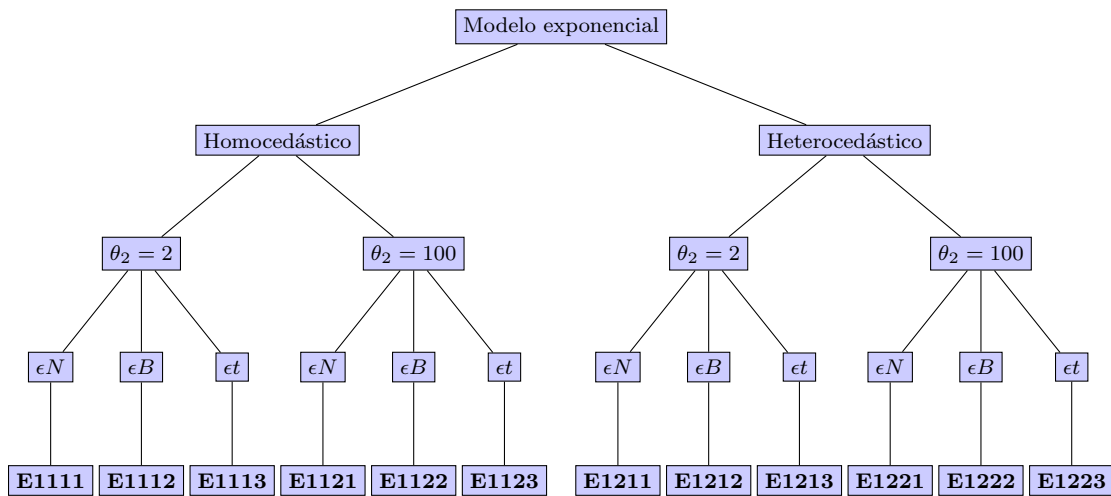


Figura 4.4: Combinación de parámetros considerados en cada escenario partiendo del modelo de covarianza exponencial.

donde  $\Gamma$  es la función gamma. En este caso  $\mathbb{E}[\mathbf{Z}] = \mathbb{E}[\boldsymbol{\epsilon}]$  y  $Var(\mathbf{Z}) = 2.67\boldsymbol{\Sigma}$  ya que  $\sigma_{\epsilon_n}^2 \approx 2.67$  considerando los parámetros dados. De la misma manera que en el caso anterior, en este caso se cumple que  $(z_n - \mathbb{E}[\epsilon_n]) / Var(z_n) \xrightarrow{D} N(0, 1)$ .

En las Figuras 4.4 y 4.5 mostramos un resumen de los escenarios propuestos para mostrar el desempeño de los predictores presentados, paramétricos y no paramétricos.

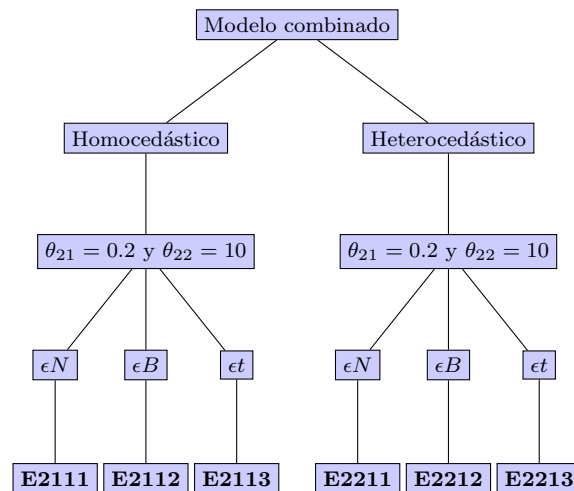


Figura 4.5: Combinación de parámetros considerados en cada escenario partiendo del modelo de covarianza combinado.

### 4.3. Estimación de parámetros y validación cruzada

Para estimar la variabilidad de los errores reportados en la Sección 4.4, del conjunto original de  $n$  observaciones construimos, *aleatoriamente*, dos muestras: la muestra de entrenamien-

to (70%) y la de prueba (30%). Con la muestra de entrenamiento estimamos todos los parámetros involucrados en los predictores, tanto paramétricos como no paramétricos, y con la muestra de prueba medimos el desempeño de los métodos calculando el error de predicción de los mismos. Dicho error será medido utilizando el **Error Medio Absoluto** (MAE por sus siglas en inglés) dado por

$$MAE = \frac{\sum_{j=1}^n |z_j - \hat{z}_j|}{n}, \quad (4.3.1)$$

donde  $\hat{z}_j$  es el valor de la predicción en sitio  $\mathbf{s}_j$ .

Para los predictores no paramétricos necesitamos estimar los parámetros de suavizado y la cantidad de vecinos más cercanos  $k$ . Para ello utilizamos la técnica de **validación cruzada de k-iteraciones**, (*k-fold cross validation, en inglés*) en la muestra de entrenamiento. Esto es, para un conjunto de posibles parámetros, dividimos la muestra de entrenamiento aleatoriamente en  $k$  grupos, tratamos al primer grupo como un conjunto de validación y entrenamos los métodos con los  $k - 1$  restantes. Luego calculamos el  $MAE_i$  sobre las observaciones del conjunto de validación. Repetimos este procedimiento  $k$  veces; cada vez, tratamos a un grupo diferente de observaciones como conjunto de validación. Este proceso da como resultado  $k$  estimaciones del error de predicción,  $MAE_1, \dots, MAE_k$ . Luego, la estimación del error por validación cruzada está dada por

$$ECVK = \frac{1}{k} \sum_{i=1}^k MAE_i.$$

Finalmente, elegimos los parámetros óptimos como aquellos que minimizan el  $ECVK$ . Para los parámetros de suavizado establecemos 40 valores y para el caso del parámetro cantidad de vecinos  $k$  establecimos solo dos opciones, 8 y 12 vecinos.

Para los predictores kriging es necesario estimar los parámetros del variograma teórico y para ello es necesario contar con el variograma empírico (1.2.5) el cual calculamos con la muestra de entrenamiento a partir de los siguientes datos:

- Distancia máxima  $D$ , hasta la cual se incluyen los pares de puntos en la estimación. Elegimos utilizar la mitad de la distancia máxima entre sitios.
- La cantidad de intervalos  $B$ . Construimos una secuencia de  $B$  números que toma como mínimo el 0 y máximo  $D$ , estos números representan los centros de cada intervalo y los extremos de los mismos quedan determinado por los puntos medios entre los centros. Elegimos utilizar 13 intervalos.
- El número mínimo de pares  $P$  para cada intervalo de distancias. Ignoramos los intervalos con un número de pares menor que este valor. En este caso  $P = 30$ .

Con esto obtuvimos 12 valores (ya que el primero no llega a la cantidad mínima de pares de puntos para su construcción) del variograma empírico  $\hat{\gamma}_j(h)$  (1.2.5) para cada  $h_j$ , los

cuales se utilizan para ajustar el semivariograma teórico. Para el caso del kriging ordinario realizamos la predicción para dos modelos de covarianza: el modelo esférico y el exponencial o el modelo esférico y el sinc según si los datos son generados a partir del modelo exponencial o el modelo combinado respectivamente. De la misma manera, para el caso del kriging universal ajustamos los parámetros para el modelo exponencial o sinc según cómo se simularon los datos. El método de estimación utilizado (Ver Sección 1.2.2.1) es mínimos cuadrados pesados (1.2.17) donde los pesos están dados por  $\frac{N(\mathbf{s}_j)}{h_j^2}$ .

Una vez que los parámetros son estimados, sobre los sitios de la muestra de prueba y utilizando la muestra de entrenamiento, realizamos la predicción y calculamos el error  $MAE$ . Repetimos este proceso 50 veces, simulando cada vez una realización diferente y obteniendo 50 errores de predicción.

En general, cuando separamos el conjunto de los datos para aplicar métodos de remuestreo y estimar el error, suponemos que los grupos obtenidos son independientes (Hastie et al., 2001). Por ello algunos autores (Rest et al., 2014; Roberts et al., 2017; Ploton et al., 2020) proponen técnicas de remuestreos espaciales para el caso de datos espaciales. Nosotros decidimos adoptar los resultados de Wadoux et al. (2021) quien muestra que, las estrategias de validación cruzada espacial tienen un desempeño deficiente y sobreestiman el error en todos los diseños de muestreo considerados en el estudio del caso presentado, por ello decidimos usar las técnicas de remuestreo clásicas.

## 4.4. Resultados

En esta sección reportamos los errores de predicción obtenidos a partir del  $MAE$ , definido en la Ecuación 4.3.1, para los 18 escenarios presentados en la Sección 4.2.

Comenzamos comparando entre procesos homocedásticos y heterocedásticos para las diferentes distribuciones del vector  $\epsilon$ . En las Figuras 4.6, 4.7 y 4.8 observamos los errores medios absolutos para las realizaciones de los procesos generados con modelos de covarianza exponencial con rangos  $\theta_2 = 2$  y  $\theta_2 = 100$  y el modelo combinado, respectivamente. En cada uno de los gráficos reportamos, además, la media y el desvío estándar (entre paréntesis) del  $MAE$ . Es importante notar que las escalas varían entre gráficos.

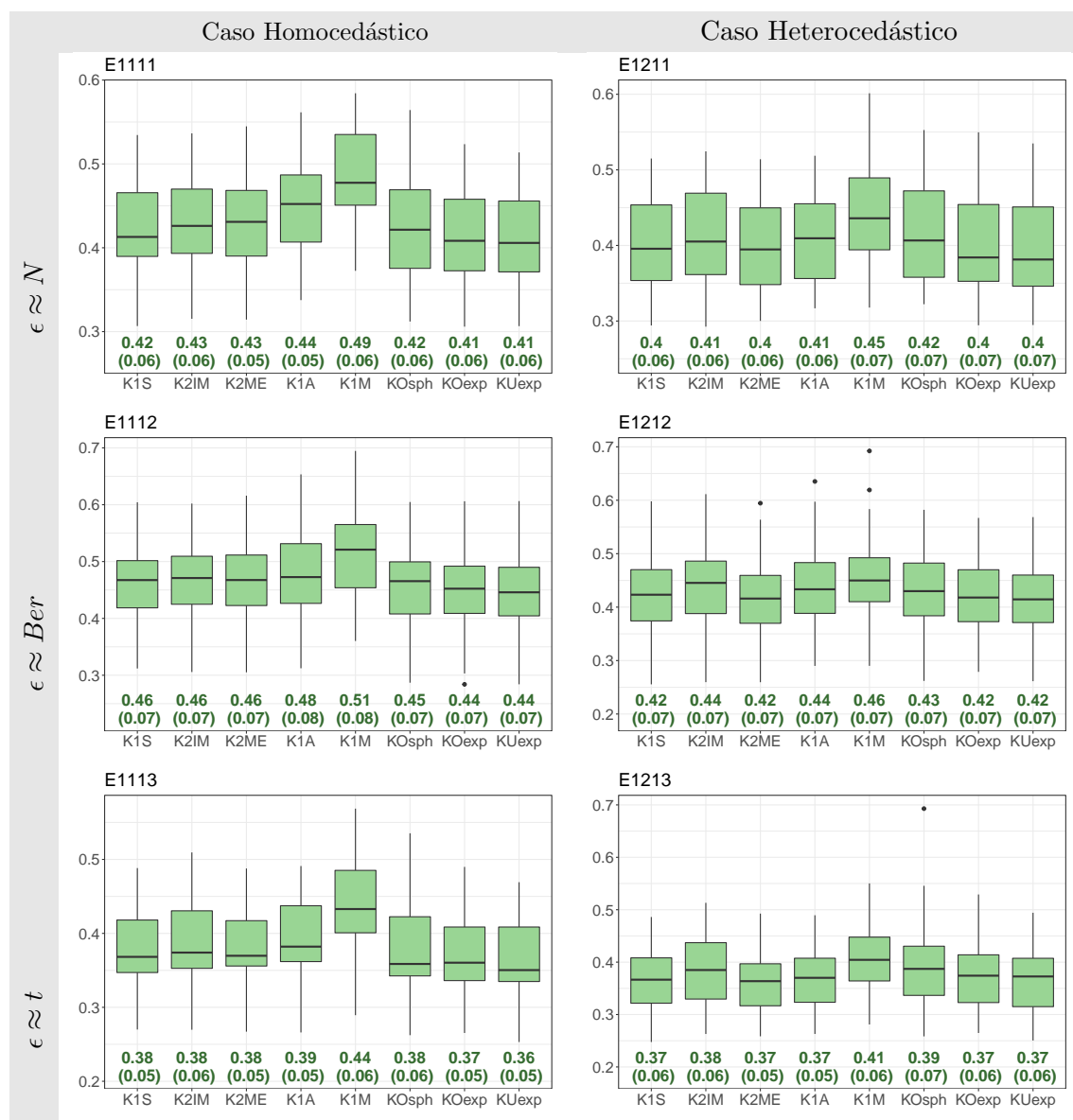


Figura 4.6: Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza exponencial y rango  $\theta_2 = 2$ . Desde arriba hacia abajo, los tres modelos considerados para el error  $\epsilon$  y de izquierda a derecha, los casos homocedásticos y heterocedásticos.

Para obtener estimaciones del error de predicción decimos realizar un remuestreo anidado. Esto quiere decir que todas las partes de la construcción del modelo (preprocesamiento y estimación de parámetros) deben incluirse en el remuestreo, es decir, repetirse para cada división entre entrenamiento y prueba. Decidimos utilizar el método **hold-out** para el remuestreo externo y **k-fold** para el interno, el cual explicamos en la Sección 4.3.

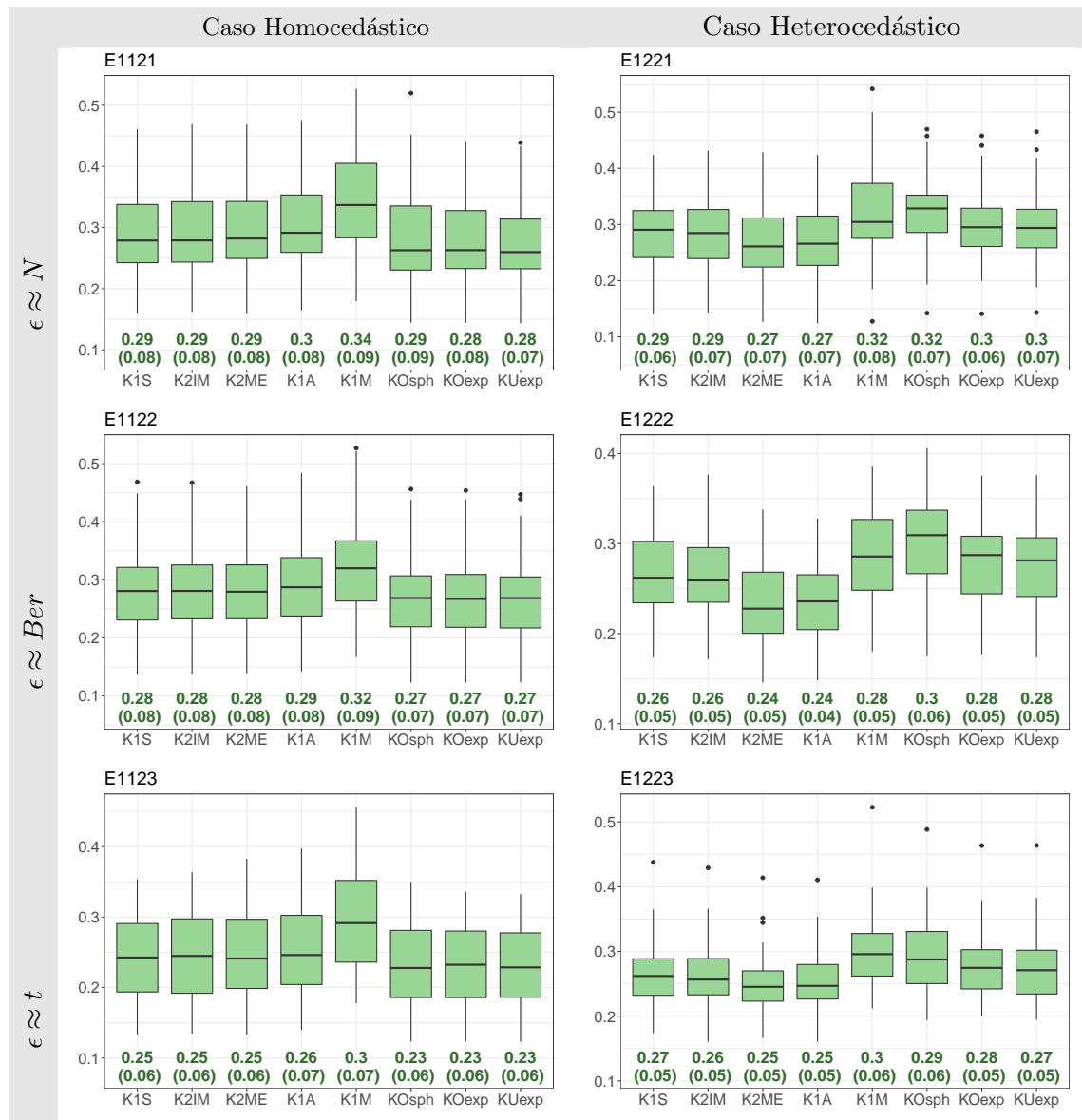


Figura 4.7: Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza exponencial y rango  $\theta_2 = 100$ . Desde arriba hacia abajo, los tres modelos considerados para el error  $\epsilon$  y de izquierda a derecha, los casos homocedásticos y heterocedásticos.

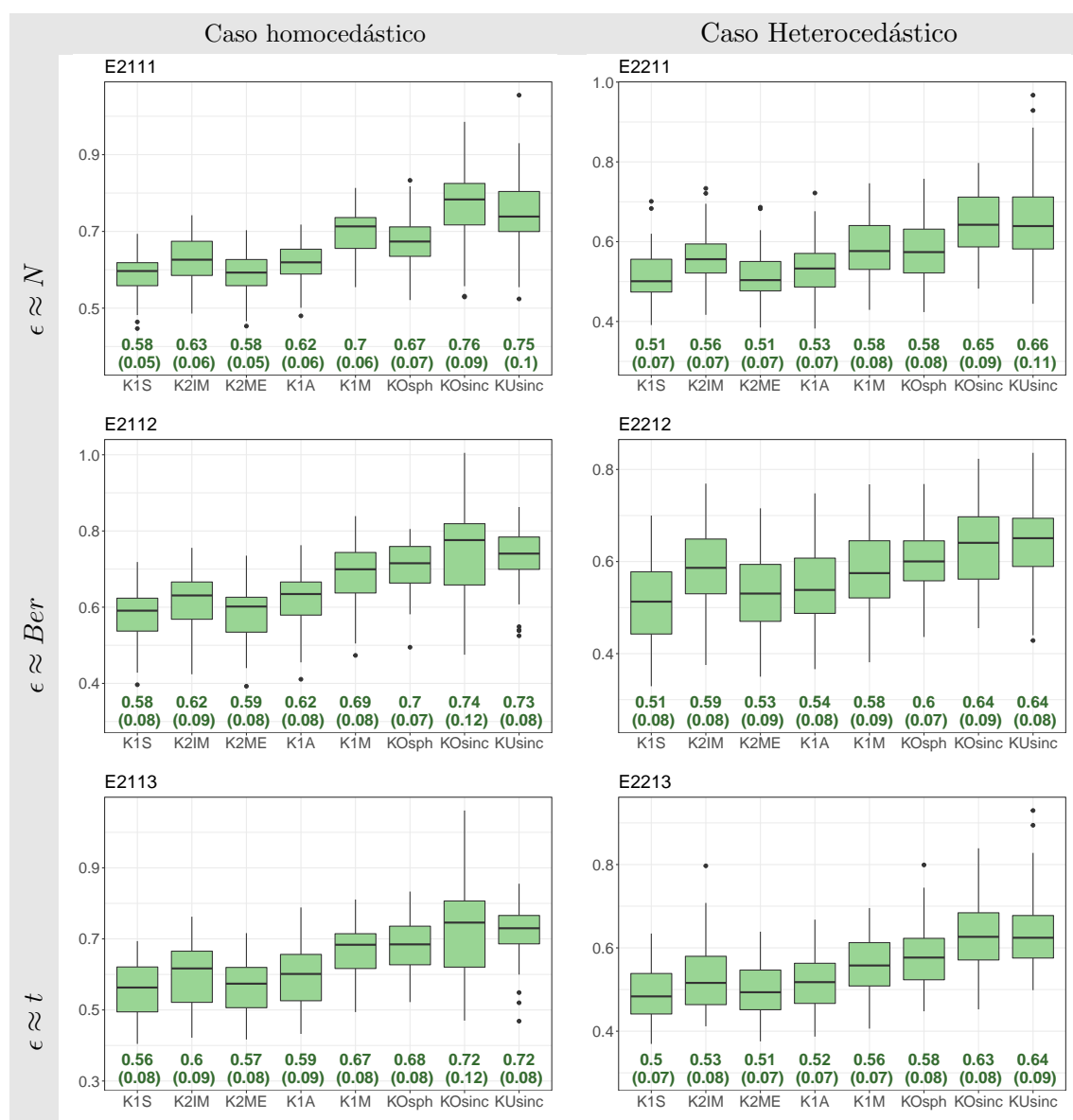


Figura 4.8: Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza combinada y rangos  $\theta_{21} = 0.2$  y  $\theta_{22} = 10$ . Desde arriba hacia abajo, los tres modelos considerados para el error  $\epsilon$  y de izquierda a derecha, los casos homocedásticos y heterocedásticos.

#### 4.4.1. Con respecto a la función de covarianza

Cuando los datos son generados con un modelo de covarianza exponencial (Figuras 4.6 y 4.7), los kriging paramétricos muestran un mejor desempeño que los no paramétricos, y dicha brecha es mayor para el caso homocedástico. Para el heterocedástico, los métodos no paramétricos muestran un desempeño más competitivo, y de hecho, en algunos casos, arrojan menores errores medios que los paramétricos. En particular esto último ocurre, cuando el  $\epsilon$  no es gaussiano y el rango es  $\theta_2 = 100$  (Figura 4.7).

Cuando utilizamos un modelo de covarianza combinado, en todos los casos obtenemos un mejor desempeño con los predictores no paramétricos propuestos en la tesis. Esta mejora en las predicciones resulta razonable dada la mayor flexibilidad de este enfoque respecto a la especificación del variograma en la estimación. En otras palabras, dado que los predictores paramétricos demandan la especificación de la estructura de covarianza para la estimación del variograma, la generación de datos resultante en este caso podría complicar la predicción del estimador al no captar la naturaleza mixta de la covarianza subyacente en los datos mientras que los enfoques no paramétricos no necesitan de ningún modelo para ello, y esta flexibilidad puede interpretarse como una ventaja ante la existencia de datos con tales características.

Como ya mencionamos en la Sección 1.2.2, la elección del modelo de covarianza a estimar es la decisión que más influye en la predicción paramétrica. Por ello, si usamos un modelo de covarianza exponencial para la simulación de los datos, estamos proponiendo un escenario que favorece a los predictores kriging clásicos. Específicamente, estos predictores funcionan mejor si el modelo de variograma ajustado coincide con el modelo de covarianza utilizado en la simulación de los datos. En cada uno de los escenarios representados en las Figuras 4.6 y 4.7 observamos que, comparando entre los kriging paramétricos, los resultados mejoran levemente si utilizamos el modelo exponencial para ajustar el variograma. En la Figura 4.8 observamos la mejora para el modelo de variograma esférico, aunque en el escenario E2113 observamos mejores resultados en cuanto a la variabilidad de los errores, considerando el modelo sinc.

Respecto al rango, en todos los casos observamos que cuando este aumenta, los errores de predicción disminuyen. Esto puede verse más puntualmente cuando comparamos los escenarios construidos a partir de un modelo exponencial de rango  $\theta_2 = 2$  y de rango  $\theta_2 = 100$  (Figuras 4.6 y 4.7) dejando fijas el resto de las variables involucradas en la simulación de los datos; esto es, comparamos el escenario E1111 con el E1121, el E1211 con el E1221, y así sucesivamente.

#### 4.4.2. Con respecto a la heterogeneidad de la varianza

El supuesto de datos homocedásticos puede ser muy fuerte o restrictivo en algunas aplicaciones. Por ejemplo, si estamos buscando predecir precios de viviendas o de lotes, resulta poco realista asumir que los precios en cuestión tendrán la misma varianza en las diferentes localizaciones, sea en un región donde habitan predominantemente individuos u hogares ricos como en aquellas donde lo hacen los más pobres. Existen numerosos contextos donde la existencia de varianza no constante caracteriza mejor a los datos, y es por ello que hemos puesto énfasis en tal característica, generando todos los escenarios tanto en contexto de datos homocedásticos como heterocedásticos.

En general observamos que los predictores no paramétricos son más competitivos fun-

cionando mejor cuando los datos espaciales son heterocedásticos. En particular esto se observa en el comportamiento de los errores del predictor de dos núcleos (K2ME) y el de un núcleo aditivo (K1A). Incluso, estas versiones no paramétricas terminan siendo los mejores predictores (i.e. de menor MAE promedio) en varias situaciones

Algo que también podemos apreciar, es que para todos los métodos, sean paramétricos o no paramétricos, en el caso heterocedástico la precisión de las predicciones es algo menor, en el sentido de que se revela una mayor varianza en el MAE resultante de cada uno de los experimentos (i.e. variando modelos de covarianza, distribución del error y rango).

#### 4.4.3. Con respecto al vector aleatorio $\epsilon$

Observando las Figuras 4.6, 4.7 y 4.8 desde arriba hacia abajo, podemos observar que el comportamiento de los errores de predicción es similar para cada una de las tres distribuciones elegidas para el vector aleatorio  $\epsilon$ . Esto puede deberse a que, tal como lo mencionamos en la Sección 4.2.3, dada la estructura del modelo (4.1.2), los elementos de dicho vector convergen a una variable con distribución normal cuando el tamaño de la muestra es suficientemente grande. Por ende podemos concluir que, dado el tamaño de la muestra que estamos simulando, la distribución del error no sería un supuesto crucial a fin de ponderar positivamente la utilización de predictores no paramétricos sobre los kriging paramétricos. De hecho en algunos casos podemos encontrar menores errores de predicción cuando el error no es normal y sigue una Distribución Bernoulli o t-Student, como es en el caso homocedástico con modelo covarianza exponencial y rango de 100.

## 4.5. Conclusiones

En el presente capítulo hemos evaluado el desempeño predictivo de los diferentes kriging paramétricos y de las alternativas no paramétricas propuestas en la tesis, con un ejercicio de simulación, contemplando diferentes escenarios factibles determinados por el modelo de covarianza, el rango, la distribución del error y la varianza del proceso.

Al comparar entre los predictores paramétricos *versus* no paramétricos, podemos concluir que en los escenarios favorables para los kriging paramétricos (esto es, en el caso homocedástico y con modelos de covarianza exponencial no combinada) estos terminan siendo un mejor predictor en términos de un menor error. En tales casos, el kriging universal muestra el mejor desempeño frente a las otras dos opciones de kriging ordinario.

Ahora bien, cuando los datos simulados son heterocedásticos, los predictores no paramétricos son más competitivos, mostrando en algunos casos mejores resultados que los kriging paramétricos. La mayor ventaja de usar predictores no paramétricos se revela en el caso

de un proceso con covarianza combinada (sea homocedástico o heterocedástico), donde la mayoría de las alternativas no paramétricas superan a los kriging paramétricos. A su vez, en tal escenario de covarianza combinada, el kriging ordinario muestra un mejor desempeño predictivo que el universal.

Entre los no paramétricos, el predictor K1M es el que muestra el peor desempeño y los que vislumbran mejor desempeño son los predictores K2ME y K1A. Todos los predictores no paramétricos contienen, de alguna manera las distancias entre sitios y con respecto a esto, un resultado que podemos observar es que son pocos los casos en los que los predictores propuestos funcionan mejor que el que solo considera las distancias entre sitios (i.e. el K1S), estos son los escenarios E1222 y E1223. Por lo tanto, dependiendo de cómo sean los datos en términos de la heterogeneidad de la varianza entre localizaciones, si esta es relativamente constante, el estimador simple de un núcleo (K1S) puede ser una gran elección para la predicción de la variable espacial de interés; en caso donde se observe que este supuesto de varianza constante difícilmente se cumpla, el estimador de dos núcleos K2ME o el de un núcleo aditivo K1A serían una buena opción, donde la incorporación entre la distancia del valor (mediano) de la variable en un entorno del punto a predecir pareciera jugar un rol relevante en predicción bajo escenarios de heterocedasticidad.

En definitiva, podemos concluir que los predictores clásicos funcionan peor en datos provenientes de procesos donde exista heterogeneidad, esto es, donde la varianza no es constante en el espacio y donde existan diferentes formas de asociación espacial, y a diferentes escalas, como en el caso del procesos simulado a partir de una función de covarianza combinada. Y aquí es donde nuestra propuesta no paramétrica cobra sentido como una alternativa superadora.

## Capítulo 5

# Aplicación a datos reales

En este capítulo aplicamos los métodos propuestos a datos reales comparando su desempeño entre ellos y con el de los predictores paramétricos clásicos. Para la aplicación nos hemos enfocado en el problema de predicción de precios de viviendas. Esta problemática es relevante tanto para la toma de decisiones en el ámbito privado (e.g. para los agentes del mercado inmobiliario o turístico), como en el ámbito del sector público, dado que, por ejemplo, un mapa de precios de inmuebles permitiría contar con una descripción de la distribución geográfica o espacial de una característica socioeconómica crucial, que puede usarse para el diseño e implementación de políticas públicas de hábitat, vivienda, infraestructura y/o desarrollo urbanístico, entre otras. Específicamente, trabajamos con tres grupos de datos, los cuales describimos y analizamos en las siguientes secciones. Además de comparar el desempeño de cada predictor presentamos mapas predictivos utilizando las estimaciones obtenidas.

### 5.1. Presentación de los datos

En primer lugar trabajamos con datos obtenidos de la competencia *Valuación de inmuebles* de 2018 en la sección *Metadata* de Fundación Sadosky (Metadata, 2018). El problema fue propuesto por el grupo de I+D de la empresa Navent en base a su experiencia con ZonaProp. Navent es una empresa que se dedica a que sus usuarios consigan empleos y hogares de la mejor manera posible. En el problema original se trata de predecir el precio de inmuebles a partir de variables como metros totales de la vivienda, antigüedad, cantidad de habitaciones, entre otras. Contamos además con la ubicación del inmueble, dada por su latitud y longitud. Puesto que vamos a evaluar métodos univariados, supondremos que además de la ubicación solo contamos con información de la variable de interés, en este caso es el *precio del inmueble*. Usando esta información queremos predecir precios en otros sitios de interés.

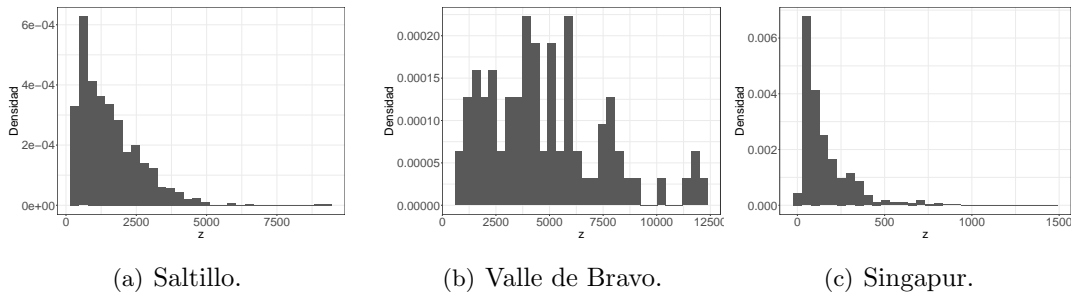


Figura 5.1: Distribución de la variable de interés sin considerar el espacio. Saltillo y Valle de Bravo en miles de pesos mexicanos y Singapur en Dólar Singapur.

	Mínimo	Q1	Mediana	Q3	Máximo	Media	Desvío	n
<b>Saltillo</b>	310	700.0	1300	2177.5	9300	1551.61	1088.09	967
<b>Valle de Bravo</b>	600	2737.5	4325	6121.5	12000	4802.93	2744.30	80
<b>Singapur</b>	14	60.0	108	200.0	1481	157.53	150.78	1326

Tabla 5.1: Resumen numérico para cada conjunto de datos. Saltillo y Valle de Bravo en miles de pesos mexicanos y Singapur en Dólar Singapurense.

Si bien el conjunto de datos original contiene observaciones de todo el país, por cuestiones de cómputo decidimos utilizar y analizar por separado los datos de la ciudad de Saltillo y la ciudad de Valle de Bravo. **Saltillo** es la cabecera del municipio que lleva el mismo nombre, el cual se localiza en el sureste del estado de Coahuila, a una altura de 1600 metros sobre el nivel del mar (coahuila.gob, 2020). Esta base contiene 967 observaciones. **Valle de Bravo** es una ciudad turística, ubicada a orillas del Lago Avándaro, al oeste de Ciudad de México. El lago está rodeado de montañas y bosque y es elegido como centro de deportes acuáticos. Este conjunto de datos es de menor tamaño y cuenta solo con 80 observaciones.

En segundo lugar trabajamos con un conjunto de datos que cuenta con información de alojamientos en la ciudad de **Singapur**, publicados en el sitio web Airbnb. Dicha plataforma digital es un medio para la oferta de alojamientos particulares, en su mayoría con fines turísticos, mediante la cual los anfitriones pueden publicitar sus propiedades. Estos datos fueron obtenidos del sitio web Kaggle<sup>1</sup>. La base consta de varias variables dentro de las cuales encontramos: tipo de hospedaje, número de comentarios, fecha del último comentario, disponibilidad, mínimo de noches, fecha del último comentario, precio por noche y la ubicación a partir de la longitud y latitud de la vivienda. En este caso la variable de interés es *el precio por noche de cada hospedaje* y la base de datos cuenta con 1326 observaciones.

<sup>1</sup><https://www.kaggle.com/datasets/jojoker/singapore-airbnb>

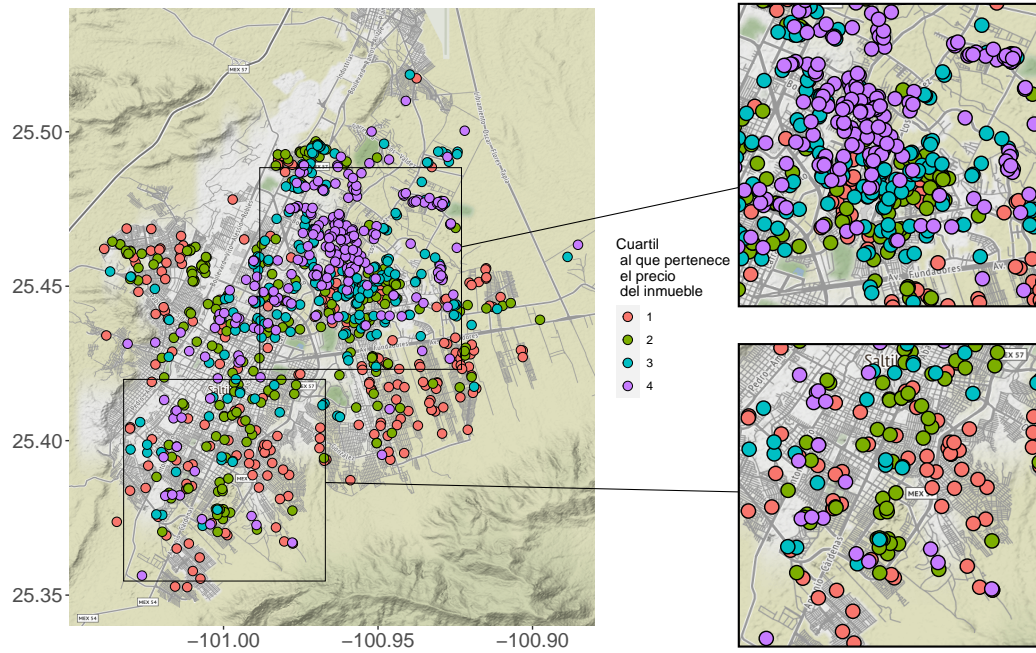


Figura 5.2: Mapa de caja: precio de inmuebles en Saltillo.

En la Figura 5.1 y en la Tabla 5.1 presentamos la distribución de la variable de interés de cada uno de los conjunto de datos elegidos y un resumen numérico respectivamente. En esta figura notamos una pronunciada asimetría a derecha tanto en los datos de Saltillo como en los de Singapur.

## 5.2. Precios de Inmuebles: Saltillo

### 5.2.1. Análisis exploratorio

En la Sección 1.3 planteamos la importancia del análisis exploratorio de datos como primera herramienta para conocer y resumir los datos de la muestra, por ello en esta sección analizamos los datos exploratoriamente. Comenzamos con la Figura 5.2 en la cual graficamos la distribución del precio del inmueble (por cuartiles) en los 967 sitios de la muestra. Como podemos observar en esta figura, la concentración de los inmuebles más caros está en el centro norte de la ciudad y los menos costosos en la periferia de la misma, de lo cual inferimos la forma acampanada de la superficie que representa.

En la Figura 5.3 observamos cómo varían los precios con respecto a cada una de las coordenadas. Vemos que los mismos tienen una distribución moderadamente simétrica en

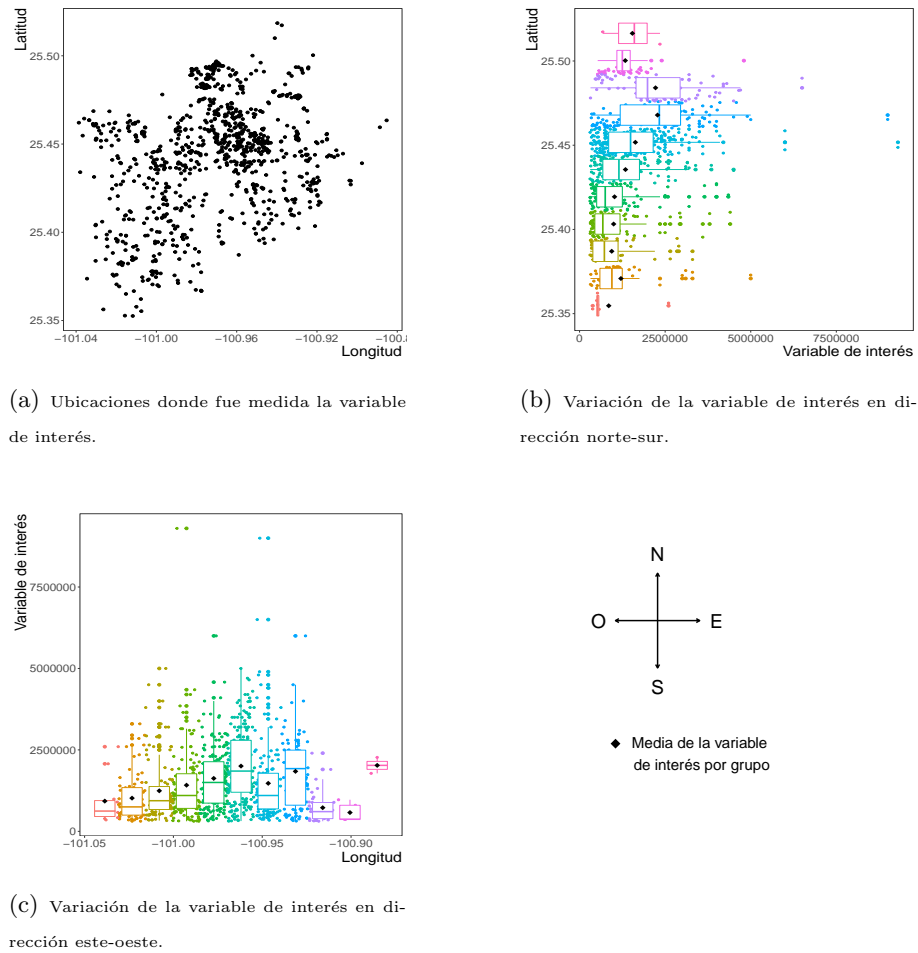


Figura 5.3: Datos Saltillo: distribución de la variable de interés en función las coordenadas geográficas.

sentido oeste-este. En otras palabras, los precios tienden a decrecer en dirección oeste y en el centro observamos mayores precios pero también mayor variabilidad. Por otro lado, en el sentido norte-sur la distribución está sesgada hacia precios mayores y encontramos una concentración de precios elevados al norte de la ciudad donde también parecería aumentar la variabilidad. La heterogeneidad de la varianza nos muestra indicios de incumplimiento del supuesto de estacionariedad; esto es, aunque no observemos una tendencia en la media sí vemos pistas de varianza no constante.

Para visualizar la dependencia espacial, en la Figura 5.4 realizamos un *gráfico de dispersión  $h$*  en el cual podemos observar cómo varía la autocorrelación a medida que consideramos mayores distancias entre los sitios. Dado que en los primeros gráficos observamos a los puntos más cerca de la recta estimada y una pendiente más pronunciada de la misma, inferimos que la presencia de dependencia espacial está presente con mayor fuerza para distancias pequeñas.

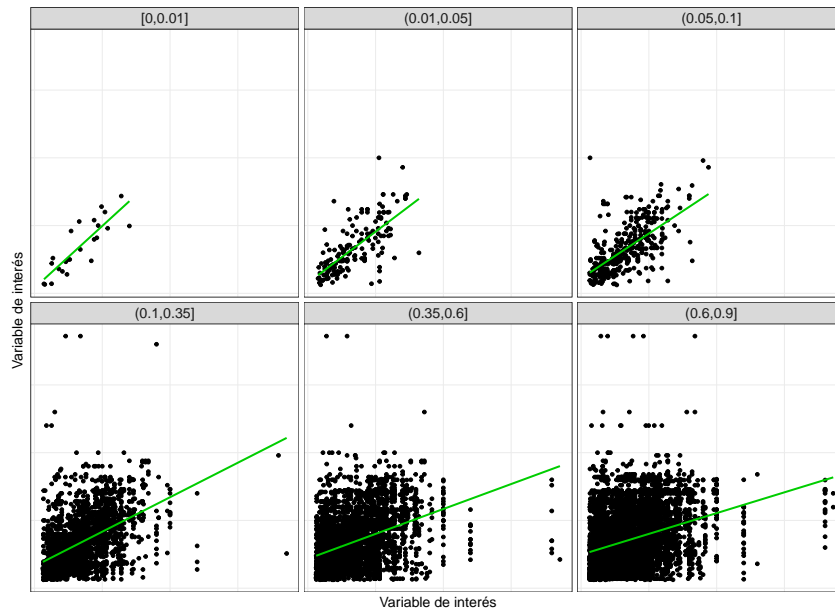


Figura 5.4: Gráfico de dispersión  $h$  del precio de inmuebles en Saltillo, México.

### 5.2.2. Resultados

Los métodos y procedimientos utilizados, tanto para la estimación del error como para la estimación de los parámetros, son los mismos que utilizamos en los estudios de simulación descritos en el capítulo anterior (Sección 4.3). En este caso también realizamos 50 repeticiones y dividimos la muestra original, aleatoriamente, en muestra de entrenamiento y muestra de prueba.

En la Figura 5.5 graficamos la distribución del  $MAE$  calculado sobre la muestra de prueba, para cada una de las 50 repeticiones y para cada método. En la Tabla 5.2 reportamos el promedio, sobre las 50 repeticiones, del  $MAE$ , de la correlación entre los valores predichos y observados (**cor\_ospred**) y de la Raíz cuadrada del Error Cuadrático Medio (**RMSE**, por su siglas en inglés) para cada método. Como podemos observar los valores obtenidos son similares, tanto en  $MAE$  promedio como en precisión, lo que nos indica que los predictores no paramétricos son competitivos sin necesidad de hacer supuestos previos como en los paramétricos. Dentro de los predictores no paramétricos, el que presenta un mejor desempeño es el incluye en los pesos el índice de Moran (**K2IM**).

En la Figura 5.6 construimos un mapa predictivo para cada método. Cabe aclarar que, para todos los conjuntos de datos considerados las predicciones fueron estandarizados con fines comparativos y el panel donde solo aparecen puntos graficamos los datos de muestra disponible también estandarizados. En general las predicciones son similares, variando el grado de suavidad de las mismas. Podemos ver que los predictores no paramétricos en general coinciden en detectar un cambio de precios (aumento) en la región del extremo suroeste, mientras que los mapas de los kriging paramétricos predicen una mayor uniformidad

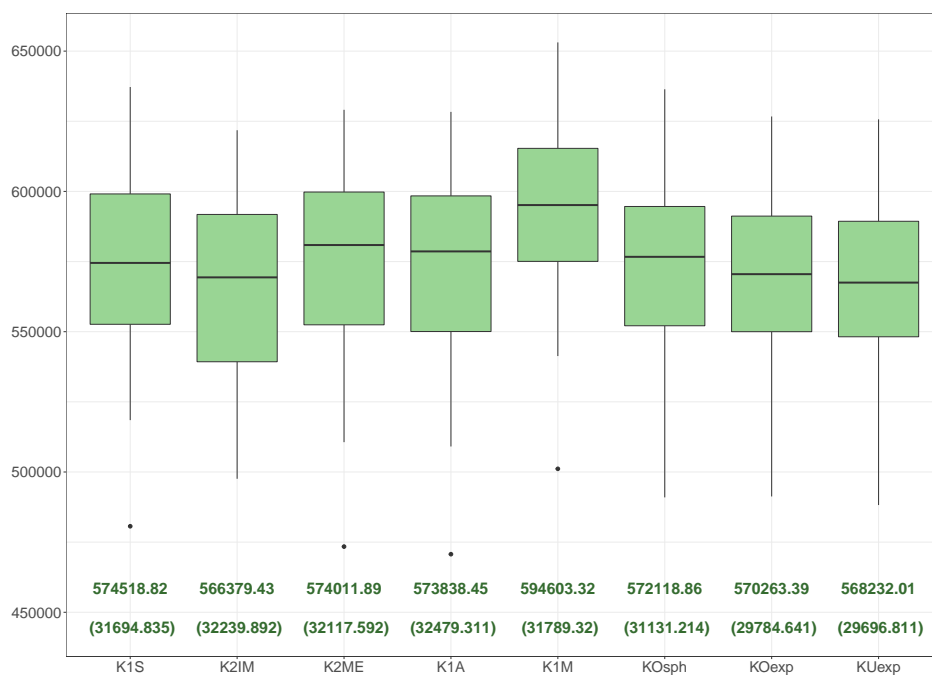


Figura 5.5: Distribución del MAE, sobre 50 repeticiones, para los datos Saltillo.

	K1S	K2IM	K2ME	K1A	K1M	KOsph	KOexp	KUexp
MAE	574.52	<b>566.38</b>	574.01	573.84	594.60	572.12	570.26	568.23
cor_obspred	0.61	0.62	0.61	0.61	0.61	0.63	<b>0.64</b>	<b>0.64</b>
RMSE	876.09	880.21	876.56	877.31	869.48	857.97	<b>849.40</b>	849.91

Tabla 5.2: Promedio del MAE, del coeficiente de correlación de Pearson y del RSME (en miles), sobre 50 repeticiones, para los datos Saltillo. Mejores resultados en negrita.

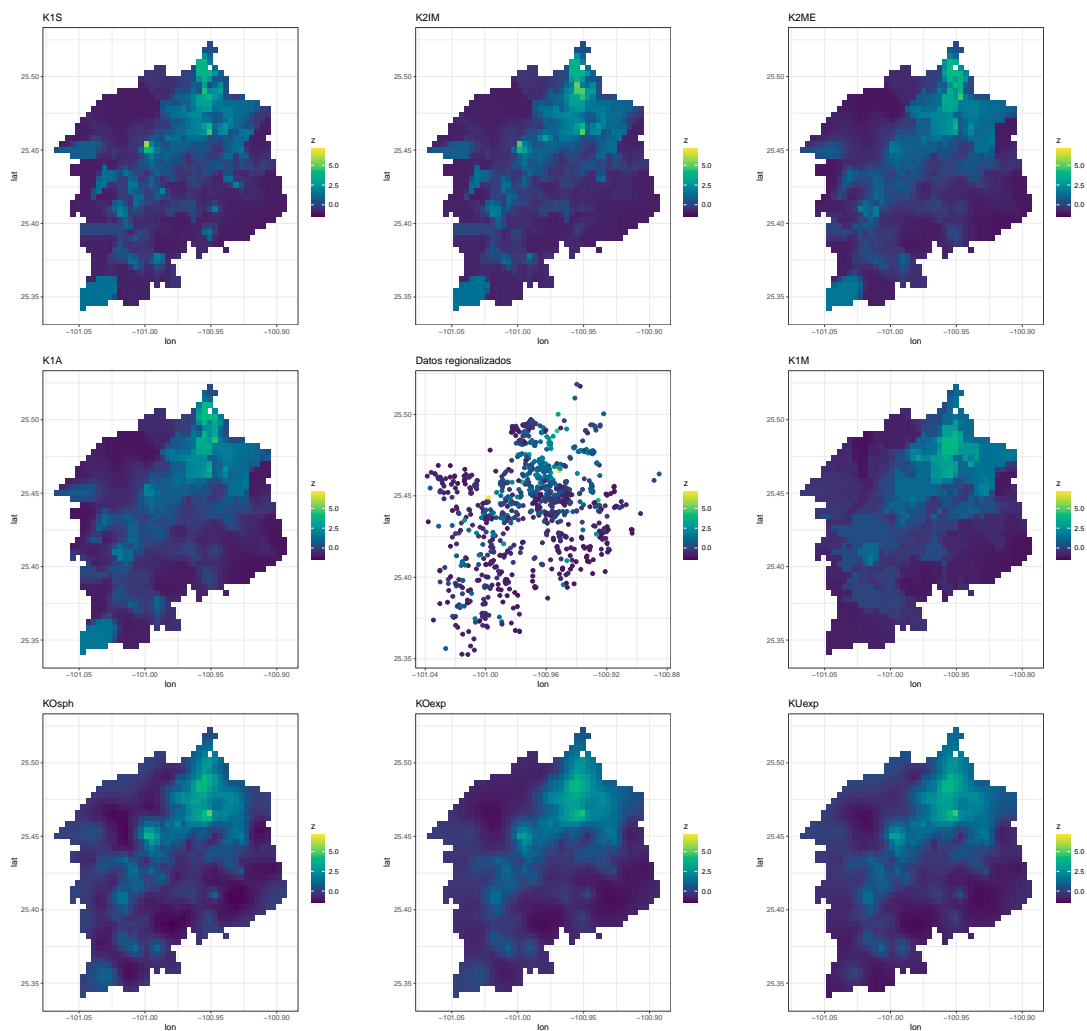


Figura 5.6: Datos Saltillo: Mapas predictivos para cada método.

en los mismos. Algo similar se observa en el extremo noreste del mapa. También se ve que algunos métodos como el K1S y el K2IM mapean con mayor dispersión de precios para dicha región.

### 5.3. Precios de Inmuebles: Valle de Bravo

#### 5.3.1. Análisis exploratorio

Los 80 datos con los que contamos en este caso se encuentran graficados en la Figura 5.7. En este caso notamos que los precios más altos se concentran al sur y los más bajos al norte, pero es más difícil determinar una tendencia en función del espacio si comparamos con el caso anterior, el cual era más evidente. En la Figura 5.8 podemos observar cómo

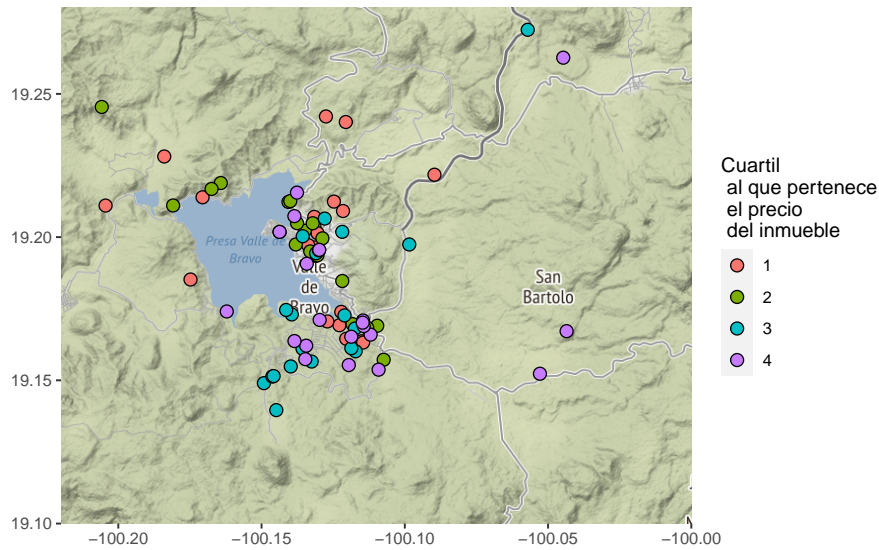


Figura 5.7: Mapa de caja: precio de inmuebles en Valle de Bravo, México.

varían los datos con respecto a cada una de las coordenadas. En el panel (b) de dicha figura, notamos una tendencia de los precios medios en dirección sur, esto es, existen locaciones más caras al sur aunque en el mismo panel observamos que la variabilidad también varía. En el sentido este-oeste (panel (c)) visualizamos una mayor concentración de los datos en el centro donde la variabilidad es elevada. En este caso, el análisis exploratorio resulta ser más complicado al tener menor cantidad de datos con respecto al caso anterior, y aunque la información obtenida sea menos confiable, también obtenemos pistas sobre la estacionariedad del proceso en cuanto a la media no constante.

La autocorrelación que observamos en el gráfico de dispersión  $h$  (Figura 5.9), solo está presente para pequeñas distancias y también es más difícil de notar por la cantidad de datos disponibles.

### 5.3.2. Resultados

En este caso observamos un peor desempeño por parte de los predictores propuestos a excepción del kriging de un núcleo Multiplicativo (K1M) el cual sigue siendo competitivo en comparación con los kriging clásicos. Con respecto a estos últimos, presentan un desempeño similar entre ellos.

Como en el caso anterior, también reportamos el promedio, sobre las 50 repeticiones realizadas, del error  $MAE$ , la correlación  $cor\_obspred$  y el error  $RMSE$ . Observamos que para todas las medidas, el que tiene el mejor desempeño es la del predictor kriging ordinario,

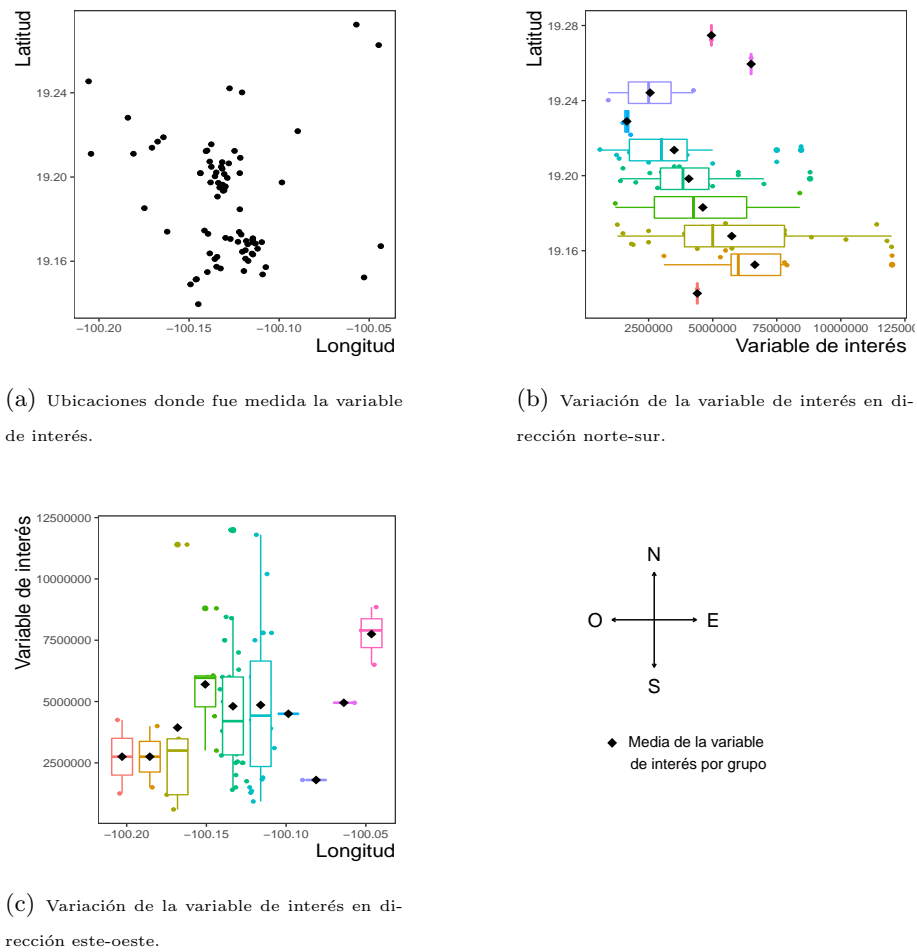


Figura 5.8: Datos Valle Bravo: Distribución de la variable de interés en función las coordenadas.

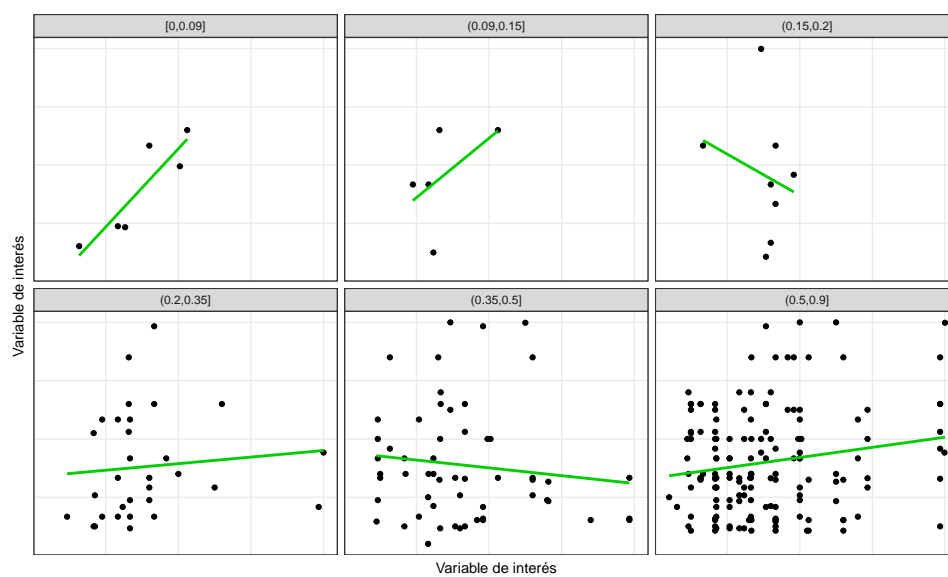


Figura 5.9: Gráfico de dispersión  $h$  del precio de inmuebles en Valle de Bravo, México.

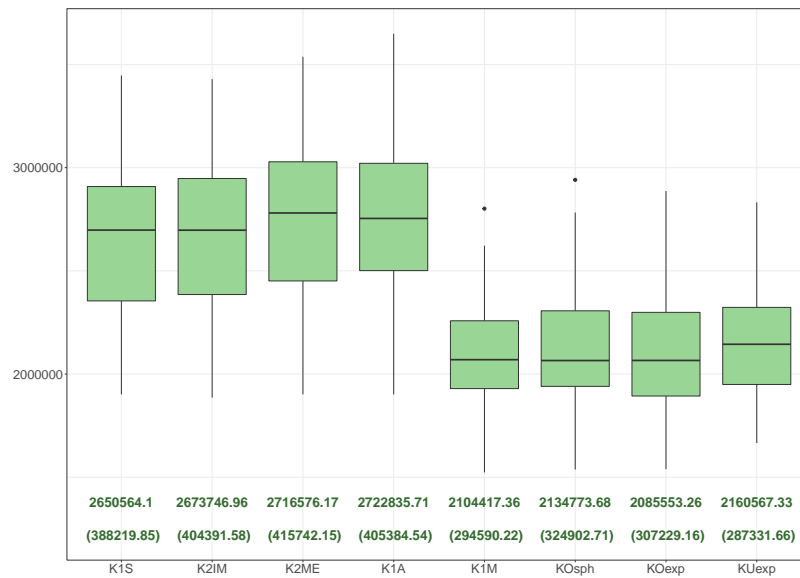


Figura 5.10: Distribución del MAE, sobre 50 repeticiones, para los datos Valle de Bravo.

	K1S	K2IM	K2ME	K1A	K1M	KOspH	KOexp	KUexp
MAE	2650.56	2673.75	2716.58	2722.84	2104.42	2134.77	<b>2085.55</b>	2160.57
cor_obsPred	0.07	0.08	0.09	0.10	0.30	0.27	<b>0.31</b>	0.29
RMSE	3478.18	3518.66	3513.53	3534.43	2657.39	2702.47	<b>2649.38</b>	2749.54

Tabla 5.3: Promedio del MAE, del coeficiente de correlación de Pearson y del RSME (en miles), para los datos Valle Bravo. Mejores resultados en negrita.

hallado a partir de un modelo de variograma exponencial, aunque en términos de precisión, predictor de un K1M tiene un mejor desempeño.

En la Figura 5.11 reportamos las predicciones sobre el área de interés y para todos los métodos estudiados. Limitamos la región a predecir a la zona donde la concentración de datos es mayor.

## 5.4. Precios de hospedajes: Singapur

### 5.4.1. Análisis exploratorio

Recordemos que en este problema la variable de interés es el precio por noche de hospedaje en algunos barrios de Singapur y que contamos con 1326 observaciones. En la Figura 5.12 observamos la distribución espacial de los puntos coloreado por cuartil al que pertenece. En dicha figura observamos los sitios de la muestra y la distribución del precio del hospedaje donde podemos ver concentraciones de valores parecidos por zona. Por ejemplo, en la zona

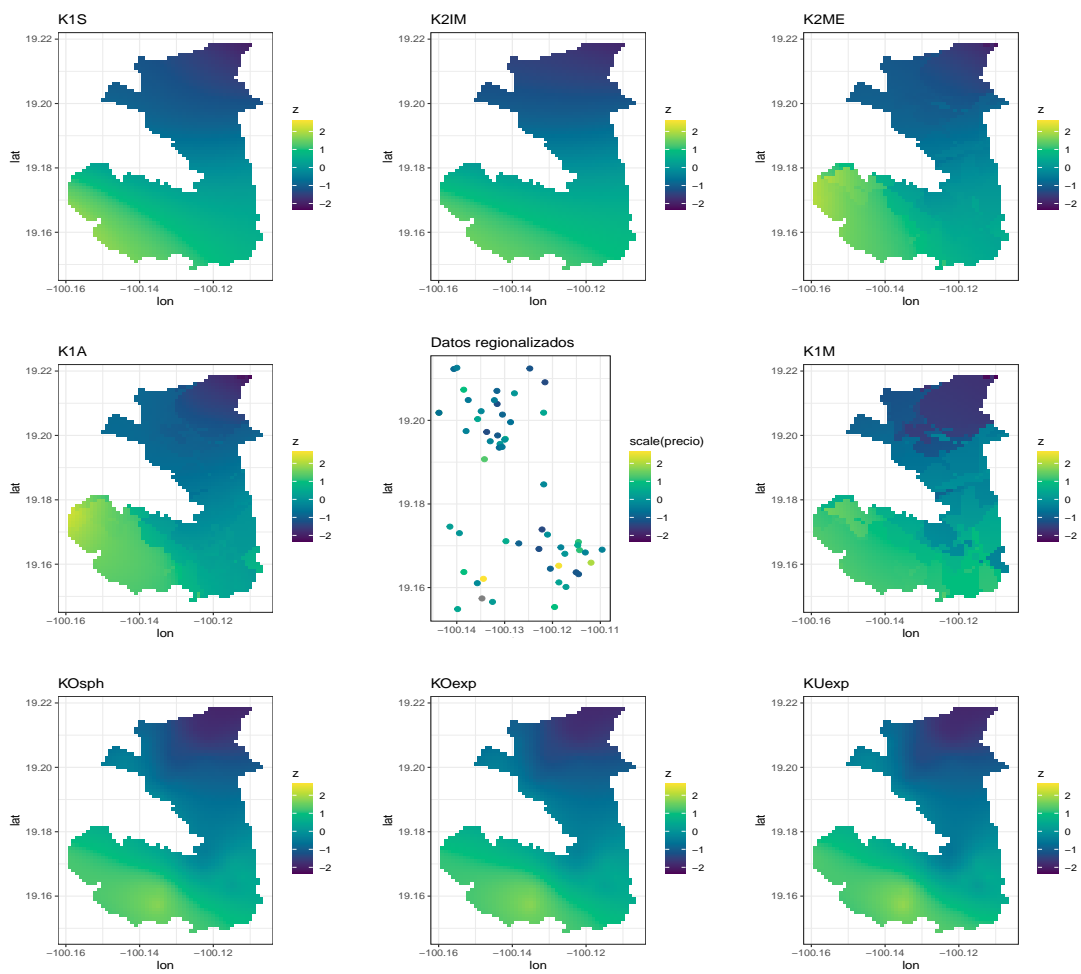


Figura 5.11: Datos Valle Bravo: Mapas predictivos para cada método.

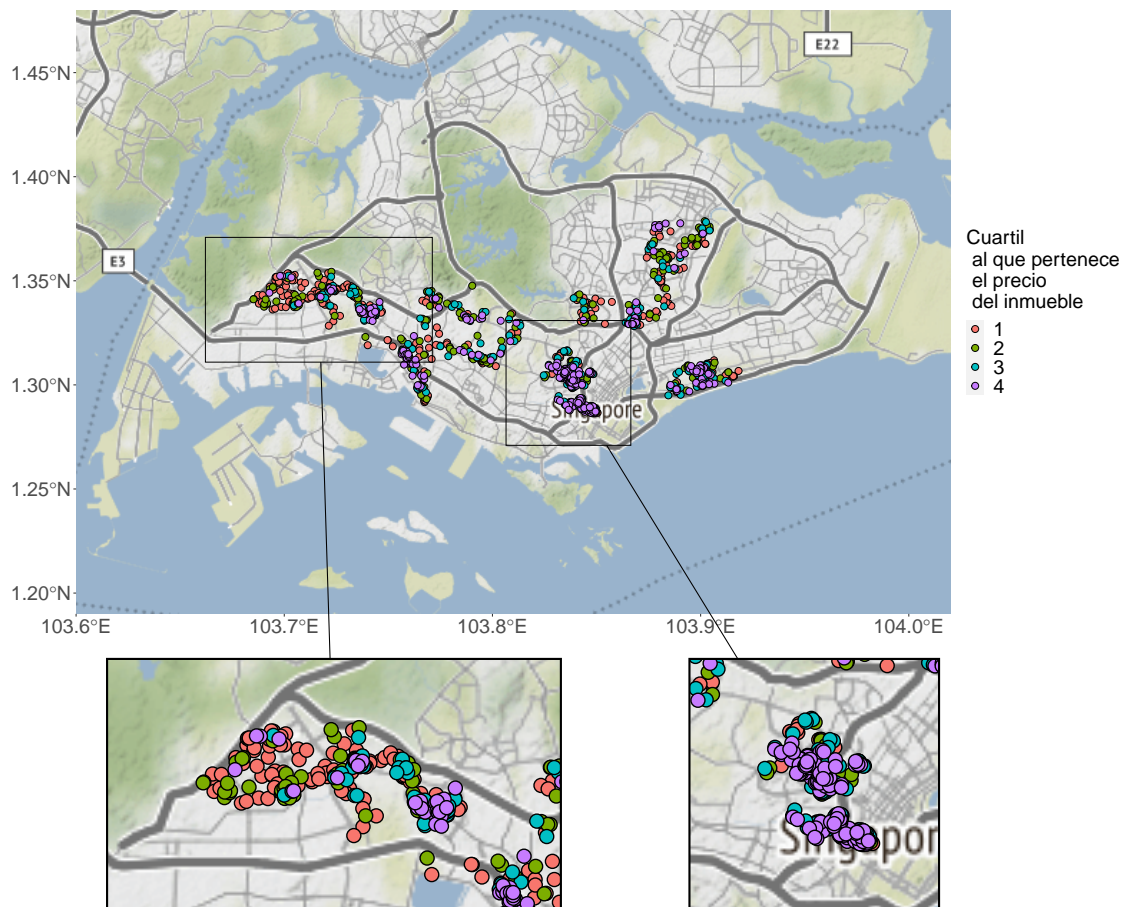


Figura 5.12: Mapa de caja: precio de hospedajes en Singapur.

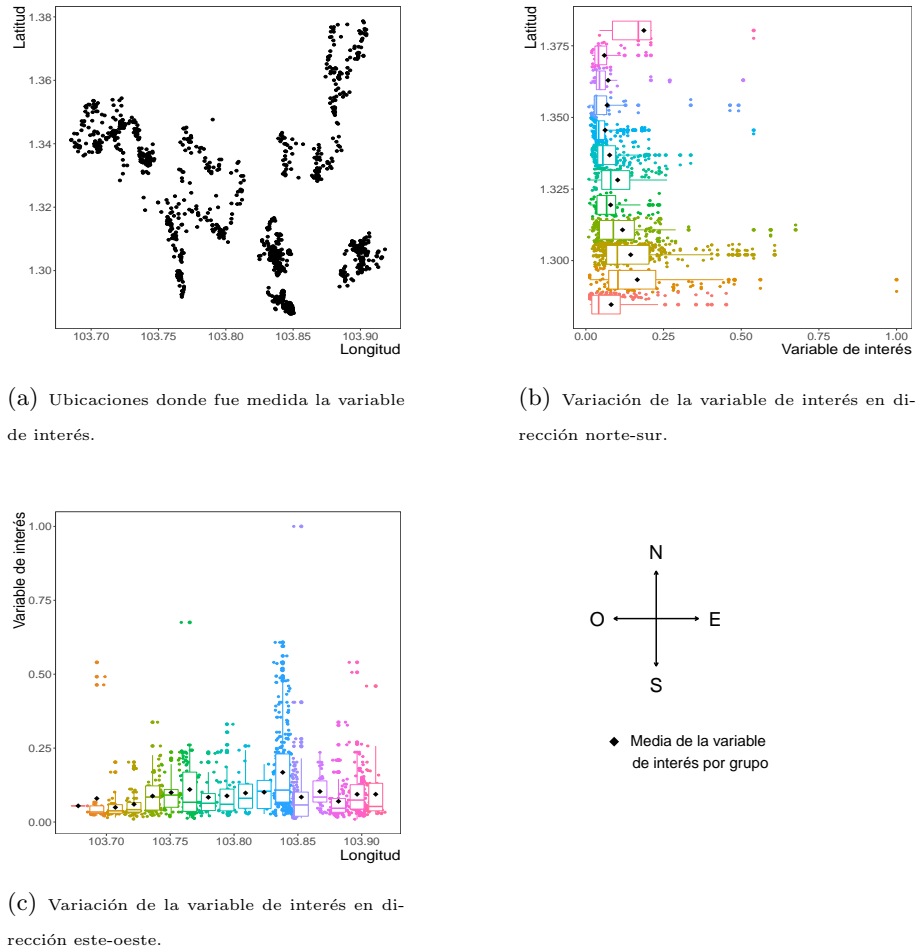


Figura 5.13: Datos Singapur: Distribución de la variable de interés en función de las coordenadas.

centro-sur y zonas cercanas a la costa, observamos valores elevados del precio. En el resto visualizamos mayor variabilidad y en dirección este-oeste vemos hospedajes cuyo precio por noche tiende a decrecer.

Al analizar la Figura 5.13 podemos observar cómo varían los datos con respecto a cada una de las coordenadas. En este caso vemos, en el panel (b), una tendencia de los precios, esto es, los precios tienden a aumentar en dirección sur. En el mismo panel notamos un aumento en la variabilidad en la misma dirección. En el sentido este-oeste (panel (c)) no observamos ninguna tendencia en el comportamiento de la media pero sí heterogeneidad en la varianza. Esto significa que en este contexto, también encontramos indicios de varianza no constante en función del espacio. Para analizar la dependencia espacial realizamos un gráfico de dispersión  $h$  (Figura 5.14) en el cual observamos cómo varía la autocorrelación a medida que consideramos mayores distancias entre los sitios. Dicha figura muestra que la presencia de autocorrelación espacial es mayor para las distancias pequeñas dado que la pendiente de la recta ajustada tiende a ser mayor para estas distancias.

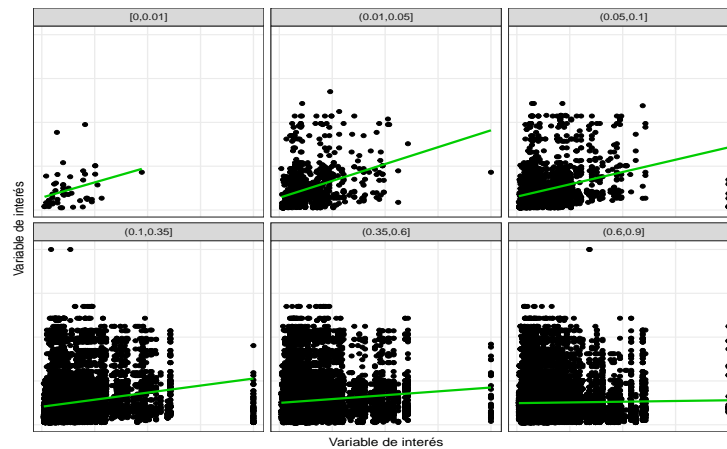


Figura 5.14: Gráfico de dispersión  $h$  del precio de hospedaje en Singapur.

	K1S	K2IM	K2ME	K1A	K1M	KOsph	KOexp	KUexp
MAE	<b>81.33</b>	83.06	81.65	81.97	87.31	105.56	102.93	98.92
cor_obspred	<b>0.52</b>	0.50	0.51	0.51	0.47	NaN	0.24	0.28
RMSE	<b>129.24</b>	134.08	130.05	130.51	133.21	150.15	147.31	143.93

Tabla 5.4: Promedio del MAE, del coeficiente de correlación de Pearson y del RSME (en miles) sobre 50 repeticiones, para los datos de Singapur. Mejores resultados en negrita.

### 5.4.2. Resultados

Los resultados obtenidos se encuentran en la Figura 5.15 y en la Tabla 5.4. Hallamos, en promedio, un mejor desempeño para los predictores no paramétricos, aunque entre ellos no observamos grandes diferencias. De los predictores clásicos, el mejor desempeño lo tiene el predictor kriging universal. En cuanto a la variabilidad de los errores, aunque notamos grandes diferencias, el K1M es el que cuenta con el mejor desempeño en este sentido. En particular observamos, tanto en el gráfico como en la tabla, el mejor desempeño para el predictor no paramétrico que solo considera las distancias (K1S).

En la Figura 5.16 reportamos las predicciones sobre el área de interés y para todos los métodos estudiados. Podemos observar cómo los métodos kriging clásicos realizan una predicción más general y los métodos propuestos captan la variabilidad de los datos teniendo en cuenta cada zona en particular. En el caso del kriging ordinario con modelo de variograma esférico, KOsph, predice el mismo valor en todos los sitios de la grilla propuesta, por ello en el mapa (Figura 5.16) no se visualiza su comportamiento y no pudimos completar la Tabla 5.4 al no poder hallar la correlación ya que la varianza de los errores, de dicho estimador, es cero.

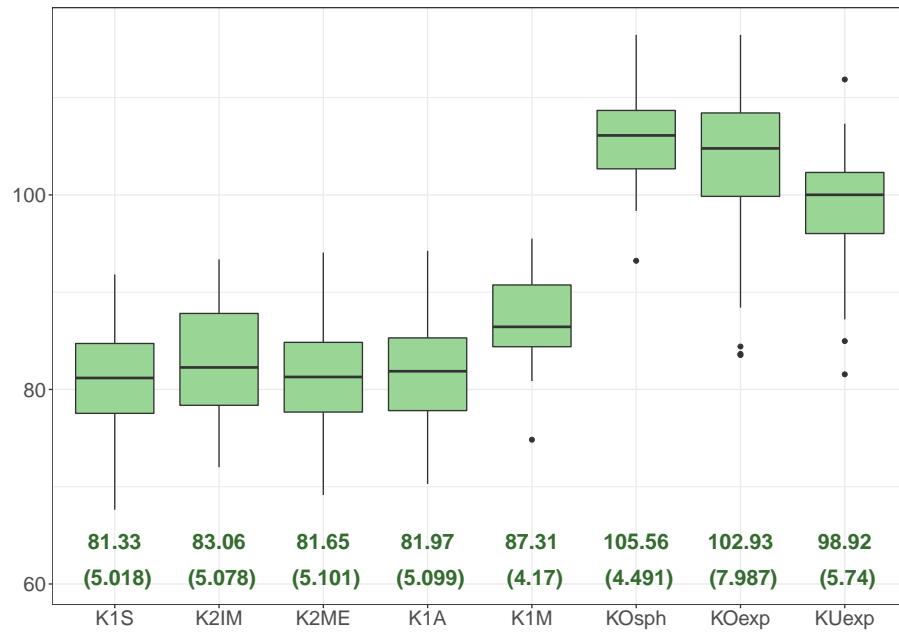


Figura 5.15: Distribución del MAE, sobre 50 repeticiones, para los datos de Singapur.

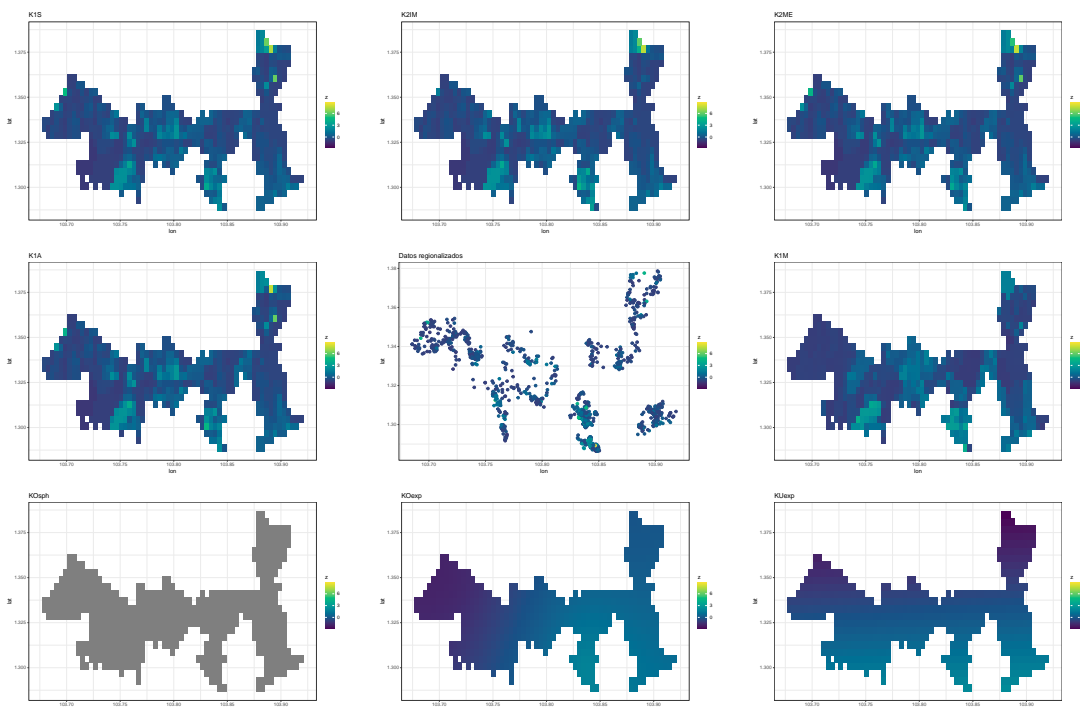


Figura 5.16: Datos de Singapur: Mapas predictivos para cada método.

## 5.5. Conclusiones

En este capítulo hemos evaluado el desempeño predictivo de los métodos propuestos considerando tres conjuntos de datos reales que se caracterizan por tener diferentes tamaño, diferente distribución en el espacio y correlación espacial a diferentes escalas. En el caso de tener una distribución de los datos sesgada y una variabilidad notoriamente heterogénea, los predictores propuestos muestran un desempeño competitivo, y hasta en algunos casos, muestran mejores resultados que los predictores clásicos.

# Conclusiones generales

Los datos espaciales, como datos correlacionados, vulneran los supuestos de independencia de los modelos utilizados por la estadística clásica. Como alternativa a dichos modelos y con fines predictivos, en este trabajo consideramos a los datos espaciales como realizaciones de un proceso estocástico espacial con el fin de brindarles un contexto aleatorio y así modelar y medir sus fuentes de variabilidad (i.e espacial e inherente).

Uno de los conceptos más importantes para medir la variabilidad espacial del fenómeno es el variograma, el cual es determinante en la definición del proceso e imprescindible para resolver el problema de la predicción a través del método kriging. Los supuestos necesarios para aplicar los predictores kriging pueden ser restrictivos y las decisiones necesarias para estimar el variograma pueden tornarse subjetivas.

Como alternativa, en la presente tesis propusimos métodos de predicción no paramétricos partiendo de la combinación lineal pesada, definida por los métodos kriging, pero con pesos obtenidos de forma no paramétrica basado en estimadores de núcleos. Cada variante de esta propuesta fue motivada por diferentes hipótesis de existencia del fenómeno de interés. En otras palabras, logramos que los predictores no solo capten la hipótesis de que valores cercanos son parecidos y por lo cual presentan mayor correlación, sino que también tengan en cuenta que pueden existir estructuras de covarianza diferentes y vecinas, provocando valores cercanos pero disímiles. Además de ello, los predictores desarrollados pueden adaptarse para considerar la mayor cantidad de información a partir de los datos disponibles.

Las diferentes versiones de los predictores presentados obtuvieron resultados competitivos al ser evaluados y comparados con alternativas paramétricas del kriging, tanto en los ejercicios de simulación como en las aplicaciones con datos reales. En los estudios de simulación propusimos diferentes escenarios contruídos al variar las principales características y propiedades, tanto del proceso como del variograma. De ello obtuvimos resultados satisfactorios en contextos donde el supuesto de varianza constante no se cumple (i.e. con datos heterocedásticos), cuando nos alejamos de la normalidad de los errores y en el caso de considerar estructuras de covarianza combinadas para definir el proceso. Con las aplicaciones también pudimos visualizar las ventajas predictivas de las metodologías no paramétricas

propuestas frente a los clásicos kriging, siendo más notorias en algunos casos que en otros donde los métodos paramétricos muestran ser una buena estrategia para predecir a pesar de su menor flexibilidad.

Específicamente, en los casos donde existe autocorrelación a grandes distancias, (modelo exponencial y rango  $\theta_2 = 100$ ) dos de los predictores que consideran la mediana del vecindario, K2ME y K1A, son los que indican un mejor desempeño, particularmente en los casos donde consideramos varianza no constante. En cambio en los casos donde la autocorrelación es local (modelo exponencial y rango  $\theta_2 = 2$ ) los predictores, anteriormente nombrados, siguen teniendo un buen comportamiento a los cuales se le suman los predictores K1S y K2IM, que consideran solamente las distancias y el Índice de Morán, respectivamente. El predictor K2IM revela buenos resultados en los casos donde asumimos varianza constante. Por último los predictores que mejor captan la variabilidad en contextos donde existen diferentes estructuras de covarianza, son los predictores K2ME y K1S los cuales presentan, además del menor MAE promedio, la menor variabilidad en los errores.

Existen diversas líneas de trabajo que permitirían extender la contribución realizada en la presente tesis, las cuales proponemos como futuras líneas de investigación. En primer lugar, un estudio de influencia sobre el desempeño predictivo, tanto de los predictores propuestos como en los predictores kriging, del tamaño de la muestra, de la asimetría de los datos y de los datos extremos incluyendo la definición, detección y análisis de estos últimos en el contexto espacial. En segundo lugar, la inclusión de modelos de variograma anisotrópicos para la generación de datos, con la respectiva evaluación comparativa de las metodologías predictivas. En tercer lugar, un análisis pormenorizado de la estimación de los parámetros de suavizado con el fin de obtener una mayor eficiencia en el cálculo de los mismos. Adicionalmente, para completar la presente propuesta metodológica en términos teóricos, creemos necesario un estudio más exhaustivo sobre las propiedades asintóticas de cada uno de los predictores no paramétricos propuestos; lo que planificamos llevar a cabo como continuidad de este trabajo. Por último, la extensión del enfoque no paramétrico en el contexto multivariado, esto es, teniendo en cuenta variables predictoras además de la localización y de la variable de interés. Tal propuesta consistiría en versiones no paramétricas y semiparamétricas del denominado predictor *cokriging*, constituyendo el eje de nuestro plan de trabajo futuro.

# Bibliografía

- Anselin, L. (2010). Local indicators of spatial association-LISA. Geographical Analysis, 27(2):93–115. 28, 29, 54, 57
- Anselin, L. y Florax, R. (2004). Advances in Spatial Econometrics. Springer. XV
- Bajat, B., Kilibarda, M., Pejović, M., y Petrović, M. S. (2017). Spatial hedonic modeling of housing prices using auxiliary maps. In Advances in Geographic Information Science, pages 97–122. Springer Berlin Heidelberg. XVI
- Banerjee, S., Carlin, B. P., y Gelfand, A. E. (2004). Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC. 29
- Biau, G. y Cadre, B. (2004). Nonparametric spatial prediction. Statistical Inference for Stochastic Processes, 7(3):327–349. V, XV, 48, 49, 50, 55
- Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. Springer New York. 46, 49
- Carbon, M., Hallin, M., y Tran, L. T. (1996). Kernel density estimation for random fields: TheL1theory. Journal of Nonparametric Statistics, 6(2-3):157–170. 49
- Carbon, M., Tran, L. T., y Wu, B. (1997). Kernel density estimation for random fields (density estimation for random fields). Statistics & Probability Letters, 36(2):115–125. 49
- Chandra, H., Aditya, K., y Sud, U. C. (2018). Localised estimates and spatial mapping of poverty incidence in the state of bihar in india - an application of small area estimation techniques. PLOS ONE, 13(6):e0198502. XVI
- Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method: Cokriging. Journal of Real Estate Research, 29(1):95–114. XVI
- Chiles, J.-P. y Delfiner, P. (2012). Geostatistics: modeling spatial uncertainty. John Wiley & Sons. 29
- coahuila.gob (2020). Datos municipios. <https://coahuila.gob.mx/micrositios/index/datos-municipios>. 78

- Cressie, N. (1985). Fitting variogram models by weighted least squares. Journal of the international Association for mathematical Geology, 17(5):563–586. 22
- Cressie, N. (1990). The origins of kriging. Mathematical Geology, 22(3):239–252. 31
- Cressie, N. (1993). Statistics for Spatial Data. John Wiley and Sons, Inc. V, XV, 7, 26, 29, 31, 38, 63
- Dabo-Niang, S., Hamdad, L., Ternynck, C., y Yao, A.-F. (2014). A kernel spatial density estimation allowing for the analysis of spatial clustering. application to monsoon asia drought atlas data. Stochastic environmental research and risk assessment, 28(8):2075–2099. 51
- Dabo-Niang, S., Ternynck, C., y Yao, A.-F. (2016). Nonparametric prediction of spatial multivariate data. Journal of Nonparametric Statistics, 28(2):428–458. XV, 45, 51, 52, 53
- Dabo-Niang, S. y Yao, A. F. (2007). Kernel regression estimation for continuous spatial processes. Mathematical Methods of Statistics, 16(4):298–317. XV, 50
- Diggle, P. (2014). Statistical Analysis of Spatial and Spatio-temporal Point Patterns, Third Edition. CRC Press. XV
- Geary, R. C. (1954). The contiguity ratio and statistical mapping. The Incorporated Statistician, 5(3):115. 28
- Gelfand, A., Fuentes, M., Guttorp, P., y Diggle, P. (2010). Handbook of Spatial Statistics. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Taylor & Francis. 31
- Griffith, D. A. y Arbia, G. (2010). Detecting negative spatial autocorrelation in georeferenced random variables. International Journal of Geographical Information Science, 24(3):417–437. 57
- Györfi, L., Härdle, W., Sarda, P., y Vieu, P., editors (1989). Nonparametric Curve Estimation from Time Series. Springer New York. 49
- Härdle, W. y Müller, M. (1997). Multivariate and semiparametric kernel regression. 47
- Hastie, T., Tibshirani, R., y Friedman, J. (2001). The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA. 70
- Journel, A. G. y Huijbregts, C. J. (1978). Mining geostatistics. Academic Press Limited. 21
- Kuntz, M. y Helbich, M. (2014). Geostatistical mapping of real estate prices: an empirical comparison of kriging and cokriging. International Journal of Geographical Information Science, 28(9):1904–1921. XVI

- Legendre, P. y Legendre, L. (2012). Numerical Ecology, Volume 24, 3rd Edition. Elsevier. XV
- Liu, R., Li, G., Wei, L., Xu, Y., Gou, X., Luo, S., y Yang, X. (2022). Spatial prediction of groundwater potentiality using machine learning methods with grey wolf and sparrow search algorithms. Journal of Hydrology, 610:127977. XV
- Matheron, G. (1963). Principles of geostatistics. Economic geology, 58(8):1246–1266. 31
- Matheron, G. (1965). Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson et CIE. 6
- Mathiassen, A. (2009). A model based approach for predicting annual poverty rates without expenditure data. Journal of Economic Inequality, 7(2):117–135. XVI
- Menezes, R., García-Soidán, P., y Ferreira, C. (2010). Nonparametric spatial prediction under stochastic sampling design. Journal of Nonparametric Statistics, 22(3):363–377. XV, 50, 51, 52
- Metadata (2018). Valuación de inmuebles. 77
- Mohamed, A. y Mohamed, A. (2009). Spatial Patterns and Geographic Determinants of Welfare and Poverty in Tunisia. Working Papers 478, Economic Research Forum. XVI
- Montero, J.-M., Fernández-Avilés, G., y Mateu, J. (2015). Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. John Wiley and Sons, Ltd. XV, 2, 8, 21, 31, 32, 66
- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1-2):17–23. 28
- Nadaraya, E. (1964). On estimating regression. Theory of Probability & Its Applications, 9:141–142. 45
- Nadaraya, E. (1965). On non nonparametric estimates for density functions and regression curves. Theory of Probability & Its Applications, 10:297–302. 45
- Norman, P. (2009). Identifying change over time in small area socio-economic deprivation. Applied Spatial Analysis and Policy, 3(2-3):107–138. XVI
- Oliver, M., editor (2010). Geostatistical Applications for Precision Agriculture. Springer Netherlands. XV
- Patterson, H. D. y Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58(3):545–554. 23
- Pham, B. T., Bui, D. T., y Prakash, I. (2018). Bagging based support vector machines for spatial prediction of landslides. Environmental Earth Sciences, 77(4). XV

- Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., et al. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature communications, 11(1):1–11. 70
- Pokhriyal, N. y Jacques, D. C. (2017). Combining disparate data sources for improved poverty prediction and mapping. Proceedings of the National Academy of Sciences, 114(46):E9783–E9792. XVI
- Rest, K. L., Pinaud, D., Monestiez, P., Chadoeuf, J., y Bretagnolle, V. (2014). Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. Global Ecology and Biogeography, 23(7):811–820. 70
- Rikken, M. y Van Rijn, R. (1993). Soil pollution with heavy metals - an inquiry into spatial variation, cost of mapping and the risk evaluation of copper, cadmium, lead and zinc in the floodplains of the meuse west of Stein. Department of Physical Geography, Utrecht University, Utrecht. V, 14
- Rizo-Decelis, L., Pardo-Igúzquiza, E., y Andreo, B. (2017). Spatial prediction of water quality variables along a main river channel, in presence of pollution hotspots. Science of The Total Environment, 605-606:276–290. XV
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guisera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., y Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8):913–929. 70
- Siabato, W. y Guszmán-Manrique, J. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. Cuadernos de Geografía: Revista Colombiana de Geografía, 28(1):1–22. 27, 28
- Ternynck, C. (2014). Spatial regression estimation for functional data with spatial dependency. Journal de la Société Française de Statistique, 155:138–160. 51
- Tikhonov, A. N. y Arsenin, V. Y. (1977). Solution of ill-posed problems. John Wiley Sons. 62
- Vasan, S. y Alcantara, A. (2016). Gis-based methods for estimating missing poverty rates and projecting future rates in census tracts. Review of Economics and Finance, 6:1–13. XVI
- Wackernagel, H. (1995). Multivariate Geostatistics. Springer. XV, 2, 20
- Wackernagel, H. (2006). Geostatistics. American Cancer Society. 1, 2, 3, 6, 10, 21, 31, 32
- Wadoux, A. M.-C., Heuvelink, G. B., de Bruin, S., y Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. Ecological Modelling, 457:109692. 70

- Watson, G. (1964). Smooth regression analysis. Sankhya Series A, 26:359–372. 45
- Webster, R. y Oliver, M. (2007). Geostatistics for Environmental Scientists. Wiley, 2th edition. XV, 7, 20, 21, 31, 32
- Yang, D. (2018). Spatial prediction using kriging ensemble. Solar Energy, 171:977–982. XV
- Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models. Biometrics, 58(1):129–136. XV

