



UNIVERSIDAD NACIONAL DE ROSARIO

Facultad de Ciencias Económicas y Estadística

Carrera de posgrado

DOCTORADO EN ESTADÍSTICA

Predicción espacial mediante modelos autorregresivos semiparamétricos

Autora: Mariel Guadalupe Lovatto

28 de septiembre de 2026

Directora: Dra. Pamela Llopp

Codirector: Dr. Rodrigo García Arancibia

Agradecimientos

En este pequeño apartado quiero expresar un sincero agradecimiento a todas y cada una de las personas que con su apoyo y ayuda desinteresada hicieron posible la realización de esta tesis.

A mis directores Pame y Rodrigo por su enseñanza y guía constante, por su tiempo, por la paciencia y la confianza. Gracias por la oportunidad de aprender de ustedes y por emprender este camino juntos.

A Pablo por su paciencia infinita, la motivación constante y por no dejarme nunca bajar los brazos.

A mi familia, por el apoyo incondicional, las oportunidades y la libertad.

A mis amigos y amigas, especialmente a mis compañeros de oficina por aguantar mis dramas y contestar cada consulta.

A la Universidad Nacional del Litoral, donde hice mi carrera de grado y donde trabajo desde que me recibí, y CONICET, por el apoyo económico que me permitió realizar la carrera que culmina con esta tesis. Específicamente agradezco a las personas que trabajan tanto en la Universidad Nacional del Litoral como en la Universidad Nacional de Rosario por su dedicación y por hacer posible la educación pública en este país a pesar de las circunstancias.

Índice general

Agradecimientos	I
Resumen	V
Introducción	VI
1. Predicción para datos espaciales	3
1.1. Los datos espaciales	4
1.2. Funciones aleatorias	5
1.2.1. Momentos	5
1.2.2. Estacionariedad	7
1.3. Modelos y métodos autorregresivos para predicción espacial	8
1.3.1. Modelos basados en Kriging	9
1.3.2. Modelos econométricos	24
2. Modelo autorregresivo semiparamétrico para predicción espacial	29
2.1. Modelo espacial semiparamétrico	29
2.2. Algoritmo de estimación y predicción	31
2.3. Resultados asintóticos	33
2.3.1. Preliminares y notación	33
2.3.2. Hipótesis generales	35
2.3.3. Resultados asociados a los parámetros lineales	37
2.3.4. Resultados asociados a la componente no paramétrica	55
3. Estudios de simulación	59
3.1. Método para simular las covariables	59
3.2. Escenarios	61
3.3. Metodología de entrenamiento, validación y evaluación de desempeño	64
3.4. Resultados	66
3.4.1. Con respecto al error de predicción	66
3.4.2. Con respecto a la estimación del parámetro β	70
4. Aplicación a datos reales	75
4.1. Metodología de entrenamiento, validación y evaluación de desempeño	75

4.2. Datos Ohio	76
4.2.1. Resultados	79
4.3. Datos Saltillo	80
4.3.1. Resultados	83
4.4. Datos Meuse	84
4.4.1. Resultados	86
4.5. Datos Jura	87
4.5.1. Resultados	89
4.6. Comparación del desempeño predictivo entre dos bloques muestrales para el conjunto Ohio	90
4.6.1. Resultados	93
Conclusiones generales	97
Bibliografía	99
A. Resultados Auxiliares	105

Resumen

La presencia de dependencia en datos plantea desafíos específicos en los problemas de predicción, ya que la información observada en una localización suele estar asociada a la de su entorno. Modelar adecuadamente esta estructura resulta clave para obtener predicciones confiables y para realizar conclusiones válidas sobre los procesos subyacentes. En este marco, los métodos geoestadísticos tradicionales han sido ampliamente utilizados por su capacidad predictiva, en contraste con los modelos econométricos espaciales, cuyo énfasis está en la interpretación de los efectos de las covariables y la modelización explícita de las interacciones entre unidades espaciales.

En esta tesis desarrollamos un modelo autorregresivo espacial de carácter semiparamétrico enfocado en la predicción, que combina una componente lineal asociada a las covariables con un término autorregresivo espacial estimado mediante técnicas de estadística no paramétrica. Esta estrategia permite representar la dependencia espacial sin recurrir a especificaciones paramétricas estrictas de la estructura de covarianza, utilizada generalmente en los métodos geoestadísticos clásicos y, al mismo tiempo, conserva la posibilidad de interpretar los efectos marginales de las covariables, aproximándose así a la riqueza explicativa de los modelos econométricos espaciales.

El trabajo aborda el estudio de las propiedades teóricas de los estimadores que intervienen en el predictor propuesto, analizando su comportamiento asintótico, en particular la normalidad asintótica y las tasas de convergencia de los parámetros lineales y de la componente no paramétrica. Asimismo, se evalúa el desempeño del modelo a través de estudios de simulación y aplicaciones a datos reales, comparándolo con metodologías geoestadísticas clásicas como así también en modelos econométricos espaciales, con el fin de examinar tanto su desempeño predictivo como su exactitud en la estimación puntual de parámetros.

Introducción

Gran parte de la literatura clásica en estadística se desarrolla bajo el supuesto de independencia entre las observaciones, lo que permite simplificar tanto la formulación de los modelos como el desarrollo teórico asociado. En este marco, la predicción de fenómenos de interés a partir de información observada constituye uno de los objetivos centrales de la estadística aplicada. La construcción de modelos estadísticos adecuados requiere no solo capturar la relación entre la variable de interés y las covariables disponibles, sino también representar de manera apropiada las estructuras presentes en los datos.

En muchos contextos aplicados, sin embargo, las observaciones exhiben distintos tipos de dependencia, cuya omisión puede conducir a estimaciones inconsistentes y a un deterioro significativo del desempeño predictivo, comprometiendo tanto la validez teórica como la utilidad práctica de los resultados obtenidos. La estructura de dependencia aparece de forma natural en numerosos contextos, entre ellos las series de tiempo, los modelos de medidas repetidas y, particularmente, los datos espaciales, en los cuales las observaciones se asocian a localizaciones geográficas específicas. En este último caso, la cercanía espacial suele inducir patrones de correlación que no pueden ser tratados adecuadamente bajo supuestos de independencia. Nuestro interés se centra en este tipo de datos y, en particular, en el problema de realizar predicciones incorporando explícitamente la dependencia asociada a la estructura espacial.

La predicción en presencia de dependencia espacial constituye un problema central en diversas áreas aplicadas (Kheir et al., 2025; Kiani et al., 2025; Cerqueti et al., 2025), donde la calidad de las estimaciones y de las predicciones depende fuertemente de la correcta modelización de dicha dependencia. En este marco, los modelos espaciales han sido ampliamente desarrollados con el objetivo de capturar estas estructuras de correlación. Entre ellos, los enfoques basados en kriging (Meng et al., 2024; Pang et al., 2023; Khan et al., 2023) son los modelos más destacados, dado su buen desempeño predictivo y su fundamentación teórica a partir de la modelización de la función de covarianza. No obstante, si bien estos métodos resultan eficaces para la predicción en sitios no observados, no permiten, en general, obtener estimaciones directamente interpretables de los efectos marginales de las covariables sobre la variable de interés.

En los últimos años, se ha observado una creciente tendencia a combinar métodos geoestadísticos clásicos con técnicas de aprendizaje automático, dando lugar a enfoques híbridos que buscan capturar simultáneamente relaciones no lineales globales y dependencia espacial residual (Fouedjio y Arya, 2024). Estas metodologías han sido aplicadas con éxito en distintos dominios, incluyendo ecología y agricultura (Shen et al., 2024), datos espaciotemporales (Jeong y Koo, 2025) y variables derivadas de sensores remotos (Tadić et al., 2024). Sin embargo, estos enfoques híbridos conservan una limitación de fondo: la ganancia en desempeño predictivo no se traduce en estimaciones interpretables de los efectos de las covariables, lo que restringe su utilidad en contextos donde interesa comprender los mecanismos subyacentes al fenómeno analizado.

En contraste, los métodos econométricos espaciales se enfocan explícitamente en la estimación e interpretación de los efectos asociados a las covariables, permitiendo cuantificar su impacto marginal sobre la variable respuesta (Vagnini et al., 2025). A través de formulaciones paramétricas que incorporan dependencia espacial, estos modelos facilitan la inferencia sobre relaciones estructurales y la evaluación de efectos directos e indirectos entre localizaciones. Esta interpretabilidad, sin embargo, se obtiene a costa de una menor flexibilidad: cuando la relación entre las variables no puede representarse adecuadamente mediante estructuras paramétricas simples, el ajuste del modelo se vuelve restrictivo.

A esta limitación se suma otra propia de los modelos econométricos espaciales: la estructura de dependencia se introduce a través de una matriz de pesos espaciales, que define el esquema de vecindad y la intensidad de interacción entre unidades geográficas. Esta matriz no se estima conjuntamente con los parámetros del modelo, sino que constituye un insumo exógeno cuya especificación condiciona los resultados; distintas definiciones de vecindad o esquemas de estandarización pueden conducir a estimaciones sustancialmente distintas. Esta sensibilidad estructural ha motivado diversos estudios orientados a la evaluación y selección óptima de la matriz de pesos (Stakhovych y Bijmolt, 2009; Duncan et al., 2017; Bauman et al., 2018), y refuerza el carácter paramétrico del modelo, que puede resultar limitante cuando la verdadera estructura espacial no se ajusta a la especificación adoptada.

Motivados por esta tensión entre flexibilidad e interpretabilidad, y a partir del estudio de la estructura estocástica subyacente a los datos espaciales, en trabajos previos desarrollamos el enfoque de kriging no paramétrico (Lovatto et al., 2022; García Arancibia et al., 2023). En esta dirección proponemos un predictor que conserva la forma funcional de los predictores de kriging, pero cuyos pesos se estiman mediante técnicas de suavizado por núcleos, evitando la especificación paramétrica explícita de la función de covarianza o de la matriz de pesos espaciales. Los resultados obtenidos evidenciaron un buen desempeño predictivo, lo que motivó la exploración de extensiones que permitieran incorporar información adicional a través de covariables.

En este contexto, en la presente tesis proponemos resolver el problema de la predicción espacial mediante un modelo autorregresivo espacial semiparamétrico, que combina

un componente lineal asociado a las covariables con un término autorregresivo espacial estimado mediante el enfoque de kriging no paramétrico desarrollado en García Arancibia et al. (2023). De este modo, el modelo captura la dependencia espacial de manera flexible sin renunciar a la interpretación directa de los efectos marginales de las covariables, posicionándose como una alternativa intermedia entre los métodos geoestadísticos puramente predictivos y los modelos econométricos espaciales tradicionales.

Uno de los objetivos de esta tesis es estudiar las propiedades teóricas de los estimadores involucrados en el predictor propuesto, así como evaluar su desempeño predictivo bajo distintos escenarios. En particular, se analizan sus propiedades asintóticas, incluyendo la normalidad asintótica y las tasas de convergencia de los parámetros lineales, junto con las tasas de convergencia correspondientes a la componente no paramétrica. Asimismo, exploramos su comportamiento empírico mediante estudios de simulación y aplicaciones a datos reales, comparando su desempeño con el de modelos autorregresivos espaciales ampliamente utilizados en la literatura.

El trabajo se organiza de la siguiente manera. En el **Capítulo 1** se presentan los conceptos básicos necesarios para el desarrollo de la tesis, incluyendo una revisión de los datos espaciales, las funciones aleatorias y las nociones de estacionariedad, así como un repaso de los principales modelos y métodos autorregresivos utilizados en la predicción espacial, con énfasis en los enfoques basados en kriging y en los modelos econométricos espaciales. El **Capítulo 2** introduce el modelo autorregresivo espacial semiparamétrico propuesto, detallando su formulación y el algoritmo de estimación necesarios para la presentación del predictor obtenido. Asimismo, se presentan los supuestos bajo los cuales se desarrolla el análisis teórico y los resultados asintóticos para los distintos componentes del modelo. En el **Capítulo 3** se desarrollan estudios de simulación destinados a evaluar el desempeño del método bajo distintos escenarios, considerando tanto errores de predicción como errores en la estimación de los parámetros. Finalmente, en el **Capítulo 4** se aplica la metodología propuesta a distintos conjuntos de datos reales, ilustrando su comportamiento en situaciones prácticas y comparando sus resultados con los de modelos alternativos. El trabajo concluye con una síntesis de los principales hallazgos y una discusión sobre posibles extensiones futuras.

Capítulo 1

Predicción para datos espaciales

Una de las herramientas fundamentales en el área de estadística es la predicción de fenómenos de interés a partir de ciertas variables observadas. Si estas variables fueron observadas o medidas en localizaciones espaciales, la predicción constituye un desafío central ya que vulnera los supuestos de independencia. La disponibilidad creciente de datos observacionales y/o experimentales de este tipo ha generado la necesidad de desarrollar métodos capaces de explicar y modelar la variabilidad de dichos fenómenos considerando la dependencia que los caracteriza. En este capítulo abordamos la definición y caracterización de este tipo de datos, explicitamos la estructura formal subyacente sobre la cual se supone que se generan los mismos y describimos las herramientas existentes para realizar dichas predicciones de manera precisa en contextos espaciales.

Comenzamos presentando y definiendo los datos espaciales y sus características, incluyendo la estructura espacial y las propiedades estadísticas relevantes, como momentos, independencia y estacionariedad. Luego, describimos los modelos que permiten capturar las particularidades de los datos espaciales, junto con los métodos de estimación específicos utilizados con fines predictivos. Se abordan los modelos basados en Kriging, ampliamente utilizados en geoestadística por su capacidad de incorporar explícitamente la correlación espacial, así como los modelos econométricos, que resultan útiles para capturar la heterogeneidad espacial y sus efectos sobre variables económicas de interés. Para cada enfoque describimos tanto los fundamentos teóricos como su implementación práctica. En el caso del Kriging, también incluimos extensiones donde los pesos se calculan mediante métodos no paramétricos; aunque el modelo subyacente es el mismo, el procedimiento de estimación varía, ofreciendo mayor flexibilidad frente a supuestos rígidos sobre la estructura de dependencia espacial.

Este enfoque integral permite combinar una comprensión conceptual del fenómeno con técnicas de modelado adecuadas, constituyendo un marco sólido para la predicción de datos espaciales.

1.1. Los datos espaciales

Los datos espaciales se caracterizan por estar vinculados a una unidad espacial específica, como coordenadas o áreas geográficas. Esto implica que, además del valor observado de la o las variables de interés, es esencial disponer de información sobre la ubicación o referencia geográfica asociada a cada observación. A partir de esta definición, surge naturalmente el concepto de **valor regionalizado** o **dato regionalizado**, que se entiende como el valor o realización de la variable de interés en el punto o lugar geográfico donde se registró.

Los datos espaciales pueden clasificarse según la forma en que se representa el dominio geográfico $D \subset \mathbb{R}^d$, es decir, la manera en que se indexan las localizaciones en el espacio (Longley, 2005). Una primera distinción se basa en si el dominio es continuo o discreto. Cuando se asume un dominio **continuo**, el número de posibles localizaciones donde se puede observar la variable es no numerable, lo que implica que dicha variable podría medirse en cualquier punto del espacio (Zhang et al., 2014). Estos datos, conocidos como **datos geoestadísticos**, se caracterizan por ser realizaciones de un campo continuo, aunque en la práctica se trabaja con una discretización del dominio mediante puntos de muestreo finitos (Fischer y Wang, 2011). El interés principal en este tipo de datos suele centrarse en la interpolación espacial.

En contraste, cuando el dominio D es **discreto**, el conjunto de localizaciones es, a lo sumo, numerable (finito o infinito numerable). En este caso, los datos corresponden a mediciones realizadas en ubicaciones específicas, sin posibilidad de observaciones intermedias. Este tipo se denomina **datos reticulares** (o *lattice data*, en inglés), y ejemplos típicos incluyen tasas de mortalidad por región, deserción escolar por institución o el color de píxeles en imágenes satelitales. Es posible transformar datos geoestadísticos a reticulares, por ejemplo, mediante un teselado del dominio continuo y utilizando los centroides de las celdas como puntos de observación. No obstante, esta aproximación puede alterar la estructura de dependencia espacial (Zhao y Wall, 2004), y el proceso inverso (interpolación desde datos reticulares) generalmente no es válido.

Existe un tercer tipo de datos espaciales, donde el dominio no es fijo sino generado por un proceso aleatorio. En este contexto se habla de **patrones puntuales** (*point pattern data*), donde las localizaciones mismas son aleatorias. Ejemplos de este tipo incluyen la ubicación de árboles, nidos, epicentros de terremotos o caídas de rayos.

En este trabajo, consideramos únicamente el análisis de datos espaciales definidos sobre un dominio D fijo. Nos enfocaremos en el estudio de varias variables medidas en ubicaciones representadas por coordenadas $\mathbf{s} \in D \subset \mathbb{R}^d$. Este enfoque se enmarca dentro del estudio de **funciones aleatorias**, que se desarrolla formalmente en la Sección 1.2.

1.2. Funciones aleatorias

Con el objetivo de enfocar nuestro análisis en un contexto estocástico, consideraremos que los datos espaciales surgen de un mecanismo aleatorio que conceptualizamos como funciones aleatorias y se definen en esta sección.

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, con Ω el espacio muestral, $\mathcal{F}$ es el σ -álgebra y $\mathbb{P}$ la medida de probabilidad y $D \subset \mathbb{R}^d$ un conjunto de localizaciones. Se llama **función aleatoria** al conjunto

$$\mathcal{Y} \doteq \{Y(\mathbf{s}) : \mathbf{s} \in D \text{ e } Y(\mathbf{s}) : \Omega \rightarrow \mathbb{R} \text{ es una variable aleatoria}\}.$$

Esto es, una función aleatoria $\mathcal{Y}$ puede entenderse como una función $\mathcal{Y} : D \rightarrow \mathbb{R}^\Omega$, donde $\mathbb{R}^\Omega$ es una colección de variables aleatorias $Y(\mathbf{s}) : \Omega \rightarrow \mathbb{R}$ y para cada $\mathbf{s} \in D$, $\mathcal{Y}(\mathbf{s}) \doteq Y(\mathbf{s})$.

Consideremos ahora p funciones aleatorias simultáneas, a las cuales llamaremos,

$$\mathcal{X}_u : D \rightarrow \mathbb{R}^\Omega, \text{ con } u = 1, \dots, p,$$

donde para cada $\mathbf{s} \in D$ y para cada $u = 1, \dots, p$ definimos $\mathcal{X}_u(\mathbf{s}) \doteq X_u(\mathbf{s})$ sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

En esta tesis, para cada localización $\mathbf{s}$ de la región de interés asumiremos que contamos con una colección de variables aleatorias $\{Y(\mathbf{s}), X_1(\mathbf{s}), \dots, X_p(\mathbf{s})\}$. En este sentido, si fijamos $\nu \in \Omega$, obtenemos una realización de las funciones aleatorias, que sigue siendo una función de $\mathbf{s}$, por lo que si la evaluamos en un conjunto de sitios de la región de interés hallamos los **datos espaciales**.

Esta forma de considerar los datos como realización de funciones aleatorias en distintos sitios, nos permite tener en cuenta diferentes fuentes de variabilidad. Por un lado, la variabilidad del fenómeno en cada sitio a partir de la variación de los $\nu \in \Omega$, esto nos permite hallar una matriz de correlación para modelar dicha variabilidad. Esta correlación se la llama intrínseca para diferenciarla de la variabilidad inducida por la cercanía. Esto es, la variación de la ubicación en el espacio de dichos objetos, nos posibilita la segunda fuente de variabilidad, en este caso también la modelamos con una función pero de los sitios $\mathbf{s}$. Estos conceptos se formalizan en la siguiente sección.

1.2.1. Momentos

Teniendo ya una definición formal de las funciones aleatorias, en esta sección nos enfocaremos en la teoría de correlación de funciones aleatorias (Yaglom, 1987). Esto significa que estudiaremos los momentos de primer y segundo orden de la función. Vale aclarar que si bien presentamos las definiciones para las funciones $\mathcal{X}_u$, para cada $u = 1, \dots, p$, las mismas valen para cualquier función aleatoria.

Llamamos primer momento de la función $\mathcal{X}_u$ a la **función media**, denotada por $\mu_u : D \rightarrow \mathbb{R}$ y definida por,

$$\mu_u(\mathbf{s}) \doteq \mathbb{E}[X_u(\mathbf{s})] \quad \text{para todo } \mathbf{s} \in D.$$

Luego, definimos el vector $\boldsymbol{\mu}(\mathbf{s})$ como

$$\boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \dots, \mu_p(\mathbf{s}))^T.$$

A su vez, definimos tres momentos de segundo orden: la varianza, la covarianza y el variograma. Llamamos **varianza**, a la función $\sigma_u^2 : D \rightarrow \mathbb{R}^+$ dada por

$$\sigma_u^2(\mathbf{s}) \doteq \mathbb{E}[(X_u(\mathbf{s}) - \mu_u(\mathbf{s}))^2].$$

Luego, definimos el siguiente vector

$$\boldsymbol{\sigma}^2(\mathbf{s}) = (\sigma_1^2(\mathbf{s}), \dots, \sigma_p^2(\mathbf{s}))^T,$$

donde ordenamos las varianzas para cada función de la colección. Además, si $\mathbb{E}[X_u(\mathbf{s})^2] < \infty$ para todo $\mathbf{s} \in D$, llamamos **función matriz de covarianza** o simplemente **matriz de covarianza** a la función $\mathbf{C} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{p \times p}$ definida por

$$\mathbf{C}(\mathbf{s}, \mathbf{t}) = \begin{pmatrix} C_{11}(\mathbf{s}, \mathbf{t}) & C_{12}(\mathbf{s}, \mathbf{t}) & \cdots & C_{1p}(\mathbf{s}, \mathbf{t}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1}(\mathbf{s}, \mathbf{t}) & C_{p2}(\mathbf{s}, \mathbf{t}) & \cdots & C_{pp}(\mathbf{s}, \mathbf{t}) \end{pmatrix},$$

donde, para todo $\mathbf{s}, \mathbf{t} \in D$,

$$C_{uv}(\mathbf{s}, \mathbf{t}) \doteq \text{Cov}(X_u(\mathbf{s}), X_v(\mathbf{t})). \quad (1.2.1)$$

De forma matricial, definiendo $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), \dots, X_p(\mathbf{s}))^T$, tenemos que

$$\mathbf{C}(\mathbf{s}, \mathbf{t}) = \mathbb{E}[\mathbf{X}(\mathbf{s})\mathbf{X}(\mathbf{t})^T] - \mathbb{E}[\mathbf{X}(\mathbf{s})]\mathbb{E}[\mathbf{X}(\mathbf{t})^T].$$

Podemos observar que los elementos diagonales $C_{uu}(\mathbf{s}, \mathbf{t}) = \text{Cov}(X_u(\mathbf{s}), X_u(\mathbf{t}))$ de la matriz de covarianza $\mathbf{C}(\mathbf{s}, \mathbf{t})$ son las **funciones de autocovarianza** de las variables aleatorias $X_u(\cdot)$, las cuales modelan la dependencia espacial de la variable $X_u(\cdot)$, medida en diferentes localizaciones. Si $\mathbf{s} = \mathbf{t}$, $C_{uu}(\mathbf{s}, \mathbf{s}) = \sigma_u^2(\mathbf{s})$. Por otro lado, la matriz de covarianza $\mathbf{C}(\mathbf{s}, \mathbf{t})$ incluye también las llamadas **funciones de covarianza cruzadas** $C_{uv}(\mathbf{s}, \mathbf{t}) = \sigma_{uv}(\mathbf{s})$ de dos funciones aleatorias $X_u(\mathbf{s})$ y $X_v(\mathbf{t})$ definidas por la ecuación (1.2.1) si $u \neq v$. En este caso modelamos la dependencia espacial entre diferentes variables medidas en diferentes localizaciones. Cuando $\mathbf{s} = \mathbf{t}$, obtenemos la **matriz de covarianza intrínseca**,

$$\mathbf{C}(\mathbf{s}, \mathbf{s}) = \text{Var}(\mathbf{X}(\mathbf{s})) \quad \text{para cada } \mathbf{s} \in D. \quad (1.2.2)$$

Por otro lado, definimos la **función matriz de correlación normalizada** o simplemente **matriz de correlación** a la función $\mathbf{R} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{p \times p}$ definida por

$$\mathbf{R}(\mathbf{s}, \mathbf{t}) = (\mathbf{I}\boldsymbol{\sigma}(\mathbf{s}))^{-1}\mathbf{C}(\mathbf{s}, \mathbf{t})(\mathbf{I}\boldsymbol{\sigma}(\mathbf{t}))^{-1},$$

donde $\mathbf{I} \in \mathbb{R}^{p \times p}$ es la matriz identidad. Los elementos de $\mathbf{R}(\mathbf{s}, \mathbf{t})$ están dados por

$$\mathbf{R}_{uv}(\mathbf{s}, \mathbf{t}) = \frac{C_{uv}(\mathbf{s}, \mathbf{t})}{\sigma_u(\mathbf{s})\sigma_v(\mathbf{t})}, \quad (1.2.3)$$

los cuales son una extensión del coeficiente de correlación de Pearson. De forma análoga la matriz de covarianza, a los elementos de la diagonal $\mathbf{R}_{uu}(\mathbf{s}, \mathbf{t})$ los llamamos **funciones de autocorrelación** y al resto $\mathbf{R}_{uv}(\mathbf{s}, \mathbf{t})$ **funciones de correlación cruzada**.

Otra forma de medir dependencia espacial es a partir de la función variograma cruzado o simplemente variograma, el cual se define como la varianza de la diferencia $X_u(\mathbf{s}) - X_v(\mathbf{t})$, esto es,

$$\gamma_{uv}(\mathbf{s}, \mathbf{t}) = \text{Var}(X_u(\mathbf{s}) - X_v(\mathbf{t})).$$

De la misma manera que el resto de momentos, lo podemos expresar como una función bivariada de las ubicaciones espaciales $\mathbf{s}$ y $\mathbf{t}$. Esto es, llamamos **función matriz de variograma** o simplemente **matriz de variograma** a la función $\mathbf{\Gamma}(\mathbf{s}, \mathbf{t}) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{p \times p}$ definida por

$$\mathbf{\Gamma}(\mathbf{s}, \mathbf{t}) = \begin{pmatrix} \gamma_{11}(\mathbf{s}, \mathbf{t}) & \gamma_{12}(\mathbf{s}, \mathbf{t}) & \cdots & \gamma_{1p}(\mathbf{s}, \mathbf{t}) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p1}(\mathbf{s}, \mathbf{t}) & \gamma_{p2}(\mathbf{s}, \mathbf{t}) & \cdots & \gamma_{pp}(\mathbf{s}, \mathbf{t}) \end{pmatrix}.$$

1.2.2. Estacionariedad

Hacer inferencia sobre la distribución de una función aleatoria a partir de una única realización, es una tarea difícil si no se imponen restricciones adicionales sobre la clase de funciones consideradas (Stein, 1999). Un supuesto ampliamente adoptado en la literatura consiste en imponer alguna forma de **estacionariedad**, entendida como la invariancia (en distribución o en momentos) frente a traslaciones de la función. Intuitivamente, la estacionariedad refleja la idea de que las propiedades de la distribución, o de los momentos de la función, son homogéneas en el espacio: observar la función en distintas regiones del dominio es, de alguna manera (ya que depende del tipo de estacionariedad), informativamente equivalente. Este supuesto permite, en particular, utilizar una única trayectoria para inferir características poblacionales, dado que valores observados en diferentes localizaciones espaciales se consideran realizaciones de la misma estructura probabilística.

Suponiendo que $D = \mathbb{R}^d$, diremos que X_u es **estrictamente estacionaria** si para todo n , $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathbb{R}^d$, $x_{1u}, \dots, x_{nu} \in \mathbb{R}$ y $\mathbf{h} \in \mathbb{R}^d$,

$$\mathbb{P}(X_u(\mathbf{s}_1 + \mathbf{h}) \leq x_{1u}, \dots, X_u(\mathbf{s}_n + \mathbf{h}) \leq x_{nu}) = \mathbb{P}(X_u(\mathbf{s}_1) \leq x_{1u}, \dots, X_u(\mathbf{s}_n) \leq x_{nu}).$$

Otra forma de definir la estacionariedad es en términos de los dos primeros momentos. Se dice que una función aleatoria X_u es **débilmente estacionaria** si su valor esperado

$\mathbb{E}[X_u(\mathbf{s})] = \mu_u$ es constante con respecto a $\mathbf{s}$ y la función de covarianza depende solo de la norma $\|\mathbf{s} - \mathbf{t}\|$, para todo u, v , esto es,

$$\text{Cov}(X_u(\mathbf{s}), X_v(\mathbf{t})) \doteq C_{uv}^*(\|\mathbf{s} - \mathbf{t}\|),$$

donde $C_{uv}^* : \mathbb{R} \rightarrow \mathbb{R}$. En este caso, todas las funciones $X_u(\mathbf{s})$, $u = 1, \dots, p$, son funciones estacionarias, y cualquier par de ellas posee la propiedad de que su función de correlación cruzada depende solo de $\|\mathbf{s} - \mathbf{t}\|$. Dos funciones aleatorias que poseen esta última propiedad se dicen *correlacionadas estacionariamente (entre sí)*. Por último diremos que, una función aleatoria X_u es **intrínsecamente estacionaria** si el valor esperado de la diferencia $\mathbb{E}[X_u(\mathbf{s}) - X_v(\mathbf{t})]$ es 0 y la función de variograma depende de la norma $\|\mathbf{s} - \mathbf{t}\|$, para todo u, v , esto es

$$\text{Var}(X_u(\mathbf{s}) - X_v(\mathbf{t})) \doteq \gamma_{uv}^*(\|\mathbf{s} - \mathbf{t}\|),$$

donde $\gamma_{uv}^* : \mathbb{R} \rightarrow \mathbb{R}$.

1.3. Modelos y métodos autorregresivos para predicción espacial

La predicción espacial consiste en hallar el valor de una variable de interés en lugares donde dicha variable no ha sido observada utilizando datos disponibles u observados en ubicaciones concretas del espacio. Para dicho procedimiento es indispensable identificar y modelar las relaciones espaciales presentes en los datos. Para ello, existen distintos modelos y métodos que difieren en su complejidad, sus supuestos y sus aplicaciones. Esta sección tiene como objetivo describir las principales metodologías utilizadas en la literatura para la predicción de datos espaciales donde se emplean modelos con términos autorregresivos.

Con el objetivo de hallar el mejor predictor de la variable respuesta, tradicionalmente se utiliza la técnica de minimizar el error cuadrático medio entre el valor real y el valor predicho. Un resultado clásico establece que, dada la información disponible, el predictor que logra ese mínimo es la esperanza condicional del valor a predecir respecto de dicha información (Ferguson, 1967).

Para clarificar esta idea, consideramos una muestra $\{Y(\mathbf{s}_i), X_1(\mathbf{s}_i), \dots, X_p(\mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{O}_n}$ donde, dado un conjunto de sitios $\mathcal{O}_n \doteq \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ con $\mathbf{s}_i \in D$, definimos la variable respuesta por

$$Y_i \doteq Y(\mathbf{s}_i),$$

y las covariables, por

$$X_{iu} \doteq X_u(\mathbf{s}_i) \quad \text{para } i = 1, \dots, n, \quad u = 1, \dots, p.$$

Para el caso de las variables X_u con $u = 1, \dots, p$, definimos para cada sitio $\mathbf{s}_i$ el vector aleatorio

$$\mathbf{X}_i \doteq \mathbf{X}(\mathbf{s}_i) = (X_{i1}, \dots, X_{ip})^T.$$

Consideremos por ahora solo la variable respuesta Y_i , más adelante veremos cómo, dependiendo de cada metodología, se utilizan las variables predictoras $\mathbf{X}_i$. Sea Y_0 , con $\mathbf{s}_0 \notin \mathcal{O}_n$, el valor de la variable que se quiere predecir e

$$\mathbf{Y} \doteq (Y_1, \dots, Y_n)^T \quad (1.3.1)$$

, el vector de los valores observados. Se sabe que, bajo el supuesto de normalidad de los datos, esto es,

$$\begin{pmatrix} Y_0 \\ \mathbf{Y} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_0 \\ \boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{0\mathbf{Y}} \\ \boldsymbol{\Sigma}_{\mathbf{Y}0} & \boldsymbol{\Sigma}_{\mathbf{Y}} \end{pmatrix} \right),$$

el predictor que minimiza el error cuadrático medio de la predicción está dado por

$$\mathbb{E}[Y_0|\mathbf{Y}] = \mu_0 + \boldsymbol{\Sigma}_{0\mathbf{Y}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}(\mathbf{Y} - \boldsymbol{\mu}). \quad (1.3.2)$$

Este predictor es llamado predictor lineal insesgado óptimo (BLUP, por sus siglas en inglés). Robinson (1991) presenta una revisión teórica sobre diferentes formas de deducirlo, ejemplos y diferentes aplicaciones.

En el contexto de la estadística espacial, diversos autores abordan el problema de la predicción desde distintas perspectivas. Algunos se apoyan explícitamente en la noción del BLUP, como los kriging y los basados en modelos mixtos (Welham et al., 2004), mientras que otros arriban a la misma expresión del predictor, o a formulaciones equivalentes, a través de enfoques alternativos, como los modelos econométricos y el Lattice Kriging. Los modelos de econometría espacial introducen la dependencia espacial a través de matrices de pesos que reflejan la interacción entre unidades geográficas. Autores como Kelejian y Prucha (2007), Kato (2008) y Goulard et al. (2017) utilizan estas ideas con el objetivo de predecir en el contexto de datos faltantes. Por otro lado, se han desarrollado enfoques no paramétricos (Menezes et al., 2010; Dabo-Niang et al., 2014; Biau y Cadre, 2004; García Arancibia et al., 2023) que combinan la lógica del kriging con técnicas de estimación de pesos mediante núcleos. Esto permite relajar supuestos paramétricos estrictos sobre la estructura de dependencia espacial en contextos de heterogeneidad.

En lo que sigue, se presentarán estos enfoques, destacando tanto su formulación como su interpretación en términos de predicción espacial, subrayando las similitudes y diferencias y su relación con el BLUP.

1.3.1. Modelos basados en Kriging

Dentro de las aplicaciones que encontramos en Robinson (1991) está al método Kriging. El autor sostiene que en el contexto de la geoestadística dicho método surge independientemente del BLUP pero que las estimaciones son esencialmente las mismas. En esta sección presentamos las características principales de este método y de los que surgen de esta idea.

1.3.1.1. Kriging

Sea $\{Y_i\}_{\mathbf{s}_i \in \mathcal{O}_n}$ una muestra resultante de una función aleatoria $Y(\mathbf{s})$ que satisface que

$$\begin{aligned}\mathbb{E}[Y(\mathbf{s})] &= \mu(\mathbf{s}), & \mathbf{s} &\in D \\ \text{Cov}(Y(\mathbf{s}), Y(\mathbf{t})) &= C(\mathbf{s}, \mathbf{t}). & \mathbf{s}, \mathbf{t} &\in D\end{aligned}$$

Sea Y_0 con $\mathbf{s}_0 \notin \mathcal{O}_n$ la variable de interés en el sitio donde queremos predecir, $\mathbf{s}_0$. El predictor kriging para dicha variable es una combinación lineal pesada de los datos observados, dada por

$$\hat{Y}_0 = \sum_{\mathbf{s}_i \in \mathcal{O}_n} \omega(\mathbf{s}_i, \mathbf{s}_0) Y_i. \quad (1.3.3)$$

Considerando el vector de pesos $\boldsymbol{\omega} \doteq (\omega_1, \dots, \omega_n)^T \in \mathbb{R}^n$, con $\omega_i \doteq \omega(\mathbf{s}_i, \mathbf{s}_0)$, podemos escribir el predictor de manera matricial como

$$\hat{Y}_0 = \boldsymbol{\omega}^T \mathbf{Y}. \quad (1.3.4)$$

Al buscar el predictor BLUP, es decir, al buscar los pesos $\boldsymbol{\omega}$ de manera tal que el predictor sea insesgado y de mínima varianza, se deduce (ver Lovatto et al., 2022, cap. 2) el siguiente sistema de $n + 1$ ecuaciones con $n + 1$ incógnitas,

$$\begin{cases} \boldsymbol{\omega}^T \mathbf{1} = 1 \\ \boldsymbol{\Sigma}_Y \boldsymbol{\omega} - \tau \mathbf{1} = \mathbf{c}_0, \end{cases}$$

donde la primera ecuación nos asegura que el predictor sea insesgado, esto es, $\mathbb{E}[\hat{Y}_0] = \mu_0$ y la segunda que la varianza del error de predicción $\sigma_e^2 \doteq \text{Var}(\hat{Y}_0 - Y_0)$ sea mínima. Para ello, si $\boldsymbol{\Sigma}_Y$ es la matriz de varianza-covarianza del vector $\mathbf{Y}$, dada por

$$\boldsymbol{\Sigma}_Y = \begin{bmatrix} C(\mathbf{s}_1, \mathbf{s}_1) & C(\mathbf{s}_1, \mathbf{s}_2) & \cdots & C(\mathbf{s}_1, \mathbf{s}_n) \\ C(\mathbf{s}_2, \mathbf{s}_1) & C(\mathbf{s}_2, \mathbf{s}_2) & \cdots & C(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(\mathbf{s}_n, \mathbf{s}_1) & C(\mathbf{s}_n, \mathbf{s}_2) & \cdots & C(\mathbf{s}_n, \mathbf{s}_n) \end{bmatrix}, \quad (1.3.5)$$

el vector $\mathbf{c}_0 = (C(\mathbf{s}_1, \mathbf{s}_0), \dots, C(\mathbf{s}_n, \mathbf{s}_0))^T$, τ es un multiplicador de Lagrange y $\mathbf{1} \in \mathbb{R}^n$ un vector de unos, resolviendo el sistema resulta

$$\boldsymbol{\omega} \doteq \boldsymbol{\Sigma}_Y^{-1} \left(\mathbf{c}_0 - \frac{\mathbf{1}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{c}_0 - 1}{\mathbf{1}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{1}} \mathbf{1} \right).$$

Una de las principales ventajas con las que cuenta kriging radica en que considera la estructura de la correlación espacial de los datos (Montero et al., 2015). Esto es, poder aplicar este método implica modelar la estructura de correlación del fenómeno a estudiar, para lo cual es necesaria la estimación de la función $C(\cdot, \cdot)$. 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.

Para obtener las predicciones $\hat{Y}_0$, mediante el cálculo de los pesos, necesitamos estimar la función de covarianza $C(\cdot, \cdot)$ de la variable objetivo con lo cual obtenemos tanto Σ_Y , mediante (1.3.5), como $\mathbf{c}_0$. Para ello, y en función del objetivo del problema, es necesario hacer supuestos sobre algunos componentes del siguiente modelo

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad (1.3.6)$$

definido para cada $\mathbf{s} \in D$, donde $\mu(\mathbf{s}) = \mathbb{E}[Y(\mathbf{s})]$ es la función media, $Z(\mathbf{s})$ es una función aleatoria con esperanza cero y función de covarianza $\text{Cov}(Z(\mathbf{s}), Z(\mathbf{t}))$, para todo $\mathbf{s}, \mathbf{t} \in D$, y $\varepsilon(\mathbf{s})$ es otra función aleatoria con esperanza cero y función de covarianza $\text{Cov}(\varepsilon(\mathbf{s}), \varepsilon(\mathbf{t})) \doteq \sigma^2$ si $\mathbf{s} = \mathbf{t}$ y cero en otro caso. El caso clásico llamado **kriging ordinario** supone que la función media es constante y $Z(\mathbf{s})$ es intrínsecamente estacionaria. Esto implica que $Y(\mathbf{s})$ es intrínsecamente estacionaria, lo cual implica que

$$\begin{aligned} \mathbb{E}[Y(\mathbf{s}) - Y(\mathbf{t})] &= 0 & \mathbf{s}, \mathbf{t} \in D, \\ \text{Var}(Y(\mathbf{s}) - Y(\mathbf{t})) &\doteq \gamma_Y^*(\|\mathbf{s} - \mathbf{t}\|) & \mathbf{s}, \mathbf{t} \in D. \end{aligned}$$

Por ejemplo, si $h \doteq \|\mathbf{s} - \mathbf{t}\|$, $\gamma_Y^*(\|\mathbf{s} - \mathbf{t}\|)$ se puede modelar con una función analítica dada por un modelo exponencial; esto es,

$$\gamma_Y^*(h; \boldsymbol{\theta}) = \begin{cases} 0, & \text{si } h = 0, \\ \theta_1 + \theta_2 \left[1 - \exp\left(-\frac{h}{\theta_3}\right) \right], & \text{si } h > 0, \end{cases} \quad (1.3.7)$$

donde $\gamma_Y^* : \mathbb{R} \rightarrow \mathbb{R}$ y $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ son el *efecto pepita*, la *meseta parcial* y el *rango*, respectivamente. El efecto pepita es la varianza del error o error de medición. La meseta parcial es la varianza de la función diferencia, con lo cual es un indicador de dependencia espacial. El rango es la distancia a partir de la cual dejamos de contar con dependencia espacial. Dichos parámetros se estiman mediante máxima verosimilitud o mínimos cuadrados generalizados. Una vez obtenidas las estimaciones, se construyen la matriz $\mathbf{\Gamma}_Y \in \mathbb{R}^{n \times n}$, cuyo ij -ésimo elemento está dado por $\gamma_Y^*(\|\mathbf{s}_i - \mathbf{s}_j\|)$, para cada $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{O}_n$, y el vector $\boldsymbol{\gamma}_{0Y} = (\gamma_Y^*(\|\mathbf{s}_1 - \mathbf{s}_0\|), \dots, \gamma_Y^*(\|\mathbf{s}_n - \mathbf{s}_0\|))^T \in \mathbb{R}^n$. Con todo esto, obtenemos el predictor de kriging, dado por

$$\begin{aligned} \hat{Y}_0 &= \boldsymbol{\omega}^T \mathbf{Y} \\ &= \left(\boldsymbol{\gamma}_{0Y} - \frac{\mathbf{1}^T \mathbf{\Gamma}_Y^{-1} \boldsymbol{\gamma}_{0Y} - 1}{\mathbf{1}^T \mathbf{\Gamma}_Y^{-1} \mathbf{1}} \mathbf{1} \right)^T \mathbf{\Gamma}_Y^{-1} \mathbf{Y} \\ &= \alpha \mathbf{1}^T \mathbf{\Gamma}_Y^{-1} \mathbf{Y} + \boldsymbol{\gamma}_{0Y}^T \mathbf{\Gamma}_Y^{-1} \mathbf{Y}, \end{aligned} \quad (1.3.8)$$

$$\text{con } \alpha \text{ a } -\frac{\mathbf{1}^T \mathbf{\Gamma}_Y^{-1} \boldsymbol{\gamma}_{0Y} - 1}{\mathbf{1}^T \mathbf{\Gamma}_Y^{-1} \mathbf{1}}.$$

Recordemos que para obtener este predictor, hemos supuesto que la función media $\mu(\mathbf{s})$ en (1.3.6) es constante en $\mathbf{s}$ y la función $Z(\mathbf{s})$ intrínsecamente estacionaria. En muchos casos, estas condiciones pueden ser restrictivas, por lo que surgen extensiones del kriging para flexibilizar estos supuestos, los cuales se describen a continuación. Para más detalles sobre este predictor consultar Lovatto et al. (2022).

1.3.1.2. Kriging con modelo de tendencia

En ocasiones, se puede incluir información auxiliar al modelo para mejorar la precisión de las predicciones y, en otras, para flexibilizar supuestos. La información mencionada puede surgir de la conexión entre la variable a predecir y otras variables, que hemos denominado covariables, o de la relación entre la variable objetivo y los lugares en los cuales ha sido muestreada. Para estos casos existe el kriging con media desconocida (Chiles y Delfiner, 1999) o kriging con un modelo de tendencia, también conocido como **external drift** (Bourennane et al., 1996), Hudson y Wackernagel (1994), Bourennane et al. (2000). Aquí pueden reconocerse al menos tres enfoques distintos, el kriging universal, que utiliza funciones de las coordenadas, el kriging con tendencia externa y el kriging de regresión que utilizan covariables.

El **kriging universal** (UK), introducido originalmente por Matheron (1969), constituye un caso particular que, a diferencia del kriging ordinario, permite considerar una función media no constante. En este enfoque, la media se modela como una función de las coordenadas espaciales, esto es, dadas $f_0, f_1, \dots, f_q$ funciones determinísticas que dependen de las coordenadas geográficas $\mathbf{s} \in D$ y $q \in \mathbb{N}$, $\mu(\mathbf{s})$ se modela como

$$\mu(\mathbf{s}) = \sum_{j=0}^q a_j f_j(\mathbf{s}), \quad (1.3.9)$$

Como consecuencia de (1.3.9), del modelo (1.3.6) e imitando la fórmula del predictor propuesto en (1.3.4), tenemos que

$$\hat{Y}_0 = \boldsymbol{\omega}^T (\mathbf{F}\mathbf{a} + \mathbf{G}),$$

donde

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

$\mathbf{a} = (1, a_1, \dots, a_q)^T$ y, a partir del modelo (1.3.6) definimos $G(\mathbf{s}_i) \doteq Z(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i)$ con $\mathbf{s}_i \in \mathcal{O}_n$ y $\mathbf{G} = (G(\mathbf{s}_1), \dots, G(\mathbf{s}_n))^T$. En este caso, solo podemos asegurar que la función G es intrínsecamente estacionaria por lo que para obtener el vector $\boldsymbol{\gamma}_{0\mathbf{G}}$ y la matriz $\boldsymbol{\Gamma}_{\mathbf{G}}$, estimamos γ_G^* , el variograma de la función error G . Luego,

$$Y_0 = \boldsymbol{\omega}^T \mathbf{Y}$$

$$= \alpha \gamma_{0\mathbf{G}}^{-1} \mathbf{Y} + \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y}, \quad (1.3.10)$$

donde α está dado por

$$\alpha = \left(\mathbf{f}_0^T - \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{F}^T \right) \left(\mathbf{F}^T \mathbf{\Gamma}_{\mathbf{Y}}^{-1} \mathbf{F} \right)^{-1} \mathbf{F}^T,$$

con $\mathbf{f}_0 \doteq (1, f_1(\mathbf{s}_0), \dots, f_q(\mathbf{s}_0))^T \in \mathbb{R}^{q+1}$. Otros autores (Wackernagel, 2006; Papritz y Stein, 1999) también coinciden en que el término **kriging universal** debe reservarse para el caso en el que sólo se usan coordenadas como información auxiliar.

Si la función media se define externamente mediante covariables, se utiliza el término **kriging con tendencia externa** (**kriging with external drift**, KED) o **external trend** (Wackernagel, 2006; Chiles y Delfiner, 1999). Este es, probablemente, el enfoque más utilizado actualmente para kriging con información auxiliar (Hengl et al., 2003). Recordemos que estamos considerando una muestra $\{Y_i, \mathbf{X}_i\}_{\mathbf{s}_i \in \mathcal{O}_n} \in \mathbb{R} \times \mathbb{R}^p$. En este caso, a diferencia de la matriz $\mathbf{F}$ construída utilizando funciones de las coordenadas, construimos la matriz de datos $\mathbb{X}_*$ a partir de la muestra de covariables $\{\mathbf{X}_i\}_{\mathbf{s}_i \in \mathcal{O}_n}$, la cual queda definida por

$$\mathbb{X}_* \doteq (\mathbf{1} \quad \mathbb{X}) \in \mathbb{R}^{n \times (p+1)},$$

donde

$$\mathbb{X} \doteq \begin{pmatrix} X_1(\mathbf{s}_1) & \dots & X_p(\mathbf{s}_1) \\ X_1(\mathbf{s}_2) & \dots & X_p(\mathbf{s}_2) \\ \vdots & \vdots & \ddots \\ X_1(\mathbf{s}_n) & \dots & X_p(\mathbf{s}_n) \end{pmatrix} \in \mathbb{R}^{n \times p}. \quad (1.3.11)$$

Operando de la misma manera que para el kriging universal, el predictor queda dado por

$$\hat{Y}_0 = \alpha \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y} + \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y}, \quad (1.3.12)$$

y

$$\alpha = \left(\mathbf{b}_0^T - \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right) \left(\mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right)^{-1} \mathbb{X}_*^T,$$

donde $\mathbf{b}_0 = (1, X_{1m}, \dots, X_{pm})^T$. Puede observarse que si $p = 1$ tanto en el caso del kriging universal (1.3.10) como en el kriging con tendencia externa (1.3.12), obtenemos el kriging ordinario (1.3.8). Por otro lado, si reescribimos el predictor, obtenemos

$$\begin{aligned} \hat{Y}_0 &= \left(\mathbf{b}_0^T - \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right) \left(\mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right)^{-1} \mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y} + \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y} \\ &= \mathbf{b}_0^T \left(\mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right)^{-1} \mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y} \\ &\quad - \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \left(\mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right)^{-1} \mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y} + \gamma_{0\mathbf{G}}^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y}. \end{aligned}$$

Luego definiendo

$$\mu_0 = \mathbf{b}_0^T \left(\mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbb{X}_* \right)^{-1} \mathbb{X}_*^T \mathbf{\Gamma}_{\mathbf{G}}^{-1} \mathbf{Y}$$

y

$$\boldsymbol{\mu} = \mathbb{X}_* \left(\mathbb{X}_*^T \boldsymbol{\Gamma}_G^{-1} \mathbb{X}_* \right)^{-1} \mathbb{X}_*^T \boldsymbol{\Gamma}_G^{-1} \mathbf{Y},$$

obtenemos una expresión similar al BLUP, dada por,

$$\begin{aligned} \hat{Y}_0 &= \mu_0 - \gamma_{0G}^T \boldsymbol{\Gamma}_G^{-1} \boldsymbol{\mu} + \gamma_{0G}^T \boldsymbol{\Gamma}_G^{-1} \mathbf{Y} \\ &= \mu_0 + \gamma_{0G}^T \boldsymbol{\Gamma}_G^{-1} (\mathbf{Y} - \boldsymbol{\mu}). \end{aligned}$$

De la misma manera, llegamos a esta forma para el caso del kriging universal.

Por otro lado, la función media y la función error G , también pueden ajustarse por separado y luego sumarse. Un procedimiento similar fue sugerido por primera vez por Ahmed y de Marsily (1987). Odeh et al. (1994, 1995) lo denominaron **kriging de regresión** (RK), mientras que Goovaerts (1999) utiliza el término **kriging después de eliminar la tendencia**. El predictor de dicha propuesta está dado por

$$\hat{Y}_0 = \mathbf{b}_0^T \mathbf{a} + \boldsymbol{\omega}^T \mathbf{G},$$

donde $\mathbf{a}$ y $\boldsymbol{\omega}$ son los pesos a estimar. Para ello, y considerando que la función error G no es observable, se plantea un método iterativo donde se estima $\mathbf{a}$ con mínimos cuadrados para obtener los residuos que estiman la función G , esto es,

$$\hat{\mathbf{G}} = \mathbf{Y} - \mathbb{X}_* \hat{\mathbf{a}},$$

luego se aplica kriging ordinario sobre dichos residuos. Con ello se estima $\boldsymbol{\Gamma}_G$ que se utiliza para reestimar $\mathbf{a}$ a partir de mínimos cuadrados generalizados.

1.3.1.3. Cokriging

Si el número de covariables es bajo y no están disponibles en todas las unidades geográficas, el predictor **co-kriging (CK)** es una alternativa para mejorar la predicción (Hengl et al., 2003). Dados $\mathbf{Y}$ y $\mathbb{X}$, definimos el predictor cokriging como

$$\hat{Y}_0 = \boldsymbol{\omega}_1^T \mathbf{Y} + \sum_{j=2}^p \boldsymbol{\omega}_j^T \mathbb{X}_{.j},$$

donde $\mathbb{X}_{.j}$ representa la columna j de la matriz $\mathbb{X}$. Bajo estacionariedad se tiene que $\mathbb{E}[\mathbf{Y}] = \mu \mathbf{1}$, $\mathbb{E}[\mathbb{X}_{.j}] = \mu_j \mathbf{1}$ y

$$\begin{aligned} \mathbb{E}[\hat{Y}_0] &= \mathbb{E}[\boldsymbol{\omega}_1^T \mathbf{Y}] + \mathbb{E}\left[\sum_{j=2}^p \boldsymbol{\omega}_j^T \mathbb{X}_{.j}\right] \\ &= \boldsymbol{\omega}_1^T \mathbb{E}[\mathbf{Y}] + \sum_{j=2}^p \boldsymbol{\omega}_j^T \mathbb{E}[\mathbb{X}_{.j}] \end{aligned}$$

$$= \mu \boldsymbol{\omega}_1^T \mathbf{1} + \sum_{j=2}^p \mu_j \boldsymbol{\omega}_j^T \mathbf{1}.$$

Luego, para que el predictor sea insesgado, esto es, $\mathbb{E}[\hat{Y}_0] = \mu$ es necesario que

$$\begin{cases} \boldsymbol{\omega}_1^T \mathbf{1} = 1 \\ \boldsymbol{\omega}_i^T \mathbf{1} = 0 \quad i = 2, \dots, p. \end{cases} \quad (1.3.13)$$

Por lo tanto, bajo las condiciones dadas en la ecuación anterior (1.3.13) obtenemos el mejor predictor lineal insesgado minimizando la varianza del error, dada por

$$\mathbb{E} \left[\left(\hat{Y}_0 - \mu \boldsymbol{\omega}_1^T \mathbf{1} - \sum_{j=2}^p \mu_j \boldsymbol{\omega}_j^T \mathbf{1} \right)^2 \right],$$

sujeta a (1.3.13). En principio, este problema no es más difícil que el kriging ordinario, excepto que hay más multiplicadores de Lagrange $m_1, \dots, m_p$. Las ecuaciones de cokriging, basadas en la covarianza, está dada por

$$\begin{aligned} \sum_{j=1}^p \sum_{i=1}^{n_j} \omega_{ij} C_{jj'}(\mathbf{s}_i, \mathbf{s}_{i'}) - m_{j'} &= C_{1j'}(\mathbf{s}_0, \mathbf{s}_{i'}) \quad i' = 1, \dots, n_j, \quad j' = 1, \dots, p \\ \sum_{i=1}^{n_1} \omega_{i1} &= 1, \quad \sum_{i=1}^{n_j} \omega_{ij} = 0, \quad j = 2, \dots, p. \end{aligned}$$

Esto es, $(n+1)p$ ecuaciones lineales con $(n+1)p$ incógnitas.

Para todos los casos descritos hasta ahora, es necesario estimar las matrices de variograma o covariograma, según corresponda. Para ello se proponen funciones paramétricas, como el ejemplo dado en (1.3.7). Estas funciones dependen de los sitios donde están medidos los datos observados y de los parámetros $\boldsymbol{\theta}$. En la práctica, dichos parámetros, llamados efecto pepita, meseta y rango, pueden estimarse de distintas maneras. Una aproximación clásica es la *estimación basada en momentos*, que ajusta un modelo teórico al variograma empírico, generalmente mediante mínimos cuadrados ponderados o métodos gráficos. Otra posibilidad es la *máxima verosimilitud* (ML) o la *verosimilitud restringida* (REML), que tratan a las observaciones como realizaciones de un proceso gaussiano y buscan los parámetros que maximizan la función de verosimilitud. También existen enfoques *bayesianos*, que incorporan información a priori y producen distribuciones posteriores de los parámetros. Cada método tiene ventajas y limitaciones: los momentos son sencillos pero menos eficientes, ML y REML son más rigurosos estadísticamente aunque computacionalmente más costosos, y los bayesianos permiten cuantificar mejor la incertidumbre a costa de mayor complejidad. Para que estas opciones cumplan con lo que proponen son necesarios diferentes supuestos. Para relajar los mismos se ha propuesto la siguiente metodología usando estadística no paramétrica.

1.3.1.4. Kriging no paramétrico

En esta sección presentamos el predictor que denominamos kriging no paramétrico, desarrollado en García Arancibia et al. (2023). Su construcción sigue la idea general del kriging (Ecuación (1.3.3)), proponiendo tres formas de estimar los pesos de forma no paramétrica. Aunque este predictor no incorpora covariables, ocupa un papel central en los resultados principales de este trabajo. La propuesta se apoya en ideas presentes en Biau (2003); Menezes et al. (2010); Dabo-Niang et al. (2016), entre otros. Sin embargo, dichos trabajos no se describen aquí en detalle porque algunos no consideran covariables y otros no incorporan términos autorregresivos, por lo que sus estructuras difieren de la que se estudia en esta sección.

Recordemos que $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ son los datos observados tal que $\mathbf{s}_i \in \mathcal{O}_n$ e Y_0 la variable que queremos predecir. Luego, de la misma manera que en (1.3.3), el predictor de Y_0 con $\mathbf{s}_0 \notin \mathcal{O}_n$ es una combinación lineal de los datos disponibles dada por

$$\hat{Y}_0 = \sum_{\mathbf{s}_i \in \mathcal{O}_n} \omega(\mathbf{s}_i, \mathbf{s}_0) Y_i.$$

Considerando diferentes formas (no paramétricas) para los pesos ω_i , obtendremos diferentes versiones de predictores que serán estudiadas a continuación. Para describir el comportamiento de dichas versiones, utilizamos el ejemplo presentado en la Figura 1.1, el cual es una realización de la función $Y_{\mathbf{s}}$ gaussiano, no estacionario, con función de covarianza exponencial, con $d = 2$ y $n = 99$. En el panel (a) de dicha figura consideramos desconocido

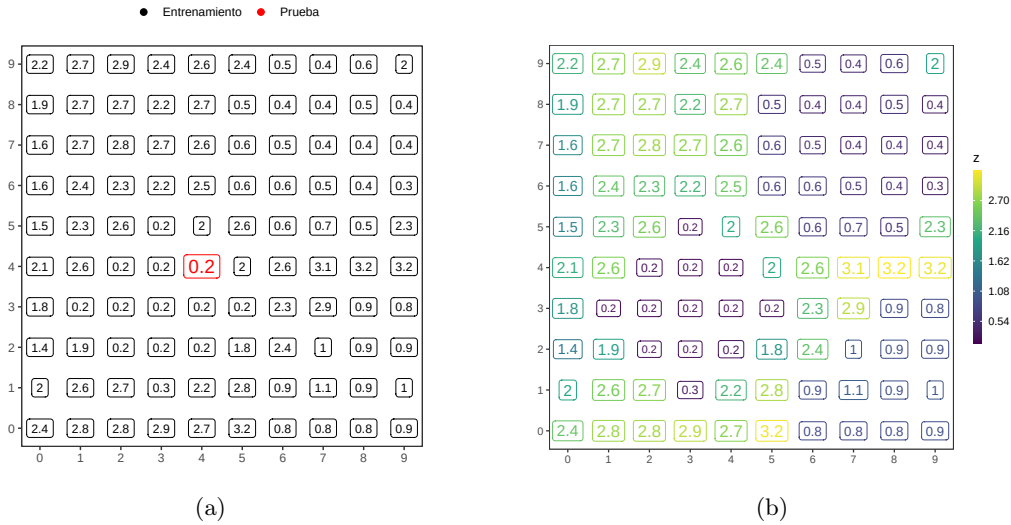


Figura 1.1: (a) Conjunto de valores regionalizados simulados a partir de una función aleatoria gaussiana $Y_{\mathbf{s}}$, no estacionario con función de covarianza exponencial para $n = 99$ y $d = 2$. (b) Comportamiento de los valores regionalizados según su ubicación.

el valor regionalizado en el sitio $\mathbf{s}_0 = (4, 4)$, es decir, suponemos que el valor $y_0 \approx 0.2$ es

desconocido y será predicho a partir del valor de la variable de interés en los sitios restantes. En el panel (b) de la misma figura agregamos una escala de colores para visualizar la distribución de tales valores regionalizados.

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 $\omega_h(\mathbf{s}_0, \mathbf{s}_i)$ consisten en un sólo núcleo que controla la distancia espacial entre los puntos muestreados y el sitio donde deseamos realizar la predicción, esto es

$$\omega_h(\mathbf{s}, \mathbf{s}_i) \doteq \frac{K_1\left(\frac{d(\mathbf{s}, \mathbf{s}_i)}{h_n}\right)}{\sum_{j=1}^n K_1\left(\frac{d(\mathbf{s}, \mathbf{s}_j)}{h_n}\right)}, \quad (1.3.14)$$

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

En la Figura 1.2 (a) mostramos el comportamiento de los pesos (1.3.14), tomando como K_1 una función de núcleo gaussiana y h el parámetro de suavizado elegido mediante validación cruzada de 5 submuestras, 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 1.2 (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.

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 los valores de la variable de interés en sitios vecinos.

Para ello necesitamos definir los sitios vecinos. Sea $k \in \mathbb{N}$ un parámetro fijo que indica la cantidad de vecinos a considerar. Para cada sitio $\mathbf{s}_i$, denotamos por $\mathbf{s}_{i(k)} \in \mathcal{O}_n$ al k -ésimo punto más cercano a $\mathbf{s}_i$, excluyendo al propio $\mathbf{s}_i$. Decimos que $\mathbf{s}_j$ es **vecino** de $\mathbf{s}_i$ si se cumple que

$$0 < d(\mathbf{s}_i, \mathbf{s}_j) \leq d(\mathbf{s}_i, \mathbf{s}_{i(k)}),$$

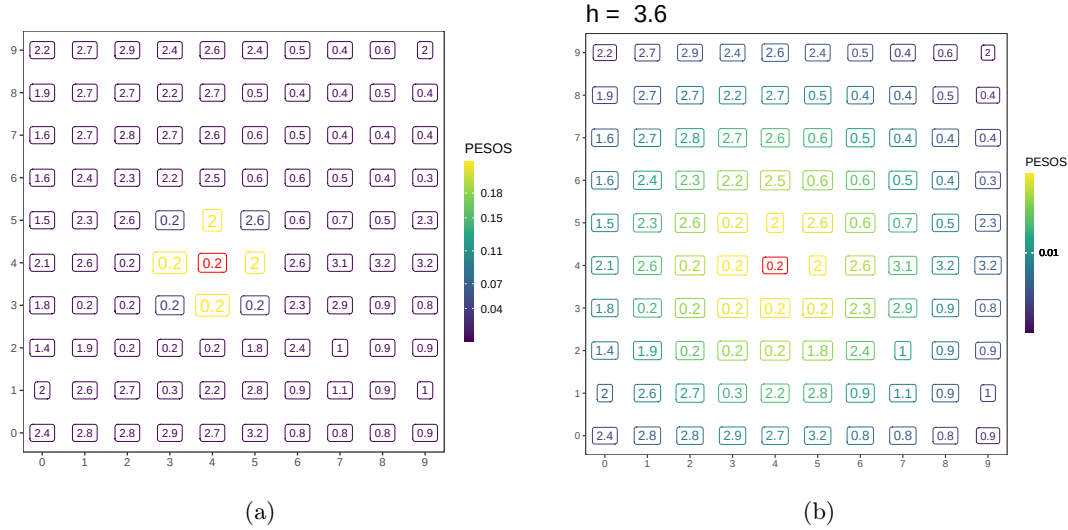


Figura 1.2: Comportamiento de los pesos del predictor K1S sobre el conjunto de valores regionalizados para: (a) el parámetro de suavizado óptimo obtenido a través de validación cruzada, (b) diferentes valores del parámetro de suavizado.

donde $d(\mathbf{s}_i, \mathbf{s}_j)$ representa la distancia euclídea entre los puntos $\mathbf{s}_i$ y $\mathbf{s}_j$. Luego, $\mathbf{T}_i \in \mathbb{R}^k$ definido por

$$\mathbf{T}_i = (T_{i(1)}, \dots, T_{i(k)}), \quad (1.3.15)$$

es el **vector de vecindad** o **vecindario**, esto es, el vector de valores de la variable a predecir medida en los k vecinos más cercanos a $\mathbf{s}_i$.

Definimos la medida de similitud,

$$d_m(\mathbf{T}_0, \mathbf{T}_i) \doteq |\text{Med}(\mathbf{T}_0) - \text{Med}(\mathbf{T}_i)|, \quad (1.3.16)$$

donde $\text{Med}(\mathbf{T}_i)$ es la mediana del vector $\mathbf{T}_i$. De esta manera, los pesos quedan dados por

$$\omega_{\mathbf{h}}(\mathbf{T}_0, \mathbf{T}_i) \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_0, \mathbf{s}_i)}{h_{1n}} \right) K_2 \left(\frac{d_m(\mathbf{T}_0, \mathbf{T}_i)}{h_{2n}} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_0, \mathbf{s}_j)}{h_{1n}} \right) K_2 \left(\frac{d_m(\mathbf{T}_0, \mathbf{T}_j)}{h_{2n}} \right)}, \quad (1.3.17)$$

donde $\mathbf{h} = \{h_{1n}, h_{2n}, k\}$. Observemos que este predictor está basado en el predictor introducido por Biau y Cadre (2004) el cual compara las distancias entre vectores de vecinos. La diferencia principal radica en que Biau y Cadre (2004) consideran la distancia entre vectores, necesitando así utilizar un núcleo multivariado. En esta nueva versión sin embargo, proponemos la distancia entre las medianas de dichos valores necesitando de esta manera un núcleo univariado.

En la Figura 1.3 (a) observamos el comportamiento de los pesos dados en la Ecuación (1.3.17) sobre los valores regionalizados. Los mismos fueron calculados tomando como K_1 y K_2 funciones núcleos gaussianas y, para la construcción del vector $\mathbf{T}$, 8-vecinos más

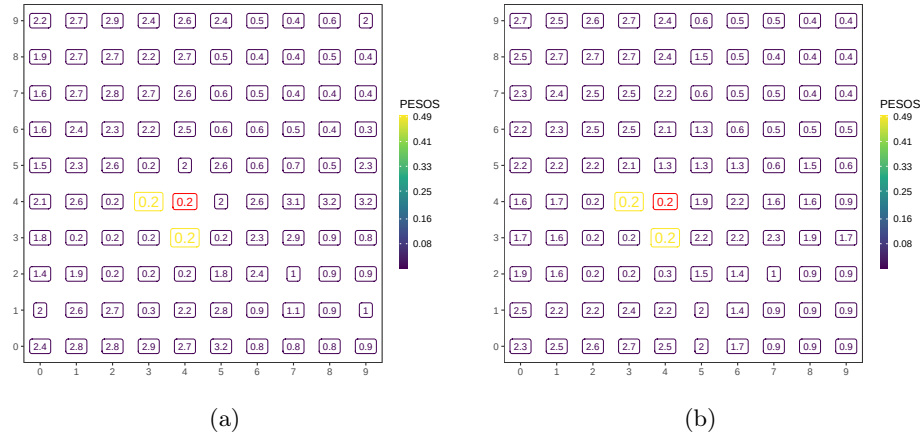


Figura 1.3: Comportamiento de los pesos del predictor K2ME sobre el conjunto de: (a) valores regionalizados, (b) medianas de los valores vecinos para cada sitio.

cercanos (lo cual se mantiene para los casos siguientes). En la Figura 1.3 (b) observamos en cada sitio s_i la mediana del vector $\mathbf{T}_i$.

En este caso los pesos más grandes están asociados a observaciones para las cuales ambas distancias son pequeñas, ya que están construidos por el producto de dos funciones núcleo. Sin embargo, basta con que una sola de las distancias sea grande (o un parámetro de suavizado pequeño) para que la función núcleo correspondiente se anule y arrastre el peso de la observación, sin importar el valor del otro núcleo. Por eso, un sitio con vecindario similar pero geográficamente alejado recibe un peso bajo, y lo mismo ocurre con un sitio cercano cuya mediana difiere considerablemente de la del sitio a predecir.

Este comportamiento puede observarse en la Figura 1.3 (a) y (b), donde solamente un par de sitios cercanos a s_0 reciben un peso alto, por estar espacialmente próximos a s_0 y tener una mediana del vecindario similar a la de éste. También se observa que el sitio ubicado en la coordenada (4, 5), a la misma distancia de s_0 que el sitio en (3, 4), recibe un peso menor debido a que la distancia entre medianas es grande.

Considerando la autocorrelación supuesta y 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—, esperamos que la mediana *sea representativa* de la distribución de la variable Y_0 : que el valor de la mediana asignado al sitio s_0 sea similar al valor, no observable, de la variable que deseamos predecir. Bajo dicha hipótesis, esta versión del predictor puede captar el comportamiento de la variable de interés incluso cuando sus valores son muy disímiles entre sitios cercanos.

En la Figura 1.3 observamos que el valor regionalizado en s_0 (panel (a)) y la mediana $Med(\mathbf{T}_0)$ (panel (b)) son iguales, lo que confirma que en este caso la hipótesis se cumple.

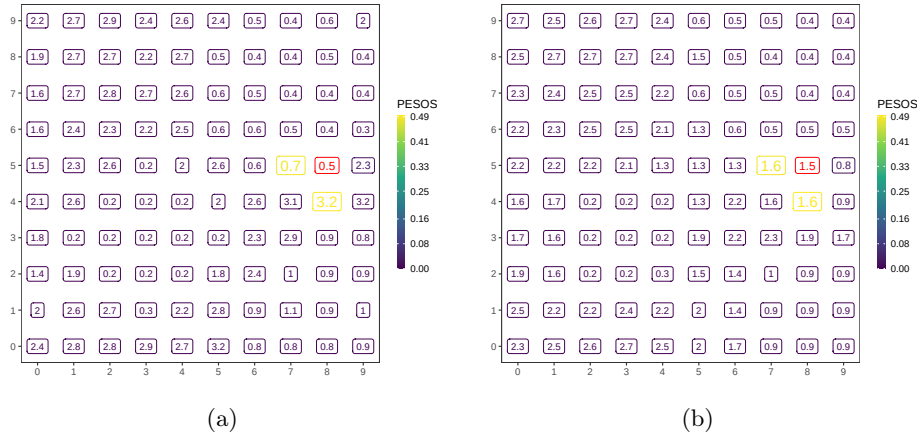


Figura 1.4: Comportamiento de los pesos del predictor $K2ME$ sobre: (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra.

La elección del sitio $\mathbf{s}_0 = (4, 4)$ como sitio a predecir no fue aleatoria sino intencional: en la Figura 1.1, donde coloreamos en función de la variable de interés, dicho sitio representa un límite en el comportamiento de la variable, ya que su vecindario es fuertemente heterogéneo.

En la Figura 1.4 (a) observamos la misma realización del proceso mostrado en la Figura 1.1, pero cambiando el sitio a predecir para ejemplificar el caso en que la mediana *no sea representativa* de la distribución de la variable Y_0 . La coordenada del nuevo sitio es $\mathbf{s}_0 = (8, 5)$, cuya particularidad es que la mediana de los valores del vecindario, $Med(\mathbf{T}_0) = 1.5$, difiere marcadamente del valor de la variable en $\mathbf{s}_0$, que es 0.5. Este tipo de comportamiento, en el que los valores vecinos a $\mathbf{s}_0$ son muy diferentes al valor en el propio sitio, se conoce como autocorrelación espacial de datos disímiles (Anselin, 2010) o autocorrelación negativa (Griffith y Arbia, 2010).

En consecuencia, en la Figura 1.4 (a) el comportamiento de los pesos no es el esperado: aparecen pesos altos asociados a valores poco similares al que queremos predecir, como el del sitio (8, 4). Esto muestra que, si bien enriquecer la versión no paramétrica con la similaridad entre vecindarios mejora su desempeño en general, los pesos se comportan de manera errática ante evidencia de autocorrelación negativa, es decir, cuando una observación está rodeada de valores muy diferentes a la propia.

Teniendo en cuenta lo anteriormente expuesto, en el próximo apartado consideramos un predictor cuyos pesos tienen en cuenta la interacción entre la distancia geográfica y la distancia entre medianas, pero considerando un solo núcleo.

Kriging de un núcleo Multiplicativo (K1M)

En dirección a lo anteriormente expuesto, introducimos en esta sección los pesos de

núcleo multiplicativo los cuales están dados por

$$\omega_{\mathbf{h}}(\mathbf{T}_0, \mathbf{T}_i) \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_0, \mathbf{s}_i) * d_m(\mathbf{T}_0, \mathbf{T}_i)}{h_n} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_0, \mathbf{s}_j) * d_m(\mathbf{T}_0, \mathbf{T}_j)}{h_n} \right)},$$

que, a diferencia de los pesos dados en la Ecuación (1.3.17), 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 y k , esto es, $\mathbf{h} = \{h_n, k\}$. Esta estrategia brinda una ventaja computacional al requerir la búsqueda de un solo parámetro de suavizado en contraposición con el de dos núcleos.

En la Figura 1.5 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 1.3 y 1.4 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 1.5 (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 1.5 (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 Y_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.

1.3.1.5. Lattice Kriging (LK)

Un método desarrollado para gestionar conjuntos de datos grandes es el **Lattice Kriging** (LK) y fue desarrollado por Nychka et al. (2015). Este modelo se centra en la capacidad de modelar estructuras complejas de dependencia espacial, en particular aquellas que presentan varias escalas de correlación espacial, mediante la separación del proceso espacial modelado en múltiples *capas*, cada una con su propia resolución espacial y escala de correlación.

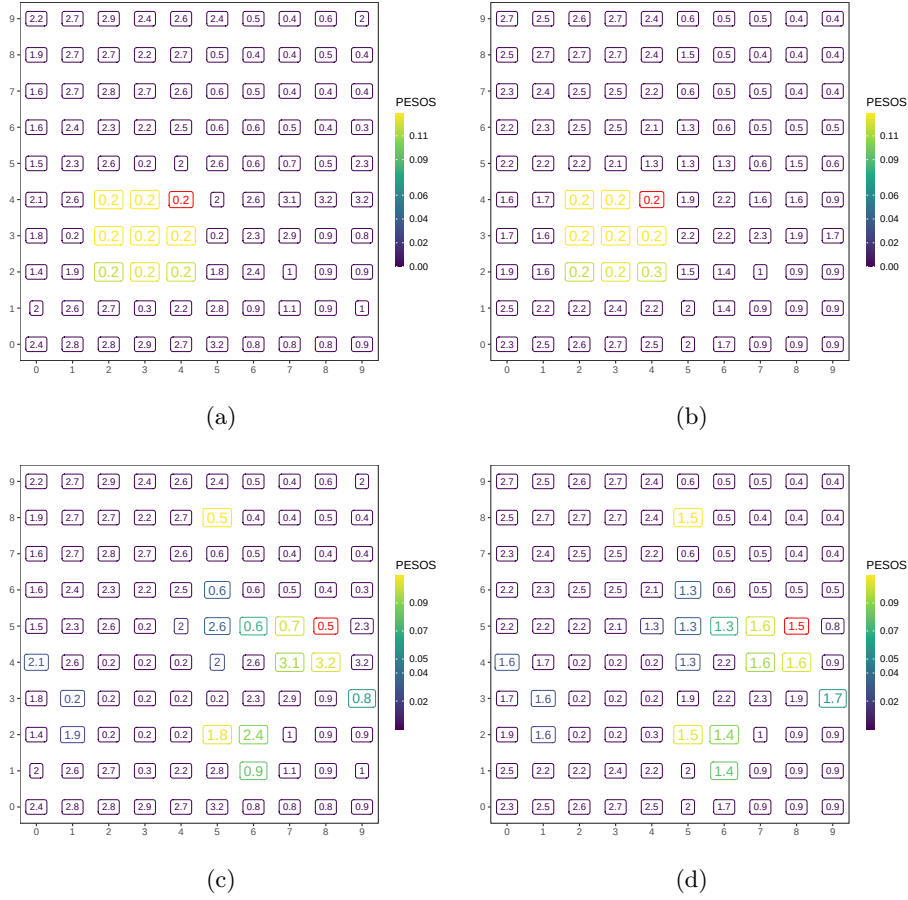


Figura 1.5: Panel superior: Comportamiento de los pesos del predictor K1M para el sitio $\mathbf{s}_0 = (4, 4)$ sobre (a) el conjunto de valores regionalizados, (b) el conjunto de medianas de los vecindarios para cada sitio de la muestra. Panel inferior: Comportamiento de los pesos del predictor K1M para el sitio $\mathbf{s}_0 = (8, 5)$ sobre (c) el conjunto de valores regionalizados, (d) el conjunto de medianas de los vecindarios para cada sitio de la muestra.

Para cada sitio $\mathbf{s}_i \in \mathcal{O}_n$, este método propone utilizar el modelo

$$Y_i = \mathbb{X}_i^T \boldsymbol{\beta} + g(\mathbf{s}_i) + \varepsilon_i. \quad (1.3.18)$$

Si consideramos el modelo dado en (1.3.6), en este contexto se modela la función media μ a partir de $\mu(\mathbf{s}_i) = X_i^T \boldsymbol{\beta}$, la función ε como $\varepsilon(\mathbf{s}_i) = \varepsilon_i$, con $\varepsilon_i \sim N(0, \sigma^2)$, independiente de g . Por último, la función Z en $\mathbf{s}_i$ se modela con la función g y el objetivo es predecir $g(\mathbf{s}_0)$ en sitios no observados $\mathbf{s}_0$. Para ello, el aporte de este enfoque radica en la estimación de la función g a partir de su descomposición en una serie de grillas con resolución espacial creciente sobre las cuales se evalúan funciones base. Formalmente,

$$g(\mathbf{s}_i) = \sum_{l=1}^L g_l(\mathbf{s}_i) = \sum_{l=1}^L \sum_{j=1}^{m(l)} c_j^l \phi_{l,j}(\mathbf{s}_i), \quad \mathbf{s}_i \in D \subset \mathbb{R}^2.$$

En esta expresión, L denota la cantidad de capas de grillas, usualmente entre 2 y 4. Las fun-

ciones $g_1, \dots, g_L$ corresponden a funciones espaciales suaves asociadas a cada grilla, representada cada una de ellas como una combinación lineal de las funciones base $\phi_{l,1}, \dots, \phi_{l,m(l)}$, cuyos coeficientes c_j^l son variables aleatorias.

Aquí las $\phi_{j,l}$ son funciones de base radiales de soporte compacto (Funciones de Wendland), organizadas en grillas de resolución creciente (cada nivel duplica la densidad de nodos). Los coeficientes $\mathbf{c}_l = (c_l^1, \dots, c_l^{m(l)})^T$ provienen de una función aleatoria gaussiana de Markov (GMRF, por su siglas en inglés) con matriz de precisión sparse, esto es,

$$\mathbf{c}_l \sim N(0, \rho \mathbf{Q}_l^{-1}), \quad \mathbf{Q}_l = \mathbf{B}_l^T \mathbf{B}_l / \alpha_l, \quad (1.3.19)$$

donde $\mathbf{B}_l$ define un modelo autorregresivo espacial sobre la grilla, α_l es un peso de nivel y ρ un parámetro de escala global. La matriz de precisión $\mathbf{Q}_l$, definida como la inversa de la matriz de covarianza, es elegida ya que es sparse, y eso reduce enormemente los costos de cálculo.

Si escribimos el modelo (1.3.18) de forma matricial, y considerando $\mathbf{g} = (g(\mathbf{s}_1), \dots, g(\mathbf{s}_n))^T$ y $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, tenemos que

$$\mathbf{Y} = \mathbb{X}\boldsymbol{\beta} + \mathbf{g} + \boldsymbol{\varepsilon}.$$

Con ello, utilizando la distribución de los coeficientes dados en (1.3.19) y considerando $\boldsymbol{\Phi}$ como la matriz cuyas columnas indexan las funciones base y cuyas filas indexan las ubicaciones de los sitios observados, esto es, $\Phi_{ij} = \phi_j(\mathbf{s}_i)$, tenemos por un lado que, $\mathbb{E}[\mathbf{Y}|\mathbb{X}] = \mathbb{X}\boldsymbol{\beta}$, ya que $\mathbb{E}[\mathbf{g} + \boldsymbol{\varepsilon}] = \mathbf{0}$ por (1.3.19) y por el supuesto de media cero de $\boldsymbol{\varepsilon}$. Por otro lado tenemos que

$$\begin{aligned} \text{Var}(\mathbf{Y}|\mathbb{X}) &= \text{Var}(\mathbf{g} + \boldsymbol{\varepsilon}) \\ &= \text{Var}(\mathbf{g}) + \text{Var}(\boldsymbol{\varepsilon}) && \text{por independencia} \\ &= \rho \boldsymbol{\Phi} \mathbf{Q}^{-1} \boldsymbol{\Phi}^T + \sigma^2 \mathbf{I}. \end{aligned}$$

Luego, tenemos que

$$\mathbf{Y}|\mathbb{X} \sim N(\mathbb{X}\boldsymbol{\beta}, \rho \boldsymbol{\Phi} \mathbf{Q}^{-1} \boldsymbol{\Phi}^T + \sigma^2 \mathbf{I}),$$

La estimación de los parámetros se realiza mediante máxima verosimilitud iterada y perfiles de verosimilitud. Para ρ y $\mathbf{Q}^{-1}$ fijos, la estimación de máxima verosimilitud de $\boldsymbol{\beta}$ coincide también con la estimación de mínimos cuadrados generalizados (GLS), dado por

$$\hat{\boldsymbol{\beta}} = (\mathbb{X}^T \mathbf{A}_\lambda^{-1} \mathbb{X})^{-1} \mathbb{X}^T \mathbf{A}_\lambda^{-1} \mathbf{Y},$$

con $\mathbf{A}_\lambda = \boldsymbol{\Phi} \mathbf{Q}^{-1} \boldsymbol{\Phi}^T + \lambda \mathbf{I}$ y $\lambda = \sigma^2 / \rho$. La predicción de la función g_l en un punto $\mathbf{s}_0$ es

$$\hat{g}_l(\mathbf{s}_0) = \sum_{j=1}^m \hat{c}_j^l \phi_{l,j}(\mathbf{s}_0), \quad \text{con} \quad \hat{\mathbf{c}}_l = \mathbf{Q}^{-1} \boldsymbol{\Phi}^T \mathbf{A}_\lambda^{-1} (\mathbf{Y} - \mathbb{X} \hat{\boldsymbol{\beta}}).$$

Es decir, un predictor lineal en las bases, pero los coeficientes se encuentran modelados por la estructura de dependencia impuesta por GMRF. Esto implica que los pesos se ajustan de forma coherente con las relaciones locales definidas por la matriz subyacente para cada capa $\mathbf{B}_l$, lo cual introduce un efecto de suavizado en la predicción y garantiza que ésta respete la estructura espacial del modelo. Entonces, el predictor queda dado por

$$\hat{\mathbf{Y}}_0 = \mathbb{X}_0 \hat{\boldsymbol{\beta}} + \boldsymbol{\Phi}_0 \mathbf{Q}^{-1} \boldsymbol{\Phi}_0^T \mathbf{A}_\lambda^{-1} (\mathbf{Y} - \mathbb{X} \hat{\boldsymbol{\beta}}).$$

Aunque el modelo presentado en esta sección no pertenece estrictamente a la clase de modelos autorregresivos, se lo incluye debido a su naturaleza semiparamétrica y a su reconocida capacidad para capturar estructuras complejas de dependencia espacial. En efecto, la combinación de un componente paramétrico asociado a las covariables con un proceso espacial latente permite modelar de manera flexible tanto los efectos sistemáticos como la correlación entre observaciones, lo cual lo convierte en una herramienta particularmente eficaz para el análisis y la predicción en contextos espaciales.

1.3.2. Modelos econométricos

En el contexto de los modelos econométricos, la predicción espacial surge con el problema del tratamiento de datos faltantes, un desafío metodológico relevante (LeSage y Pace, 2004; Kelejian y Prucha, 2007; Kato, 2008, 2013; Goulard et al., 2017). En particular, cuando el modelo incluye errores espaciales estructurados, como los modelos autorregresivos condicionales y simultáneos, la estimación de parámetros en presencia de datos ausentes requiere técnicas específicas que aprovechen la estructura de dependencia espacial.

Dada la muestra $\{Y_i\}_{\mathbf{s}_i \in \mathcal{O}_n}$, se define el modelo autorregresivo simultáneo mediante

$$Y_i = \mu_i + \sum_{\substack{\mathbf{s}_j \in \mathcal{O}_n \\ \mathbf{s}_j \neq \mathbf{s}_i}} b_{ij} (Y_j - \mu_j) + \varepsilon_i,$$

donde $\varepsilon_i \sim \mathcal{N}(0, \lambda_i)$, $\mathbb{E}(Y_i) = \mu_i$, y b_{ij} son constantes conocidas o desconocidas, con $b_{ii} = 0$ para $i = 1, \dots, n$. Este modelo se llama *simultáneo* porque se modelan simultáneamente las Y_i con el resto del ellas y los términos de error ε_i estarán correlacionados con $\{Y_j : j \neq i\}$. Si n es finito, consideramos la matriz $\mathbf{B}$ que contiene los b_{ij} , el vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$, $\mathbf{I}_n$ la matriz identidad de dimensión n e $\mathbf{Y}$ el vector de valores de la muestra, podemos escribir el modelo de forma matricial,

$$\mathbf{Y} = (\mathbf{I}_n - \mathbf{B})^{-1} \boldsymbol{\varepsilon} + \boldsymbol{\mu}.$$

Luego la distribución conjunta de $\mathbf{Y}$ está dada por

$$\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, (\mathbf{I}_n - \mathbf{B})^{-1} \boldsymbol{\Lambda} (\mathbf{I}_n - \mathbf{B}^T)^{-1}).$$

Con ello, y usando que $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ tenemos que la covarianza entre $\mathbf{Y}$ y $\boldsymbol{\varepsilon}$ está dada por

$$\begin{aligned}\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{Y}) &= \mathbb{E}[\boldsymbol{\varepsilon}\mathbf{Y}^T] \\ &= \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T(\mathbf{I}_n - \mathbf{B}^T)^{-1} + \boldsymbol{\varepsilon}\boldsymbol{\mu}^T] \\ &= \boldsymbol{\Lambda}(\mathbf{I}_n - \mathbf{B}^T)^{-1},\end{aligned}$$

y como $\boldsymbol{\Lambda}(\mathbf{I}_n - \mathbf{B}^T)^{-1}$ no es diagonal, probamos que $\boldsymbol{\varepsilon}$ e $\mathbf{Y}$ están correlacionadas.

Por otro lado, para poder definir el modelo autorregresivo condicional partimos de suponer que

$$Y_i | Y_{-i} \sim \mathcal{N}\left(\mu_i + \sum_{\substack{\mathbf{s}_j \in \mathcal{C}_n \\ \mathbf{s}_j \neq \mathbf{s}_i}} c_{ij}(Y_j - \mu_j), \tau_i^2\right), \quad (1.3.20)$$

donde $Y_{-i} = \{Y_j : j \neq i\}$, $\mathbb{E}(Y_i) = \mu_i$, τ_i^2 es la varianza condicional, y c_{ij} son constantes conocidas o desconocidas, en particular $c_{ii} = 0$, $i = 1, \dots, n$. De la misma manera que en el caso anterior $\mathbf{C}$ es la matriz cuyos elementos son las constantes c_{ij} y $\mathbf{M} = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_n^2)$, y por el teorema de factorización (Cressie, 1991, 6.4), se obtiene que

$$\mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}, (\mathbf{I}_n - \mathbf{C})^{-1}\mathbf{M}).$$

A partir de (1.3.20), podemos escribir el modelo de forma matricial como

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{C}(\mathbf{Y} - \boldsymbol{\mu}) + \boldsymbol{\nu},$$

donde $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)^T$ tal que $\mathbb{E}[\nu_i | Y_{-i}] = 0$ y $\text{Var}(\nu_i | Y_{-i}) = \tau_i^2$, para $i = 1, \dots, n$. Luego con $\boldsymbol{\nu} = (\mathbf{I} - \mathbf{C})(\mathbf{Y} - \boldsymbol{\mu})$, la varianza de $\boldsymbol{\nu}$ está dada por

$$\begin{aligned}\text{Var}(\boldsymbol{\nu}) &= (\mathbf{I} - \mathbf{C})\text{Var}(\mathbf{Y})(\mathbf{I} - \mathbf{C})^T \\ &= (\mathbf{I} - \mathbf{C})(\mathbf{I} - \mathbf{C})^{-1}\mathbf{M}(\mathbf{I} - \mathbf{C})^T \\ &= \mathbf{M}(\mathbf{I} - \mathbf{C}^T),\end{aligned}$$

y la covarianza

$$\begin{aligned}\text{Cov}(\boldsymbol{\nu}, \mathbf{Y}) &= \mathbb{E}[\boldsymbol{\nu}\mathbf{Y}^T] \\ &= \mathbb{E}[\boldsymbol{\nu}\boldsymbol{\nu}^T(\mathbf{I} - \mathbf{C}^T)^{-1} + \boldsymbol{\nu}\boldsymbol{\mu}^T] \\ &= \mathbf{M}(\mathbf{I} - \mathbf{C}^T)(\mathbf{I} - \mathbf{C}^T)^{-1} \\ &= \mathbf{M}.\end{aligned}$$

Luego como $\mathbf{M}$ es diagonal, ν_i es independiente de $\{Y_j, i \neq j\}$.

La estructura de $\mathbf{B}$ y $\mathbf{C}$ usualmente se especifica en función de la estructura de las unidades geográficas. Una forma común de construir $\mathbf{B}$ o $\mathbf{C}$ es usando un parámetro que escala una matriz de vecindad $\mathbf{W}$ definida por el investigador, la cual indica si las unidades son vecinas o no.

Retomando el objetivo de obtener predicciones, Griffith et al. (1989a) propusieron un enfoque iterativo. Dada una muestra $\{Y_i, \mathbb{X}_i\} \in \mathbb{R} \times \mathbb{R}^p$, $\mathbf{Y}_0$ el valor a predecir e $\mathbf{Y}$ y $\mathbb{X}$ los valores observados. Consideran $\boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta}$, $\mathbf{M} = \boldsymbol{\Lambda} = \sigma^2 \mathbf{I}$ y proponen

$$\mathbf{Y} \sim \mathcal{N}(\mathbb{X}\boldsymbol{\beta}, \sigma^2 \mathbf{V}).$$

Si $\mathbf{V} = (\mathbf{I} - \tau \mathbf{W})^{-1}$, es un modelo condicional y si $\mathbf{V} = [(\mathbf{I} - \rho \mathbf{W})(\mathbf{I} - \rho \mathbf{W}^T)]^{-1}$, es una modelo simultáneo. A partir de esto, estiman $\hat{\boldsymbol{\beta}} = (\mathbb{X}\mathbb{X}^T)^{-1}\mathbb{X}^T \mathbf{Y}$, hallan $\hat{\mathbf{Y}}_0 = \mathbb{X}_0 \hat{\boldsymbol{\beta}}$ y completan el conjunto de datos haciendo $\hat{\mathbf{Y}}^T = (\mathbf{Y}^T, \hat{\mathbf{Y}}_0^T)$. Una vez obtenidos los datos, estiman $\mathbf{V}$ mediante máxima verosimilitud y reestiman $\hat{\boldsymbol{\beta}} = (\mathbb{X}\hat{\mathbf{V}}^{-1}\mathbb{X}^T)^{-1}\mathbb{X}^T \hat{\mathbf{V}}^{-1} \mathbf{Y}$. Por último, usando el BLUP, actualizan los valores a predecir

$$\hat{\mathbf{Y}}_0 = \mathbb{X}_0 \hat{\boldsymbol{\beta}} + \hat{\mathbf{V}}_{0\mathbf{Y}} \hat{\mathbf{V}}_{\mathbf{Y}}^{-1} (\mathbf{Y} - \mathbb{X} \hat{\boldsymbol{\beta}}), \quad (1.3.21)$$

donde $\hat{\mathbf{V}} = \begin{pmatrix} \hat{\mathbf{V}}_{\mathbf{Y}} & \hat{\mathbf{V}}_{\mathbf{Y}0} \\ \hat{\mathbf{V}}_{0\mathbf{Y}} & \hat{\mathbf{V}}_{00} \end{pmatrix}$. Repiten el proceso con los nuevos valores predichos hasta alcanzar la convergencia.

LeSage y Pace (2004) abordan esta problemática en el contexto de modelos de error espacial (SEM), de autocorrelación espacial (SAR) y del modelo de Durbin (SDM). Para estimar los parámetros de dichos modelos, maximizan la verosimilitud de la distribución condicional de los datos observados, completados con la predicción de los valores faltantes obtenida mediante el BLUP, dada en (1.3.21). A diferencia de Griffith et al. (1989b), obtienen las estimaciones de $\boldsymbol{\beta}$, ρ y σ^2 utilizando una versión del algoritmo EM, en lugar de un procedimiento iterativo directo.

Kelejian y Prucha (2007) enfocan su estudio en la predicción dentro del modelo SDM, comparando predictores clásicos con nuevas propuestas menos eficientes con respecto al error cuadrático medio, pero más intuitivas. Considera diferentes escenarios donde varía la información conocida.

Por su parte, Kato (2008, 2013) examina el problema desde una perspectiva de simulación, analizando el rendimiento del BLUP bajo distintos modelos espaciales y distintas configuraciones del proceso generador de datos (incluso cuando el modelo está mal especificado). En particular, compara el desempeño del BLUP cuando se parte de un modelo geoestadístico frente a un modelo SEM o SDM, y analiza su desempeño tanto en predicción como en estimación.

En Goulard et al. (2017), retoman el concepto de BLUP y evidencian los retos que supone aplicar este directamente a predicciones simultáneas fuera de muestra en múltiples lugares. En respuesta a esto, sugieren tácticas *casi óptimas*, que son más fáciles de manejar desde el punto de vista computacional, y explican la manera en que se vinculan estas con algoritmos EM. La contribución fue doble; por una parte, compilaron las diversas fórmulas de predicción que existían en la literatura sobre modelos SAR; por otra parte, realizaron simulaciones con el fin de cotejar el rendimiento de los predictores óptimos, casi óptimos y

clásicos, subrayando en qué situaciones es más ventajoso cada uno. Hasta le incorporaron extensiones a la perspectiva que Kelejian y Prucha (2007) propusieron en un principio, aunque lo hicieron en un contexto más general y realista, con varias predicciones fuera de muestra y con aplicaciones prácticas.

Capítulo 2

Modelo autorregresivo semiparamétrico para predicción espacial

En este capítulo presentamos el modelo autorregresivo semiparamétrico para predicción espacial desarrollado en esta tesis y probamos las propiedades asintóticas de los estimadores de sus parámetros. Este modelo permite combinar flexibilidad en la captación de la dependencia espacial con la interpretabilidad propia de los modelos paramétricos.

El modelo se basa en el propuesto por Aneiros-Pérez y Vieu (2008) y la estimación del componente autorregresivo no paramétrico se basa en una noción de vecindario y se inspira en los resultados de Dabo-Niang et al. (2016) y García Arancibia et al. (2023), utilizando núcleos en los pesos que permiten controlar tanto la geometría espacial de los datos como la similitud entre vecindarios.

En las secciones siguientes describimos en detalle el modelo propuesto, el procedimiento de estimación y los resultados teóricos obtenidos. En particular, para el estimador paramétrico se establece un resultado de normalidad asintótica y, bajo hipótesis adicionales, la convergencia casi segura que permite caracterizar su tasa de convergencia. Asimismo, para el estimador no paramétrico se obtienen tasas de convergencia bajo condiciones regulares.

2.1. Modelo espacial semiparamétrico

En Aneiros-Pérez y Vieu (2008), los autores proponen un modelo autorregresivo semiparamétrico para series de tiempo, el cual predice la variable de interés en función de un período específico de la serie (modelado como dato funcional) y de covariables. Estos períodos pueden representar, por ejemplo, días o años. En este modelo, la componente

no paramétrica es la que los autores consideran autorregresiva y se define en función de datos funcionales. Estructuralmente, nuestra propuesta es similar pero la diferencia radica en que, en el contexto de datos espaciales, no disponemos del orden que caracteriza a las series de tiempo. Una aproximación espacial a la idea de período temporal podría ser suponer que la función aleatoria presenta el mismo comportamiento en diferentes zonas de la región de interés, lo cual no es tan fácilmente identificable ni demasiado realista. Además, no siempre contamos con observaciones medidas en grillas regulares, lo cual es una ventaja en los datos medidos en el tiempo. Por esta razón, nuestra propuesta consiste en reemplazar lo que los autores utilizan como períodos anteriores de la variable a predecir, en la componente autorregresiva, por vecindarios.

Para ello, sea $\{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}_{\mathbf{s}_i \in \mathbb{Z}^d}$ una muestra de vectores aleatorios de $(p + k + 1)$ variables definidas sobre algún espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde $Y_i \in \mathbb{R}$, $\mathbf{X}_i \in \mathbb{R}^p$ y $\mathbf{T}_i \in \mathbb{R}^k$. Consideramos que la variable de interés está observada en los sitios $\mathbf{s} = (i_1, \dots, i_d) \in \mathbb{Z}^d$ los cuales forman parte de una grilla definida por $\mathcal{D}_{\mathbf{n}} = \{\mathbf{s}, 1 \leq i_l \leq n_l, l = 1, \dots, d\}$ donde n_l es la cantidad de puntos en cada dirección, $\mathbf{n} = (n_1, \dots, n_d)$ y definimos $\hat{n} \doteq n_1 \times \dots \times n_d$ como el tamaño de la muestra. Por simplicidad asumimos que $n = n_1 = n_2 = \dots = n_d$ y que $\hat{n} \rightarrow \infty$ si $n \rightarrow \infty$. Si $\mathbf{s}_0 \in \mathcal{D}_{\mathbf{n}}$ entonces $\mathcal{O}_{\mathbf{n}} = \mathcal{D}_{\mathbf{n}} \setminus \{\mathbf{s}_0\}$. Para simplificar la notación, en lo que sigue prescindiremos del subíndice en la notación de las muestras consideraremos que $\{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}$ es la muestra de vectores en los sitios observados $\mathcal{D}_{\mathbf{n}}$.

En este contexto de datos dependientes, Y_i es una variable aleatoria no solo asociada al conjunto de covariables, X_{ij} ($j = 1, \dots, p$) sino también a ella misma mediante el vector $\mathbf{T}_i$ definido en (1.3.15). A partir de las definiciones anteriores, modelamos dichas relaciones mediante el modelo semiparamétrico autorregresivo, dado por

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + r(\mathbf{T}_i) + \varepsilon_i, \quad \forall \mathbf{s}_i \in \mathcal{O}_{\mathbf{n}} \quad (2.1.1)$$

donde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p$ es un vector de parámetros desconocidos, $r : \mathbb{R}^k \rightarrow \mathbb{R}$ una función suave desconocida y $\varepsilon_i \in \mathbb{R}$ son errores aleatorios, idénticamente distribuidos con $\mathbb{E}[\varepsilon_i | \mathbf{X}_i, \mathbf{T}_i] = 0$.

Observación 2.1.1. *El modelo presentado en (2.1.1) es una adaptación semiparamétrica y autorregresiva del modelo presentado en (1.3.6). Es decir, modelamos la función media $\mu(\mathbf{s})$ con una función paramétrica lineal en los parámetros y la función aleatoria $Z(\mathbf{s})$ con una función no paramétrica autorregresiva.*

Para obtener las predicciones, entonces, necesitamos estimar el vector de parámetros $\boldsymbol{\beta}$ y la función r , lo cual explicamos en la próxima sección.

2.2. Algoritmo de estimación y predicción

En esta sección explicitamos el algoritmo de estimación para las componentes del modelo (2.1.1) mediante una adaptación del método de Aneiros-Pérez y Vieu (2008) al caso de datos espaciales.

La esperanza condicional del modelo está dada por,

$$\begin{aligned}\mathbb{E}[Y_i|\mathbf{T}_i] &= \mathbb{E}[\mathbf{X}_i^T|\mathbf{T}_i]\boldsymbol{\beta} + r(\mathbf{T}_i) + \mathbb{E}[\varepsilon_i|\mathbf{T}_i] \\ &= \mathbb{E}[\mathbf{X}_i^T|\mathbf{T}_i]\boldsymbol{\beta} + r(\mathbf{T}_i),\end{aligned}\tag{2.2.1}$$

ya que hemos supuesto que $\mathbb{E}[\varepsilon_i|\mathbf{X}_i, \mathbf{T}_i] = 0$ lo cual implica que $\mathbb{E}[\varepsilon_i|\mathbf{T}_i] = 0$. Considerando las ideas de Robinson (1988), restamos (2.2.1) de (2.1.1) y obtenemos

$$Y_i - \mathbb{E}[Y_i|\mathbf{T}_i] = (\mathbf{X}_i - \mathbb{E}[\mathbf{X}_i|\mathbf{T}_i])^T\boldsymbol{\beta} + \varepsilon_i.\tag{2.2.2}$$

Para lo que sigue es necesario estimar las esperanzas condicionales. Para ello, sea $\mathbf{t} \doteq \mathbf{T}(\mathbf{s})$ el valor de la función $\mathbf{T}$ en $\mathbf{s}$. Con esto, para cada $\mathbf{h}$ fijo, consideramos

$$\hat{\mathbb{E}}[Y_i|\mathbf{T}_i = \mathbf{t}] = \sum_{\substack{\mathbf{s}_k \in \mathcal{O}_{\mathbf{n}} \\ \mathbf{s}_k \neq \mathbf{s}}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_k) Y_k,\tag{2.2.3}$$

y

$$\hat{\mathbb{E}}[X_{ij}|\mathbf{T}_i = \mathbf{t}] = \sum_{\substack{\mathbf{s}_k \in \mathcal{O}_{\mathbf{n}} \\ \mathbf{s}_k \neq \mathbf{s}}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_k) X_{kj}.\tag{2.2.4}$$

En esta tesis, los pesos se calcularán a partir de los propuestos en la Sección 1.3.1.4. En particular, en este capítulo trabajamos con el peso, dado en la ecuación (2.2.5):

$$\omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_k) \doteq \frac{K_1\left(\frac{d(\mathbf{s}, \mathbf{s}_k)}{h_{1\mathbf{n}}}\right) K_2\left(\frac{d_m(\mathbf{t}, \mathbf{T}_k)}{h_{2\mathbf{n}}}\right)}{\sum_{\substack{\mathbf{s}_j \in \mathcal{O}_{\mathbf{n}} \\ \mathbf{s}_j \neq \mathbf{s}}} K_1\left(\frac{d(\mathbf{s}, \mathbf{s}_j)}{h_{1\mathbf{n}}}\right) K_2\left(\frac{d_m(\mathbf{t}, \mathbf{T}_j)}{h_{2\mathbf{n}}}\right)},\tag{2.2.5}$$

con d_m , la medida de similaridad, definida en (1.3.16) y $\mathbf{h} = \{h_{1\mathbf{n}}, h_{2\mathbf{n}}, k\}$. Además consideraremos que $\omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_k) = 0$ para todo $\mathbf{t} = \mathbf{T}_k$.

Si definimos la matriz $\mathbf{W}_{\mathbf{h}} \in \mathbb{R}^{\hat{n} \times \hat{n}}$ como la matriz de pesos tal que $(\mathbf{W}_{\mathbf{h}})_{ij} = \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_j)$ podemos escribir (2.2.3) de forma matricial como sigue

$$\hat{\mathbb{E}}[\mathbf{Y}_i|\mathbf{T}_i] = (\mathbf{W}_{\mathbf{h}}\mathbf{Y})_i,\tag{2.2.6}$$

con $(\mathbf{W}_{\mathbf{h}}\mathbf{Y})_i$ el i -ésimo elemento del vector $\mathbf{W}_{\mathbf{h}}\mathbf{Y} \in \mathbb{R}^{\hat{n}}$. De la misma manera, podemos escribir (2.2.4) como

$$\hat{\mathbb{E}}[X_{ij}|\mathbf{T}_i] = (\mathbf{W}_{\mathbf{h}}\mathbb{X})_{ij},\tag{2.2.7}$$

con $(\mathbf{W}_{\mathbf{h}}\mathbb{X})_{ij}$ el ij -ésimo elemento de la matriz $\mathbf{W}_{\mathbf{h}}\mathbb{X} \in \mathbb{R}^{\hat{n} \times p}$.

A partir de estas dos estimaciones definimos,

$$\tilde{Y}_i \doteq Y_i - \hat{\mathbb{E}}[Y_i | \mathbf{T}_i] = Y_i - (\mathbf{W}_h \mathbf{Y})_i$$

y, considerando $(\mathbf{W}_h \mathbb{X})_i$ como la fila i de la matriz $\mathbf{W}_h \mathbb{X}$, definimos

$$\tilde{\mathbf{X}}_i \doteq \mathbf{X}_i - \hat{\mathbb{E}}[\mathbf{X}_i | \mathbf{T}_i] = \mathbf{X}_i - (\mathbf{W}_h \mathbb{X})_i,$$

de esta manera, usando (2.2.2) obtenemos que

$$\tilde{Y}_i = \tilde{\mathbf{X}}_i^T \boldsymbol{\beta} + \varepsilon_i.$$

De igual forma que definimos $\mathbf{Y}$ en (1.3.1) y $\mathbb{X}$ en (1.3.11), definimos $\tilde{\mathbf{Y}} \doteq (\tilde{Y}_1, \dots, \tilde{Y}_{\hat{n}})^T \in \mathbb{R}^{\hat{n}}$ y $\tilde{\mathbb{X}} \doteq (\tilde{\mathbf{X}}_1, \dots, \tilde{\mathbf{X}}_{\hat{n}})^T \in \mathbb{R}^{\hat{n} \times p}$. Por lo tanto, si la matriz $\tilde{\mathbb{X}}^T \tilde{\mathbb{X}}$ es invertible, el estimador de mínimos cuadrados de $\boldsymbol{\beta}$ está dado por

$$\hat{\boldsymbol{\beta}}_h = (\tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbf{Y}}. \quad (2.2.8)$$

Observación 2.2.1. *El procedimiento de estimación se basa en remover, tanto de la variable respuesta como de las covariables, la información explicable a partir de $\mathbf{T}_i$. En consecuencia, los coeficientes $\boldsymbol{\beta}$ miden la relación lineal entre las covariables X_i y la variable de interés Y_i que no puede ser explicada por la información espacial local contenida en los vecinos. Es decir, $\boldsymbol{\beta}$ cuantifica el efecto de X_i sobre Y_i una vez descontada la dependencia espacial inducida por los valores observados de Y en sitios cercanos.*

Ahora, para estimar la función r , a partir de la ecuación (2.2.1) se obtiene la siguiente expresión

$$\hat{r}_h(\mathbf{T}_i) = \hat{\mathbb{E}}[Y_i | \mathbf{T}_i] - \hat{\mathbb{E}}[\mathbf{X}_i | \mathbf{T}_i] \hat{\boldsymbol{\beta}}_h.$$

Definiendo $(\mathbf{W}_h)_i$ como la fila i de la matriz $\mathbf{W}_h$ y reemplazando, en la ecuación anterior, las estimaciones halladas en (2.2.6), (2.2.7) y en (2.2.8), obtenemos

$$\begin{aligned} \hat{r}_h(\mathbf{T}_i) &= \hat{\mathbb{E}}[Y_i | \mathbf{T}_i] - \hat{\mathbb{E}}[\mathbf{X}_i | \mathbf{T}_i] \hat{\boldsymbol{\beta}}_h \\ &= (\mathbf{W}_h \mathbf{Y})_i - (\mathbf{W}_h \mathbb{X})_i \hat{\boldsymbol{\beta}}_h \\ &= (\mathbf{W}_h)_i \mathbf{Y} - (\mathbf{W}_h)_i \mathbb{X} \hat{\boldsymbol{\beta}}_h \\ &= (\mathbf{W}_h)_i (\mathbf{Y} - \mathbb{X} \hat{\boldsymbol{\beta}}_h). \end{aligned} \quad (2.2.9)$$

Finalmente, usando el modelo propuesto, la predicción de la variable de interés Y_0 , en el sitio $\mathbf{s}_0 \in \mathcal{O}_n$ está dada por

$$\begin{aligned} \hat{Y}_0 &= \hat{\mathbb{E}}[Y_0 | \mathbf{X}_0, \mathbf{T}_0] \\ &= \mathbf{X}_0^T \hat{\boldsymbol{\beta}}_h + \hat{r}(\mathbf{T}_0) \\ &= \mathbf{X}_0^T \hat{\boldsymbol{\beta}}_h + (\mathbf{W}_h)_0 (\mathbf{Y} - \mathbb{X} \hat{\boldsymbol{\beta}}_h). \end{aligned} \quad (2.2.10)$$

donde $\mathbf{X}_0 \in \mathbb{R}^p$ es el vector las covariables medidas en $\mathbf{s}_0$ y $\mathbf{T}_0 \in \mathbb{R}^k$ es el vecindario de $\mathbf{s}_0$.

Por último resta estimar el vector de parámetros $\mathbf{h}$, esto es, los parámetros de suavizado y la cantidad de vecinos más cercanos k . Para ello utilizamos la técnica de validación cruzada de k -submuestras, (k -fold cross validation, en inglés) el cual detallamos en los capítulos siguientes.

Observación 2.2.2. *De forma similar a los métodos presentados en la Sección 1.3 también arribamos en (2.2.10) a una expresión similar del BLUP (1.3.2), donde lo que varía con respecto a dichos predictores es que no necesitamos modelar ni estimar la matriz de varianza y covarianza.*

Observación 2.2.3. *En el caso de no considerar covariables, el predictor propuesto en (2.2.10) queda dado por*

$$\hat{\mathbb{E}}[Y_0|\mathbf{T}_0] = \hat{r}(\mathbf{T}_0),$$

cuya estimación queda resuelta con lo dado en (2.2.6) desarrollado en García Arancibia et al. (2023) y descrito en la Sección 1.3.1.4.

2.3. Resultados asintóticos

En esta sección estudiamos el comportamiento asintótico de los estimadores $\hat{\beta}_{\mathbf{h}}$ y $\hat{r}$ definidos en (2.2.8) y (2.2.9), respectivamente.

2.3.1. Preliminares y notación

Aneiros-Pérez y Vieu (2008) demostraron la normalidad asintótica y obtuvieron las tasas de convergencia tanto del estimador del parámetro lineal, definido como (2.2.8) y de la componente no paramétrica, definida de forma análoga a (2.2.9). Una diferencia relevante con respecto al enfoque adoptado en esta tesis radica en que sus estimaciones no paramétricas se construyen a partir de una única función núcleo, la cual depende exclusivamente de una variable $\mathbf{T}_i$. Los autores suponen que $\mathbf{T}_i$ toma valores en un espacio semi-métrico abstracto $\mathcal{H}$, provisto de una semi-métrica asociada $d_{\mathcal{H}}(\cdot, \cdot)$.

Por otro lado, Dabo-Niang et al. (2016) proponen un predictor no paramétrico para datos espaciales cuyos pesos se construyen a partir de dos funciones núcleo: una asociada a las distancias entre sitios espaciales y otra vinculada a las distancias entre las variables predictoras. Bajo supuestos de dependencia débil, formalizados mediante una condición de α -mixing, las autoras establecen la consistencia uniforme casi completa y la normalidad asintótica del predictor. A diferencia del enfoque desarrollado en esta tesis, el predictor propuesto no incorpora términos autorregresivos espaciales. En particular, se trata de un modelo de regresión no paramétrico clásico en el cual la variable $\mathbf{T}_i$ corresponde alternativamente a las covariables o a la variable respuesta, sin considerar la inclusión de ambas dentro del modelo.

La metodología presentada en esta tesis integra y extiende los enfoques anteriores. Previo a su desarrollo, se formulan algunas consideraciones preliminares. Concretamente, para estimar las esperanzas condicionales (2.2.3) y (2.2.4) proponemos estimadores del tipo Nadaraya-Watson, con pesos dados por (2.2.5) cuyo numerador está dado por

$$K_{1,h_{1\mathbf{n}}}(d(\mathbf{s}, \mathbf{s}_i)) K_2 \left(\frac{d_m(\mathbf{t}, \mathbf{T}_i)}{h_{2\mathbf{n}}} \right),$$

donde los núcleos K_1 y K_2 están definidos sobre $\mathbb{R}$, y las secuencias $h_{1\mathbf{n}}$ y $h_{2\mathbf{n}}$ son parámetros de suavizado que tienden a cero, las cuales supondremos que $\hat{n}h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \rightarrow \infty$. Además, definimos

$$K_{1,h_{1\mathbf{n}}}(d(\mathbf{s}, \mathbf{s}_i)) = K_1 \left(\frac{d(\frac{\mathbf{s}}{h_{1\mathbf{n}}}, \frac{\mathbf{s}_i}{h_{1\mathbf{n}}})}{h_{1\mathbf{n}}} \right) \quad \text{con} \quad \frac{\mathbf{s}}{h_{1\mathbf{n}}} = \left(\frac{i_1}{n}, \frac{i_2}{n}, \dots, \frac{i_d}{n} \right)$$

y cada sitio fijo $\mathbf{s}$, definimos el **tamaño del vecindario** como

$$k_{\mathbf{n}} = \sum_{\mathbf{s}_i \in \mathcal{D}_{\mathbf{n}}} \mathbb{I}_{\{d(\mathbf{s}, \mathbf{s}_i) \leq r_{\mathbf{n}}\}},$$

donde $r_{\mathbf{n}} > 0$ es un **radio de vecindad** creciente y $d(\cdot)$ denota, como antes, la distancia euclídea. La cantidad $k_{\mathbf{n}}$ representa el número de ubicaciones incluidas en la predicción local cerca de $\mathbf{s}$.

De este modo los sitios que verifican

$$\frac{d(\frac{\mathbf{s}}{h_{1\mathbf{n}}}, \frac{\mathbf{s}_i}{h_{1\mathbf{n}}})}{h_{1\mathbf{n}}} \leq 1$$

son exactamente aquellos tales que $d(\mathbf{s}_0, \mathbf{s}_i) \leq r_{\mathbf{n}}$, con

$$r_{\mathbf{n}} = nh_{1\mathbf{n}}.$$

Además supondremos que

$$r_{\mathbf{n}} \rightarrow \infty, \quad \text{y por lo tanto} \quad (nh_{1\mathbf{n}})^d = \hat{n}h_{1\mathbf{n}}^d \rightarrow \infty,$$

lo cual garantiza que el número de observaciones efectivamente utilizadas en la predicción local aumenta sin acotarse.

Desde el punto de vista geométrico, sobre una grilla regular en $\mathbb{Z}^d$, el número de puntos contenidos en una bola euclídea $B(r)$ de radio r satisface

$$\#(\mathbb{Z}^d \cap B(r)) = \text{Vol}_d(B(r)) + O(r^{d-1}),$$

resultado clásico conocido como el problema del círculo de Gauss, donde $\#$ representa el cardinal del conjunto. En consecuencia,

$$k_{\mathbf{n}} = O(r_{\mathbf{n}}^d).$$

Más aún, asumimos la expansión asintótica

$$k_{\mathbf{n}} = C_d r_{\mathbf{n}}^d + O(r_{\mathbf{n}}^\beta), \quad 0 < \beta < d, \quad (2.3.1)$$

donde C_d es el volumen de la bola unitaria en $\mathbb{R}^d$. Esta condición expresa que el número de vecinos crece esencialmente al ritmo del volumen del vecindario, con una corrección de orden menor.

Estas hipótesis permiten controlar el compromiso sesgo–varianza del predictor propuesto: el crecimiento de $k_{\mathbf{n}}$ asegura una reducción de la varianza, mientras que el decrecimiento de $h_{1\mathbf{n}}$ controla el sesgo inducido por el suavizado espacial.

2.3.2. Hipótesis generales

Los resultados asintóticos asociados al predictor propuesto se obtienen bajo las siguientes hipótesis sobre las funciones de regresión y de densidades, las funciones núcleos, los anchos de banda y la condición de dependencia.

H1. Sea $\mathbf{T} \in \mathcal{C} \subset \mathbb{R}^k$, con $\mathcal{C}$ compacto tal que

$$\mathcal{C} \subset \bigcup_{k=1}^{\tau_{\mathbf{n}}} B(z_k, l_{\mathbf{n}}),$$

donde $\tau_{\mathbf{n}} l_{\mathbf{n}}^k = C$, con C constante, $\tau_{\mathbf{n}} \rightarrow \infty$ y $l_{\mathbf{n}} \rightarrow 0$ cuando $n \rightarrow \infty$ y $B(\mathbf{t}', h) = \{\mathbf{t}' \in \mathbb{R}^k; d_m(\mathbf{t}', \mathbf{t}) < h\}$, donde $d_m(\mathbf{t}', \mathbf{t}) = |\text{Med}(\mathbf{t}') - \text{Med}(\mathbf{t})|$.

H2. Las funciones núcleo K_1 y K_2 tienen soporte en $[0, 1]$, son Lipschitz continuas en $[0, \infty)$, y existen constantes k_i , con $i = 1, 2$, tales que para todo $u \in [0, 1]$,

$$-K_1'(u) > k_1 > 0 \text{ y } -K_2'(u) > k_2 > 0.$$

Además, para todo $u \in [0, 1]$, $K_i(u) > 0$ con $i = 1, 2$.

H3. La densidad marginal f de $\mathbf{T}$ con respecto a la medida de Lebesgue en $\mathbb{R}^k$ existe, es Lipschitz y satisface que

$$\inf_{\mathbf{t} \in \mathcal{C}} f(\mathbf{t}) \geq \delta > 0.$$

La densidad conjunta $f_{\mathbf{T}_i, \mathbf{T}_j}$ de $(\mathbf{T}_i, \mathbf{T}_j)$ con respecto a la medida de Lebesgue en $\mathbb{R}^{2k}$ existe, es acotada y, para todo $\mathbf{t}_1, \mathbf{t}_2 \in \mathbb{R}^d$, existe una constante $C > 0$ tal que

$$\left| f_{\mathbf{T}_i, \mathbf{T}_j}(\mathbf{t}_1, \mathbf{t}_2) - f_{\mathbf{T}_i}(\mathbf{t}_1) f_{\mathbf{T}_j}(\mathbf{t}_2) \right| < C,$$

para todo $\mathbf{t}_1, \mathbf{t}_2$ y $\mathbf{s}_i \neq \mathbf{s}_j$.

- H4. (i) Las densidades condicionales $f_{\mathbf{T}_i, \mathbf{T}_j | Y_i, Y_j}$ de $(\mathbf{T}_i, \mathbf{T}_j)$ dado (Y_i, Y_j) y las densidades condicionales $f_{\mathbf{T}_i | Y_j}$ de $\mathbf{T}_i$ dado Y_j existen y verifican que

$$f_{\mathbf{T}_i, \mathbf{T}_j | Y_i, Y_j}(\mathbf{t}_1, \mathbf{t}_2 | y_1, y_2) \leq C, \quad f_{\mathbf{T}_i | Y_j}(\mathbf{t}_1 | y_1) \leq C$$

para todo $y_1, y_2, \mathbf{t}_1, \mathbf{t}_2, \mathbf{s}_i, \mathbf{s}_j$.

- (ii) $\sup_i \mathbb{E}[|Y_i|^r] < \infty$, $\sup_{\mathbf{q}} \int |y|^r f_{\mathbf{T}_i, Y_i}(\mathbf{q}, y) dy < C$ para algún $r > 4$, y para todo $\mathbf{s}_i$. Sea $g_2(\mathbf{q}) = \text{Var}(Y_i | \mathbf{T}_i = \mathbf{q})$, independiente de $\mathbf{s}_i$ y continua en algún entorno de $\mathbf{t}$, de modo que $\sup_{\mathbf{q}, \|\mathbf{q} - \mathbf{t}\| \leq h} |g_2(\mathbf{q}) - g_2(\mathbf{t})| = o(1)$ cuando $h \rightarrow 0$. Además, con m un modelo no paramétrico puro, $g_r(\mathbf{q}) = \mathbb{E}[|Y_i - m(\mathbf{t})|^r | \mathbf{T}_i = \mathbf{q}]$ es independiente de $\mathbf{s}_i$ y continua en algún entorno de $\mathbf{t}$. Asimismo, $g(\mathbf{q}_1, \mathbf{q}_2, \mathbf{t}) = \mathbb{E}[(Y_i - m(\mathbf{t}))((Y_j - m(\mathbf{t}))) | \mathbf{T}_i = \mathbf{q}_1, \mathbf{T}_j = \mathbf{q}_2]$, $\forall \mathbf{s}_i \neq \mathbf{s}_j$, no depende de $\mathbf{s}_i, \mathbf{s}_j$, y es continua en algún entorno de $(\mathbf{t}, \mathbf{t})$.

Análogamente, las condiciones (i) y (ii), se cumplen para cada una de las variables del vector $\mathbf{X}_i$ en lugar de Y_i .

- H5. Sea $g_j(\mathbf{t}) = \mathbb{E}[X_{ij} | \mathbf{T}_i = \mathbf{t}]$ y $r(\mathbf{t}) = \mathbb{E}[Y_i - \mathbf{X}_i^T \boldsymbol{\beta} | \mathbf{T}_i = \mathbf{t}]$ con $1 \leq i \leq \hat{n}, 1 \leq j \leq p$. Suponemos que todas las funciones a estimar son Lipschitz, esto es $\forall (\mathbf{t}_1, \mathbf{t}_2) \in \mathcal{C} \times \mathcal{C}$, $\forall h \in \{r, g_1, \dots, g_p\}$,

$$|h(\mathbf{t}_1) - h(\mathbf{t}_2)| \leq C \|\mathbf{t}_1 - \mathbf{t}_2\|,$$

para algún $C < \infty$.

- H6. Suponemos que $\{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}_{i=1}^{\hat{n}}$ es conjuntamente estrictamente estacionaria y proviene de alguna función fuertemente mixing cuyo coeficiente mixing $\alpha(r)$ verifica que,

$$\alpha(r) \leq Cr^{-b}, \quad \text{para algún } b > 4.5d.$$

- H7. Sea $\boldsymbol{\eta}_i \doteq (\eta_{i1}, \dots, \eta_{ip})^T$ donde $\eta_{ij} = X_{ij} - \mathbb{E}(X_{ij} | \mathbf{T}_i)$ para $j = 1, \dots, p$, consideramos que para cada $i = 1, \dots, n$

$$\boldsymbol{\eta}_i \text{ es independiente de } \varepsilon_i \quad (i = 1, \dots, n).$$

- H8. Sea

$$\mathbf{V}_\varepsilon = \mathbb{E}(\varepsilon \varepsilon^T) \quad \text{con} \quad \varepsilon^T = (\varepsilon_1, \dots, \varepsilon_{\hat{n}}) \quad \text{y} \quad \boldsymbol{\eta}^T = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{\hat{n}}).$$

suponemos que

$$\mathbf{B} = \mathbb{E}[\boldsymbol{\eta}_1 \boldsymbol{\eta}_1^T] \quad \text{y} \quad \mathbf{C} = \lim_{n \rightarrow \infty} n^{-1} \mathbb{E}[\boldsymbol{\eta}^T \mathbf{V}_\varepsilon \boldsymbol{\eta}]$$

son matrices definidas positivas.

Observación 2.3.1. Las hipótesis H1-H5 del Teorema A.0.5 de Dabo-Niang et al. (2016) están directamente aseguradas por las hipótesis H2-H5 de esta tesis. De manera particular, la hipótesis H5 y los supuestos sobre los parámetros de suavizado (explícitos en el enunciado de los resultados obtenidos) permiten verificar las hipótesis H6 y H7 de dicho teorema.

Además, $H8$, así como la condición para aplicar la Observación 4 del trabajo de las autoras, quedan garantizadas por el supuesto de estacionariedad estricta en $H6$. Finalmente, la compacidad del conjunto $\mathcal{C}$, establecida en $H1$, nos permite emplear la medida de similitud dada por la diferencia entre las medianas y , de esta forma, mantener el resultado obtenido por las autoras.

2.3.3. Resultados asociados a los parámetros lineales

El siguiente teorema es una adaptación del Teorema 1 de Aneiros-Pérez y Vieu (2008) al contexto espacial.

Teorema 2.3.2. *Bajo las hipótesis en $H1$ - $H8$ y si, además, cuando $\hat{n} \rightarrow \infty$*

$$\begin{aligned} \blacksquare \hat{n} h_{2\mathbf{n}}^4 &\rightarrow 0, \\ \blacksquare \frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4}-\frac{1}{r}}} &\rightarrow 0, \end{aligned}$$

entonces

$$\sqrt{\hat{n}} (\hat{\beta}_{\mathbf{h}} - \beta) \xrightarrow{d} N(0, \mathbf{A}), \quad \text{donde } \mathbf{A} = \mathbf{B}^{-1} \mathbf{C} \mathbf{B}^{-1},$$

y

$$\limsup_{\hat{n} \rightarrow \infty} \left(\frac{\hat{n}}{2 \log \log \hat{n}} \right)^{\frac{1}{2}} |\hat{\beta}_{hj} - \beta_j| = a_{jj}^{\frac{1}{2}} \text{ c.s. donde } a_{jj} = A_{jj}. \quad (2.3.2)$$

Antes de la demostración del teorema, probaremos algunos lemas necesarios para su prueba.

Lema 2.3.3. *Sea $\{Z(s) : s \in \mathbb{Z}^d\}$ una función aleatoria con coeficiente de mixing espacial definido, para $r \geq 0$, por*

$$\alpha(r) \doteq \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \sigma(Z(s) : s \in S), B \in \sigma(Z(s) : s \in S'), \text{dist}(S, S') \geq r \right\},$$

donde $S, S' \subset \mathbb{Z}^d$ son conjuntos finitos y

$$\text{dist}(S, S') \doteq \inf \{\|s - s'\| : s \in S, s' \in S'\}.$$

Suponemos que:

1. Existen constantes $C > 0$ y $b > 1$ tales que, para todo $r > 0$,

$$\alpha(r) \leq C r^{-b}.$$

2. Suponemos que el dominio $\mathcal{D}_{\mathbf{n}} \subset \mathbb{Z}^d$, para constantes $0 < c_L \leq C_L$, satisface que

$$c_L \hat{n}^{1/d} \leq \mathcal{D}_{\mathbf{n}} \leq C_L \hat{n}^{1/d},$$

o, equivalentemente, $\mathcal{D}_{\mathbf{n}} \asymp \hat{n}^{1/d}$

3. Suponemos que los conjuntos $E, E' \subset \mathbb{Z}^d$ tienen densidad acotada, es decir, que existe una constante $c_d > 0$ tal que, para toda bola $B(\mathbf{x}, r) \subset \mathbb{Z}^d$ con $r \geq 1$,

$$\#(E \cap B(\mathbf{x}, r)) \leq c_d r^d \quad y \quad \#(E' \cap B(\mathbf{x}, r)) \leq c_d r^d.$$

Entonces, para cualquier $\varepsilon > 0$ fijo, existe una constante C' tal que, para todo par de subconjuntos $E, E' \subset \mathcal{D}_{\mathbf{n}}$ con $\text{dist}(E, E') \geq \varepsilon \hat{n}^{1/d}$, se cumple

$$\alpha(\sigma(E), \sigma(E')) \leq C' \hat{n}^{-b/d}.$$

Demostración. Sean E, E' conjuntos de $\mathcal{D}_{\mathbf{n}}$ y $\varepsilon > 0$ fijo. La condición 3 excluye concentraciones locales de puntos, mientras que la extensión del dominio $\mathcal{D}_{\mathbf{n}}$, del orden $\hat{n}^{1/d}$ (condición 2) garantiza la existencia de subconjuntos separados a distancia del mismo orden.

Luego, aplicando directamente la hipótesis 1. con $r = \text{dist}(E, E') \geq \varepsilon \hat{n}^{1/d}$, tenemos que

$$\alpha(\sigma(E), \sigma(E')) \leq C(\text{dist}(E, E'))^{-b} \leq C(\varepsilon \hat{n}^{1/d})^{-b} = C' \hat{n}^{-b/d},$$

lo que prueba el lema. □

Observación 2.3.4. El lema establece que, bajo un decaimiento polinómico del mixing y un crecimiento regular del dominio, dos subconjuntos de sitios separados una distancia del orden $\hat{n}^{1/d}$ dentro del dominio presentan una dependencia débil la cual decrece sin que existan efectos locales de agrupamiento de puntos.

Observación 2.3.5. El Lema 2.3.3 conecta la hipótesis H6 con la hipótesis sobre el coeficiente mixing del Lema A.0.2, asegurándonos su aplicación.

Lema 2.3.6. Sea $\{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}$ una secuencia que satisfacen las hipótesis H1, H4, H5, H6 y H7. Supongamos que los datos satisfacen el modelo

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + r(\mathbf{T}_i) + \varepsilon_i,$$

donde

$$\mathbb{E}[\varepsilon_i | \mathbf{X}_i, \mathbf{T}_i] = 0. \tag{2.3.3}$$

Si $\boldsymbol{\eta}_i = \mathbb{E}[\mathbf{X}_i - \mathbb{E}[\mathbf{X}_i | \mathbf{T}_i]]$, entonces, las secuencias $\{\varepsilon_i\}$, $\{\boldsymbol{\eta}_i\}$ y $\{\boldsymbol{\eta}_i \varepsilon_i\}$ tienen media cero, son estrictamente estacionarias y fuertemente mixing, con coeficientes mixing acotados por $\alpha(r)$. Además, existen $r > 4$, $\delta > 0$ y una constante $C > 0$ tales que

$$\begin{aligned} \sup_i \mathbb{E}[|\varepsilon_i|^r] &\leq C < \infty, \\ \sup_i \mathbb{E}[\|\boldsymbol{\eta}_i\|^r] &\leq C < \infty, \end{aligned} \tag{2.3.4}$$

y

$$\mathbb{E}[\|\boldsymbol{\eta}_i \varepsilon_i\|^{2+\delta}] \leq \infty,$$

respectivamente.

Demostración. En primer lugar, probaremos que las secuencias tienen media nula: para ε_i , por (2.3.3) tenemos que

$$\mathbb{E}[\varepsilon_i] = \mathbb{E}[\mathbb{E}[\varepsilon_i | \mathbf{X}_i, \mathbf{T}_i]] = 0; \quad (2.3.5)$$

mientras que para η_{ij} tenemos que

$$\mathbb{E}[\eta_{ij}] = \mathbb{E}[X_{ij} - \mathbb{E}[X_{ij} | \mathbf{T}_i]] = 0. \quad (2.3.6)$$

Luego, de (2.3.5), (2.3.6) y de la hipótesis H8, que nos asegura que $\boldsymbol{\eta}_i$ y ε_i son independientes, tenemos que $\mathbb{E}[\boldsymbol{\eta}_i \varepsilon_i] = \mathbf{0}$.

Por definición, la secuencia $\{\boldsymbol{\eta}_i\}$, es estacionaria, ya que tanto $\mathbf{X}_i$ como $\mathbb{E}(\mathbf{X}_i | \mathbf{T}_i)$ son funciones medibles de $\{(\mathbf{X}_i, \mathbf{T}_i)\}$. De manera análoga, el proceso $\{\varepsilon_i\}$, es estacionario por ser una función medible de $\{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}$, y con el mismo argumento e independencia, la secuencia $\{\boldsymbol{\eta}_i \varepsilon_i\}$ es estacionaria.

Por otra parte, por hipótesis H6 tenemos que, conjuntamente, $\{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}$ es fuertemente mixing, con coeficientes $\alpha(r) \leq Cr^{-b}$. Dado que tanto ε_i como η_{ij} son funciones medibles de $\{(Y_i, \mathbf{X}_i, \mathbf{T}_i)\}$, notando que $\mathbb{E}(X_{ij} | \mathbf{T}_i)$ depende únicamente de $\mathbf{T}_i$, tanto ε_i , como η_{ij} heredan esta propiedad de mixing con coeficientes menores que $\alpha(r)$ por la inclusión de los σ -álgebras. Esto es, sea $\{\mathbf{Z}_i\} = \{(Y_i, \mathbf{X}_i^T, \mathbf{T}_i^T)\}$, entonces $\varepsilon_i = g(\mathbf{Z}_i)$ es una función medible de $\{\mathbf{Z}_i\}$. Para todo conjunto $U \subset \mathbb{Z}^d$ tenemos

$$\sigma(\varepsilon_i : \mathbf{s}_i \in U) = \sigma(g(\mathbf{Z}_i) : \mathbf{s}_i \in U).$$

Como g es medible, para todo boreliano B se cumple

$$[\varepsilon_i \in B] = [g(\mathbf{Z}_i) \in B] = [\mathbf{Z}_i \in g^{-1}(B)],$$

y por lo tanto $\sigma(\varepsilon_i) \subset \sigma(\mathbf{Z}_i)$ para cada $\mathbf{s}_i$. Sean ahora $U, V \subset \mathbb{Z}^d$ tales que $\text{dist}(U, V) \geq r$. Dado que

$$\sigma(\varepsilon_i : \mathbf{s}_i \in U) \subset \sigma(\mathbf{Z}_i : \mathbf{s}_i \in U) \quad \text{y} \quad \sigma(\varepsilon_j : \mathbf{s}_j \in V) \subset \sigma(\mathbf{Z}_j : \mathbf{s}_j \in V),$$

el supremo que define $\alpha_\varepsilon(r)$ se toma sobre un subconjunto de los eventos utilizados para definir $\alpha_{\mathbf{Z}}(r)$. Por lo tanto,

$$\alpha_\varepsilon(r) \leq \alpha_{\mathbf{Z}}(r).$$

Un argumento idéntico se aplica a la sucesión $\{\boldsymbol{\eta}_i\}$, así como al producto $\{\boldsymbol{\eta}_i \varepsilon_i\}$, ya que ambos son funciones medibles de $\{\mathbf{Z}_i\}$.

En lo que respecta a los momentos, de la hipótesis H4(ii) se deduce la existencia de un momento de orden $r > 4$ uniforme en i , es decir,

$$\sup_i \mathbb{E}[|Y_i|^r] < \infty, \quad \sup_i \mathbb{E}[|X_{ij}|^r] < \infty, \quad j = 1, \dots, p. \quad (2.3.7)$$

Como $r(\cdot)$ es Lipschitz (H5) y $\mathbf{T}_i$ toma valores en un conjunto compacto $\mathcal{C}$ (H1), tenemos que $r(\cdot)$ es uniformemente acotada en $\mathcal{C}$, esto es,

$$\sup_{\mathbf{t} \in \mathcal{C}} |r(\mathbf{t})| < \infty.$$

En consecuencia,

$$\sup_i \mathbb{E} [|r(\mathbf{T}_i)|^r] < \infty. \quad (2.3.8)$$

Por lo tanto, usando que

$$|\varepsilon_i|^r = |Y_i - \mathbf{X}_i^T \beta - r(\mathbf{T}_i)|^r \leq C_r \left(|Y_i|^r + |\mathbf{X}_i^T \beta|^r + |r(\mathbf{T}_i)|^r \right),$$

para alguna constante $C_r > 0$, y (2.3.7) y (2.3.8), obtenemos

$$\sup_i \mathbb{E} [|\varepsilon_i|^r] \leq C_r \left(\sup_i \mathbb{E} [|Y_i|^r] + \sup_i \mathbb{E} [|\mathbf{X}_i^T \beta|^r] + \sup_i \mathbb{E} [|r(\mathbf{T}_i)|^r] \right) < \infty. \quad (2.3.9)$$

De manera similar, para $\eta_{kj} = X_{ij} - \mathbb{E}[X_{ij} | \mathbf{T}_i]$, la desigualdad de Jensen implica que

$$|\mathbb{E}[X_{ij} | \mathbf{T}_i]|^r \leq \mathbb{E}[|X_{ij}|^r | \mathbf{T}_i] \quad \text{c.s.},$$

tomando esperanza en cada lado y considerando que el momento r -ésimo de X_{ij} es finito, (por (2.3.7)) tenemos que

$$\mathbb{E} [|\mathbb{E}[X_{ij} | \mathbf{T}_i]|^r] \leq \mathbb{E} [|X_{ij}|^r] < \infty.$$

Esto garantiza, junto con (2.3.7), que

$$\sup_i \mathbb{E} [|\eta_{ij}|^r] \leq \sup_i \mathbb{E} [|X_{ij}|^r] - \sup_i \mathbb{E} [|\mathbb{E}[X_{ij} | \mathbf{T}_i]|^r] < \infty, \quad (2.3.10)$$

.

También, necesitamos probar que existe $\delta > 0$, tal que $\mathbb{E} [\|\boldsymbol{\eta}_i \varepsilon_i\|^{2+\delta}] < \infty$. Para lo cual, sabemos que para $r > 4$, $\mathbb{E} [|\varepsilon_i|^r] < \infty$ y que $\mathbb{E} [|\eta_{ij}|^r] < \infty$ uniformemente en i , probado en (2.3.9) y (2.3.10), respectivamente. Ahora, tomando $0 < \delta < \frac{r}{2} - 2$ y como $\varepsilon_i \in \mathbb{R}$, tenemos que

$$\|\boldsymbol{\eta}_i \varepsilon_i\|^{2+\delta} = |\varepsilon_i|^{2+\delta} \|\boldsymbol{\eta}_i\|^{2+\delta}.$$

Dado que la dimensión de $\boldsymbol{\eta}_i$ es fija, por equivalencia de normas existe una constante $C > 0$ tal que

$$\|\boldsymbol{\eta}_i\|^{2+\delta} \leq C \sum_{k=1}^p |\eta_{ik}|^{2+\delta}.$$

Por lo tanto,

$$\mathbb{E} \|\boldsymbol{\eta}_i \varepsilon_i\|^{2+\delta} \leq C \sum_{k=1}^p \mathbb{E} (|\varepsilon_i|^{2+\delta} |\eta_{ik}|^{2+\delta}).$$

Aplicando la desigualdad de Hölder con exponentes 2 y 2, se obtiene

$$\mathbb{E} (|\varepsilon_i|^{2+\delta} |\eta_{ik}|^{2+\delta}) \leq (\mathbb{E} |\varepsilon_i|^{2(2+\delta)})^{1/2} (\mathbb{E} |\eta_{ik}|^{2(2+\delta)})^{1/2}.$$

Como definimos $r > 2(2+\delta)$, los momentos del lado derecho están uniformemente acotados en i y k . En consecuencia,

$$\mathbb{E} \left[\|\boldsymbol{\eta}_i \varepsilon_i\|^{2+\delta} \right] < \infty.$$

□

Lema 2.3.7. *Supongamos que H1 - H6 se cumplen, entonces*

$$\max_{1 \leq i, j \leq n} |\omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_j)| = O((\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1}) \text{ c.s.,}$$

para $\omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_j)$ definido en (2.2.5).

Demostración. Consideramos los pesos definidos en (2.2.5). Definimos

$$\omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) \doteq \frac{\Delta_i(\mathbf{t})}{\hat{q}(\mathbf{t})}, \quad (2.3.11)$$

con

$$\Delta_i(\mathbf{t}) = \frac{1}{a_{\mathbf{s}, \mathbf{n}} h_{2\mathbf{n}}^k} K_{1, h_{1\mathbf{n}}} (d(\mathbf{s}, \mathbf{s}_i)) K_2 \left(\frac{d_m(\mathbf{t}, \mathbf{T}_i)}{h_{2\mathbf{n}}} \right)$$

y

$$\hat{q}(\mathbf{t}) = \frac{1}{a_{\mathbf{s}, \mathbf{n}} h_{2\mathbf{n}}^k} \sum_{j=1}^{\hat{n}} K_{1, h_{1\mathbf{n}}} (d(\mathbf{s}, \mathbf{s}_j)) K_2 \left(\frac{d_m(\mathbf{t}, \mathbf{T}_j)}{h_{2\mathbf{n}}} \right).$$

donde $a_{\mathbf{n}} \doteq a_{\mathbf{s}, \mathbf{n}} = \sum_{j=1}^{\hat{n}} K_{1, h_{1\mathbf{n}}} (d(\mathbf{s}, \mathbf{s}_j))$. Para demostrar el lema, procederemos en dos etapas. En primer lugar, acotaremos superiormente el numerador por la expresión objetivo, es decir, por $(\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1}$. A continuación, demostraremos que el denominador está acotado inferiormente por una constante positiva. La combinación de estas dos cotas concluirá la demostración.

Por hipótesis H2 sabemos que para todo $u \in [0, 1]$, existen $0 < m_i < M_i < \infty$ tales que para $i = 1, 2$,

$$m_i \leq K_i(u) \leq M_i. \quad (2.3.12)$$

Luego, para todo $\mathbf{s} \in \mathcal{D}_{\mathbf{n}}$, podemos acotar $a_{\mathbf{n}}$ como sigue

$$m_1 \sum_{\mathbf{s}_i \in \mathcal{D}_{\mathbf{n}}} \mathbb{I}_{\{d(\mathbf{s}, \mathbf{s}_i) \leq n h_{1\mathbf{n}}\}} \leq a_{\mathbf{n}} \leq M_1 \sum_{\mathbf{s}_i \in \mathcal{D}_{\mathbf{n}}} \mathbb{I}_{\{d(\mathbf{s}, \mathbf{s}_i) \leq n h_{1\mathbf{n}}\}}. \quad (2.3.13)$$

Para asegurar la primera etapa, probaremos que $\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \Delta_i(\mathbf{t})$ está acotado por una constante. Esto es, considerando (2.3.12), tenemos que, para todo $\mathbf{t} \in \mathcal{C}$,

$$\frac{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}{a_{\mathbf{n}} h_{2\mathbf{n}}^k} K_{1, h_{1\mathbf{n}}} (d(\mathbf{s}, \mathbf{s}_i)) K_2 \left(\frac{d_m(\mathbf{t}, \mathbf{T}_i)}{h_{2\mathbf{n}}} \right) \leq \frac{\hat{n} h_{1\mathbf{n}}^d}{a_{\mathbf{n}}} M_1 M_2.$$

Ahora utilizando (2.3.13) y el supuesto (2.3.1), $r_{\mathbf{n}} = n h_{1\mathbf{n}}$, $\beta < d$ y $\hat{n} = n^d$, aseguramos que

$$m_1 C_d \hat{n} h_{1\mathbf{n}}^d + O((n h_{1\mathbf{n}})^\beta) \leq a_{\mathbf{n}} \leq M_1 C_d \hat{n} h_{1\mathbf{n}}^d + O((n h_{1\mathbf{n}})^\beta).$$

Esto implica que cuando $\hat{n}$ tiende a ∞ ,

$$\frac{\hat{n}h_{1n}^d}{a_n} \rightarrow \frac{1}{C_d'} \quad (2.3.14)$$

con C_d' entre $m_1 C_d$ y $M_1 C_d$ ya que $\frac{O((nh_{1n})^\beta)}{\hat{n}h_{1n}^d} \rightarrow 0$ pues $\beta < d$. Luego, por (2.3.14) para el numerador tenemos que

$$\frac{1}{a_n h_{2n}^k} K_{1,h_{1n}}(d(s_0, s_i)) K_2 \left(\frac{d_m(t, T_i)}{h_{2n}} \right) = O \left(\frac{1}{\hat{n}h_{1n}^d h_{2n}^k} \right) \quad \text{c.s.} \quad (2.3.15)$$

para todo s, s_i, t, T_i .

Para completar la prueba queda demostrar que el denominador $\hat{q}(t)$ está acotado por una constante. Para ello, en la demostración del Teorema A.0.5 (el cumplimiento de las hipótesis se justifica en la Observación 2.3.1) Dabo-Niang et al. (2016) prueban que $\hat{q}(t)$ converge uniformemente y casi seguro a $q(t)$, esto es, existe un conjunto Ω_0 con $\mathbb{P}(\Omega_0) = 1$ tal que, para todo $\omega \in \Omega_0$, existe $\hat{n}_0(\omega)$ con

$$\sup_{t \in \mathcal{C}} |\hat{q}(t) - q(t)| \leq \frac{\delta}{2}, \quad \text{para todo } \hat{n} \geq \hat{n}_0(\omega).$$

Como además, por hipótesis H3 se cumple que

$$\inf_{t \in \mathcal{C}} |q(t)| \geq \delta > 0,$$

y usando que

$$|q(t)| \leq |\hat{q}(t)| + |-\hat{q}(t) + q(t)|,$$

obtenemos que, para todo $t \in \mathcal{C}$, $\hat{n} \geq \hat{n}_0(\omega)$ y $\omega \in \Omega_0$,

$$|\hat{q}(t)| \geq |q(t)| - |\hat{q}(t) - q(t)| \geq \frac{\delta}{2}.$$

En consecuencia, para n suficientemente grande,

$$\inf_{t \in \mathcal{C}} |\hat{q}(t)| \geq \frac{\delta}{2} \quad \text{c.s.}$$

Esto implica que

$$\left| \frac{1}{\hat{q}(t)} \right| = \frac{1}{\frac{1}{a_n h_{2n}^k} \sum_{i=1}^n K_{1,h_{1n}}(d(s, s_i)) K_2 \left(\frac{d_m(t, T_i)}{h_{2n}} \right)} \leq \frac{1}{\delta/2}. \quad (2.3.16)$$

A partir de (2.3.15) y de (2.3.16) y para n suficientemente grande, tenemos que

$$\begin{aligned} \omega_h(t, T_i) &= \frac{\frac{1}{a_n h_{2n}^k} K_{1,h_{1n}}(d(s, s_i)) K_2 \left(\frac{d_m(t, T_i)}{h_{2n}} \right)}{\hat{q}(t)} \\ &= \frac{\frac{1}{a_n h_{2n}^k} K_{1,h_{1n}}(d(s, s_i)) K_2 \left(\frac{d_m(t, T_i)}{h_{2n}} \right)}{\frac{1}{a_n h_{2n}^k} \sum_{i=1}^n K_{1,h_{1n}}(d(s, s_i)) K_2 \left(\frac{d_m(t, T_i)}{h_{2n}} \right)} \leq C' (\hat{n}h_{1n}^d h_{2n}^k)^{-1}, \end{aligned}$$

con lo cual, queda probado el lema. $\square$

Lema 2.3.8. *Bajo las hipótesis H1-H6, y si la muestra $\{(X_{i1}, \dots, X_{ip}, \mathbf{T}_i)\}_{i=1}^{\hat{n}}$ proviene de una función estacionaria α -mixing cuyo coeficiente satisface $\alpha(r) \leq cr^{-b}$, con $b > 1$, y, existe $r > 4$ tal que $\max_{1 \leq i \leq n} \mathbb{E}[|X_{ij}|^r] \leq C < \infty$, para $j = 1, \dots, p$. Además, si $h_{2n} \rightarrow 0$ y*

$$\frac{\log \hat{n}}{\hat{n} h_{1n}^d h_{2n}^k} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty,$$

entonces se tiene que

$$\hat{n}^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \longrightarrow \mathbf{B} \quad \text{c.s.}$$

Demostración. Como $\tilde{\mathbb{X}} = \mathbb{X} - \mathbf{W}_h \mathbb{X}$, tenemos que el elemento kj de dicha matriz está dado por

$$\begin{aligned} \tilde{X}_{kj} &= X_{kj} - \sum_{l=1}^{\hat{n}} \omega_h(\mathbf{T}_k, \mathbf{T}_l) X_{lj} \\ &= X_{kj} - \mathbb{E}[X_{kj} | \mathbf{T}_k] + \mathbb{E}[X_{kj} | \mathbf{T}_k] - \sum_{l=1}^{\hat{n}} \omega_h(\mathbf{T}_k, \mathbf{T}_l) X_{lj} \\ &= \eta_{kj} - \bar{g}_j(\mathbf{T}_k), \end{aligned}$$

donde $\eta_{kj} = X_{kj} - \mathbb{E}[X_{kj} | \mathbf{T}_k]$ definido en la hipótesis H7 y $\bar{g}_j(\mathbf{T}_k)$ está definido por

$$\bar{g}_j(\mathbf{T}_k) \doteq \mathbb{E}[X_{kj} | \mathbf{T}_k] - \sum_{l=1}^{\hat{n}} \omega_h(\mathbf{T}_k, \mathbf{T}_l) X_{lj}. \quad (2.3.17)$$

Luego el elemento js de $\hat{n}^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbb{X}}$ puede escribirse como

$$\begin{aligned} \left(\hat{n}^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \right)_{js} &= \hat{n}^{-1} \sum_{k=1}^{\hat{n}} \tilde{X}_{kj} \tilde{X}_{ks} \\ &= \hat{n}^{-1} \sum_{k=1}^{\hat{n}} (\eta_{kj} - \bar{g}_j(\mathbf{T}_k)) (\eta_{ks} - \bar{g}_s(\mathbf{T}_k)) \\ &= \hat{n}^{-1} \left(\sum_{k=1}^{\hat{n}} \eta_{kj} \eta_{ks} - \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k) \eta_{ks} - \sum_{k=1}^{\hat{n}} \bar{g}_s(\mathbf{T}_k) \eta_{kj} \right. \\ &\quad \left. + \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k) \bar{g}_s(\mathbf{T}_k) \right), \end{aligned} \quad (2.3.18)$$

Aplicamos el Teorema A.0.3 a la sucesión $Y_k = \eta_{kj} \eta_{ks}$. Dado que $\{\eta_k\}$ es estrictamente estacionaria, (por el Lema 2.3.6) cualquier función medible de sus componentes también lo es; en particular, $\{Y_k\}$ es estrictamente estacionaria. Además, como $\{\eta_k\}$ es fuertemente mixing con coeficientes $\alpha(r)$, la sucesión $\{Y_k\}$, al ser función medible de $\{\eta_k\}$, hereda la propiedad de mixing fuerte con coeficientes no mayores. Por otra parte, queremos probar que $\mathbb{E}[|\eta_{ij} \eta_{is}| \log^+ |\eta_{ij} \eta_{is}|] < \infty$. Para ello, probaremos en primer lugar que $\mathbb{E}[|\eta_{ij} \eta_{is}|^{r/2}] < \infty$,

esto es, para índices j, s fijos, tenemos que cada componente está acotada por la norma del vector, es decir,

$$|\eta_{ij}| \leq \|\boldsymbol{\eta}_i\|, \quad |\eta_{is}| \leq \|\boldsymbol{\eta}_i\|.$$

Por lo tanto,

$$|\eta_{ij}\eta_{is}| \leq \|\boldsymbol{\eta}_i\|^2.$$

Elevando ambos lados a la potencia $r/2$ y tomando esperanza, tenemos que

$$\mathbb{E} \left[|\eta_{ij}\eta_{is}|^{r/2} \right] \leq \mathbb{E} [\|\boldsymbol{\eta}_i\|^r].$$

Por último, por (2.3.4) (Lema 2.3.6), existe $r > 4$, tal que

$$\mathbb{E} [\|\boldsymbol{\eta}_i\|^r] \leq C < \infty.$$

En consecuencia,

$$\mathbb{E} \left[|\eta_{ij}\eta_{is}|^{r/2} \right] < \infty. \quad (2.3.19)$$

Con esto, podemos asegurar que $\mathbb{E} [|\eta_{ij}\eta_{is}| \log^+ |\eta_{ij}\eta_{is}|] < \infty$, ya que, considerando $\ell = \frac{r}{2} > 1$ (pues $r > 4$), sabemos que para todo $\ell > 1$ existe una constante $C_\ell > 0$ tal que, para todo $|\eta_{ij}\eta_{is}| \geq 0$,

$$|\eta_{ij}\eta_{is}| \log^+ |\eta_{ij}\eta_{is}| \leq C_\ell (1 + |\eta_{ij}\eta_{is}|^\ell).$$

En efecto, para $|\eta_{ij}\eta_{is}| \leq 1$ el lado izquierdo es cero, y para $|\eta_{ij}\eta_{is}| > 1$, $\log |\eta_{ij}\eta_{is}| = o(|\eta_{ij}\eta_{is}|^{\ell-1})$, por lo que $|\eta_{ij}\eta_{is}| \log |\eta_{ij}\eta_{is}| \leq C_\ell |\eta_{ij}\eta_{is}|^\ell$ para alguna constante adecuada. Tomando esperanza, y considerando (2.3.19), obtenemos

$$\mathbb{E} [|\eta_{ij}\eta_{is}| \log^+ |\eta_{ij}\eta_{is}|] \leq C_\ell (1 + \mathbb{E} [|\eta_{ij}\eta_{is}|^\ell]) < \infty.$$

Esto completa las pruebas de las hipótesis del Teorema A.0.3, el cual nos asegura que

$$\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \eta_{kj} \eta_{ks} \longrightarrow \mathbf{B}_{js} \quad \text{c.s.}$$

En particular,

$$\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \eta_{ks}^2 \longrightarrow \mathbf{B}_{ss} \quad \text{c.s.},$$

y por lo tanto,

$$\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \eta_{ks}^2 = O(1) \quad \text{c.s.} \quad (2.3.20)$$

Consideremos ahora que en (2.3.17) el peso está dado por (2.2.5), que $\mathbf{T}_k$ son las covariables y que cada X_{kj} representa la variable a predecir. Bajo las hipótesis H1–H6 y, por la Observación 2.3.1, aplicamos directamente la Observación 4 de Dabo-Niang et al. (2016), de donde se sigue que, casi seguramente,

$$\max_{1 \leq j \leq p} \max_{1 \leq i \leq \hat{n}} |\bar{g}_{jh}(T_i)| = O(h_{2\mathbf{n}}) + O\left(\frac{\log \hat{n}}{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}\right) \quad (2.3.21)$$

Por último, analizamos el resto de los términos de (2.3.18). Comenzamos con el término

$$\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k) \eta_{ks}.$$

Aplicando la desigualdad de Cauchy-Schwarz tenemos que

$$\left| \hat{n}^{-1} \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k) \eta_{ks} \right| \leq \left(\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k)^2 \right)^{1/2} \left(\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \eta_{ks}^2 \right)^{1/2}.$$

Por hipótesis sabemos que $h_{2n} \rightarrow 0$ y $\frac{\log \hat{n}}{\hat{n} h_{1n}^d h_{2n}^k} \rightarrow 0$ cuando $\hat{n}$ tiende a infinito. Esto, junto con (2.3.21), nos aseguran que para $\hat{n} \rightarrow \infty$,

$$\max_k \bar{g}_j(\mathbf{T}_k)^2 \rightarrow 0,$$

luego

$$\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k)^2 \leq \max_k \bar{g}_j(\mathbf{T}_k)^2 \rightarrow 0. \quad (2.3.22)$$

Finalmente, usando (2.3.20) y (2.3.22), tenemos que

$$\hat{n}^{-1} \sum_{k=1}^{\hat{n}} \bar{g}_j(\mathbf{T}_k) \eta_{ks} \rightarrow 0.$$

Los mismos argumentos se aplican a los términos restantes, lo que concluye la prueba del lema. □

Demostración del Teorema 2.3.2. Recordemos que $\hat{\beta}_{\mathbf{h}} = (\tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbf{Y}}$ donde $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{W}_{\mathbf{h}} \mathbf{Y}$ y $\tilde{\mathbb{X}} = \mathbb{X} - \mathbf{W}_{\mathbf{h}} \mathbb{X}$. Con ello tenemos que

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{\mathbf{h}} - \beta) &= \sqrt{n}((\tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbf{Y}} - \beta) \\ &= \sqrt{n}(\tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1} (\tilde{\mathbb{X}}^T \tilde{\mathbf{Y}} - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \beta) \\ &= \sqrt{n}(\tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1} (\tilde{\mathbb{X}}^T (\mathbf{Y} - \mathbf{W}_{\mathbf{h}} \mathbf{Y}) - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \beta). \end{aligned} \quad (2.3.23)$$

Además, definiendo $\mathbf{r} \doteq (r_1, \dots, r_{\hat{n}})^T$ con $r_i = r(\mathbf{T}_i)$ y $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{\hat{n}})$, podemos escribir el modelo (2.1.1) de forma matricial como sigue

$$\mathbf{Y} = \mathbb{X} \beta + \mathbf{r} + \boldsymbol{\varepsilon}.$$

Reemplazando el modelo en (2.3.23) tenemos que

$$\begin{aligned} \tilde{\mathbb{X}}^T (\mathbf{Y} - \mathbf{W}_{\mathbf{h}} \mathbf{Y}) - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \beta &= \tilde{\mathbb{X}}^T \mathbf{Y} - \tilde{\mathbb{X}}^T \mathbf{W}_{\mathbf{h}} \mathbf{Y} - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \beta \\ &= \tilde{\mathbb{X}}^T (\mathbb{X} \beta + \mathbf{r} + \boldsymbol{\varepsilon}) - \tilde{\mathbb{X}}^T \mathbf{W}_{\mathbf{h}} (\mathbb{X} \beta + \mathbf{r} + \boldsymbol{\varepsilon}) - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \beta \\ &= \tilde{\mathbb{X}}^T \mathbb{X} \beta + \tilde{\mathbb{X}}^T \mathbf{r} + \tilde{\mathbb{X}}^T \boldsymbol{\varepsilon} - \tilde{\mathbb{X}}^T \mathbf{W}_{\mathbf{h}} \mathbb{X} \beta \end{aligned}$$

$$\begin{aligned}
& -\tilde{\mathbb{X}}^T \mathbf{W}_h \mathbf{r} - \tilde{\mathbb{X}}^T \mathbf{W}_h \boldsymbol{\varepsilon} - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \boldsymbol{\beta} \\
& = \tilde{\mathbb{X}}^T \mathbb{X} \boldsymbol{\beta} - \tilde{\mathbb{X}}^T \mathbf{W}_h \mathbb{X} \boldsymbol{\beta} - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \boldsymbol{\beta} \\
& \quad + \tilde{\mathbb{X}}^T \mathbf{r} - \tilde{\mathbb{X}}^T \mathbf{W}_h \mathbf{r} \\
& \quad + \tilde{\mathbb{X}}^T \boldsymbol{\varepsilon} - \tilde{\mathbb{X}}^T \mathbf{W}_h \boldsymbol{\varepsilon}.
\end{aligned} \tag{2.3.24}$$

Luego, como

$$\begin{aligned}
\tilde{\mathbb{X}}^T \mathbb{X} \boldsymbol{\beta} - \tilde{\mathbb{X}}^T \mathbf{W}_h \mathbb{X} \boldsymbol{\beta} &= \tilde{\mathbb{X}}^T (\mathbb{X} - \mathbf{W}_h \mathbb{X}) \boldsymbol{\beta} \\
&= \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \boldsymbol{\beta},
\end{aligned}$$

en (2.3.24) tenemos que

$$\tilde{\mathbb{X}}^T \mathbb{X} \boldsymbol{\beta} - \tilde{\mathbb{X}}^T \mathbf{W}_h \mathbb{X} \boldsymbol{\beta} - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \boldsymbol{\beta} = 0.$$

Por otro lado, definiendo $\tilde{\mathbf{r}} \doteq \mathbf{r} - \mathbf{W}_h \mathbf{r}$, resulta que

$$\tilde{\mathbb{X}}^T \mathbf{r} - \tilde{\mathbb{X}}^T \mathbf{W}_h \mathbf{r} = \tilde{\mathbb{X}}^T \tilde{\mathbf{r}}.$$

Con estas dos igualdades, tenemos que

$$\tilde{\mathbb{X}}^T (\mathbf{Y} - \mathbf{W}_h \mathbf{Y}) - \tilde{\mathbb{X}}^T \tilde{\mathbb{X}} \boldsymbol{\beta} = \tilde{\mathbb{X}}^T \tilde{\mathbf{r}} + \tilde{\mathbb{X}}^T \boldsymbol{\varepsilon} - \tilde{\mathbb{X}}^T \mathbf{W}_h \boldsymbol{\varepsilon}.$$

Finalmente en (2.3.23), resulta

$$\begin{aligned}
\sqrt{\hat{n}}(\hat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}) &= \sqrt{\hat{n}}(\tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1}(\tilde{\mathbb{X}}^T \tilde{\mathbf{r}} - \tilde{\mathbb{X}}^T \mathbf{W}_h \boldsymbol{\varepsilon} + \tilde{\mathbb{X}}^T \boldsymbol{\varepsilon}) \\
&\doteq (\hat{n}^{-1} \tilde{\mathbb{X}}^T \tilde{\mathbb{X}})^{-1} \hat{n}^{-\frac{1}{2}} (\mathbf{S}_{n1} - \mathbf{S}_{n2} + \mathbf{S}_{n3}).
\end{aligned} \tag{2.3.25}$$

Para el estudio del comportamiento asintótico de $\mathbf{S}_{n1}$, $\mathbf{S}_{n2}$ y $\mathbf{S}_{n3}$, vectores en $\mathbb{R}^p$, primero trabajaremos en la descomposición del elemento ij de la matriz $\tilde{\mathbb{X}}$ presente en todos los términos. Para ello, sumando y restando $\tilde{g}_j(\mathbf{T}_i) \doteq \mathbb{E}[X_{ij}|\mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[X_{kj}|\mathbf{T}_k]$ y utilizando $\eta_{ij} = X_{ij} - \mathbb{E}[X_{ij}|\mathbf{T}_i]$ definida en H7, tenemos que

$$\begin{aligned}
\tilde{X}_{ij} &= X_{ij} - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) X_{kj} \\
&= \mathbb{E}[X_{ij}|\mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[X_{kj}|\mathbf{T}_k] + X_{ij} - \mathbb{E}[X_{ij}|\mathbf{T}_i] \\
&\quad - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) X_{kj} + \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[X_{kj}|\mathbf{T}_k] \\
&= \mathbb{E}[X_{ij}|\mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[X_{kj}|\mathbf{T}_k] + X_{ij} - \mathbb{E}[X_{ij}|\mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \\
&= \tilde{g}_j(\mathbf{T}_i) + \eta_{ij} - \sum_{k=1}^{\hat{n}} \omega_h(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj}.
\end{aligned} \tag{2.3.26}$$

Con esto, podemos escribir el elemento j -ésimo del vector $\mathbf{S}_{n1} = \tilde{\mathbf{X}}^T \tilde{\mathbf{r}} = \sum_{i=1}^{\hat{n}} \tilde{\mathbf{X}}_i \tilde{r}(\mathbf{T}_i)$ como

$$\begin{aligned} S_{n1,j} &= \sum_{i=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_i) \tilde{r}(\mathbf{T}_i) + \sum_{i=1}^{\hat{n}} \eta_{ij} \tilde{r}(\mathbf{T}_i) - \sum_{i=1}^{\hat{n}} \tilde{r}(\mathbf{T}_i) \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right) \\ &\doteq S_{n1,j1} + S_{n1,j2} - S_{n1,j3}. \end{aligned} \quad (2.3.27)$$

De la misma manera, podemos escribir el elemento j -ésimo del vector $\mathbf{S}_{n2} = \tilde{\mathbf{X}}^T \mathbf{W}_{\mathbf{h}} \boldsymbol{\varepsilon} = \sum_{i=1}^{\hat{n}} \tilde{\mathbf{X}}_i \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k \right)$ como

$$\begin{aligned} S_{n2,j} &= \sum_{i=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_i) \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k \right) + \sum_{i=1}^{\hat{n}} \eta_{ij} \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k \right) \\ &\quad - \sum_{i=1}^{\hat{n}} \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_l) \varepsilon_l \right) \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right) \\ &\doteq S_{n2,j1} + S_{n2,j2} - S_{n2,j3}, \end{aligned} \quad (2.3.28)$$

y el elemento j -ésimo del vector $\mathbf{S}_{n3} = \tilde{\mathbf{X}}^T \boldsymbol{\varepsilon} = \sum_{i=1}^{\hat{n}} \tilde{\mathbf{X}}_i \varepsilon_i$ como

$$\begin{aligned} S_{n3,j} &= \sum_{i=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_i) \varepsilon_i + \sum_{i=1}^{\hat{n}} \eta_{ij} \varepsilon_i - \sum_{i=1}^{\hat{n}} \varepsilon_i \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right) \\ &\doteq S_{n3,j1} + S_{n3,j2} - S_{n3,j3}. \end{aligned} \quad (2.3.29)$$

En lo que sigue aplicaremos el Lema A.0.2 sobre $\{\varepsilon_k\}$ y $\{\eta_{kj}\}$; para ello, a continuación verificamos que ambos procesos satisfacen sus hipótesis. A partir de las hipótesis H1, H4, H5, H6 y H7 y el Lema 2.3.6 tenemos que las secuencias $\{\varepsilon_i\}$, $\{\eta_i\}$ tienen media cero, son estrictamente estacionarias y fuertemente mixing, con coeficientes mixing acotados por $\alpha(r)$. Además, el Lema 2.3.6 también nos asegura que existe $r > 4$ y una constante $C > 0$ tales que

$$\begin{aligned} \max_{1 \leq i \leq n} \mathbb{E} [|\varepsilon_i|^r] &\leq C < \infty, \\ \max_{1 \leq i \leq n} \mathbb{E} [\|\eta_i\|^r] &\leq C < \infty. \end{aligned}$$

Por otro lado, el Lema 2.3.3 asegura que

$$\sum_{\hat{n}=1}^{\infty} \hat{n}^{\frac{5+4\gamma}{4(1-\gamma)}} \alpha(\hat{n}) \leq \sum_{\hat{n}=1}^{\infty} \hat{n}^{\frac{5+4\gamma}{4(1-\gamma)}} \hat{n}^{-b/d},$$

además, por la hipótesis H6, $b > 4.5d$ y eligiendo $0.5 < \gamma < 1 - \frac{9}{4b}$, la serie resulta convergente. Por lo tanto, ambos procesos satisfacen todas las condiciones requeridas para la aplicación del Lema A.0.2. Además, por el Lema 2.3.7 sabemos que

$$\max_{1 \leq i, j \leq \hat{n}} |\omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_j)| = O((\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1}), \text{ c.s.}$$

Con esto tenemos que si en el Lema A.0.2, tomamos $a_{ik} = \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k)$ con $a_{\mathbf{n}} = (\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1}$, $V_k = \varepsilon_k$ y $0.5 < \gamma < 1 - 9/(4b)$, tenemos que

$$\max_i \left| \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k \right| = O \left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n} \right) \quad \text{c.s.} \quad (2.3.30)$$

y, bajo las mismas condiciones tomando $V_k = \eta_{kj}$ tenemos que

$$\max_i \left| \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right| = O \left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n} \right) \quad \text{c.s.}, \quad (2.3.31)$$

ambos para $r > 4$.

Por otro lado, usando la igualdad dada en (2.3.26) y la definición de η_{kj} , tenemos que

$$\begin{aligned} \tilde{g}_j(\mathbf{T}_i) &= \tilde{X}_{ij} - \eta_{ij} + \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \\ &= X_{ij} - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) X_{kj} - X_{ij} + \mathbb{E}[X_{ij} | \mathbf{T}_i] + \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \\ &= \mathbb{E}[X_{ij} | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) X_{kj} + \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \\ &= \bar{g}_j(\mathbf{T}_i) + \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj}, \end{aligned}$$

con $\bar{g}(\mathbf{T}_i)$ definido en (2.3.17). Finalmente, aplicando lo obtenido en (2.3.21) y (2.3.31) tenemos que,

$$\max_{i,j} |\tilde{g}_j(\mathbf{T}_i)| = O(h_{2\mathbf{n}}) + O \left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n} \right) \quad \text{c.s.} \quad (2.3.32)$$

De igual manera, necesitamos acotar el máximo $|\tilde{r}(\mathbf{T}_i)|$, para ello recordemos que $\tilde{\mathbf{r}} = \mathbf{r} - \mathbf{W}_{\mathbf{h}} \mathbf{r}$, por lo que el elemento i -ésimo del vector $\tilde{\mathbf{r}}$ está dado por

$$\tilde{r}(\mathbf{T}_i) = r(\mathbf{T}_i) - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) r(\mathbf{T}_k).$$

Además, por un lado, considerando la definición del modelo (2.1.1), tenemos que $r(\mathbf{T}_k) = Y_k - \mathbf{X}_k^T \boldsymbol{\beta} - \varepsilon_k$. Por otro lado, considerando la esperanza condicional (2.2.1) $r(\mathbf{T}_k) = \mathbb{E}[Y_i - \mathbf{X}_i^T \boldsymbol{\beta} | \mathbf{T}_i]$. Con ello, tenemos que la ecuación anterior puede escribirse como

$$\begin{aligned} \tilde{r}(\mathbf{T}_i) &= r(\mathbf{T}_i) - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) r(\mathbf{T}_k) \\ &= \mathbb{E}[Y_i - \mathbf{X}_i^T \boldsymbol{\beta} | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) (Y_k - \mathbf{X}_k^T \boldsymbol{\beta} - \varepsilon_k) \\ &= \mathbb{E}[Y_i | \mathbf{T}_i] - \mathbb{E}[\mathbf{X}_i^T \boldsymbol{\beta} | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) Y_k \end{aligned}$$

$$+ \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \mathbf{X}_k^T \boldsymbol{\beta} + \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k,$$

luego, sumando y restando $\boldsymbol{\beta}^T \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[\mathbf{X}_k | \mathbf{T}_k]$, obtenemos

$$\tilde{r}(\mathbf{T}_i) = \mathbb{E}[Y_i | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) Y_k \quad (2.3.33)$$

$$+ \boldsymbol{\beta}^T \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \left(\mathbf{X}_k - \mathbb{E}[\mathbf{X}_k | \mathbf{T}_k] \right) \quad (2.3.34)$$

$$- \boldsymbol{\beta}^T \left(\mathbb{E}[\mathbf{X}_i | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[\mathbf{X}_k | \mathbf{T}_k] \right) \quad (2.3.35)$$

$$+ \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k. \quad (2.3.36)$$

Si en (2.3.33), los pesos $\omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k)$ están dado por (2.2.5), $\mathbf{T}_k$ son las covariables y cada Y_k representa la variable a predecir, bajo las hipótesis H1–H6, y, por la Observación 2.3.1, aplicamos directamente la Observación 4 de Dabo-Niang et al. (2016), obteniendo los mismos órdenes que en (2.3.21), esto es,

$$\max_{1 \leq i \leq \hat{n}} \left| \mathbb{E}[Y_i | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) Y_k \right| = O(h_{2\mathbf{n}}) + O\left(\frac{\log \hat{n}}{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}\right). \quad (2.3.37)$$

En (2.3.34), usando (2.3.31), para cada j -ésimo elemento del vector $\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) (\mathbf{X}_k - \mathbb{E}[\mathbf{X}_k | \mathbf{T}_k])$ tenemos que

$$\begin{aligned} \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) (\mathbf{X}_k - \mathbb{E}[\mathbf{X}_k | \mathbf{T}_k]) \right)_j &= \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \\ &\leq \max_i \left| \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right| \\ &= O\left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n}\right) \quad \text{c.s.} \end{aligned} \quad (2.3.38)$$

En (2.3.35), usando (2.3.32), tenemos que

$$\begin{aligned} \left(\mathbb{E}[\mathbf{X}_i | \mathbf{T}_i] - \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \mathbb{E}[\mathbf{X}_k | \mathbf{T}_k] \right)_j &= \tilde{g}_j(\mathbf{T}_i) \\ &\leq \max_{i,j} |\tilde{g}_{j\mathbf{h}}(\mathbf{T}_i)| \\ &= O(h_{2\mathbf{n}}) + O\left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n}\right) \quad \text{c.s.} \end{aligned} \quad (2.3.39)$$

Por último, para (2.3.36), obtuvimos la cota en (2.3.30). Con esto, (2.3.39), (2.3.38) y (2.3.37), tenemos que

$$\max_i |\tilde{r}_{\mathbf{h}}(\mathbf{T}_i)| = O(h_{2\mathbf{n}}) + O\left(\sqrt{\frac{\log \hat{n}}{n h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}}\right) + O\left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n}\right)$$

$$= O(h_{2\mathbf{n}}) + O\left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-1/2+1/r} \log \hat{n}\right) \quad \text{c.s.} \quad (2.3.40)$$

A continuación, analizamos los nueve términos, obtenidos en (2.3.27), (2.3.28) y (2.3.29), que componen el elemento j -ésimo de los vectores $\mathbf{S}_{n1}$, $\mathbf{S}_{n2}$ y $\mathbf{S}_{n3}$, respectivamente. Para $S_{n1,j1}$, consideramos los resultados hallados en (2.3.40), (2.3.32) y acotando por el máximo en i , tenemos que

$$\begin{aligned} \hat{n}^{-\frac{1}{2}} S_{n1,j1} &= \hat{n}^{-\frac{1}{2}} \sum_{i=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_i) \tilde{r}_h(\mathbf{T}_i) \\ &\leq \hat{n}^{-\frac{1}{2}} \hat{n} \max_i |\tilde{g}_j(\mathbf{T}_i)| \max_i |\tilde{r}(\mathbf{T}_i)| \\ &= \hat{n}^{\frac{1}{2}} \left(O(h_{2\mathbf{n}}^2) + O\left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-2} \hat{n}^{\frac{2}{r}-1} \log^2 \hat{n}\right) \right) \\ &= O(\hat{n}^{\frac{1}{2}} h_{2\mathbf{n}}^2) + O\left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-2} \hat{n}^{\frac{2}{r}-1+\frac{1}{2}} \log^2 \hat{n}\right) \\ &= O\left((\hat{n} h_{2\mathbf{n}}^4)^{\frac{1}{2}}\right) + O\left(\left(\frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4}-\frac{1}{r}}}\right)^2\right) \quad \text{c.s.} \end{aligned} \quad (2.3.41)$$

Para el segundo término de $S_{n1,j}$, volvemos a aplicar el Lema A.0.2, tomando $a_{ik} = \tilde{r}_h(T_k)$ con $a_{\mathbf{n}} = h_{2\mathbf{n}} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-\frac{1}{2}+\frac{1}{r}} \log \hat{n}$ (probado en (2.3.40)), $V_k = \eta_{kj}$ y $0.5 < \gamma < 1 - 9/(4b)$, obteniendo

$$\begin{aligned} \hat{n}^{-\frac{1}{2}} S_{n1,j2} &= \hat{n}^{-\frac{1}{2}} \sum_{k=1}^{\hat{n}} \tilde{r}(T_k) \eta_{kj} \\ &= \hat{n}^{-\frac{1}{2}} \max_i \left| \sum_{k=1}^{\hat{n}} \tilde{r}(T_k) \eta_{kj} \right| \\ &= \hat{n}^{-\frac{1}{2}} O\left(h_{2\mathbf{n}} \hat{n}^{\frac{1}{2}+\frac{1}{r}} \log \hat{n} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{\frac{2}{r}} \log^2 \hat{n}\right) \\ &= O\left(h_{2\mathbf{n}} \hat{n}^{\frac{1}{r}} \log \hat{n} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{\frac{2}{r}-\frac{1}{2}} \log^2 \hat{n}\right) \\ &= O\left((h_{2\mathbf{n}}^4 \hat{n})^{\frac{1}{4}} \frac{\log \hat{n}}{\hat{n}^{\frac{1}{4}-\frac{1}{r}}} + \left(\frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} \hat{n}^{\frac{1}{4}-\frac{1}{r}}}\right)^2\right) \quad \text{c.s.} \end{aligned} \quad (2.3.42)$$

En el siguiente caso, acotamos por los máximos y aplicamos lo obtenido en (2.3.31) y (2.3.40), esto es,

$$\begin{aligned} \hat{n}^{-\frac{1}{2}} S_{n1,j3} &= \hat{n}^{-\frac{1}{2}} \sum_{i=1}^{\hat{n}} \tilde{r}(\mathbf{T}_i) \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right) \\ &\leq \hat{n}^{-\frac{1}{2}} \hat{n} \max_i |\tilde{r}(\mathbf{T}_i)| \max_i \left| \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right| \\ &= \hat{n}^{\frac{1}{2}} O\left(h_{2\mathbf{n}} (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-\frac{1}{2}+\frac{1}{r}} \log \hat{n} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-2} \hat{n}^{-1+\frac{2}{r}} \log^2 \hat{n}\right) \\ &= O\left((h_{2\mathbf{n}}^4 \hat{n})^{\frac{1}{4}} \frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4}-\frac{1}{r}}} + \left(\frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4}-\frac{1}{r}}}\right)^2\right) \quad \text{c.s.} \end{aligned} \quad (2.3.43)$$

Mediante un razonamiento similar al usado para obtener (2.3.43), pero utilizando los resultados obtenidos en (2.3.30) y (2.3.32) en lugar de (2.3.31) y (2.3.40), respectivamente, se tiene que

$$\begin{aligned}
\hat{n}^{-\frac{1}{2}} S_{n2,j1} &= \hat{n}^{-\frac{1}{2}} \sum_{i=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_i) \left(\sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k \right) \\
&\leq \hat{n}^{-\frac{1}{2}} \hat{n} \max_i |\tilde{g}_j(\mathbf{T}_i)| \max_i \left| \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \varepsilon_k \right| \\
&= \hat{n}^{\frac{1}{2}} O \left((h_{2\mathbf{n}}^d h_{1\mathbf{n}}^k)^{-1} \hat{n}^{-\frac{1}{2} + \frac{1}{r}} \log \hat{n} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-2} \hat{n}^{-1 + \frac{2}{r}} \log^2 \hat{n} \right) \\
&= O \left((h_{2\mathbf{n}}^4 \hat{n})^{\frac{1}{4}} \frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4} - \frac{1}{r}}} + \left(\frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \right)^2 \right) \quad \text{c.s.} \quad (2.3.44)
\end{aligned}$$

Por otro lado, considerando $a_{ik} = \sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \varepsilon_l$ con $\mathbf{a}_{\mathbf{n}} = (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-\frac{1}{2} + \frac{1}{r}} \log \hat{n}$, $V_k = \eta_{kj}$ y $0.5 < \gamma < 1 - 9/(4b)$ en el Lema A.0.2, obtenemos

$$\begin{aligned}
\hat{n}^{-\frac{1}{2}} S_{n2,j2} &= \hat{n}^{-\frac{1}{2}} \sum_{k=1}^{\hat{n}} \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \varepsilon_l \right) \eta_{kj} \\
&= \hat{n}^{-\frac{1}{2}} \max_i \left| \sum_{k=1}^{\hat{n}} \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \varepsilon_l \right) \eta_{kj} \right| \\
&= \hat{n}^{-\frac{1}{2}} O \left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{\frac{2}{r}} \log^2 \hat{n} \right) \\
&= O \left(\left(\frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \right)^2 \right) \quad \text{c.s.} \quad (2.3.45)
\end{aligned}$$

En el siguiente término acotamos por el máximo y, utilizando lo obtenido en (2.3.30) y (2.3.31), tenemos que

$$\begin{aligned}
\hat{n}^{-\frac{1}{2}} S_{n2,j3} &= \hat{n}^{-\frac{1}{2}} \sum_{k=1}^{\hat{n}} \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \eta_{lj} \right) \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \varepsilon_l \right) \\
&\leq \hat{n}^{-\frac{1}{2}} \hat{n} \max_i \left| \sum_{k=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_k) \eta_{kj} \right| \max_i \left| \sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_i, \mathbf{T}_l) \varepsilon_l \right| \\
&= \hat{n}^{\frac{1}{2}} O \left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-2} \hat{n}^{-1 + \frac{2}{r}} \log^2 \hat{n} \right) \\
&= O \left(\left(\frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \right)^2 \right) \quad \text{c.s.} \quad (2.3.46)
\end{aligned}$$

Usando el mismo razonamiento utilizado para obtener (2.3.42), esto es, volvemos a aplicar el Lema A.0.2, tomando $a_{ik} = \tilde{g}_j(\mathbf{T}_k)$ con $\mathbf{a}_{\mathbf{n}} = h_{2\mathbf{n}} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-\frac{1}{2} + \frac{1}{r}} \log \hat{n}$ (probado en (2.3.32)), $V_k = \epsilon_k$ y $0.5 < \gamma < 1 - 9/(4b)$, con ello, tenemos que

$$\hat{n}^{-\frac{1}{2}} S_{n3,j1} = \hat{n}^{-\frac{1}{2}} \sum_{k=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_k) \varepsilon_k$$

$$\begin{aligned}
&= \hat{n}^{-\frac{1}{2}} \max_i \left| \sum_{k=1}^{\hat{n}} \tilde{g}_j(\mathbf{T}_k) \varepsilon_k \right| \\
&= \hat{n}^{-\frac{1}{2}} O \left(h_{2\mathbf{n}} \hat{n}^{\frac{1}{2} + \frac{1}{r}} \log \hat{n} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{\frac{2}{r}} \log^2 \hat{n} \right) \\
&= O \left(h_{2\mathbf{n}} \hat{n}^{\frac{1}{r}} \log \hat{n} + (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{\frac{2}{r} - \frac{1}{2}} \log^2 \hat{n} \right) \\
&= O \left((h_{2\mathbf{n}}^4 \hat{n})^{\frac{1}{4}} \frac{\log \hat{n}}{\hat{n}^{\frac{1}{4} - \frac{1}{r}}} + \left(\frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \right)^2 \right) \quad \text{c.s.} \quad (2.3.47)
\end{aligned}$$

Si utilizamos el razonamiento con el que obtuvimos (2.3.45) tomando $a_{ik} = \sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \eta_{lj}$ con $a_{\mathbf{n}} = (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{-\frac{1}{2} + \frac{1}{r}} \log \hat{n}$, (probado en (2.3.31)), $V_k = \varepsilon_k$ y $0.5 < \gamma < 1 - 9/(4b)$ tenemos que

$$\begin{aligned}
\hat{n}^{-\frac{1}{2}} S_{n3,j3} &= \hat{n}^{-\frac{1}{2}} \sum_{k=1}^{\hat{n}} \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \eta_{lj} \right) \varepsilon_k \\
&= \hat{n}^{-\frac{1}{2}} \max_i \left| \sum_{k=1}^{\hat{n}} \left(\sum_{l=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{T}_k, \mathbf{T}_l) \eta_{lj} \right) \varepsilon_k \right| \\
&= \hat{n}^{-\frac{1}{2}} O \left((h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{-1} \hat{n}^{\frac{2}{r}} \log^2 \hat{n} \right) \\
&= O \left(\left(\frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \right)^2 \right) \quad \text{c.s.} \quad (2.3.48)
\end{aligned}$$

Para probar que los órdenes hallados convergen a cero, usaremos las hipótesis del Teorema:

$$\hat{n} h_{2\mathbf{n}}^4 \rightarrow 0 \quad \text{y} \quad \frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \rightarrow 0, \quad (2.3.49)$$

cuando $n \rightarrow \infty$. En primer lugar, bajo dichas hipótesis, es directo que, $\hat{n}^{-\frac{1}{2}} S_{n1,j1}$, $\hat{n}^{-\frac{1}{2}} S_{n1,j3}$, y $\hat{n}^{-\frac{1}{2}} S_{n2,j1}$, dadas en (2.3.41), (2.3.43), (2.3.44), respectivamente, convergen a cero cuando $n \rightarrow \infty$. Por otro lado, como $h_{1\mathbf{n}}$ y $h_{2\mathbf{n}}$ son los parámetros de suavizado, que tienden a cero, lo que para un $\hat{n}$ suficientemente grande implica que $(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} > (h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)$, lo que nos asegura que

$$\frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k)^{\frac{1}{2}} \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \leq \frac{\log \hat{n}}{(h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k) \hat{n}^{\frac{1}{4} - \frac{1}{r}}}. \quad (2.3.50)$$

De (2.3.49) y (2.3.50) tenemos que $\hat{n}^{-\frac{1}{2}} S_{n2,j2}$, $\hat{n}^{-\frac{1}{2}} S_{n2,j3}$ y $\hat{n}^{-\frac{1}{2}} S_{n3,j3}$, dados en, (2.3.45), (2.3.46), (2.3.48) convergen a cero cuando $n \rightarrow \infty$. Por último, como $r > 4$, (por hipótesis H4) para $\hat{n}$ suficientemente grande tenemos que,

$$\frac{\log \hat{n}}{\hat{n}^{\frac{1}{4} - \frac{1}{r}}} \rightarrow 0,$$

esto nos asegura, junto con (2.3.49) y (2.3.50), que $\hat{n}^{-\frac{1}{2}} S_{n1,j2}$ y $\hat{n}^{-\frac{1}{2}} S_{n3,j1}$ obtenidos en (2.3.42) y (2.3.47) convergen a cero. Considerando estos tres casos obtenemos que, si $\hat{n} \rightarrow \infty$

$$\hat{n}^{-\frac{1}{2}} (S_{n1,j1} + S_{n1,j2} + S_{n1,j3} - S_{n2,j1} - S_{n2,j2} - S_{n2,j3} + S_{n3,j1} + S_{n3,j3}) \rightarrow 0,$$

equivalente a

$$S_{n1,j1} + S_{n1,j2} + S_{n1,j3} - S_{n2,j1} - S_{n2,j2} - S_{n2,j3} + S_{n3,j1} + S_{n3,j3} = o(\hat{n}^{\frac{1}{2}}),$$

y por lo tanto,

$$\mathbf{S}_{n1} - \mathbf{S}_{n2} + \mathbf{S}_{n3} = S_{n3,j2} + o(\hat{n}^{\frac{1}{2}}) = \sum_{i=1}^{\hat{n}} \boldsymbol{\eta}_i \varepsilon_i + o(\hat{n}^{\frac{1}{2}}) \quad \text{c.s.} \quad (2.3.51)$$

Ahora, mediante el Teorema A.0.4 queremos probar que

$$\hat{n}^{-1/2} S_{n3,j2} = \hat{n}^{-1/2} \sum_{i=1}^{\hat{n}} \boldsymbol{\eta}_i \varepsilon_i \xrightarrow{d} N(0, \mathbf{C}), \quad (2.3.52)$$

donde $\mathbf{C} = \lim_{\hat{n} \rightarrow \infty} \hat{n}^{-1} \mathbb{E} [\boldsymbol{\eta}^T \mathbf{V}_\varepsilon \boldsymbol{\eta}]$ definido en H8. Para ello en lo que sigue, probaremos el cumplimiento de las hipótesis de dicho Teorema. En primer lugar necesitamos que $\{\boldsymbol{\eta}_i \varepsilon_i\}$ tenga esperanza cero, sea estacionaria, α -mixing y que exista δ de forma tal que $\mathbb{E} [\|\boldsymbol{\eta}_i \varepsilon_i\|^{2+\delta}] < \infty$. Lo cual queda asegurado por el Lema 2.3.6. Además, queremos probar que la serie $\sum_{n=1}^{\infty} \alpha(\hat{n})^{\frac{\delta}{2+\delta}}$ es convergente para algún $\delta > 0$. Esto queda probado, ya que, por la hipótesis H6 y el Lema 2.3.3, sabemos que

$$\sum_{\hat{n}=1}^{\infty} \alpha(\hat{n})^{\frac{\delta}{2+\delta}} \leq \sum_{\hat{n}=1}^{\infty} \left(\hat{n}^{-\frac{b}{d}} \right)^{\frac{\delta}{2+\delta}},$$

por lo que, eligiendo $\frac{2}{b-d} < \delta < r-2$ y utilizando $b > 4.5d$ (H6), que nos asegura que δ es positivo, tenemos que la serie es convergente. Por lo que (2.3.52) es válido.

Ahora, a partir de (2.3.51) y (2.3.52) tenemos que

$$\hat{n}^{-\frac{1}{2}} (\mathbf{S}_{n1} - \mathbf{S}_{n2} + \mathbf{S}_{n3}) \xrightarrow{d} N(0, \mathbf{C}). \quad (2.3.53)$$

Finalmente, usando (2.3.53), que

$$\hat{n}^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \longrightarrow \mathbf{B} \quad \text{c.s.}, \quad (2.3.54)$$

obtenido en el Lema 2.3.8, y aplicando el Teorema de Slutsky en (2.3.25), tenemos que

$$\sqrt{\hat{n}} (\hat{\boldsymbol{\beta}}_{\mathbf{h}} - \boldsymbol{\beta}) = (\hat{n}^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \hat{n}^{-\frac{1}{2}} (\mathbf{S}_{n1} - \mathbf{S}_{n2} + \mathbf{S}_{n3}) \xrightarrow{d} N(0, \mathbf{B}^{-1} \mathbf{C} \mathbf{B}^{-1}),$$

lo cual prueba la primera parte del Teorema.

Para probar la segunda parte del Teorema, dada en (2.3.2), esto es,

$$\limsup_{\hat{n} \rightarrow \infty} \left(\frac{\hat{n}}{2 \log \log \hat{n}} \right)^{\frac{1}{2}} |\hat{\boldsymbol{\beta}}_{hj} - \boldsymbol{\beta}_j| = a_{jj}^{\frac{1}{2}} \quad \text{c.s.},$$

donde $a_{jj} = \mathbf{A}_{jj}$ y $\mathbf{A} = \mathbf{B}^{-1} \mathbf{C} \mathbf{B}^{-1}$, consideramos la descomposición en (2.3.25). Luego de (2.3.54) y (2.3.51), tenemos que

$$\hat{\boldsymbol{\beta}}_{\mathbf{h}} - \boldsymbol{\beta} = \left(\hat{n}^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \right)^{-1} \hat{n}^{-1} (\mathbf{S}_{n1} - \mathbf{S}_{n2} + \mathbf{S}_{n3})$$

$$= (\mathbf{B}^{-1} + o(1)) \left(\hat{n}^{-1} \sum_{i=1}^{\hat{n}} \boldsymbol{\eta}_i \varepsilon_i + o(\hat{n}^{-1/2}) \right) \quad \text{c.s.} \quad (2.3.55)$$

Ahora, teniendo en cuenta lo que queremos probar y considerando que $\mathbf{b}_j$ es la j -ésima fila de la matriz $\mathbf{B}^{-1}$, estudiaremos el término

$$\left(\frac{\hat{n}}{2 \log \log \hat{n}} \right)^{\frac{1}{2}} \hat{n}^{-1} \mathbf{b}_j \cdot \sum_{i=1}^{\hat{n}} \boldsymbol{\eta}_i \varepsilon_i = \left(\frac{1}{2 \hat{n} \log \log \hat{n}} \right)^{-\frac{1}{2}} \mathbf{b}_j \cdot \sum_{i=1}^{\hat{n}} \boldsymbol{\eta}_i \varepsilon_i.$$

Para ello, usaremos el Teorema A.0.1, tomando $V_i = \mathbf{b}_j \cdot \boldsymbol{\eta}_i \varepsilon_i$. Usando el Lema 2.3.6 (aplicable bajo H1, H4, H5, H6 y H7), y que $\mathbf{b}_j$ es no aleatoria y no depende de i , tenemos que $\{V_i\}$ tiene media cero, es estrictamente estacionaria, α -mixing y $\mathbb{E}[|V_i|^{2+\delta}] < \infty$. Por otro lado, por el Lema 2.3.3, sabemos que

$$\sum_{n=1}^{\infty} \alpha(\hat{n})^{\frac{\delta'}{2+\delta'}} \leq \sum_{n=1}^{\infty} \left(\hat{n}^{-\frac{b}{d}} \right)^{\frac{\delta'}{2+\delta'}},$$

además, por la hipótesis H6, $b > 4.5d$ y eligiendo $\frac{2}{b-d} < \delta' < \delta < r-2$, la serie resulta convergente.

Sea $S_n = \sum_{i=1}^{\hat{n}} V_i$, también tenemos que probar que la varianza asintótica de S_n es positiva. Por un lado, considerando $\boldsymbol{\eta}^T = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)$ y $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ definidas en H7 y que $\boldsymbol{\eta}$ es independiente de $\boldsymbol{\varepsilon}$, tenemos que

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^n \mathbf{b}_j \cdot \boldsymbol{\eta}_i \varepsilon_i \right) &= \mathbf{b}_j \cdot \text{Var} \left(\sum_{i=1}^n \boldsymbol{\eta}_i \varepsilon_i \right) \mathbf{b}_j^T \\ &= \mathbf{b}_j \cdot \text{Var} (\boldsymbol{\eta}^T \boldsymbol{\varepsilon}) \mathbf{b}_j^T \\ &= \mathbf{b}_j \cdot \mathbb{E} [\boldsymbol{\eta}^T \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \boldsymbol{\eta}] \mathbf{b}_j^T \\ &= \mathbf{b}_j \cdot \mathbb{E} [\mathbb{E} [\boldsymbol{\eta}^T \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \boldsymbol{\eta} | \boldsymbol{\eta}]] \mathbf{b}_j^T \\ &= \mathbf{b}_j \cdot \mathbb{E} [\boldsymbol{\eta}^T \mathbb{E} [\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T | \boldsymbol{\eta}] \boldsymbol{\eta}] \mathbf{b}_j^T \\ &= \mathbf{b}_j \cdot \mathbb{E} [\boldsymbol{\eta}^T \mathbb{E} [\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T] \boldsymbol{\eta}] \mathbf{b}_j^T, \end{aligned}$$

dividiendo por $\hat{n}$, obtenemos que

$$\frac{1}{\hat{n}} \text{Var} (S_n) = \mathbf{b}_j \cdot \left(\frac{1}{\hat{n}} \mathbb{E} [\boldsymbol{\eta}^T V_{\varepsilon} \boldsymbol{\eta}] \right) \mathbf{b}_j^T. \quad (2.3.56)$$

Por hipótesis H8, el límite

$$\mathbf{C} = \lim_{\hat{n} \rightarrow \infty} \frac{1}{\hat{n}} \mathbb{E} [\boldsymbol{\eta}^T V_{\varepsilon} \boldsymbol{\eta}]$$

existe y es definida positiva. En consecuencia, por continuidad de las formas cuadráticas,

$$\lim_{\hat{n} \rightarrow \infty} \frac{1}{\hat{n}} \text{Var} (S_n) = \mathbf{b}_j \cdot \mathbf{C} \mathbf{b}_j^T.$$

Como $\mathbf{b}_j$ es una fila de $\mathbf{B}^{-1}$ podemos asegurar que $\mathbf{b}_j \neq 0$, luego tenemos que $\mathbf{b}_j^T \mathbf{C} \mathbf{b}_j > 0$ lo que implica que la varianza de S_n es estrictamente positiva para n suficientemente grande. Con esto, queda demostrado que se cumplen todas las hipótesis del Lema.

Además, usando (2.3.56) tenemos que

$$\begin{aligned} s_n^2 &\doteq \mathbb{E}[S_{\mathbf{n}}^2] = \text{Var} \left(\sum_{i=1}^{\hat{n}} \mathbf{b}_{j \cdot} \boldsymbol{\eta}_i \varepsilon_i \right) \\ &= \hat{n} \mathbf{b}_{j \cdot} \mathbf{C} \mathbf{b}_{j \cdot}^T + o(\hat{n}) \\ &= \hat{n} \mathbf{b}_{j \cdot} \mathbf{C} \mathbf{b}_{j \cdot}^T (1 + o(1)) \\ &= \hat{n} a_{jj} (1 + o(1)), \end{aligned}$$

y el Teorema A.0.1 nos asegura que

$$\limsup_{\hat{n} \rightarrow \infty} \left(\frac{1}{2\hat{n} \log \log \hat{n}} \right)^{1/2} \left| \sum_{i=1}^{\hat{n}} \mathbf{b}_{j \cdot} \boldsymbol{\eta}_i \varepsilon_i \right| = (a_{jj})^{1/2} \quad \text{c.s.} \quad (2.3.57)$$

Luego, (2.3.55) y (2.3.57) prueban la segunda parte del Teorema. $\square$

2.3.4. Resultados asociados a la componente no paramétrica

Teorema 2.3.9. *Bajo las hipótesis H1 - H7 y si, además, cuando $\hat{n} \rightarrow \infty$*

- $\hat{n} h_{2\mathbf{n}}^4 \rightarrow 0$,
- $\frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \rightarrow 0$,

tenemos que

$$\sup_{\mathbf{t} \in \mathcal{C}} |\hat{r}(\mathbf{t}) - r(\mathbf{t})| = O(h_{2\mathbf{n}}) + O \left(\sqrt{\frac{\log \hat{n}}{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}} \right) \quad \text{c.s.}$$

Demostración. A partir de la definición del estimador de la regresión dado en (2.2.10) y la del modelo dado en (2.1.1), podemos escribir

$$\begin{aligned} \hat{r}(\mathbf{t}) &= \sum_{i=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) (Y_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{\mathbf{h}}) \\ &= \sum_{i=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) (\mathbf{X}_i^T \boldsymbol{\beta} + r(\mathbf{T}_i) + \varepsilon_i - \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_{\mathbf{h}}) \\ &= \sum_{i=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) (r(\mathbf{T}_i) + \varepsilon_i) - \sum_{i=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) \mathbf{X}_i^T (\hat{\boldsymbol{\beta}}_{\mathbf{h}} - \boldsymbol{\beta}), \end{aligned}$$

donde, el primer término corresponde a un estimador no paramétrico puro. Definimos $\hat{r}_{\mathbf{h}}^*(\mathbf{t}) = \sum_{i=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) (r(\mathbf{T}_i) + \varepsilon_i)$ y usando, bajo las hipótesis H1–H6, y, por la Observación 2.3.1, la observación 4 del Teorema A.0.5 de Dabo-Niang et al. (2016) tenemos que

$$\sup_{\mathbf{t} \in \mathcal{C}} |\hat{r}_{\mathbf{h}}^*(\mathbf{t}) - r(\mathbf{t})| = O(h_{2\mathbf{n}}) + O \left(\sqrt{\frac{\log \hat{n}}{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}} \right) \quad \text{c.s.} \quad (2.3.58)$$

Por otro lado, a partir del resultado obtenido en (2.3.2) sabemos que para cada $j = 1, \dots, p$,

$$\left| \hat{\beta}_{\mathbf{h}j} - \beta_j \right| = O\left(\sqrt{\frac{\log \log \hat{n}}{\hat{n}}}\right) \quad \text{c.s.} \quad (2.3.59)$$

Como los pesos son no negativos (H2) y satisfacen $\sum_{i=1}^n \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) = 1$, tenemos por desigualdad triangular

$$\left\| \sum_{i=1}^n \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) \mathbf{X}_i \right\| \leq \sum_{i=1}^n \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) \|\mathbf{X}_i\|.$$

Por la definición de los pesos en (2.3.11) tenemos que

$$\Delta_i(\mathbf{t}) = \frac{1}{a_{\mathbf{s}, \mathbf{n}} h_{2\mathbf{n}}^k} K_{1, h_{1\mathbf{n}}} (d(\mathbf{s}, \mathbf{s}_i)) K_2 \left(\frac{d_m(\mathbf{t}, \mathbf{T}_i)}{h_{2\mathbf{n}}} \right)$$

Dado que los núcleos tienen soporte compacto (H2), sólo contribuyen los índices tales que

$$d(\mathbf{s}, \mathbf{s}_i) \leq C n h_{1\mathbf{n}}, \quad d_m(\mathbf{t}, \mathbf{T}_i) \leq C h_{2\mathbf{n}}.$$

Sea $\mathcal{I}_{\mathbf{n}}(\mathbf{t})$ dicho conjunto de índices activos, tenemos que su cardinal es del orden

$$|\mathcal{I}_{\mathbf{n}}(\mathbf{t})| = O\left(n h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k\right) \quad \text{c.s.}$$

Además del Lema 2.3.7, tenemos que

$$\omega_{\mathbf{n}}(\mathbf{t}, \mathbf{T}_i) = O\left(\frac{1}{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}\right) \quad \text{c.s.}$$

uniformemente en $\mathbf{t}$ e $\mathbf{s}_i$. En consecuencia,

$$\sum_{i=1}^{\hat{n}} \omega_{\mathbf{n}}(\mathbf{t}, \mathbf{T}_i) \|\mathbf{X}_i\| = O\left(\frac{1}{\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}\right) \sum_{i \in \mathcal{I}_{\mathbf{n}}(\mathbf{t})} \|\mathbf{X}_i\|.$$

Como $E\|\mathbf{X}\| < \infty$ y la sucesión es α -mixing, la ley fuerte de los grandes números (Teorema A.0.3, aplicable bajo los mismos argumentos que en (2.3.21)) implica

$$\frac{1}{|\mathcal{I}_{\mathbf{n}}(\mathbf{t})|} \sum_{\mathbf{s}_i \in \mathcal{I}_{\mathbf{n}}(\mathbf{t})} \|\mathbf{X}_i\| \rightarrow E\|\mathbf{X}\| \quad \text{c.s.}$$

uniformemente en $\mathbf{t}$. Dado que $\hat{n} h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \rightarrow \infty$, ya que por hipótesis $\frac{\log \hat{n}}{h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k \hat{n}^{\frac{1}{4} - \frac{1}{r}}} \rightarrow 0$, concluimos que

$$\sup_{\mathbf{t} \in \mathcal{C}} \sum_{i=1}^{\hat{n}} \omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) \|\mathbf{X}_i\| = O(1) \quad \text{c.s.}$$

Este resultado, combinado con la tasa de convergencia de $\hat{\beta}_{\mathbf{h}}$ en (2.3.59), y lo obtenido en (2.3.58) nos permite obtener la siguiente cota

$$\sup_{\mathbf{t} \in \mathcal{C}} |\hat{r}_{\mathbf{h}}(\mathbf{t}) - r(\mathbf{t})| \leq \sup_{\mathbf{t} \in \mathcal{C}} |\hat{r}_{\mathbf{h}}^*(\mathbf{t}) - r(\mathbf{t})| + \sup_{\mathbf{t} \in \mathcal{C}} \sum_{i=1}^{\hat{n}} \|\omega_{\mathbf{h}}(\mathbf{t}, \mathbf{T}_i) \mathbf{X}_i\| \|\hat{\beta}_{\mathbf{h}} - \beta\|$$

$$\begin{aligned}
&= O(h_{2\mathbf{n}}) + O\left(\sqrt{\frac{\log \hat{n}}{\hat{n}h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}}\right) + O\left(\sqrt{\frac{\log \log \hat{n}}{\hat{n}}}\right) \quad \text{c.s.}, \\
&= O(h_{2\mathbf{n}}) + O\left(\sqrt{\frac{\log \hat{n}}{\hat{n}h_{1\mathbf{n}}^d h_{2\mathbf{n}}^k}}\right).
\end{aligned}$$

donde $\|\cdot\|$ denota la norma euclidiana. Este último resultado concluye esta prueba.

□

Capítulo 3

Estudios de simulación

Con el objetivo de evaluar el desempeño de la metodología propuesta bajo condiciones controladas, en este capítulo se presentan distintos estudios de simulación orientados a analizar el comportamiento de los predictores considerados en diversos escenarios. Estos estudios se basan en la generación de datos sintéticos y ponen especial énfasis en el efecto que ejercen tanto la estructura de correlación espacial como la configuración de la grilla de sitios sobre la calidad de la predicción. El propósito principal es cuantificar la exactitud en la estimación de los parámetros del término paramétrico y analizar la distribución de los errores de predicción que se obtienen. Los resultados de estas simulaciones posibilitarán obtener conclusiones fundamentadas acerca de las limitaciones, fortalezas y condiciones de aplicabilidad de los procedimientos desarrollados.

El propósito principal es cuantificar la exactitud en la estimación de los parámetros del modelo y analizar la distribución de los errores de predicción resultantes. Los resultados obtenidos permiten extraer conclusiones fundamentadas acerca de las fortalezas, limitaciones y condiciones de aplicabilidad de los procedimientos desarrollados.

3.1. Método para simular las covariables

Para simular los datos, partimos de una estructura de covarianza espacial (1.2.1) y una estructura para su covarianza intrínseca (1.2.2) y simulamos realizaciones de las covariables.

Recordemos que dada una muestra $\{\mathbf{X}_i\}_{\mathbf{s}_i \in \mathcal{O}_n}$, tenemos la matriz de covariables dada por $\mathbb{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$, donde cada columna $\mathbb{X}_{\cdot u} \in \mathbb{R}^{n \times 1}$, con $u = 1, \dots, p$, es una de las covariables. Para simular las realizaciones de las mismas, utilizamos la descomposición de Cholesky de la matriz de covarianza (Cressie, 1993). A partir de las n localizaciones definimos el vector de esperanzas de cada covariable como

$$\mathbb{E}[\mathbb{X}_{\cdot u}] = (\mu_u(\mathbf{s}_1), \dots, \mu_u(\mathbf{s}_n))^T \doteq \boldsymbol{\mu}_u,$$

y la matriz de covarianzas $\Sigma_{uu} \in \mathbb{R}^{n \times n}$, simétrica y definida positiva, tal que su ij -ésimo elemento esté dado por

$$C_{uu}(\mathbf{s}_i, \mathbf{s}_j) = \text{Cov}(X_u(\mathbf{s}_i), X_u(\mathbf{s}_j)) \doteq (\Sigma_{uu})_{ij}.$$

Utilizando la descomposición de Cholesky tenemos que

$$\Sigma_{uu} = \mathbf{L}_u \mathbf{L}_u^T,$$

donde $\mathbf{L}_u$ es una matriz de $n \times n$, triangular inferior. A partir de dicha matriz $\mathbf{L}_u$ modelamos la u -ésima covariable como

$$\mathbb{X}_{\cdot u} = \boldsymbol{\mu}_u + \mathbf{L}_u \boldsymbol{\epsilon}_u, \quad (3.1.1)$$

donde $\boldsymbol{\epsilon}_u = (\epsilon(\mathbf{s}_1), \dots, \epsilon(\mathbf{s}_n))^T$ es un vector aleatorio de variables no correlacionadas e idénticamente distribuidas con esperanza $\mathbb{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ y $\text{Var}(\boldsymbol{\epsilon}) = \mathbf{I}_n$. Esto nos asegura que $\mathbb{E}[\mathbb{X}_{\cdot u}] = \boldsymbol{\mu}_u$ y

$$\text{Var}(\mathbb{X}_{\cdot u}) = \text{Var}(\mathbf{L}_u \boldsymbol{\epsilon}_u) = \mathbf{L}_u \text{Var}(\boldsymbol{\epsilon}_u) \mathbf{L}_u^T = \Sigma_{uu}.$$

Con el objetivo de controlar además la correlación intrínseca entre las covariables, seguimos las ideas de Oliver (2003) donde, además de la matriz de autocovarianzas Σ_{uu} , utilizan la matriz de autocovarianzas Σ_{vv} , con $u, v = 1, \dots, p$, definidas en (1.2.1).

En consecuencia, los datos se construyen de manera tal que, para todo $u, v = 1, \dots, p$, se verifique que

- Si u es impar, hallamos la covariable $\mathbb{X}_{\cdot u}$ siguiendo (3.1.1),

$$\mathbb{X}_{\cdot u} = \boldsymbol{\mu}_u + \mathbf{L}_u \boldsymbol{\epsilon}_u. \quad (3.1.2)$$

- Si $v = u + 1$ obtenemos la covariable de la siguiente forma

$$\mathbb{X}_{\cdot v} = \boldsymbol{\mu}_v + \mathbf{L}_v \left(\rho_{uv} \boldsymbol{\epsilon}_u + \sqrt{1 - \rho_{uv}^2} \boldsymbol{\epsilon}_v \right), \quad (3.1.3)$$

donde $\boldsymbol{\epsilon}_v$ y $\boldsymbol{\epsilon}_u$ son variables aleatorias independientes con distribución normal, media cero y matriz de varianza-covarianza identidad. Por otro lado, la correlación entre X_{iu} y X_{jv} está dada por $\rho_{uv} = \frac{C_{uv}(\mathbf{s}_i, \mathbf{s}_j)}{\sigma_u(\mathbf{s}_i)\sigma_v(\mathbf{s}_j)}$, la cual es constante con respecto a $\mathbf{s}_i$ y $\mathbf{s}_j$. Además, esta forma de definir $\mathbb{X}_{\cdot v}$ explicita el mecanismo mediante el cual se introduce dependencia lineal entre ambas a través del parámetro ρ_{uv} . En particular, $\mathbb{X}_{\cdot v}$ se construye como una combinación de dos componentes: una parte proporcional a $\boldsymbol{\epsilon}_u$, que induce correlación con $\mathbb{X}_{\cdot u}$, y otra parte independiente de $\boldsymbol{\epsilon}_u$, asociada a $\boldsymbol{\epsilon}_v$. El coeficiente ρ_{uv} controla la intensidad de esta dependencia, de modo que cuando $\rho_{uv} = 0$ ambos vectores resultan no correlacionados, y cuando $|\rho_{uv}| = 1$ la variabilidad de $\mathbb{X}_{\cdot v}$ queda completamente explicada por $\mathbb{X}_{\cdot u}$. El factor $\sqrt{1 - \rho_{uv}^2}$ asegura que la

varianza total de $\mathbb{X}_v$ se mantenga constante al variar el grado de correlación. Esto último quiere decir que si hallamos la varianza de $\mathbb{X}_v$ obtenemos que

$$\begin{aligned}\text{Var}(\mathbb{X}_v) &= \text{Var}\left(\mathbf{L}_v \left(\rho_{uv}\boldsymbol{\epsilon}_u + \sqrt{1 - \rho_{uv}^2}\boldsymbol{\epsilon}_v\right)\right) \\ &= \mathbf{L}_v \text{Var}\left(\rho_{uv}\boldsymbol{\epsilon}_u + \sqrt{1 - \rho_{uv}^2}\boldsymbol{\epsilon}_v\right) \mathbf{L}_v^T \\ &= \mathbf{L}_v \rho_{uv}^2 \text{Var}(\boldsymbol{\epsilon}_u) + (1 - \rho_{uv}^2) \text{Var}(\boldsymbol{\epsilon}_v) \mathbf{L}_v^T \\ &= \mathbf{L}_v \mathbf{L}_v^T.\end{aligned}$$

De esta forma, de (3.1.2) y (3.1.3), y utilizando que los errores son no correlacionados obtenemos que la covarianza cruzada de $\mathbb{X}_u$ y $\mathbb{X}_v$ es

$$\begin{aligned}\text{Cov}(\mathbb{X}_u, \mathbb{X}_v) &= \mathbb{E}[(\mathbb{X}_u - \boldsymbol{\mu}_u)(\mathbb{X}_v - \boldsymbol{\mu}_v)^T] \\ &= \mathbb{E}\left[\mathbf{L}_u \boldsymbol{\epsilon}_u \left(\mathbf{L}_v \left(\rho_{uv}\boldsymbol{\epsilon}_u + \sqrt{1 - \rho_{uv}^2}\boldsymbol{\epsilon}_v\right)\right)^T\right] \\ &= \mathbb{E}\left[\mathbf{L}_u \boldsymbol{\epsilon}_u \left(\rho_{uv}\boldsymbol{\epsilon}_u^T \mathbf{L}_v^T + \sqrt{1 - \rho_{uv}^2}\boldsymbol{\epsilon}_v^T \mathbf{L}_v^T\right)\right] \\ &= \rho_{uv} \mathbf{L}_u \mathbb{E}[\boldsymbol{\epsilon}_u \boldsymbol{\epsilon}_u^T] \mathbf{L}_v^T + \mathbf{L}_u \sqrt{1 - \rho_{uv}^2} \mathbb{E}[\boldsymbol{\epsilon}_u \boldsymbol{\epsilon}_v^T] \mathbf{L}_v^T \\ &= \rho_{uv} \mathbf{L}_u \mathbf{L}_v^T.\end{aligned}$$

- En el caso que u y v no cumplan con las condiciones explicitadas, esto es, que v sea par y u impar o $v - u > 1$, la covarianza $\text{Cov}(\mathbb{X}_u, \mathbb{X}_v)$ es la matriz de ceros.

3.2. Escenarios

Dado $D = [0, 1]^2$, generamos grillas regulares, irregulares y con focos, que podemos observar en la Figura 3.1 y se describen a continuación.

Grilla regular. Bajo esta estructura, la mayor parte de las localizaciones ($\approx 0.7n$) se disponen siguiendo un arreglo sistemático (una grilla) sobre el dominio D , de manera que las distancias entre puntos vecinos resultan aproximadamente constantes. Dado que un arreglo sistemático no permite ajustar de manera exacta el número de puntos generados a un valor arbitrario n , las localizaciones restantes se completan mediante puntos ubicados de forma completamente aleatoria sobre D (ver estructura irregular), de modo de garantizar exactamente n localizaciones totales conservando el carácter predominantemente regular del patrón.

Grilla irregular. Bajo esta estructura, las n localizaciones $\mathbf{s}_i = (s_{i1}, s_{i2})^T$ se distribuyen de manera independiente y uniforme sobre el dominio D :

$$s_{i1} \sim U[0, 1], \quad s_{i2} \sim U[0, 1], \quad i = 1, \dots, n,$$

lo que corresponde a una realización de aleatoriedad espacial completa, es decir, un proceso puntual homogéneo sobre D sin estructura de dependencia espacial.

Grilla con focos. Bajo esta estructura, las localizaciones se distribuyen alrededor de M focos (o centros de agregación) $\mathbf{s}_1^*, \dots, \mathbf{s}_M^* \in D$, cuyas coordenadas se generan de manera independiente y uniforme sobre D . Del total de n localizaciones, una proporción $\pi \in (0, 1]$ se asigna a los focos, en partes iguales:

$$n_f = \left\lfloor \frac{n\pi}{M} \right\rfloor$$

puntos por foco. Para el foco j -ésimo, cada uno de sus n_f puntos $\mathbf{s}_i = (s_{i1}, s_{i2})^T$ se genera, con $\mathbf{s}_j^* = (s_{j1}^*, s_{j2}^*)^T$, como

$$s_{i1} \sim N(s_{j1}^*, \sigma^2), \quad s_{i2} \sim N(s_{j2}^*, \sigma^2), \quad i = 1, \dots, n_f,$$

de manera independiente entre sí, con $\sigma = 0.035$. Las $n_b = n - M n_f$ localizaciones restantes se distribuyen de forma dispersa sobre todo el dominio, con coordenadas

$$s_{i1} \sim U[0, 1], \quad s_{i2} \sim U[0, 1],$$

representando así una componente de “ruido” espacial sin asociación con los focos. En los escenarios simulados se utilizó $M = 4$ focos y $\pi = 0.8$.

Para el tamaño muestral, consideramos $n = 100, 625, 1089$. Para asegurarnos de que obtenemos realizaciones de diferente tamaño pero de las mismas funciones aleatorias, calculamos una realización de tamaño $n = 1936$ y luego seleccionamos submuestras de tamaño 100, 625 y 1089 de dicha realización. Para ello, subdividimos el cuadrado original en cuadrados anidados más pequeños. El cuadrado más pequeño con 100 datos, está completamente incluido en un cuadrado intermedio que contiene 625 datos, y este a su vez, está contenido en el cuadrado de mayor tamaño con 1089 datos.

Para simular las covariables usamos el algoritmo descrito en la sección anterior, consideramos $p = 8$ covariables. Para cada ϵ_u y ϵ_v , consideramos n realizaciones de una variable con distribución normal estándar. Si u es par y $v = u + 1$, tenemos que $\rho_{uv} \sim U[-1, 1]$. En cuanto a las funciones medias, consideramos funciones constantes en $\mathbf{s}_i \in \mathbb{R}^2$ para todas las covariables pero diferentes entre sí. Esto es $\mu_u(\mathbf{s}_i) = c_u$ con $c_u \sim U[0.5, 2]$ para cada $u = 1, \dots, p$. Por último, para las matrices de varianza-covarianza Σ_{uu} utilizamos la definición de matriz de autocorrelación (1.2.3). Recordemos que el elemento uv de dicha matriz, evaluada en $(\mathbf{s}_i, \mathbf{s}_j)$, está dada por

$$\mathbf{R}_{uv}(\mathbf{s}_i, \mathbf{s}_j) = \frac{C_{uv}(\mathbf{s}_i, \mathbf{s}_j)}{\sigma_u(\mathbf{s}_i)\sigma_v(\mathbf{s}_j)}.$$

Además, proponemos $\sigma_u = \sqrt{\text{Var}(X_u(\mathbf{s}))}$ constante en $\mathbf{s}$, para cada $u = 1, \dots, p$. Luego como $C_{uv}(\mathbf{s}_i, \mathbf{s}_j) = (\Sigma_{uv})_{ij}$ tenemos que

$$(\Sigma_{uu})_{ij} = \sigma_u^2 \mathbf{R}_{uu}(\mathbf{s}_i, \mathbf{s}_j).$$

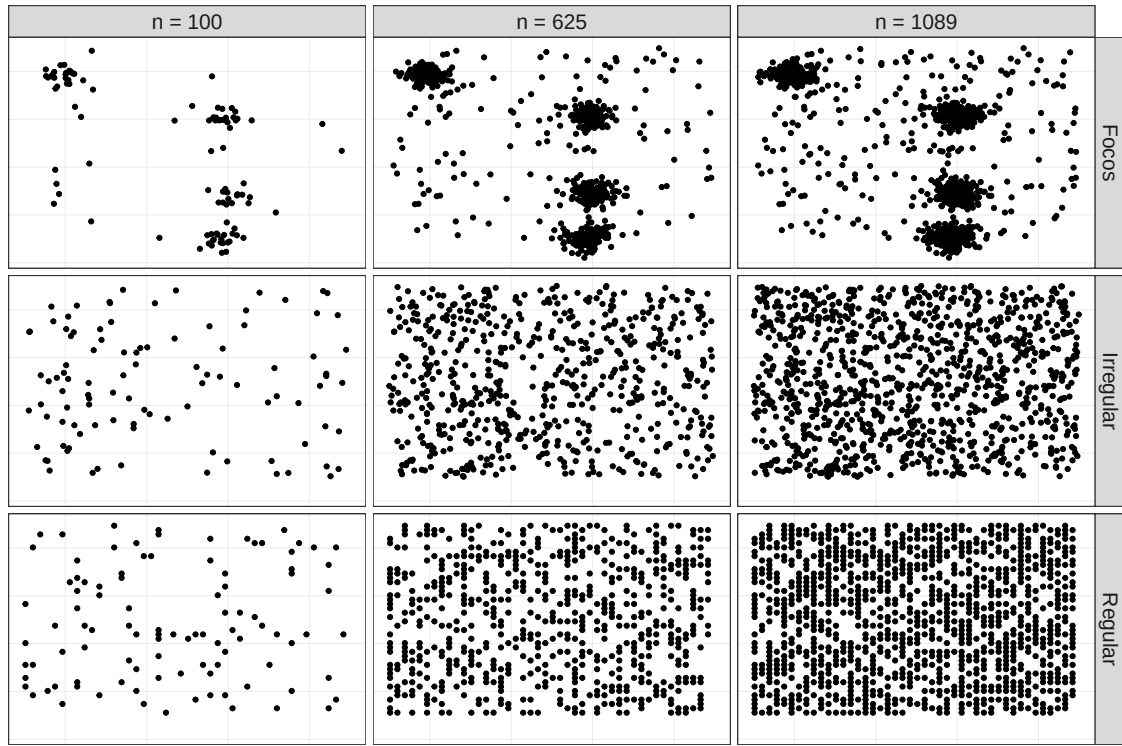


Figura 3.1: Variación de la grilla en función del tamaño y la estructura.

donde consideramos $\sigma_u^2 = 1$ para cada $u = 1, \dots, p$. Por último, describimos las diferentes propuestas para las matrices de autocorrelación $\mathbf{R}_{uu}(\mathbf{s}_i, \mathbf{s}_j)$. Para ello consideramos que, para cada $i, j = 1, \dots, n$, el elemento de dicha matriz está dado por

$$(\mathbf{R}_{uv})_{ij} \doteq R(\|\mathbf{s}_i - \mathbf{s}_j\|),$$

donde $R : \mathbb{R}^+ \rightarrow \mathbb{R}$ es una función determinada de autocorrelación. De esta manera y con $h = \|\mathbf{s}_i - \mathbf{s}_j\|$, vamos a considerar cuatro funciones diferentes de autocorrelación R , para así construir cuatro matrices asignadas de forma aleatoria a las p covariables.

Las funciones elegidas fueron, un *modelo esférico con efecto pepita*,

$$R(h) = \begin{cases} 1, & \text{si } h = 0, \\ \frac{m_1}{m_1 + m_2} \left[1 - \left(1.5 \left(\frac{h}{r} \right) + 0.5 \left(\frac{h}{r} \right)^3 \right) \right], & \text{si } 0 < h \leq r, \\ 0, & \text{si } h > r, \end{cases}$$

con la meseta parcial $m_1 = 1$, el efecto pepita $m_2 = 0.1$ y el rango $r = 0.5$. El mismo *modelo esférico pero sin efecto pepita*, esto es con $m_2 = 0$; un *modelo gaussiano con efecto pepita* dado por

$$R(h) = \begin{cases} 1, & \text{si } h = 0, \\ \frac{m_1}{m_1 + m_2} \left[\exp \left(-\frac{h}{r} \right) \right], & \text{si } h > r, \end{cases}$$

donde la meseta $m_1 = 1$, el efecto pepita $m_2 = 0.1$ y el rango $r = 0.5$ y por último un *modelo seno cardinal* (sinc), un caso particular del modelo J-Bessel para el cual $\alpha = \frac{1}{2}$, el

cual está dado por

$$R(h) = \begin{cases} 1, & \text{si } h = 0, \\ \frac{m_1}{m_1+m_2} \left[\left(\frac{2\theta_{21}}{h} \right)^{\frac{1}{2}} \Gamma\left(\frac{3}{2}\right) J_{\frac{1}{2}}\left(\frac{h}{\theta_{21}}\right) \right], & \text{si } h \neq 0, \end{cases}$$

donde la meseta $m_1 = 1$, el efecto pepita $m_2 = 0.1$, el rango $r = 0.05$ y Γ es la función gamma y $J_{1/2}$ es la función J-bessel de primer tipo de orden $1/2$.

Para generar las realizaciones de la variable respuesta $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$, utilizamos las covariables obtenidas usando la información detallada en el párrafo anterior, el algoritmo definido en la Sección 3.1 y el siguiente modelo,

$$\begin{aligned} \mathbf{Y} &= \mathbb{X}\boldsymbol{\beta} + \rho\mathbf{V}\mathbf{Y} + \boldsymbol{\varepsilon} \\ &= (\mathbf{I} - \rho\mathbf{V})^{-1}(\mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}), \end{aligned} \tag{3.2.1}$$

donde $\mathbf{V} \in \mathbb{R}^{n \times n}$ es la matriz de vecindarios generada a partir de los pesos dados en (1.3.14) con $h_1 = 0.5$. Los errores $\boldsymbol{\varepsilon} = (\varepsilon(\mathbf{s}_1), \dots, \varepsilon(\mathbf{s}_n))^T$ son n realizaciones de una normal estándar y el vector $\boldsymbol{\beta} \in \mathbb{R}^p$, son p realizaciones de una variable aleatoria con distribución normal de media 0 y desvío 10.

En este modelo, el parámetro ρ mide la intensidad con la que influye el término autorregresivo en la respuesta. Esto es, ρ es el parámetro de dependencia espacial (o de rezago espacial) asociado a la matriz $\mathbf{V}$. Dicho parámetro mide en qué medida el valor de $\mathbf{Y}$ en cada sitio está explicado por los valores de $\mathbf{Y}$ en los sitios vecinos, donde la noción de vecindad y la intensidad de la interacción están determinadas por $\mathbf{V}$. Cuando $\rho = 0$, el término $\rho\mathbf{V}\mathbf{Y}$ desaparece y el modelo se reduce a una regresión lineal estándar sin interacción espacial. A medida que $|\rho|$ aumenta, el desempeño de la variable de interés en un sitio deja de ser puramente local y pasa a depender cada vez más del entorno espacial. Consideramos tres posibles valores para el parámetro ρ , estos valores son $\rho = 0$, $\rho = 0.6$ y $\rho = 0.9$. El objetivo es estudiar el comportamiento de las predicciones en términos de la influencia de dicho término en la variable respuesta.

Por último, todos los conjuntos de datos fueron previamente estandarizados, es decir, a cada variable se le restó su media y se la dividió por su desvío estándar; en el caso del conjunto de prueba, la estandarización se realizó utilizando exclusivamente los parámetros (media y desvío) hallados a partir del conjunto de entrenamiento.

3.3. Metodología de entrenamiento, validación y evaluación de desempeño

Para medir el desempeño de los métodos proponemos dos caminos. Por un lado medir la capacidad predictiva de los mismos y por otro medir la precisión en la estimación puntual de los parámetros lineales a aquellos que corresponda.

Para estimar la variabilidad de los errores reportados en la Sección 3.4, del conjunto original de n observaciones construimos, aleatoriamente, dos muestras: la muestra de entrenamiento (70 %) y la de prueba (30 %) (Wadoux et al., 2021). 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 la **raíz cuadrada del error cuadrático medio** (RMSE por sus siglas en inglés) dado por

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (Y_j - \hat{Y}_j)^2}{n}}, \quad (3.3.1)$$

donde $\hat{Y}_j$ es el valor de la predicción en el sitio $\mathbf{s}_j$. Por otro lado, para medir el desempeño de los métodos mediante su capacidad para la estimación del parámetro lineal, medimos el error utilizando la distancia de las estimaciones con respecto a los valores reales de los parámetros β (MAE, por su siglas en inglés), esto es,

$$MAE_{\beta} = \frac{1}{p} \sum_{i=1}^p |\beta_i - \hat{\beta}_i|.$$

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-submuestras**, (*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 $RMSE_i$ 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, $RMSE_1, \dots, RMSE_k$. Luego, la estimación del error por validación cruzada está dada por

$$E_{CVK} = \frac{1}{k} \sum_{i=1}^k RMSE_i.$$

Finalmente, elegimos los parámetros óptimos como aquellos que minimizan el E_{CVK} . Para los parámetros de suavizado implementamos una versión personalizada de validación cruzada K-fold, con el objetivo de reducir el tiempo computacional, esto es, refinamos la búsqueda de dichos parámetros ajustando el rango del intervalo alrededor del mejor valor encontrado en cada iteración.

Repetimos este proceso 50 veces para obtener 50 errores de predicción. Todos los análisis, simulaciones y aplicaciones (de este capítulo y del siguiente) presentados fueron realizados utilizando el software R, versión 4.3.3.¹

¹El código con las funciones necesarias para implementar la metodología propuesta se encuentra disponible en el siguiente enlace.

3.4. Resultados

En pos de organizar los resultados obtenidos, dividimos el análisis en dos partes. En primer lugar, comparamos el desempeño general de los métodos con respecto al tamaño de la muestra, a la estructura de la grilla utilizada para generar los datos y al parámetro ρ , definido en (3.2.1), que controla la influencia del término autorregresivo para cada tipo de grilla. En segundo lugar comparamos el desempeño de la estimación del parámetro lineal.

Los métodos se presentan en el siguiente orden; los cuatro primeros son los predictores semiparamétricos propuestos que varían en el término no paramétrico. Recordemos que el KS1 cuenta, en dicho término, solo las distancias geográficas, el K1ME, solo las distancias entre medianas de la respuesta, el K2ME que está formado por dos núcleos, uno para cada distancia y, por último, K1M considera las dos distancias pero un solo núcleo evaluado en el producto de dichas distancias. A ellos los comparamos con el kriging con tendencia externa, KED, el lattice kriging, LK y el modelo econométrico usado para generar los datos, SAR, y por último, la regresión clásica por mínimos cuadrados ordinarios, OLS.

3.4.1. Con respecto al error de predicción

Analizamos, entonces, la distribución de los errores de predicción (RMSE) mediante un diagrama de caja para cada método y cada escenario, además, se incluyen la media y el desvío estándar, calculados sobre las 50 repeticiones. En las siguientes figuras, si bien representan la misma cantidad, para asegurar la visualización, la escala del eje y no es la misma en todos los casos.

3.4.1.1. Grilla regular

En cada panel de la Figura 3.2 observamos, para el caso de la grilla regular, las nueve combinaciones de los tres valores diferentes de ρ , ($\rho = 0, 0.6$ y 0.9) y para cada tamaño muestral propuestos, esto es, $n = 100, 625$ y 1089 .

En general, observamos que a medida que aumenta el tamaño de la grilla, el RMSE disminuye, siendo esto más evidentes para la dispersión que para los valores centrales. Este comportamiento resulta particularmente marcado al pasar de $n = 100$ a tamaños de grilla mayores; en cambio, entre $n = 625$ y $n = 1089$ no se observan diferencias relevantes en los resultados y, en algunos casos, incluso se registra un leve aumento del RMSE. .

La influencia del término autorregresivo (ρ) parece ser más notoria en tamaños muestrales pequeños. En la primera columna ($n = 100$), al aumentar ρ de 0 a 0.6 o de 0 a 0.9, se observa una mejora en el desempeño de casi todo los métodos pero más marcada en los semiparamétricos, K1M y K2ME y el LK. En cambio, para $n = 625$ y 1089 , el efecto

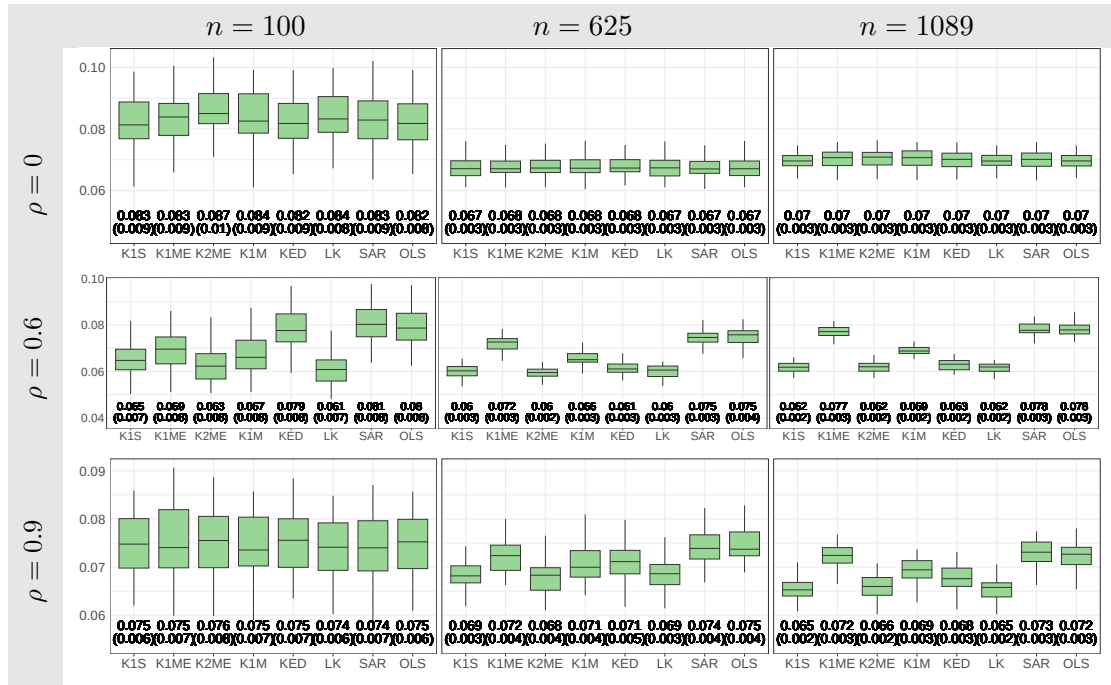


Figura 3.2: Distribución del RMSE para los datos bajo la grilla regular.

de ρ se vuelve casi imperceptible: los métodos tienden a mostrar un RMSE muy similar. En particular, en el caso de $\rho = 0.6$, se observa una mayor variabilidad en el desempeño entre métodos. Esto podría explicarse por el hecho de que, en este escenario, tanto las covariables como la estructura de dependencia espacial aportan información relevante para la predicción. A diferencia de los extremos, donde para $\rho = 0$ la información proviene exclusivamente de las covariables, y para $\rho = 0.9$ domina fuertemente el componente autorregresivo, el caso intermedio representa una situación en la que la contribución relativa de ambas fuentes de información está más equilibrada. Esto puede favorecer o perjudicar a los métodos según su capacidad para integrar eficazmente estos dos aspectos. Ahondaremos en este análisis en siguientes secciones. En términos generales, podemos ver que el desempeño de las metodologías propuestas, exceptuando el caso de inexistencia de término autorregresivo, el K2ME muestra el mejor desempeño predictivo, con resultados muy comparables a LK. También, el K1S muestra resultados competitivos en varios de los escenarios propuestos bajo una grilla regular.

3.4.1.2. Grilla irregular

La Figura 3.3, muestra la distribución de los errores bajo una grilla irregular. En los nueve casos, los métodos semiparamétricos se comportan de forma similar si miramos el desempeño relativo entre ellos. Los mejores resultados lo muestran los que en los pesos cuentan con la distancia geográfica, ya que al contar con más datos tenemos más información cerca y como el núcleo que considera distancias geográficas tiene toda la influencia en

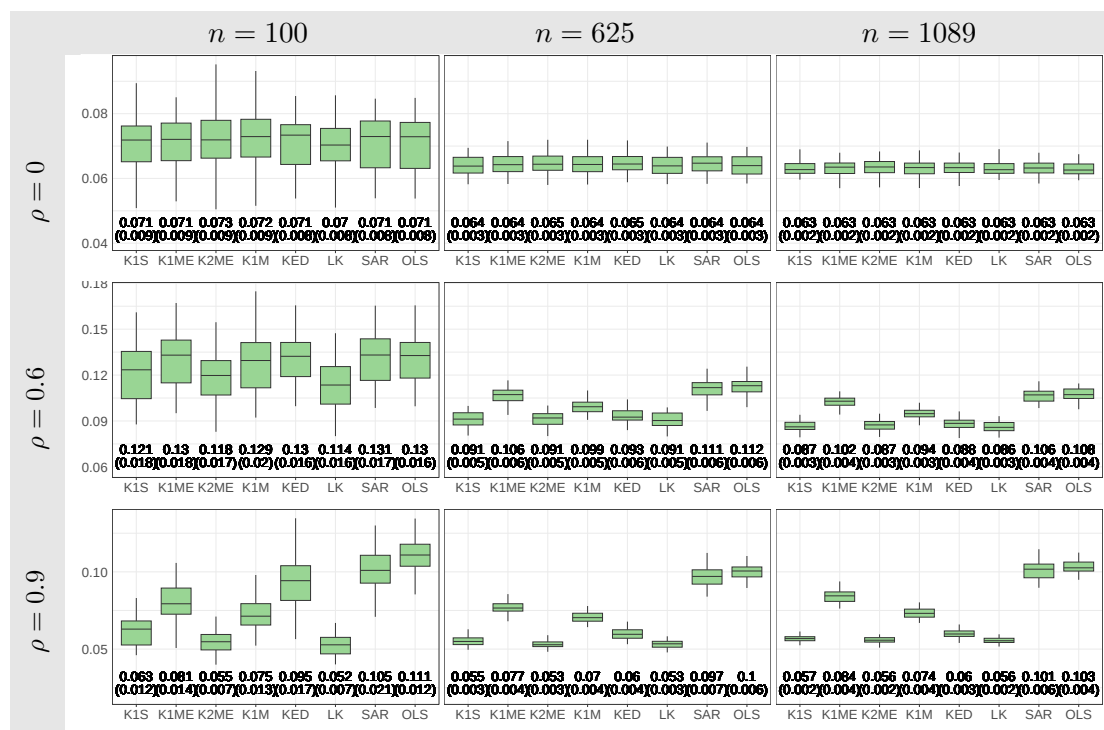


Figura 3.3: Distribución del RMSE para la grilla irregular.

K1S y mucha influencia en el K2ME, mejora el desempeño. En particular, el K2ME y el LK, reportan el mejor desempeño en casi todos los casos en los que existe autocorrelación. Esto quiere decir que tiene un rol más relevante la dependencia espacial, lo cual hace que el predictor estimado por OLS deje de ser suficiente, ya que observamos sus errores más alejados de los semiparamétricos. Lo mismo le ocurre al SAR, el cual tiene un rendimiento pobre en casi todos los escenarios donde $\rho \neq 0$. En comparación con el escenario de grilla regular (Figura 3.2), hay más variabilidad entre métodos para el caso de mayor influencia del término autorregresivo ($\rho = 0.9$). Con respecto a los valores de RMSE observamos valores más elevados, para todos predictores en los escenarios con $\rho = 0.6$. Si $\rho = 0.9$, predictores como el K1ME, el SAR y el OLS también aumentan el error. Esto sugiere que el diseño de grilla irregular introduce mayor complejidad en la estimación, posiblemente por una menor capacidad de capturar eficazmente la estructura de correlación espacial entre puntos cercanos.

Para $\rho = 0$, donde no hay dependencia espacial, las diferencias entre esquemas de las grillas son menos marcadas. En ese caso, los métodos muestran un desempeño relativamente similar bajo ambos diseños, lo cual es esperable, dado que la estructura espacial no contribuye a la predicción y el diseño de la grilla tiene un impacto limitado.

En escenarios con dependencia espacial ($\rho > 0$) y tamaños muestrales altos, los métodos semiparamétricos tienden a mostrar mejor desempeño, con menores valores de RMSE y menor variabilidad. A medida que aumenta ρ , los métodos basados en kriging (paramé-

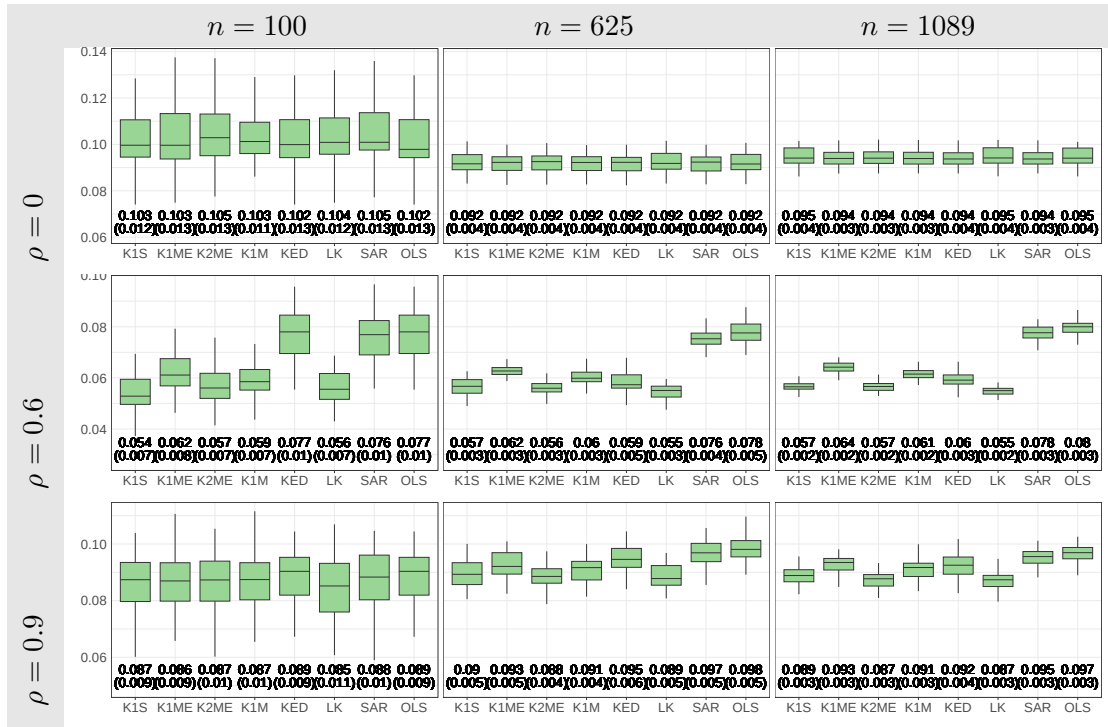


Figura 3.4: Distribución del RMSE para la grilla con focos.

tricos) también mejoran su rendimiento relativo, aunque con mayor sensibilidad al tipo de grilla. El modelo OLS, que no incorpora información espacial, se mantiene sistemáticamente con errores más altos en escenarios con $\rho > 0$, reflejando su incapacidad para captar la dependencia espacial subyacente. El SAR, aún incluyendo un término autorregresivo, su desempeño predictivo resulta muy similar al OLS.

3.4.1.3. Grilla con focos

Por último, en la Figura 3.4, observamos los mismos nueve escenarios para el tipo de grilla con focos. En este caso, al igual que en el escenario de grilla regular, encontramos mayor variabilidad para $\rho = 0.6$. En dicho escenario vemos que los métodos semiparamétricos junto con el LK son los que mejor desempeño presentan, disminuyendo la diferencia con el resto a medida que aumenta el tamaño de la muestra.

La comparación entre los tres esquemas de grilla, regular (Figura 3.2), irregular (Figura 3.3) y con focos (Figura 3.4), nos permite analizar cómo interactúan el diseño de la grilla, la dependencia espacial y método en el desempeño predictivo, medido a través del RMSE. En términos generales, observamos que el error disminuye al aumentar el tamaño muestral (n), independientemente del diseño o del valor de ρ , lo cual es consistente con una mejora en la precisión al disponer de más información.

Sin embargo, el diseño de la grilla impacta en la magnitud del RMSE, especialmente

cuando existe dependencia espacial ($\rho > 0$). La grilla irregular tiende a producir errores más altos y más dispersos, particularmente para tamaños muestrales pequeños ($n = 100$), mientras que la grilla con focos parece facilitar una mejor captación de la estructura espacial, reflejándose en menores valores de error, especialmente en presencia de autocorrelación elevada ($\rho \geq 0.6$). Esto tiene sentido ya que son los escenarios donde contamos con mayor información, esto es, muchas observaciones agrupadas y muy correlacionadas.

En cuanto a los métodos, en los tres diseños de grilla, los métodos semiparamétricos muestran un comportamiento favorable: presentan menores valores del error, menor dispersión, y una mayor estabilidad ante cambios en n y ρ . Esta ventaja es especialmente clara en escenarios con dependencia espacial y mayor tamaño muestral.

Estos resultados sugieren que los métodos semiparamétricos logran integrar de forma robusta la información proveniente de las covariables y la estructura espacial, adaptándose mejor a diferentes condiciones de la grilla. Por tanto, constituyen una alternativa flexible, en especial en situaciones donde la estructura de correlación y la calidad del diseño de la grilla donde se miden los datos no son totalmente controlables.

3.4.2. Con respecto a la estimación del parámetro β

En este apartado analizamos el desempeño de los métodos de predicción estudiados, mediante su capacidad para la estimación del parámetro lineal del modelo propuesto. Solo comparamos los que en su construcción consideran dicha estimación, estos son, los semiparamétricos propuestos en este trabajo, el lattice kriging, el modelo SAR y OLS.

La Figura 3.5 muestra cómo varía el promedio de las diferencias entre los coeficientes estimados y los verdaderos, bajo distintos escenarios a medida que aumenta el tamaño muestral. Se analizaron los tres tipos de grilla (grilla regular, irregular y con focos) combinados con los tres niveles de autocorrelación ($\rho = 0$, $\rho = 0.6$, $\rho = 0.9$), y en dicha figura se compararon solo cinco métodos, K1S, K2ME, LK, OLS y SAR para hacer posible una mejor visualización y, en la Tabla 3.1 incluimos los métodos K1M y K1ME. En dicha tabla reportamos las medias y desvíos estándar promedio del MAE_β para cada escenario.

En la mayoría de los casos, al incrementarse el tamaño muestral, el error tiende a disminuir, lo cual confirma el comportamiento esperable, donde más datos permiten estimaciones más precisas. Este patrón se observa de manera consistente en todos los métodos, aunque con diferencias en la magnitud de la mejora.

Entre los métodos comparados, los semiparamétricos (K1S, K2ME) se destacan por su desempeño estable y poco sensible a los cambios en los escenarios. Su error disminuye sistemáticamente con el tamaño muestral, y en la mayoría de los escenarios se mantienen por debajo del resto. En contraste, el modelo OLS muestra un comportamiento deficiente:

Cuadro 3.1: MAE_β sobre las 50 corridas para cada combinación de ρ , n y tipo de grilla.

ρ	Predictor	Regular			Irregular			Focos		
		$n = 100$	$n = 625$	$n = 1089$	$n = 100$	$n = 625$	$n = 1089$	$n = 100$	$n = 625$	$n = 1089$
0	K1M	0.83 (0.19)	0.16 (0.05)	0.13 (0.04)	1.11 (0.47)	0.25 (0.1)	0.18 (0.06)	0.61 (0.24)	0.23 (0.08)	0.24 (0.08)
	K1ME	0.82 (0.19)	0.17 (0.06)	0.14 (0.05)	1.02 (0.41)	0.27 (0.11)	0.19 (0.07)	0.6 (0.25)	0.22 (0.07)	0.24 (0.08)
	K1S	0.81 (0.18)	0.53 (0.24)	0.34 (0.14)	1.16 (0.52)	0.54 (0.21)	0.37 (0.13)	0.72 (0.26)	0.62 (0.27)	0.47 (0.17)
	K2ME	0.81 (0.2)	0.2 (0.05)	0.14 (0.04)	1.33 (0.48)	0.26 (0.08)	0.2 (0.07)	0.7 (0.24)	0.25 (0.08)	0.24 (0.08)
	LK	0.85 (0.18)	0.54 (0.23)	0.34 (0.14)	1.18 (0.39)	0.54 (0.2)	0.37 (0.13)	0.65 (0.26)	0.63 (0.27)	0.46 (0.17)
	OLS	0.82 (0.18)	0.53 (0.24)	0.34 (0.14)	1.02 (0.41)	0.54 (0.21)	0.36 (0.12)	0.6 (0.24)	0.61 (0.27)	0.47 (0.17)
0.6	SAR	0.82 (0.19)	0.15 (0.03)	0.13 (0.03)	1.03 (0.42)	0.3 (0.1)	0.2 (0.06)	0.62 (0.25)	0.25 (0.07)	0.22 (0.08)
	K1M	0.6 (0.15)	0.41 (0.06)	0.47 (0.04)	0.92 (0.31)	0.52 (0.14)	0.44 (0.09)	0.77 (0.2)	0.55 (0.08)	0.56 (0.07)
	K1ME	0.63 (0.2)	0.64 (0.07)	0.77 (0.05)	0.73 (0.23)	0.64 (0.16)	0.6 (0.11)	0.91 (0.28)	0.73 (0.09)	0.83 (0.08)
	K1S	0.59 (0.17)	0.46 (0.17)	0.38 (0.15)	1.26 (0.37)	0.6 (0.2)	0.47 (0.17)	0.64 (0.18)	0.54 (0.13)	0.4 (0.1)
	K2ME	0.53 (0.13)	0.25 (0.08)	0.18 (0.05)	1 (0.33)	0.32 (0.07)	0.25 (0.06)	0.8 (0.25)	0.32 (0.08)	0.29 (0.06)
	LK	0.58 (0.1)	0.61 (0.24)	0.46 (0.16)	0.98 (0.34)	0.61 (0.17)	0.45 (0.15)	0.91 (0.14)	0.46 (0.14)	0.34 (0.1)
0.9	OLS	1.3 (0.12)	1.13 (0.24)	1.02 (0.19)	1.09 (0.26)	1.03 (0.18)	0.87 (0.15)	2.48 (0.15)	2.46 (0.19)	2.32 (0.12)
	SAR	1.16 (0.23)	0.73 (0.05)	0.79 (0.04)	0.96 (0.28)	0.77 (0.14)	0.71 (0.1)	1.81 (0.29)	1.76 (0.09)	1.63 (0.06)
	K1M	0.87 (0.18)	0.73 (0.08)	0.67 (0.06)	1.21 (0.3)	0.9 (0.13)	0.93 (0.1)	0.68 (0.25)	0.47 (0.05)	0.53 (0.06)
	K1ME	0.93 (0.19)	0.86 (0.08)	0.84 (0.05)	1.41 (0.33)	1.05 (0.12)	1.31 (0.06)	0.72 (0.19)	0.69 (0.09)	0.86 (0.08)
	K1S	0.8 (0.16)	0.43 (0.14)	0.3 (0.09)	0.77 (0.19)	0.68 (0.19)	0.49 (0.17)	0.6 (0.22)	0.45 (0.18)	0.4 (0.13)
	K2ME	0.75 (0.2)	0.29 (0.06)	0.2 (0.05)	0.68 (0.13)	0.35 (0.08)	0.23 (0.06)	0.69 (0.23)	0.37 (0.1)	0.31 (0.07)
0.9	LK	0.7 (0.15)	0.48 (0.12)	0.36 (0.1)	0.72 (0.11)	0.54 (0.19)	0.42 (0.15)	0.85 (0.15)	0.57 (0.2)	0.41 (0.17)
	OLS	0.93 (0.14)	0.9 (0.1)	0.85 (0.07)	3.39 (0.2)	3.18 (0.27)	3.13 (0.23)	1.17 (0.15)	1.03 (0.11)	1.16 (0.1)
	SAR	0.95 (0.18)	0.89 (0.07)	0.87 (0.05)	1.46 (0.27)	1.34 (0.08)	1.4 (0.05)	0.96 (0.17)	0.9 (0.09)	1.08 (0.06)

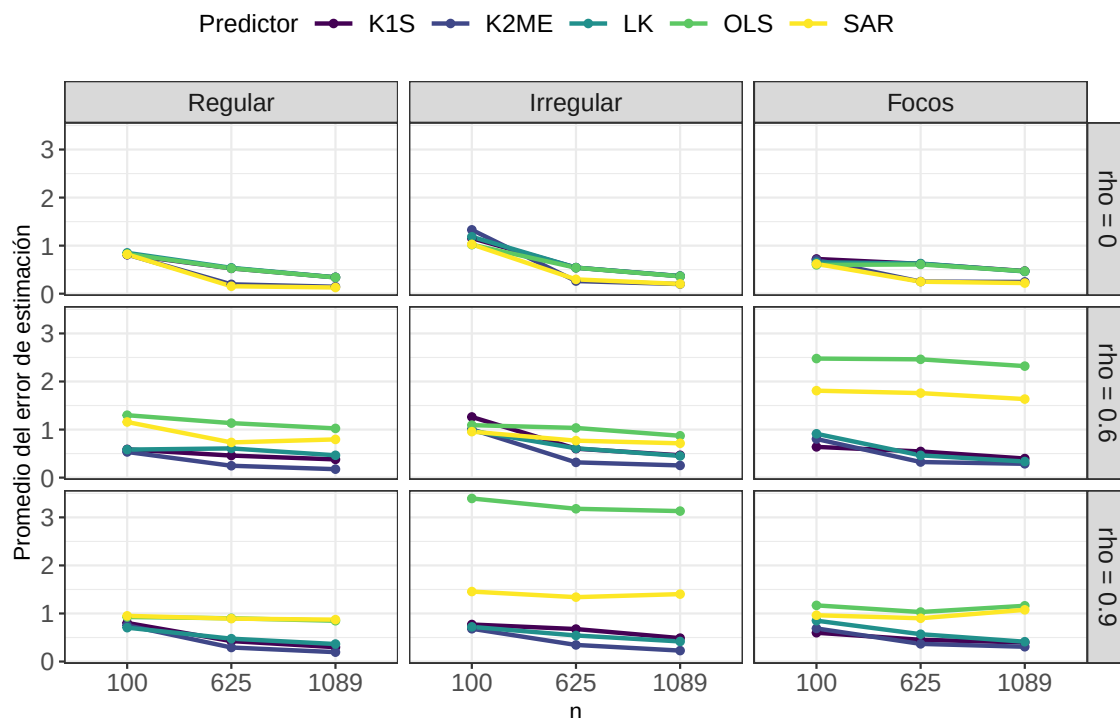


Figura 3.5: Promedio del MAE_{β} para distintos métodos predictivos.

no solo parte de errores altos en muestras pequeñas, sino que incluso con muestras grandes su desempeño sigue siendo pobre, sobre todo en presencia de alta autocorrelación.

Los estimadores semiparamétricos (K1S y K2ME) exhiben, en términos generales, el desempeño más consistente: presentan descensos marcados entre $n = 100$ y $n = 625$ y una estabilización posterior en niveles de error comparativamente bajos. El método LK también mejora con el tamaño muestral, aunque su rendimiento inicial resulta menos favorable, especialmente en escenarios con $\rho = 0$. Tanto OLS como SAR muestran errores notablemente más elevados a lo largo de todas las configuraciones, y si bien ambos métodos se benefician del aumento en n , la magnitud de la mejora es sensiblemente menor, lo que los posiciona como las alternativas poco convenientes en este contexto.

Las diferencias entre tipos de grilla son visibles pero no determinantes: la grilla irregular tiende a generar los niveles de errores más altos, particularmente cuando ρ es intermedio, mientras que en la grilla regular observamos un decrecimiento más suaves y estables.

Cuando $\rho = 0$, todos los métodos presentan errores promedios comparables, porque la estructura espacial no domina el proceso y la ventaja de los modelos flexibles no es tan marcada. A medida que la dependencia crece, los métodos semiparamétricos y, en menor medida, el predictor LK son los que mejor capturan la estructura espacial mejorando así el desempeño de la estimación. Por el contrario, OLS y SAR muestran una disminución del error bastante más lenta y permanecen persistentemente por encima del resto. Esto genera

que la separación entre grupos de métodos se haga más visible conforme aumenta ρ .

A partir de la Tabla 3.1 podemos observar que, para $\rho = 0$, el método K1S presenta errores sistemáticamente más elevados, en particular bajo la grilla regular y la grilla con focos, lo que indica un desempeño menos estable en comparación con los restantes enfoques semiparamétricos. En contraste, K1M, K1ME y K2ME muestran comportamientos similares, con diferencias reducidas una vez alcanzados tamaños muestrales intermedios.

Cuando la dependencia espacial es moderada ($\rho = 0.6$), el comportamiento de los errores se vuelve más heterogéneo. El incremento del tamaño muestral no siempre se traduce en una reducción monótona del error, observándose incluso aumentos locales, especialmente para K1ME bajo grilla regular y con focos. En este contexto, K2ME se destaca por presentar los menores errores promedio y una disminución más clara al aumentar n . Por su parte, K1S continúa mostrando un desempeño relativamente más débil.

En escenarios de alta dependencia espacial ($\rho = 0.9$), las diferencias entre métodos se intensifican. En general, los errores de estimación promedio son más elevados y la reducción asociada al aumento del tamaño muestral resulta más lenta e irregular. En particular, K1ME presenta un comportamiento poco estable bajo la grilla irregular y con focos, con incrementos del error para $n = 1089$, lo que sugiere posibles problemas de variabilidad o de adaptación del método en presencia de fuerte autocorrelación espacial. En contraste, K2ME mantiene una tendencia decreciente más consistente y alcanza los valores más bajos de error en prácticamente todos los esquemas de las grillas, mientras que K1M exhibe un comportamiento intermedio.

En conjunto, los resultados indican que no existe un método claramente dominante en todos los escenarios analizados. Sin embargo, a medida que aumenta la dependencia espacial, las diferencias entre predictores adquieren mayor relevancia que el mero incremento del tamaño muestral. En particular, K2ME exhibe un comportamiento más estable frente a variaciones en ρ y en el tipo de grilla, mientras que K1S y, en ciertos casos, K1ME presentan un desempeño más inestable.

En resumen, el análisis indica que los métodos semiparamétricos logran un equilibrio favorable entre flexibilidad, precisión y capacidad predictiva. Estos métodos aprovechan la información espacial disponible y muestran un desempeño creciente a medida que aumenta el tamaño muestral. En contraste, enfoques como OLS, que ignoran la estructura de dependencia espacial, presentan fallas sistemáticas en escenarios donde dicha estructura es relevante. Asimismo, el diseño de las grillas resulta un factor determinante: configuraciones más homogéneas o centradas conducen a una estimación más precisa de los parámetros del modelo.

Capítulo 4

Aplicación a datos reales

Este capítulo presenta la aplicación de los métodos propuestos a datos reales, contrastando su desempeño tanto entre ellos como frente a predictores paramétricos clásicos y a aquellos basados en kriging. Para la aplicación nos hemos enfocado en el problema de predicción de precios de viviendas, concentración de metales y puntajes de desempeño escolar. Específicamente, trabajamos con 4 grupos de datos y, para cada uno de ellos, presentamos sus características y describimos mediante un análisis exploratorio. Por otro lado, presentamos los resultados de aplicarle los diferentes predictores, ya analizados en el capítulo anterior, y estudiamos el desempeño predictivo de cada uno y entre ellos.

4.1. Metodología de entrenamiento, validación y evaluación de desempeño

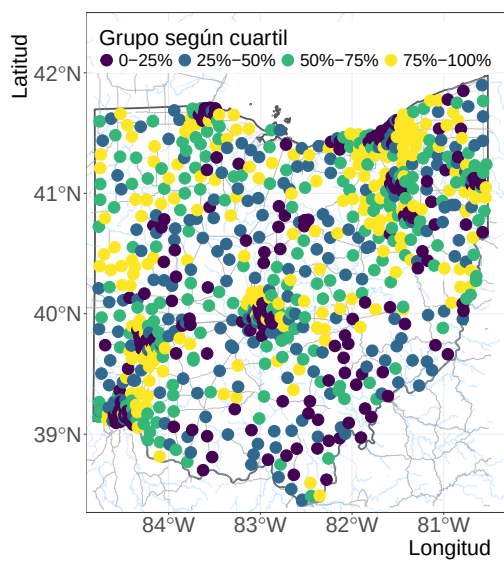
En este capítulo, consideramos el mismo procedimiento que para los datos simulados, esto es, para cada conjunto de datos construimos, aleatoriamente, dos muestras: la muestra de entrenamiento (70 %) y la de prueba (30 %). Con la muestra de entrenamiento estimamos todos los parámetros involucrados en los predictores, tanto paramétricos como no paramétricos. Para éstos últimos, y con la muestra de entrenamiento utilizamos validación cruzada de 5 submuestras, para estimar los parámetros de suavizado. Por último, 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 RMSE definido en (3.3.1). Cabe destacar que, de la misma manera que en las simulaciones, todos los conjuntos de datos fueron previamente estandarizados, es decir, a cada variable se le restó su media y se la dividió por su desvío estándar; en el caso del conjunto de prueba, la estandarización se realizó utilizando exclusivamente los parámetros (media y desvío) hallados a partir del conjunto de entrenamiento.

4.2. Datos Ohio

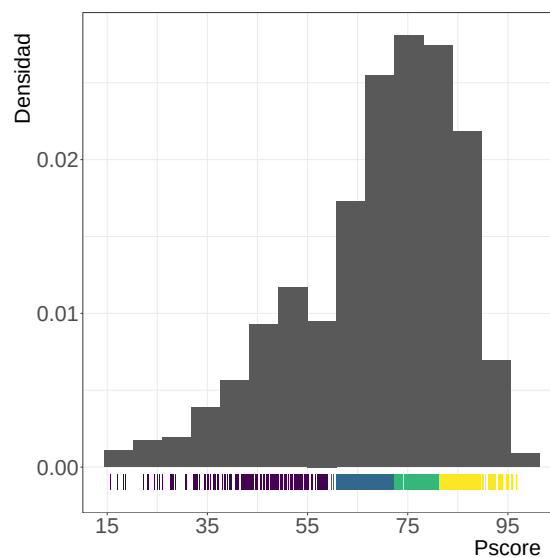
En primer lugar, trabajamos con un conjunto de datos de 1965 escuelas primarias del estado de Ohio correspondientes al año 2001–2002 (LeSage y Pace, 2009). El objetivo es predecir un índice que representa el desempeño promedio estudiantil (**pscore**) en cada distrito. Como variables explicativas se consideran indicadores del sistema educativo y del contexto socioeconómico. En particular, se incluyen la matrícula escolar (**enroll**), el número de docentes (**teachers**), los años promedio de experiencia docente (**experience**), el salario docente promedio (**salary**), la razón alumnos–docente (**ppupil**) y el gasto por alumno (PPS) desagregado en instrucción (**pinstruct**), operaciones edilicias (**pbuilding**), administración (**padminist**), apoyo al alumnado (**ppsupport**) y apoyo al personal (**pssupport**). Asimismo, se incorporan las proporciones del gasto total destinadas a instrucción (**instructp**), operaciones edilicias (**buildingp**), administración (**administp**), apoyo al alumnado (**psupportp**) y apoyo al personal (**ssupportp**). El contexto socioeconómico se caracteriza mediante el ingreso per cápita de la zona (**logpincome**), la proporción de población no blanca (**nonwhite**), el porcentaje de población en situación de pobreza (**poverty**), la proporción de personas que residen en la misma vivienda desde hace al menos cinco años (**samehouse**) y la proporción de alumnos que asisten a escuelas públicas (**public**). Finalmente, se consideran variables asociadas al nivel educativo de la población mayor de 25 años, medidas a través de los porcentajes de graduados de secundaria (**highsh**), formación técnica o terciaria (**assoc**), universitaria (**college**), posgrado (**grad**) y formación profesional (**prof**).

En este conjunto de datos, las coordenadas espaciales están asociadas a códigos postales (ZIP), expresados en latitud y longitud, habiendo en algunos casos más de una escuela en un mismo punto espacial. En tales casos, promediamos las variables para obtener una única observación por código postal de distrito. De esta manera, obtenemos una muestra de 799 observaciones espaciales. En este caso, el mapa en la Figura 4.1 (a) sugiere una distribución espacial ampliamente dispersa, con una marcada superposición de los distintos grupos entre cuartiles a lo largo de todo el dominio. No se observan áreas extensas dominadas de manera consistente por valores altos o bajos del puntaje, sino más bien una alternancia fina entre rangos, lo que indica una organización espacial débil a gran escala. Si bien los valores correspondientes a los grupos de valores altos aparecen repartidos en todo el territorio, sin conformar núcleos claramente delimitados, podemos observar diferentes agrupamientos de puntos sobre la región de interés. Esto lo podemos observar en los paneles (c) y (d) de la Figura 4.1 en los que vemos que en el centro de los agrupamientos encontramos valores bajos del puntaje y a medida que nos alejamos del centro ese valor disminuye. Este ejemplos inspira la grilla con focos como alternativa en el análisis realizado en el apartado de simulaciones.

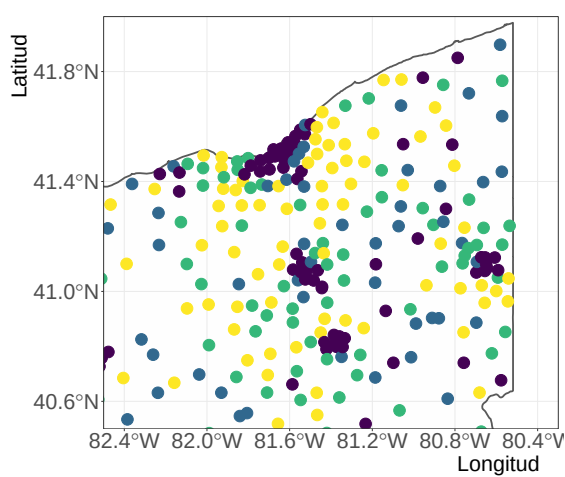
El histograma muestra una distribución unimodal, relativamente concentrada en valores medios–altos del puntaje, con una ligera asimetría hacia la izquierda. Esta forma sugiere



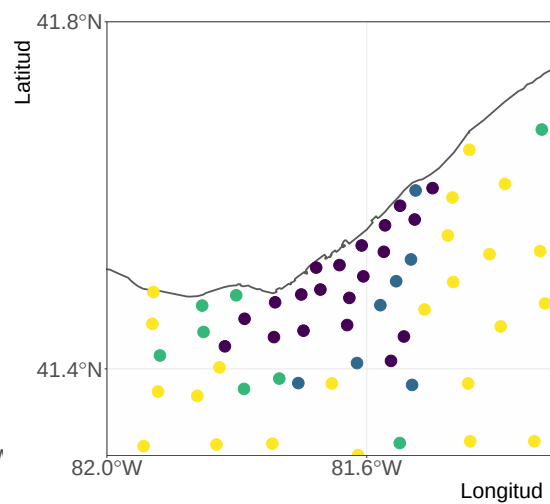
(a) Boxmap.



(b) Histograma.



(c) Zoom del Boxmap.



(d) Zoom del Boxmap.

Figura 4.1: Distribución de la variable p_score de la base Ohio.



Figura 4.2: Correlación intrínseca para los datos Ohio.

una variabilidad moderada, sin presencia marcada de valores extremos. En conjunto, ambos gráficos son coherentes en indicar que, si bien existe dispersión espacial, la variable no presenta contrastes espaciales pronunciados a nivel global pero a nivel local podemos observar cierta tendencia.

En el gráfico de correlaciones, Figura 4.2, solo incorporamos las covariables más correlacionadas con la variable respuesta. En primer lugar, el desempeño estudiantil (pscore) aparece negativamente correlacionado con variables asociadas a características estructurales, en particular con la proporción de población no blanca (nonwhite) y con el nivel de pobreza (poverty). Estas correlaciones son relativamente altas en valor absoluto, lo que sugiere que el rendimiento promedio está fuertemente condicionado por el contexto socio-económico del distrito, más que por decisiones presupuestarias puntuales. En contraste, pscore muestra una correlación positiva con el ingreso per cápita (logpincome) y con el

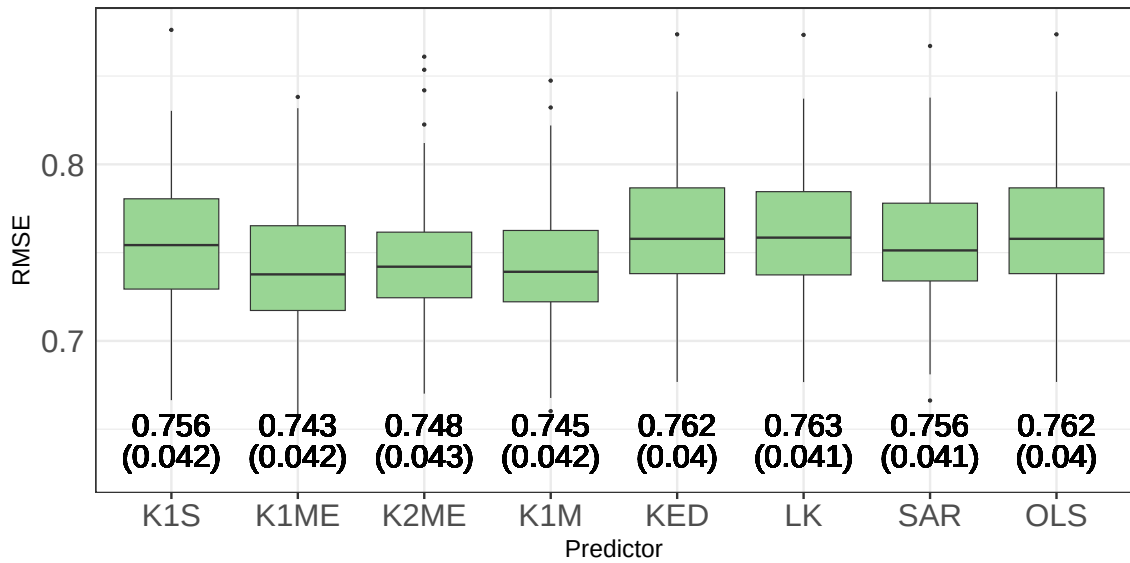


Figura 4.3: Distribución del RMSE para los datos Ohio.

nivel educativo de la población adulta, especialmente con el porcentaje de graduados universitarios (college), lo que refuerza la idea de un gradiente socioeconómico persistente. En cuanto a las variables de gasto, se observa una estructura de correlación bastante marcada entre las distintas partidas presupuestarias. Por ejemplo, los gastos en apoyo al alumnado (ppsupport) y al personal (pssupport), así como sus proporciones sobre el gasto total, están fuertemente correlacionados entre sí, lo cual es esperable dado que responden a decisiones presupuestarias conjuntas y a restricciones contables. Sin embargo, su relación directa con pscore es más débil y, en algunos casos, incluso negativa, lo que sugiere que mayores asignaciones a estos rubros no se traducen necesariamente en mejores resultados promedio, al menos de manera lineal. También es relevante la alta correlación entre pobreza (poverty) y proporción de población no blanca (nonwhite), así como la correlación negativa de ambas con logpincome. Esta colinealidad indica que muchas de las variables explicativas están capturando dimensiones similares del contexto socioeconómico, lo que tiene implicancias metodológicas claras: parte de la relación observada entre pscore y ciertos gastos puede estar mediada, o incluso enmascarada, por estas condiciones estructurales. En conjunto, el gráfico refuerza la idea de que el desempeño estudiantil está más estrechamente vinculado al perfil socioeconómico y educativo del distrito que a la mera magnitud o composición del gasto escolar.

4.2.1. Resultados

En el caso de los datos de Ohio, los valores de RMSE se concentran en torno a 0.74–0.76, si bien los resultados son comparables, es importante destacar que nuestros métodos logran un mejor desempeño predictivo, presentando errores incluso menores que los obtenidos con el LK, método que ya había demostrado un rendimiento altamente competitivo en las simu-

laciones. Esta cercanía sugiere que, para Ohio, la estructura del problema permite que tanto enfoques flexibles como modelos más simples capturen una fracción importante de la variabilidad de la respuesta. Desde el análisis exploratorio, esto es coherente con la presencia de asociaciones claras entre la variable respuesta y las covariables, junto con una organización espacial que no es completamente aleatoria, pero tampoco dominada por un patrón global fuerte. El método KED presenta una mediana algo más elevada y una dispersión ligeramente mayor, lo que puede interpretarse como una mayor sensibilidad a observaciones extremas o a configuraciones particulares del muestreo. No obstante, esta diferencia no es sustancial y no altera el orden general de desempeño. El método LK muestra resultados muy cercanos a los enfoques kernel, reforzando la idea de que la incorporación de suavizado local es suficiente para capturar la variabilidad relevante del proceso. Por su parte, los modelos SAR y OLS no muestran un deterioro marcado del desempeño, a diferencia de lo observado en otros conjuntos de datos. Esto sugiere que, en Ohio, la dependencia espacial global y las relaciones lineales con las covariables ofrecen una aproximación razonable al proceso subyacente. En conjunto, los resultados indican que, para este conjunto de datos, la elección del método tiene un impacto limitado sobre el RMSE, y que las ganancias de complejidad son moderadas, en línea con lo anticipado por el análisis exploratorio.

4.3. Datos Saltillo

En segundo 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. Nuestro objetivo es predecir el precio de los inmuebles a partir de las variables que capturan los atributos físicos del inmueble (modelo hedónico): metros totales, metros cubiertos, antigüedad, cantidad de habitaciones, de garages y de baños. Contamos además con la ubicación del inmueble, dada por su latitud y longitud.

Si bien el conjunto de datos original contiene observaciones de todo el país, decidimos utilizar y analizar los datos de la ciudad de Saltillo. Esta base contiene 967 observaciones. **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).

Los gráficos presentados en la Figura 4.4 permiten caracterizar de manera complementaria la variabilidad espacial y marginal de la variable precio. En el boxmap (Figura (a)) observamos una marcada heterogeneidad en el espacio, sin una distribución estricta por grupo cuartílico aunque sí con ciertos patrones. Los valores correspondientes a los grupos superiores tienden a concentrarse en áreas específicas del territorio, lo que sugiere la presencia de zonas con mayor valorización relativa, mientras que los cuartiles inferiores aparecen

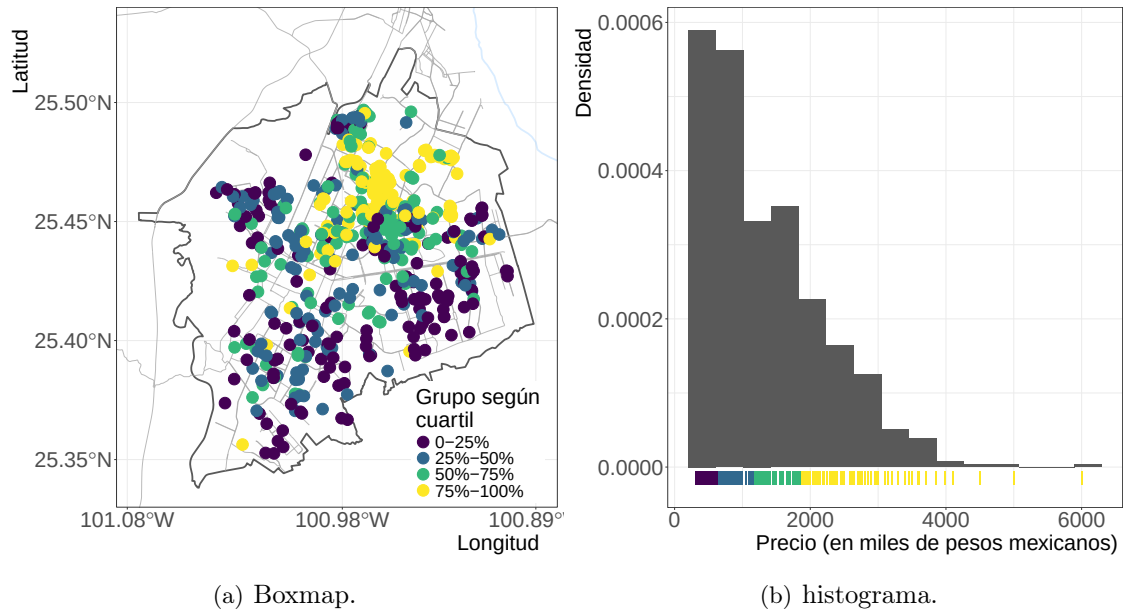


Figura 4.4: Distribución de la variable Precio (en miles de pesos mexicanos) de la base Saltillo.

más dispersos y con mayor presencia en sectores periféricos. Esta superposición parcial entre grupos indica que la estructura espacial no responde a una estructura espacial simple, sino a un entramado más complejo en el que coexisten valores altos y bajos en proximidad geográfica, lo que deja entrever posibles efectos locales o discontinuidades espaciales que no quedarían capturadas por un análisis puramente global.

El histograma muestra una distribución claramente asimétrica a la derecha, con una alta concentración de observaciones en los rangos de precios más bajos y sugiriendo algunos datos extremos. Esto sugiere que la mayoría de las unidades analizadas se sitúan en niveles de precio relativamente moderados, mientras que los precios altos son menos frecuentes pero influyen de manera sustantiva en la forma de la distribución. La ausencia de simetría y la presencia de valores extremos hacen poco plausible el supuesto de normalidad, necesario para algunos métodos (LK, OLS, SAR). En conjunto, ambos gráficos apuntan a una distribución no homogénea, tanto en el espacio como en el rango de los precios, lo que justifica explorar modelos que contemplen explícitamente la variabilidad espacial y la naturaleza fuertemente sesgada de la variable.

En la Figura 4.5 observamos el gráfico de correlaciones el cual muestra que el precio del inmueble está fuertemente asociado con las variables que capturan el tamaño en metros. En particular, se observan correlaciones altas y positivas entre el precio y los metros cubiertos y metros totales, lo que indica que la superficie es uno de los principales determinantes del valor del inmueble. La magnitud de estas asociaciones sugiere que buena parte de la variabilidad del precio está explicada por diferencias en el tamaño de las viviendas.

Asimismo, el precio presenta correlaciones positivas relevantes con el número de baños

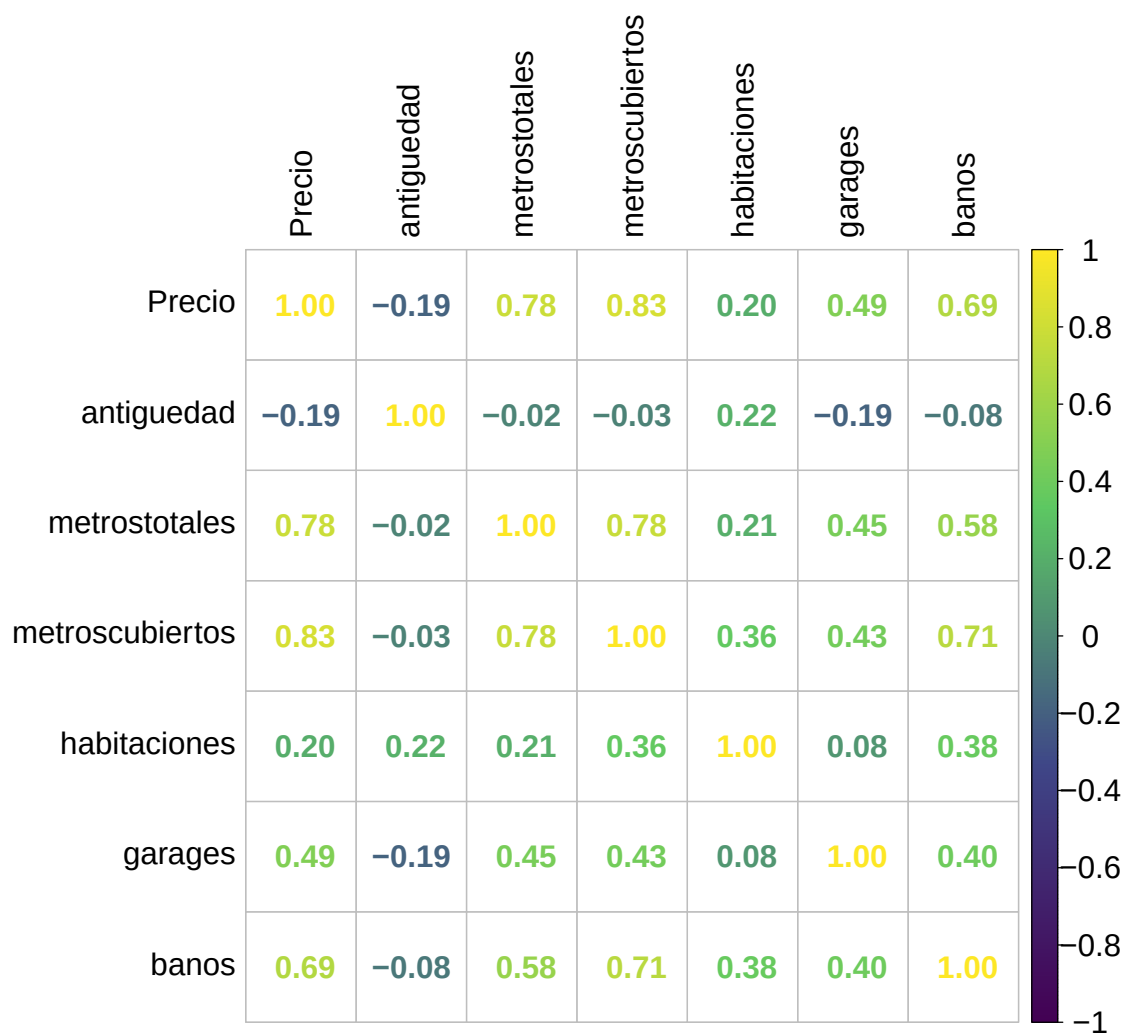


Figura 4.5: Correlación intrínseca para los datos Saltillo.

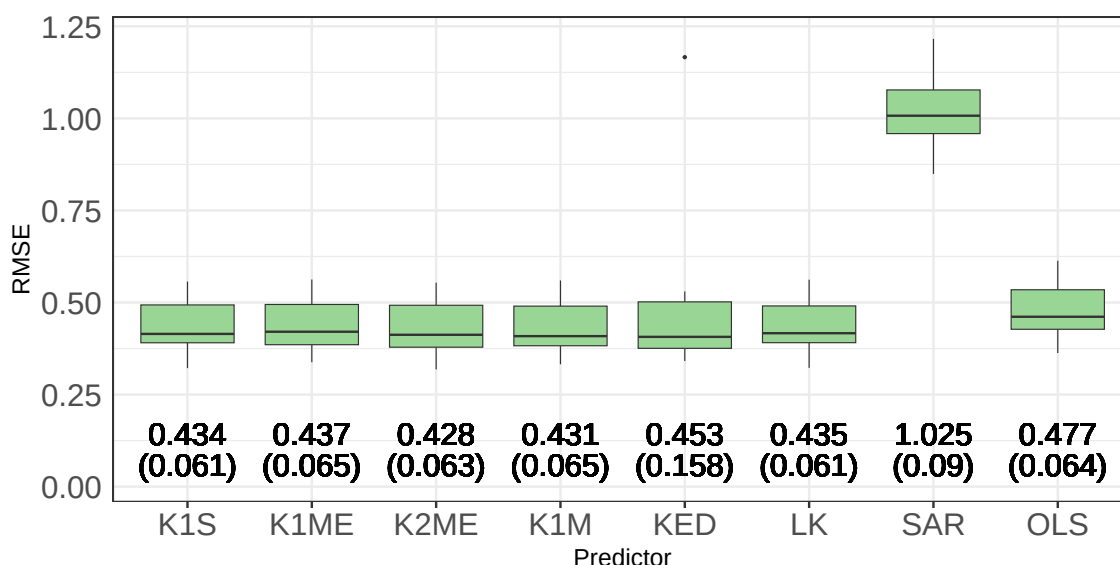


Figura 4.6: Distribución de RMSE para los datos Saltillo.

y de garajes, lo que refuerza la interpretación de estas variables como indicadores del nivel o la calidad del inmueble. No se trata de efectos marginales aislados, sino de atributos que tienden a incrementarse conjuntamente con la superficie y, en consecuencia, con el precio. El número de habitaciones también se asocia positivamente con el precio, aunque con menor intensidad, lo que sugiere que este efecto está en parte explicado por las medidas de superficie.

En contraste, la antigüedad muestra una correlación negativa con el precio, indicando que, en promedio, los inmuebles más antiguos tienden a presentar valores de mercado más bajos. La intensidad de esta correlación es pequeña lo cual sugiere que la antigüedad influye, pero no domina la formación del precio frente a variables como superficie o equipamiento.

4.3.1. Resultados

Recordemos con esta base de datos contamos con 967 observaciones y 6 covariables. En la Figura 4.6 observamos los resultados que muestran algunas diferencias claras y otras de menor magnitud en el desempeño predictivo de los distintos métodos, medido a través del RMSE. En particular, los métodos semiparamétricos presentan valores de RMSE muy similares entre sí y tienden a registrar RMSE ligeramente menores que los enfoques paramétricos. Las medias se ubican en torno a 0.43–0.45, con dispersiones similares, lo que indica un desempeño estable frente a las distintas particiones o repeticiones del experimento. Este comportamiento puede vincularse directamente con la estructura observada en el análisis exploratorio. Por un lado, la distribución del precio mostraba una variabilidad considerable, con asimetría y ausencia de patrones espaciales simples.

Por otra parte, el mapa exhibe una marcada superposición espacial de valores, sin

patrones de variación definidos ni regiones homogéneas de gran extensión. En ese contexto, resulta razonable que métodos flexibles, capaces de capturar relaciones locales y no lineales entre el precio y las covariables, logren un mejor ajuste que modelos que imponen una estructura funcional rígida. El método LK se ubica en un nivel de desempeño comparable al de los enfoques semiparamétricos, lo que sugiere que, aun bajo una formulación más estructurada, la incorporación de suavizado local permite adaptarse adecuadamente a la heterogeneidad presente en los datos.

En contraste, el modelo SAR exhibe un RMSE sustancialmente mayor, lo que indica un ajuste consistentemente pobre. En el caso del método KED se observa una dispersión ligeramente mayor del RMSE en comparación con los restantes enfoques basados en núcleos. Este incremento en el desvío está asociado a la presencia de un valor atípico, que introduce una variabilidad adicional en el error de predicción. Si bien la mediana del RMSE se mantiene en un nivel comparable al de los otros métodos, la sensibilidad del procedimiento a observaciones extremas se refleja en una cola superior más pronunciada de la distribución. Finalmente, el modelo OLS presenta un desempeño intermedio, peor que los métodos flexibles pero claramente superior al SAR. Esto es consistente con el análisis de correlaciones, que mostraba relaciones fuertes entre el precio y variables de tamaño y calidad del inmueble, lo que permite a un modelo lineal explicar una parte relevante de la variabilidad.

4.4. Datos Meuse

En tercer lugar trabajamos con los datos clásicos del río Meuse (Rikken y Van Rijn, 1993). Este conjunto de datos proporciona, entre otras variables, mediciones de diferentes tipos de metales en la capa superior del suelo en una llanura aluvial del río Meuse, cerca del pueblo de Stein (Países Bajos), medida en diferentes localizaciones espaciales, las cuales están identificadas con coordenadas geográficas (latitud y longitud). En general, este conjunto de datos se utiliza para predecir el valor de alguno de esos metales en una localización espacial donde el mismo no fue medido; en este caso, queremos estudiar el comportamiento de la concentración de zinc en función de la concentración de plomo y cadmio, para poder predecirlo en sitios no muestrados. Particularmente, estos datos son utilizados para ejemplificar el método de predicción espacial clásico, *Kriging* (1.3.1). En la Figura 4.7 el Boxmap muestra una organización espacial más estructurada que en el ejemplo anterior. Los grupos cuartílicos de concentración más elevada tienden a alinearse siguiendo el río Meuse, por lo que podemos asociar una clara influencia del espacio en los valores de la variable objetivo. En contraste, los grupos cuartílicos inferiores aparecen con mayor frecuencia alejados del río indicando una diferenciación espacial más marcada entre zonas de baja y alta concentración. Esto implica una tendencia creciente en dirección noroeste. Aún así, persiste cierto solapamiento entre grupos.

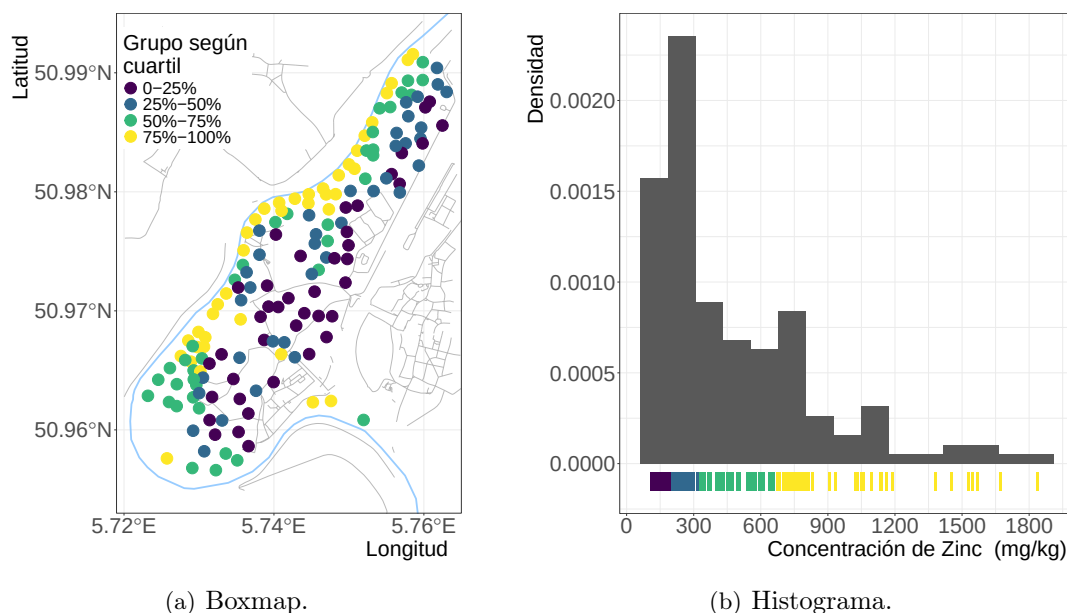


Figura 4.7: Distribución de la variable zinc de la base de datos Meuse.

El histograma muestra nuevamente una distribución asimétrica a la derecha, con una fuerte acumulación de observaciones en concentraciones bajas y una cola que se extiende hacia valores elevados. Sin embargo, a diferencia de una distribución donde casi todas las observaciones se acumulan cerca del mínimo y algunos valores extremos, aquí el histograma muestra un volumen no despreciable de datos en valores intermedios. Esta forma de la distribución es coherente con el patrón espacial observado: la franja de valores altos corresponde a un subconjunto relativamente acotado de observaciones, mientras que la mayor parte del área presenta concentraciones bajas o moderadas. En conjunto, ambos gráficos indican que la variable exhibe una estructura espacial más organizada, superpuesta a una distribución marginal claramente sesgada, lo que refuerza la necesidad de considerar simultáneamente la dependencia espacial y la asimetría de la variable en los análisis posteriores.

La Figura 4.8 presenta la estructura de correlación entre zinc, lead y cadmio en los datos de Meuse, considerando al zinc como variable respuesta. Se observan correlaciones positivas muy elevadas entre el zinc y las dos covariables, en particular con lead y cadmio, lo que indica que las concentraciones de estos metales tienden a variar de manera conjunta en el área de estudio. Esta asociación sugiere la presencia de procesos geoquímicos comunes o de fuentes compartidas de contaminación que afectan simultáneamente a los tres elementos.

La correlación casi perfecta entre zinc y lead, junto con la también elevada asociación con cadmio, pone de manifiesto una fuerte dependencia lineal entre las variables. Desde un punto de vista exploratorio, esto refuerza la idea de que plomo y cadmio contienen información altamente relevante para explicar la variabilidad del zinc. Sin embargo, tam-

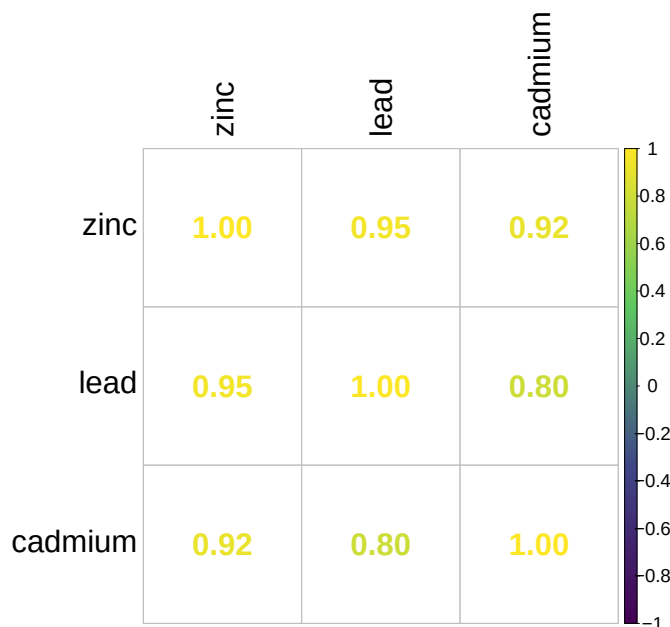


Figura 4.8: Correlación intrínseca para los datos Meuse.

bién advierte que estas covariables no aportan información independiente entre sí, ya que mantienen una correlación interna muy alta.

4.4.1. Resultados

En este caso contamos con 155 observaciones y dos covariables. En la Figura 4.9, los resultados muestran un comportamiento sensiblemente distinto al observado para Saltillo, con desempeños más homogéneos entre los métodos considerados. Los valores de RMSE se concentran en un rango estrecho, con medias muy próximas entre sí, lo que indica que ningún enfoque se impone de manera clara sobre el resto en términos de precisión predictiva. Esta cercanía entre distribuciones sugiere que, para este conjunto de datos, las diferencias entre métodos son mucho menos pronunciadas y no presentan el carácter sistemático observado previamente.

Este resultado es coherente con el análisis exploratorio, donde la variable respuesta presenta una estructura espacial más organizada y una dependencia espacial más marcada, pero, sobre todo, una relación clara con las covariables. En este contexto, tanto los métodos que incorporan explícitamente la dependencia espacial como aquellos más flexibles logran capturar adecuadamente la variabilidad del proceso, lo que se refleja en desempeños similares. A diferencia del caso de Saltillo, donde la heterogeneidad local y la ausencia de un patrón global penalizaban a ciertos modelos, aquí la información espacial resulta menos conflictiva con los supuestos de los distintos enfoques.

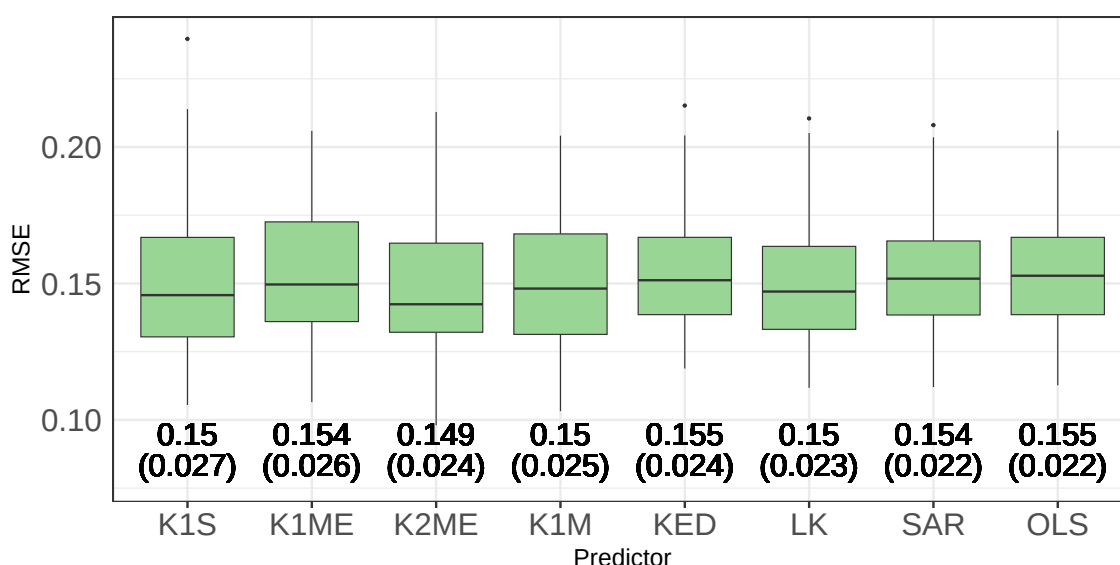


Figura 4.9: Distribución del RMSE para los datos Meuse.

4.5. Datos Jura

En cuarto lugar, trabajamos con un conjunto de datos Jura. Los datos fueron recolectados por el Instituto Federal Suizo de Tecnología en Lausana. Nuestro objetivo es predecir las concentraciones de cadmio, usando las concentraciones de otros metales presentes en la muestras como cobalto (Co), cromo (Cr), cobre (Cu), níquel (Ni), plomo (Pb) y zinc (Zn) en una región en el Jura suizo. Todas las concentraciones de metales fueron medidas en 359 ubicaciones. En este caso, el mapa (Figura 4.10 (a)) muestra una distribución espacial más fragmentada y menos alineada con una estructura dominante clara. Los distintos grupos entre cuartiles aparecen ampliamente intercalados en el territorio, sin que se identifiquen zonas extensas ocupadas de manera predominante por un único rango de concentraciones. Si bien pueden reconocerse agrupamientos locales de valores altos o bajos, estos no se traducen en patrones espaciales continuos ni en una variación ordenada en el espacio, sino que se manifiestan a escalas más pequeñas. Esta distribución sugiere una fuerte variabilidad local y una dependencia espacial débil o, al menos, no fácilmente capturable.

El histograma presenta una distribución unimodal (Figura 4.10) con asimetría a la derecha, concentrando la mayor parte de las observaciones en valores bajos y medios de concentración, y solo unos pocos valores elevados. En conjunto, ambos gráficos indican que la variabilidad de la concentración de cadmio está dominada principalmente por fluctuaciones locales, más que por contrastes espaciales de gran escala, lo que sugiere un escenario en el que los efectos espaciales se manifiestan principalmente a distancias cortas.

La Figura 4.11 resume las correlaciones entre las concentraciones de los distintos metales, considerando a la concentración de cadmio (Cd) como variable respuesta. En primer lugar, se observa que el cadmio presenta correlaciones positivas de magnitud moderada con

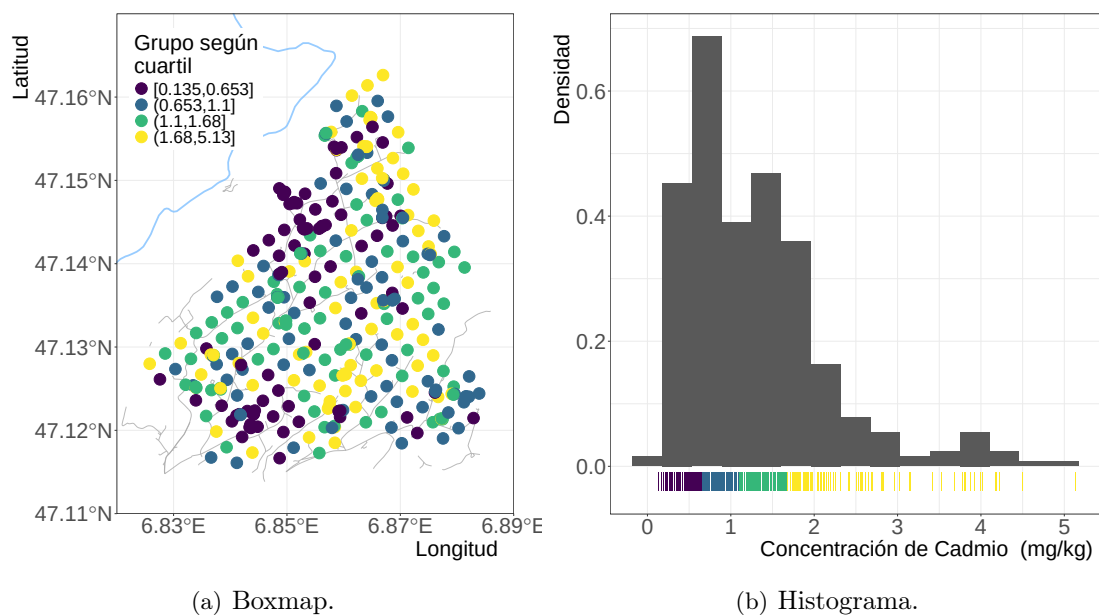


Figura 4.10: Distribución de la variable Concentración de Cadmio (mg/kg) de la base Jura.

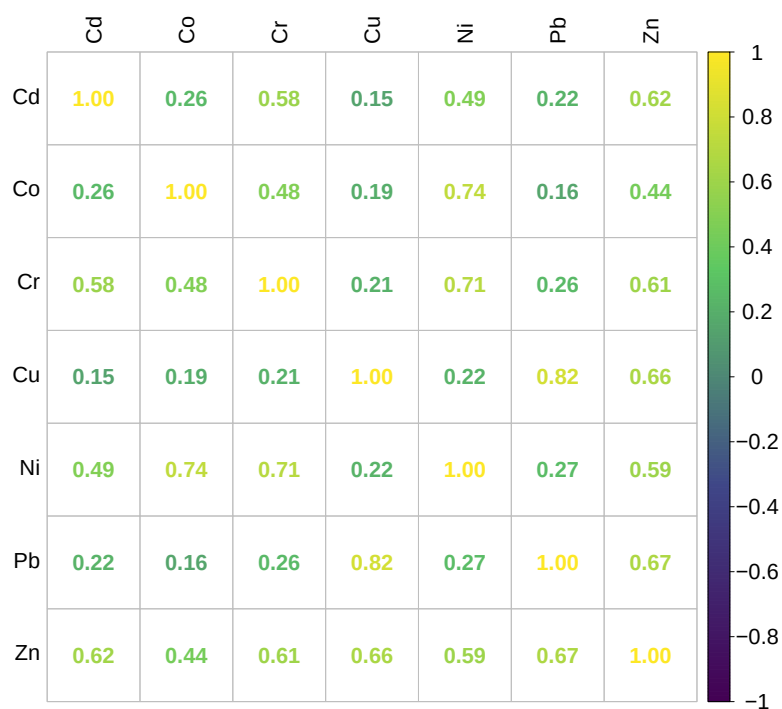


Figura 4.11: Correlación intrínseca para los datos Jura.

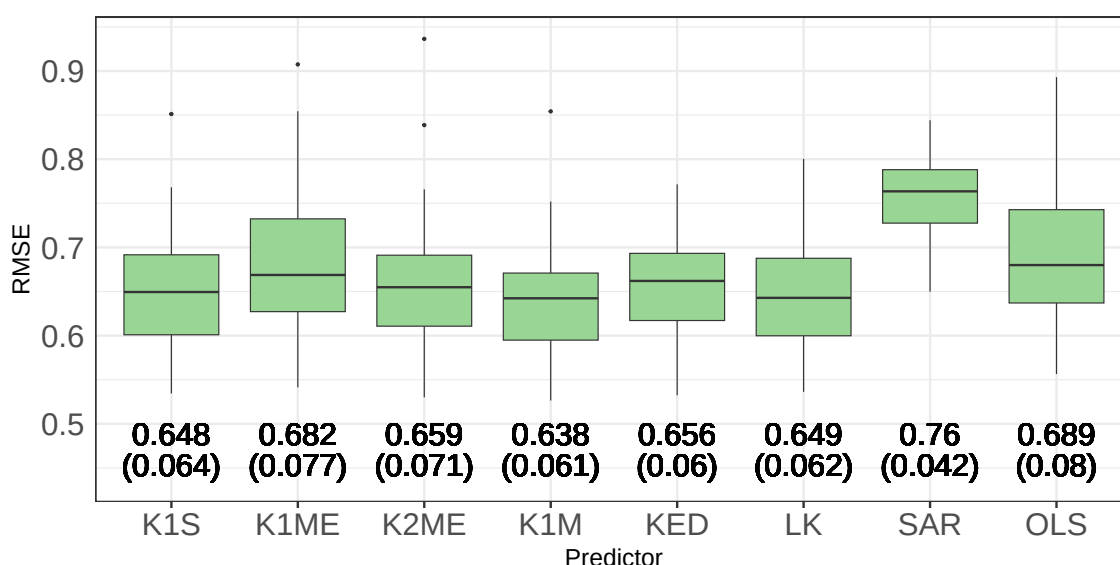


Figura 4.12: Distribución del RMSE de los datos Jura.

varias de las restantes variables, en particular con Zn y Cr, y en menor medida con Ni. Esto indica que las áreas con concentraciones elevadas de estos metales, en promedio, también tienden a presentar concentraciones altas de cadmio, lo que sugiere la posible existencia de fuentes comunes o de procesos geoquímicos compartidos que afectan simultáneamente a varios elementos.

Las correlaciones entre las covariables muestran además una estructura interna marcada. Destacan las asociaciones elevadas entre Cu y Pb, así como entre Ni y Cr, lo que apunta a agrupamientos de metales que varían de forma conjunta. Esta organización no es aleatoria y sugiere que el sistema está gobernado por uno o más factores latentes que explican gran parte de la variabilidad conjunta, más que por relaciones independientes entre pares de variables.

4.5.1. Resultados

En este caso contamos con 259 observaciones y dos covariables. Los resultados observados en la Figura 4.12 muestran nuevamente diferencias apreciables en el desempeño predictivo de los métodos, aunque con un patrón distinto al observado en Saltillo y Meuse. Entre los semiparamétricos se destacan el K1S, K1M, los cuales presentan los valores de RMSE más bajos, con medianas ubicadas en torno a 0.63–0.65, lo que indica una mayor capacidad para adaptarse a la estructura del proceso subyacente. Dentro de este grupo, K1M se destaca levemente.

El método LK presenta un desempeño comparable al de los enfoques semiparamétricos, lo que sugiere que la incorporación de suavizado local es suficiente para capturar gran parte de la variabilidad del proceso, aun bajo una formulación más estructurada. En contraste,

el modelo SAR exhibe el mayor RMSE y una dispersión levemente más baja, lo que indica un desempeño inferior.

El modelo OLS muestra un desempeño intermedio, mejor que SAR pero claramente por detrás de los métodos semiparamétricos. Esto sugiere que, si bien las covariables explican una parte sustancial de la variabilidad del cadmio, las relaciones no son completamente lineales ni homogéneas en el espacio. En conjunto, los resultados para Jura refuerzan la conclusión general de que, en presencia de heterogeneidad espacial y ausencia de una estructura global clara, los métodos flexibles tienden a ofrecer ventajas predictivas frente a enfoques paramétricos globales.

4.6. Comparación del desempeño predictivo entre dos bloques muestrales para el conjunto Ohio

En esta sección comparamos el comportamiento de los métodos cuando la validación se realiza por bloques no aleatorios, en contraposición con la validación aleatoria utilizada en las secciones anteriores.

En particular, analizamos dos bloques muestrales de prueba, que difieren en la localización espacial y en las características distribucionales. Para cada combinación de muestra y predictor se calcularon los errores de predicción al cuadrado de cada punto y se resumieron mediante su media (MSE) y su desvío (reportados entre paréntesis). Este enfoque permite comparar, por un lado, el desempeño relativo de los distintos métodos para cada muestra de entrenamiento y prueba y, por otro, evaluar cómo las diferencias en la estructura de dichas muestras influyen en la precisión y estabilidad de las predicciones.

En la Figura 4.13, graficamos los conjuntos de entrenamiento y de prueba de la **Muestra 1** (panel superior) y la **Muestra 2** (panel inferior). Al comparar los conjuntos de prueba desde el punto de vista espacial (paneles correspondientes a prueba), observamos que presentan patrones de distribución espacial diferentes. La primera muestra de prueba (panel (b)) exhibe una dispersión de valores por cuartiles caracterizada, relativamente, por valores mayores a la media. Por otro lado, la segunda muestra de prueba (panel (d)) posee mayor variabilidad, con valores máximos (cuartiles 75-100 %) concentrados en la zona norte de la región observada, mientras que los valores mínimos (cuartiles 0-25 %) se agrupan en el sureste, con una separación geográfica muy clara entre los grupos de alto y bajo valor.

Por otro lado, en la Figura 4.14, observamos la distribución de la variable respuesta *Pscore*. Observamos una diferencia clara en la estructura de la distribución: en la **Muestra 1**, la distribución se encuentra fuertemente concentrada en un rango relativamente acotado de valores, principalmente entre 60 y 90, con muy escasa presencia de observaciones bajas y una forma global unimodal y regular. En contraste, la **Muestra 2** presenta una distribu-

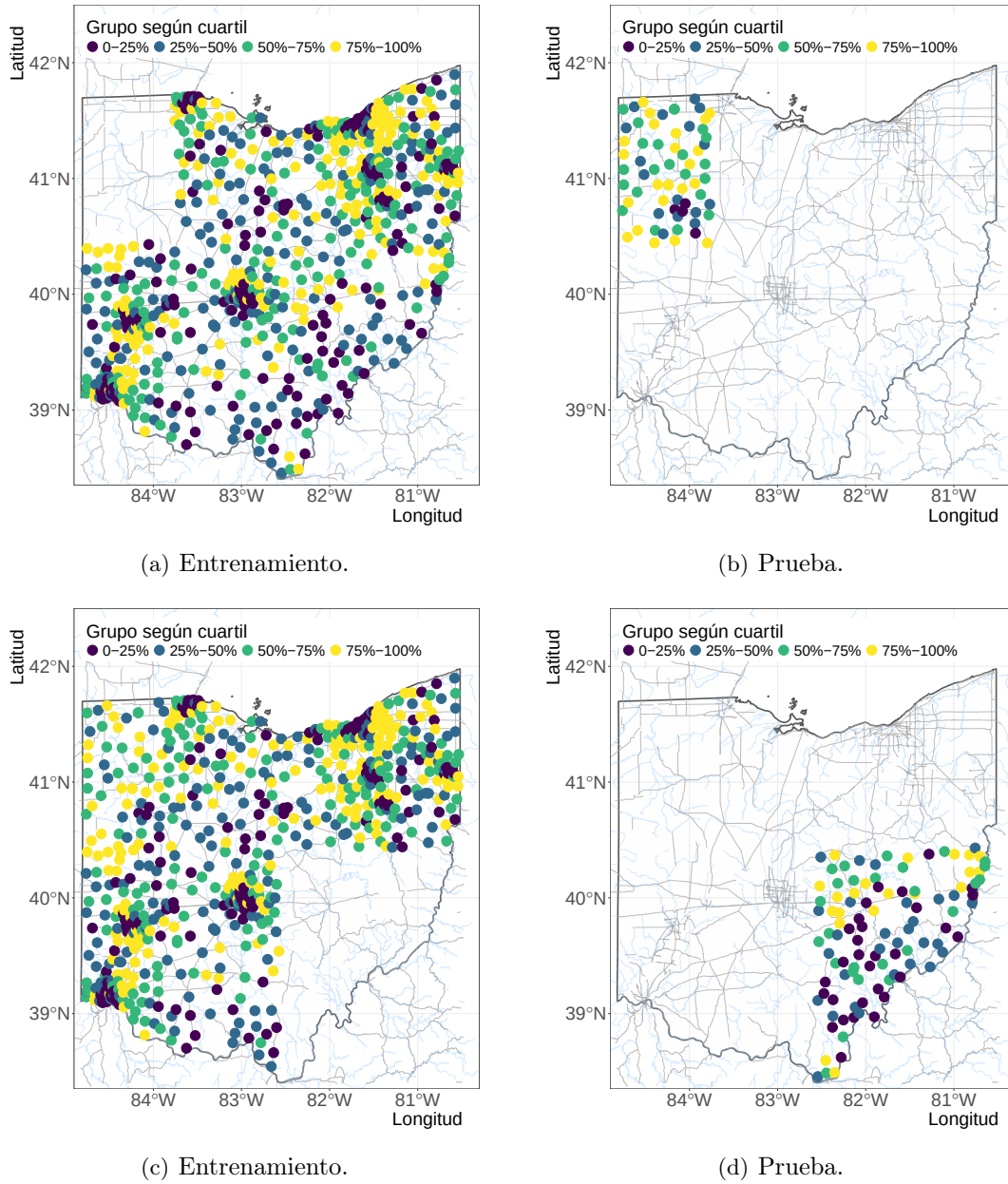


Figura 4.13: Distribución espacial de las muestras de entrenamiento y prueba para Ohio.

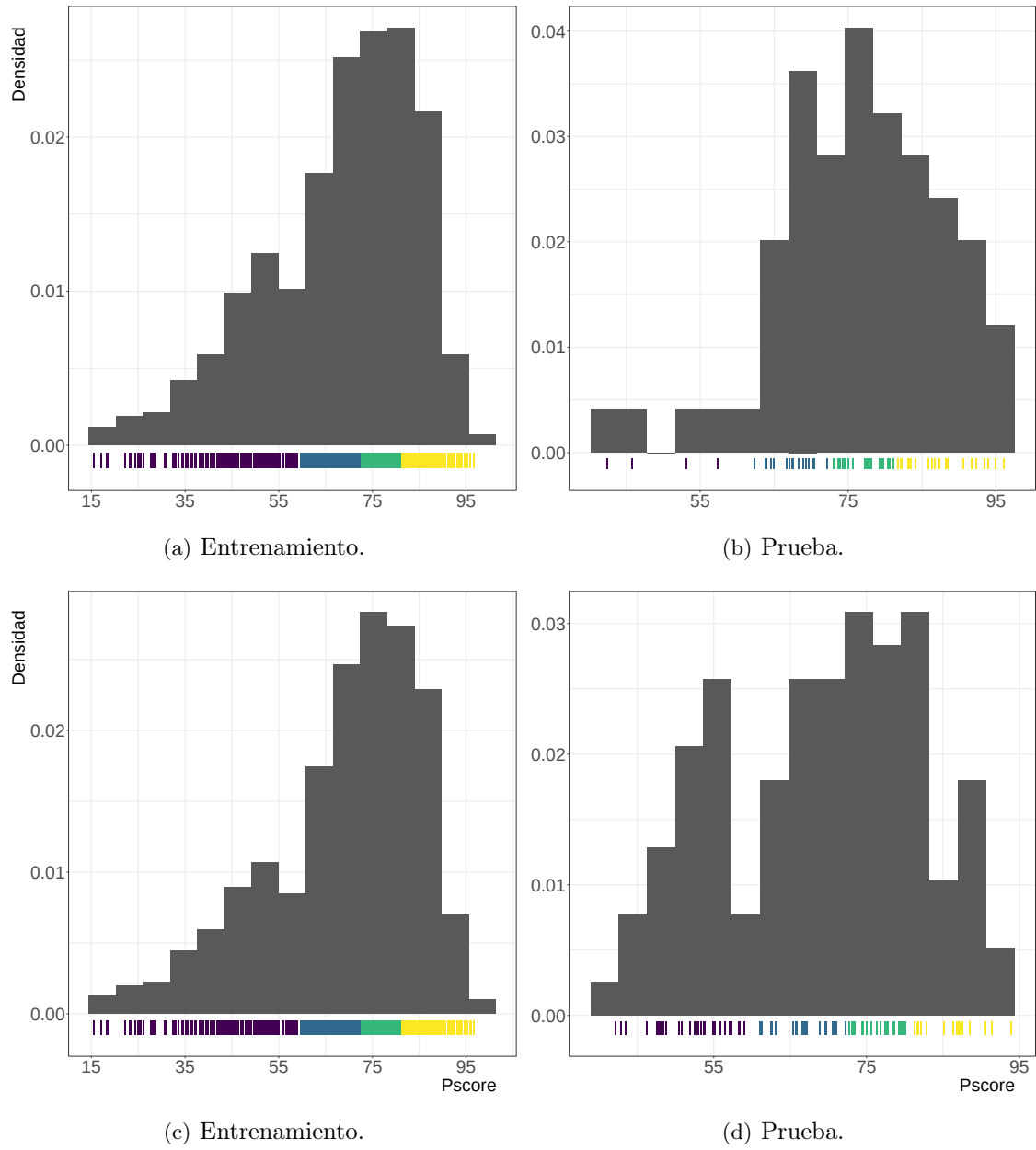


Figura 4.14: Distribución espacial de las muestras de entrenamiento y prueba para Ohio. Panel superior: Muestra 1, panel inferior: Muestra 2.

Muestra	K1S	K1ME	K2ME	K1M	KED	LK	SAR	OLS
Muestra 1	0.54 (0.71)	0.46 (0.56)	0.46 (0.57)	0.45 (0.56)	0.52 (0.64)	0.56 (0.7)	0.51 (0.65)	0.52 (0.64)
Muestra 2	0.71 (1.24)	0.65 (0.91)	0.61 (0.81)	0.63 (0.87)	0.71 (1.25)	0.7 (1.25)	0.7 (1.24)	0.71 (1.25)

Cuadro 4.1: MSE para cada método y cada muestra. Entre paréntesis se informa el desvío estándar de los errores al cuadrado.

ción más extendida hacia valores bajos, con mayor heterogeneidad y una forma claramente más irregular, evidenciada por la presencia de múltiples picos y una densidad menos concentrada. Esta mayor complejidad estructural de la Muestra 2 configura un escenario de predicción más exigente, ya que la variabilidad de la respuesta es considerablemente mayor que en el caso de la Muestra 1.

4.6.1. Resultados

En la Tabla 4.1 se reportan los MSE y, entre paréntesis, los desvíos de los errores al cuadrado para cada muestra. Dentro de cada muestra observamos ventajas consistentes de los métodos K1M, K1ME y K2ME, sobre el resto de predictores. En Muestra 1, estos métodos presentan los menores niveles de error, con valores de RMSE muy similares entre sí y claramente inferiores a los del resto de los predictores, lo que indica una mejora moderada pero sostenida en la precisión de las predicciones. En cambio, predictores como K1S y LK exhiben errores más elevados, mientras que SAR, OLS y KED se ubican en un nivel intermedio. Este mismo patrón se mantiene en Muestra 2, donde K2ME logra el menor RMSE, seguido de cerca por K1M y K1ME, confirmando su mejor desempeño relativo incluso en un escenario más exigente. Aunque el error global aumenta en esta muestra para todos los métodos, la jerarquía entre predictores se conserva, lo que sugiere que los enfoques K1M, K1ME y K2ME ofrecen un comportamiento más robusto respecto de la variación de la muestra a predecir y sistemáticamente más preciso dentro de cada conjunto de datos.

En la Figura 4.15 y 4.16 observamos las predicciones por cuartiles, para cada método, de los datos de la Muestra 1 y Muestra 2, respectivamente. En este contexto, los métodos K1M, K1ME y K2ME reproducen de manera consistente la ubicación relativa de varios puntos pertenecientes al grupo cuartílico superior y conservan en buena medida la estructura general observada. K1S presenta un patrón algo más uniforme, con menor contraste entre grupo cuartílicos. El método LK mantiene una distribución espacial coherente con los datos, aunque con cierta regularización en algunas zonas. Por su parte, OLS y KED tienden a mostrar una mayor homogeneidad en la asignación de grupos cuartílicos, reduciendo la variabilidad local, mientras que SAR refleja parcialmente la dependencia espacial



Figura 4.15: Distribución de las predicciones de la variable *p_score* de la Muestra 1 la base Ohio.

sin generar contrastes tan marcados. En conjunto, la Muestra 1 evidencia un desempeño visual relativamente similar entre varios métodos, lo que sugiere que la estructura espacial de esta muestra no impone grandes desafíos en términos de diferenciación local del puntaje. En la Figura 4.16, la comparación gráfica entre los predictores y los datos reales muestra diferencias claras en la capacidad para reproducir la heterogeneidad espacial observada. Los datos reales presentan zonas bien definidas de valores altos y una marcada variabilidad local. En este contexto, K2ME y, en menor medida, K1S, K1M y K1ME, logran capturar con mayor nitidez los focos de alto puntaje y las transiciones entre grupos cuartílicos, reproduciendo mejor los contrastes espaciales. El método LK muestra un comportamiento intermedio: consigue reflejar parte de la estructura espacial y algunas zonas intensas, aunque con un grado de suavizado mayor que K2ME, lo que atenúa ciertos extremos. Los métodos globales o de estructura más rígida, como OLS, SAR y KED, tienden a generar superficies un tanto más uniformes. En conjunto, la evidencia visual sugiere que los métodos con mayor capacidad de adaptación local, los semiparamétricos, reproducen con mayor fidelidad la estructura espacial de la Muestra 2, mientras que los enfoques más globales privilegian regularidad a costa de la precisión puntual.

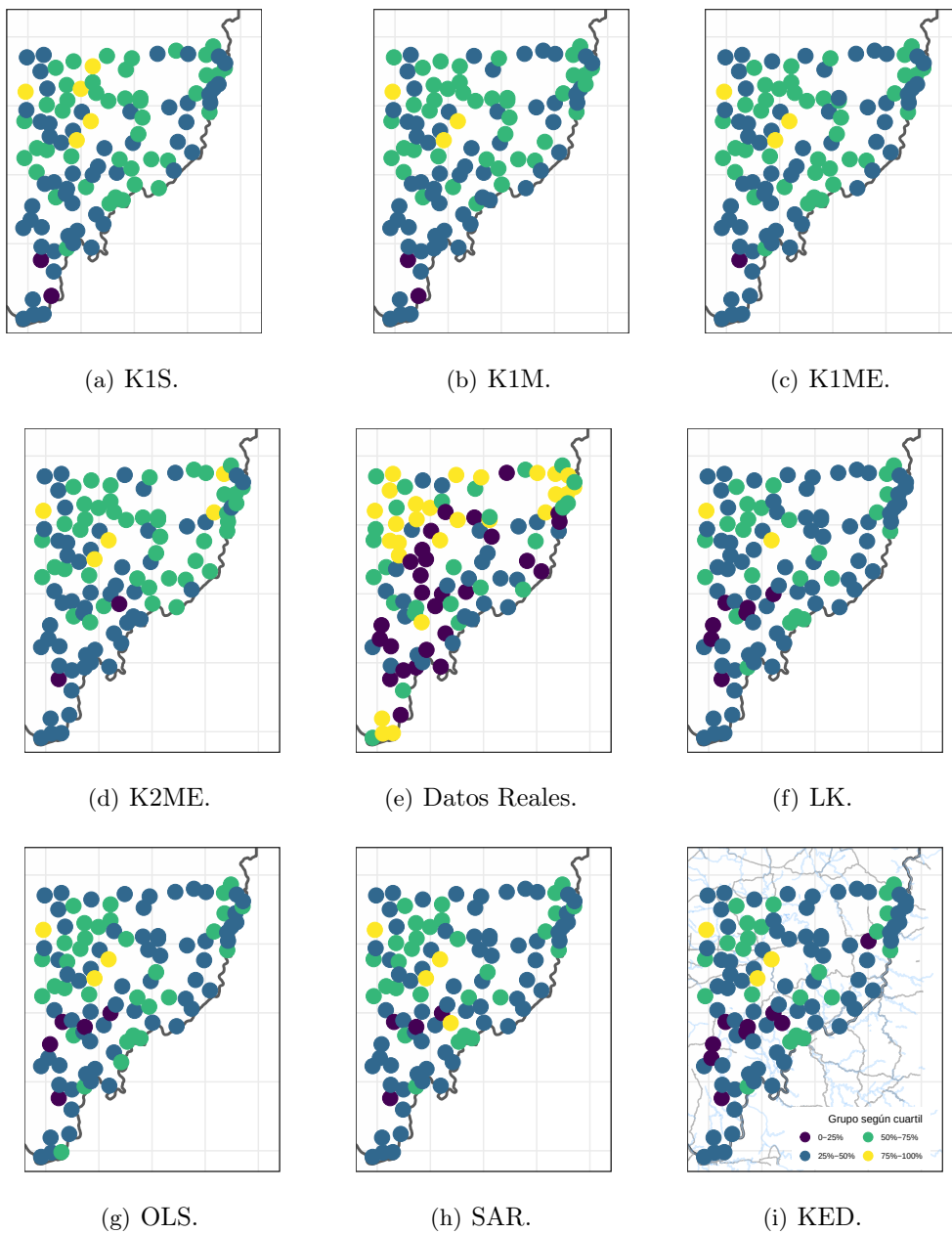


Figura 4.16: Distribución de las predicciones de la variable p_score de la Muestra 2 la base Ohio.

Conclusiones generales

En esta tesis se estudió un modelo autorregresivo espacial de tipo semiparamétrico orientado a problemas de predicción. El enfoque adoptado permitió integrar, dentro de una misma formulación, una estructura lineal interpretable y una componente no paramétrica flexible, preservando al mismo tiempo la dependencia espacial inherente a los datos.

Desde el punto de vista teórico, se establecieron resultados asintóticos para los estimadores asociados tanto a los parámetros lineales como a la componente no paramétrica, bajo un conjunto de hipótesis que, si bien son estándar en la literatura, no están exentas de restricciones. Estos resultados aportan respaldo formal al procedimiento de estimación propuesto, y a las predicciones que dichas estimaciones permitieron, además de caracterizar su comportamiento en muestras grandes. Concretamente hemos probado la normalidad asintótica del parámetro lineal, cuyo resultado nos permitió obtener las tasas de convergencia tanto para dicho parámetro como para la componente no paramétrica. A partir de ello, podemos asegurar la convergencia del predictor, que constituye uno de los objetivos de esta tesis.

Los estudios de simulación realizados permitieron complementar el análisis teórico y evaluar el desempeño del modelo en escenarios controlados. Los resultados sugieren que el enfoque semiparamétrico logra un desempeño competitivo en términos de error de predicción frente a métodos paramétricos tradicionales, especialmente en situaciones donde la especificación funcional es más compleja o parcialmente desconocida. Al mismo tiempo, observamos que el grado de dependencia espacial y la distribución de los puntos en el espacio influyen de manera no trivial en la calidad de las predicciones, lo que refuerza la necesidad de cautela al interpretar los resultados y de un análisis exploratorio adecuado previo a la aplicación del modelo.

La aplicación a datos reales mostró que la metodología propuesta es operativamente viable y capaz de adaptarse a distintos contextos empíricos. En estos casos, el modelo semiparamétrico permitió capturar patrones que no siempre son recogidos por enfoques puramente paramétricos, sin evidenciar pérdidas en el desempeño predictivo. Sin embargo, las mejoras observadas no son uniformes en todos los conjuntos de datos, lo que sugiere que el beneficio del enfoque depende fuertemente de las características específicas del problema

considerado.

En conjunto, los resultados obtenidos indican que el modelo autorregresivo espacial semiparamétrico y el algoritmo de estimación propuestos en esta tesis constituyen una alternativa válida dentro del conjunto de herramientas disponibles para la predicción espacial. Su principal aporte radica en ofrecer mayor flexibilidad cuando la estructura funcional subyacente es incierta y evitar la estimación tanto de estructuras de covarianza como de matrices de vecindario.

Como líneas futuras de investigación, resulta natural considerar extensiones del modelo a contextos no estacionarios, a estructuras de dependencia más generales o flexibilizar los supuestos sobre la estructura de los sitios de muestreo, modelándolos como aleatorios. Asimismo, queda pendiente la propuesta, y posterior análisis teórico, de un estimador de la varianza de los parámetros de la componente lineal que permita realizar inferencia sobre los mismos, ya sea mediante la derivación de un estimador clásico consistente o mediante una aproximación bootstrap de la varianza adaptada a estructuras de dependencia.

Bibliografía

- Ahmed, S. y de Marsily, G. (1987). Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. Water Resources Research, 23(9):1717–1737.
- Aneiros-Pérez, G. y Vieu, P. (2008). Nonparametric time series prediction: A semi-functional partial linear modeling. Journal of Multivariate Analysis, 99:834–857.
- Anselin, L. (2010). Local indicators of spatial association-LISA. Geographical Analysis, 27(2):93–115.
- Bauman, D., Drouet, T., Fortin, M.-J., y Dray, S. (2018). Optimizing the choice of a spatial weighting matrix in eigenvector-based methods. Ecology, 99(10):2159–2166.
- Biau, G. (2003). Spatial kernel density estimation. Mathematical Methods of Statistics, 12:371–390.
- Biau, G. y Cadre, B. (2004). Nonparametric spatial prediction. Statistical Inference for Stochastic Processes, 7(3):327–349.
- Bourennane, H., King, D., Chéry, P., y Bruand, A. (1996). Improving the kriging of a soil variable using slope gradient as external drift. European Journal of Soil Science, 47(4):473–483.
- Bourennane, H., King, D., y Couturier, A. (2000). Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. Geoderma, 97(3–4):255–271.
- Cerqueti, R., Maranzano, P., y Mattera, R. (2025). Spatially-clustered spatial autoregressive models with application to agricultural market concentration in europe. Journal of Agricultural, Biological and Environmental Statistics, pages 1–35.
- Chiles, J.-P. y Delfiner, P. (1999). Geostatistics: modeling spatial uncertainty. John Wiley & Sons.
- coahuila.gob (2020). Datos municipios. <https://coahuila.gob.mx/micrositios/index/datos-municipios>.

- Cressie, N. (1993). Statistics for Spatial Data. John Wiley and Sons, Inc.
- 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.
- Dabo-Niang, S., Ternynck, C., y Yao, A.-F. (2016). Nonparametric prediction of spatial multivariate data. Journal of Nonparametric Statistics, 28(2):428–458.
- Duncan, E. W., White, N. M., y Mengersen, K. (2017). Spatial smoothing in bayesian models: a comparison of weights matrix specifications and their impact on inference. International journal of health geographics, 16(1):47.
- Fischer, M. M. y Wang, J. (2011). Spatial data analysis: models, methods and techniques. Springer Science & Business Media.
- Fouedjio, F. y Arya, E. (2024). Locally varying geostatistical machine learning for spatial prediction. Artificial Intelligence in Geosciences, 5:100081.
- García Arancibia, R., Llop Orzan, P. N., y Lovatto, M. G. (2023). Nonparametric prediction for univariate spatial data: methods and applications. Papers in Regional Science.
- Goovaerts, P. (1999). Using elevation to aid the geostatistical mapping of rainfall erosivity. Catena, 34(3-4):227–242.
- Goulard, M., Laurent, T., y Thomas-Agnan, C. (2017). About predictions in spatial autoregressive models: Optimal and almost optimal strategies. Spatial Economic Analysis, 12(2-3):304–325.
- 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.
- Griffith, D. A., Bennett, R. J., y Haining, R. P. (1989a). Statistical analysis of spatial data in the presence of missing observations: a methodological guide and an application to urban census data. Environment and Planning A, 21(11):1511–1523.
- Griffith, D. A., Bennett, R. J., y Haining, R. P. (1989b). Statistical analysis of spatial data in the presence of missing observations: a methodological guide and an application to urban census data. Environment and Planning A, 21(11):1511–1523.
- Hengl, T., Heuvelink, G. B., y Stein, A. (2003). Comparison of kriging with external drift and regression-kriging. Technical note, International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands.

- Hudson, G. y Wackernagel, H. (1994). Mapping temperature using kriging with external drift: Theory and an example from scotland. International Journal of Climatology, 14(1):77–91.
- Ibragimov, I. A. (1962). Some limit theorems for stationary processes. Theory of Probability & Its Applications, 7(4):349–382.
- Jeong, M. y Koo, H. (2025). Evaluating spatio-temporal kriging with machine learning considering the sources of spatio-temporal variation. ISPRS International Journal of Geo-Information, 14(6):224.
- Kato, T. (2008). A further exploration into the robustness of spatial autocorrelation specifications. Journal of Regional Science, 48(3):615–639.
- Kato, T. (2013). Usefulness of the information contained in the prediction sample for the spatial error model. Journal of Real Estate Finance and Economics, 47:169–195.
- Kelejian, H. H. y Prucha, I. R. (2007). The relative efficiencies of various predictors in spatial econometric models containing spatial lags. Regional Science and Urban Economics, 37(3):363–374.
- Khan, M., Almazah, M. M., Ellahi, A., Niaz, R., Al-Rezami, A., y Zaman, B. (2023). Spatial interpolation of water quality index based on ordinary kriging and universal kriging. Geomatics, Natural Hazards and Risk, 14(1):2190853.
- Kheir, A. M., Govind, A., Nangia, V., El-Maghraby, M. A., Elnashar, A., Ahmed, M., Aboelsoud, H., Gamal, R., y Feike, T. (2025). Hybridization of process-based models, remote sensing, and machine learning for enhanced spatial predictions of wheat yield and quality. Computers and Electronics in Agriculture, 234:110317.
- Kiani, A., Motamedvaziri, B., Khaleghi, M. R., y Ahmadi, H. (2025). Spatial prediction of flood susceptible areas using machine learning methods in the shahkhor watershed of kermanshah province. Earth Science Informatics, 18(1):20.
- LeSage, J. y Pace, R. K. (2009). Introduction to Spatial Econometrics. Chapman and Hall/CRC.
- LeSage, J. P. y Pace, R. K. (2004). Models for spatially dependent missing data. Journal of Real Estate Finance and Economics, 29(2):233–254.
- Longley, P. (2005). Geographic information systems and science. John Wiley & Sons.
- Lovatto, M. G., García Arancibia, R., y Llop Orzan, P. N. (2022). Kriging semiparamétrico para datos univariados. Master's thesis, Universidad Nacional de Rosario, Argentina.
- Matheron, G. (1969). Le krigeage universel, volume 1 of Cahiers du Centre de Morphologie Mathématique. École des Mines de Paris, Fontainebleau, France.

- Menezes, R., García-Soidán, P., y Ferreira, C. (2010). Nonparametric spatial prediction under stochastic sampling design. Journal of Nonparametric Statistics, 22(3):363–377.
- Meng, D., Yang, H., Yang, S., Zhang, Y., De Jesus, A. M., Correia, J., Fazeres-Ferradosa, T., Macek, W., Branco, R., y Zhu, S.-P. (2024). Kriging-assisted hybrid reliability design and optimization of offshore wind turbine support structure based on a portfolio allocation strategy. Ocean Engineering, 295:116842.
- Metadata (2018). Valuación de inmuebles.
- Montero, J.-M., Fernández-Avilés, G., y Mateu, J. (2015). Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. John Wiley and Sons, Ltd.
- Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., y Sain, S. (2015). A multiresolution gaussian process model for the analysis of large spatial datasets. Journal of Computational and Graphical Statistics, 24(2):579–599.
- Odeh, I. O. A., McBratney, A. B., y Chittleborough, D. J. (1994). Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma, 63(3-4):197–214.
- Odeh, I. O. A., McBratney, A. B., y Chittleborough, D. J. (1995). Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma, 67(3-4):215–226.
- Oliver, D. S. (2003). Gaussian cosimulation: Modelling of the cross-covariance. Mathematical Geology, 35(6):681–698.
- Oodaira, H. y ichi Yoshihara, K. (1971). The law of the iterated logarithm for stationary processes satisfying mixing conditions. KODAI MATHEMATICAL SEMINAR REPORTS, 23(3):311–334.
- Pang, Y., Wang, Y., Lai, X., Zhang, S., Liang, P., y Song, X. (2023). Enhanced kriging leave-one-out cross-validation in improving model estimation and optimization. Computer Methods in Applied Mechanics and Engineering, 414:116194.
- Papritz, A. y Stein, A. (1999). Spatial prediction by linear kriging. In Stein, A., van der Meer, F., y Gorte, B., editors, Spatial statistics for remote sensing, pages 83–113. Kluwer Academic Publishers, Dordrecht.
- 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 oodplains of the meuse west of Stein. Department of Physical Geography, Utrecht University, Utrecht.
- Robinson, G. K. (1991). That blup is a good thing: The estimation of random effects. Statistical Science, 6(1):15–51.

- Robinson, P. M. (1988). Root-n-consistent semiparametric regression. Econometrica, 56(4):931–954.
- Shen, L., LaRue, E., Fei, S., y Zhang, H. (2024). Spatial prediction of plant invasion using a hybrid of machine learning and geostatistical method. Ecology and Evolution, 14:e11605.
- Stakhovych, S. y Bijmolt, T. H. (2009). Specification of spatial models: A simulation study on weights matrices. Papers in Regional Science, 88(2):389–409.
- Stein, M. L. (1999). Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.
- Tadić, J. M., Ilić, V., Ilić, S., Pavlović, M., y Tadić, V. (2024). Hybrid machine learning and geostatistical methods for gap filling and predicting solar-induced fluorescence values. Remote Sensing, 16(10):1707.
- Vagnini, C., Canal Vieira, L., Longo, M., y Mura, M. (2025). Regional drivers of industrial decarbonisation: a spatial econometric analysis of 238 eu regions between 2008 and 2020. Regional Studies, 59(1):2380369.
- Wackernagel, H. (2006). Geostatistics. American Cancer Society.
- 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.
- Walk, H. (2010). Strong laws of large numbers and nonparametric estimation. In Devroye, L., Karasözen, B., Kohler, M., y Korn, R., editors, Recent Developments in Applied Probability and Statistics, pages 1–40. Springer, Berlin.
- Welham, S., Cullis, B., Gogel, B., Gilmour, A., y Thompson, R. (2004). Prediction in linear mixed models. Australian & New Zealand Journal of Statistics, 46(3):325–347.
- Yaglom, A. M. (1987). Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results, volume 131. Springer.
- Zhang, J., Atkinson, P., y Goodchild, M. F. (2014). Scale in spatial information and analysis. CRC Press.
- Zhao, Y. y Wall, M. M. (2004). Investigating the use of the variogram for lattice data. Journal of Computational and Graphical Statistics, 13(3):719–738.

Apéndice A

Resultados Auxiliares

A continuación introducimos algunos resultados necesarios para probar nuestro resultado principal.

Teorema A.0.1 (Teorema 5, Odaira y Yoshihara (1971)). *Sea $\{V_k\}$ una función aleatoria real, estrictamente estacionaria, de media cero y fuertemente mixing, tal que*

$$E|V_1|^{2+\delta} < \infty \quad y \quad \sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta'}{2+\delta}} < \infty \quad (0 < \delta' < \delta),$$

y

$$Var(V_1) + 2 \sum_{k=1}^{\infty} Cov(V_1, V_{1+k}) > 0.$$

Sea $S_n = \sum_{k=1}^n V_k$ y $s_n^2 = E(S_n^2)$. Entonces,

$$\limsup_{n \rightarrow \infty} \frac{S_n}{(2s_n^2 \log \log s_n^2)^{1/2}} = 1 \quad c.s.$$

Lema A.0.2 (Aneiros-Pérez y Vieu (2008)). *Sea $\{V_k\}$ un proceso real, estacionario, fuertemente mixing y de media cero, tal que para algún $r > 4$,*

$$\max_{1 \leq k \leq n} E|V_k|^r < C < \infty.$$

Sea $\{a_{ik}, i, k = 1, \dots, \hat{n}\}$ una sucesión de números positivos tal que $\max_{1 \leq i, k \leq \hat{n}} |a_{ik}| = O(a_n)$. Si, además,

$$\sum_{n=1}^{\infty} n^{\frac{5+4\gamma}{4(1-\gamma)}} \alpha(n) < \infty \quad (0.5 < \gamma < 1),$$

entonces,

$$\max_{1 \leq i \leq \hat{n}} \left| \sum_{k=1}^{\hat{n}} a_{ik} V_k \right| = O\left(a_n \hat{n}^{1/2+1/r} \log \hat{n}\right) \quad c.s.$$

Teorema A.0.3 (Walk (2010)). *Sean $\{Y_n\}_{n \geq 1}$ variables aleatorias reales idénticamente distribuidas. Si $\mathbb{E}(|Y_1| \log^+ |Y_1|) < \infty$ e $\{Y_n\}$ es α -mixing con*

$$\alpha_n = O(n^{-b}) \quad \text{para algún } b > 0,$$

entonces

$$\frac{1}{n} \sum_{i=1}^n Y_i \longrightarrow \mathbb{E}(Y_1) \quad c.s.$$

Teorema A.0.4 (Teorema 1.7, Ibragimov (1962)). Sea $\{Y_j\}$ un proceso aleatorio estacionario fuertemente α -mixing, tal que

$$\mathbb{E}|Y_j|^{2+\delta} < \infty$$

para algún $\delta > 0$, y además

$$\sum_{j=1}^{\infty} (\alpha(j))^{\delta/(2+\delta)} < \infty.$$

Entonces,

$$\sigma^2 = \mathbb{E}(Y_0 - \mathbb{E}Y_0)^2 + 2 \sum_{j=1}^{\infty} \mathbb{E}[(Y_0 - \mathbb{E}Y_0)(Y_j - \mathbb{E}Y_j)] < \infty.$$

Además, si $\sigma \neq 0$ y $\mathbb{E}[Y_j] = 0$, entonces

$$\mathbb{P}\left(\frac{Y_1 + \cdots + Y_n}{\sigma\sqrt{n}} < z\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-u^2/2} du.$$

Teorema A.0.5 (Teorema 3.1 Dabo-Niang et al. (2016)). Si se cumplen las siguientes hipótesis

- H1: La función de densidad f y la función de regresión r son continuas en D y se cumple que $\inf_D f > 0$.
- H2: La función de densidad f y la función de regresión r son Lipschitz.
- H3: Las funciones $K_1(\cdot)$ y $K_2(\cdot)$ son núcleos integrables y acotados en $\mathbb{R}$. Además, el núcleo $K_1(\cdot)$ satisface ciertas condiciones de Lipschitz.
- H4: Existen constantes C_{1i} y C_{2i} con $0 < C_{1i} < C_{2i} < \infty$, para $i = 1, 2$, tales que

$$C_{11}\mathbf{1}_{[0,1]}(s's) \leq K_1(s) \leq C_{21}\mathbf{1}_{[0,1]}(s's) \quad \text{para } s \in \mathbb{R}^d,$$

$$C_{12}\mathbf{1}_{[0,1]}(t) \leq K_2(t) \leq C_{22}\mathbf{1}_{[0,1]}(t) \quad \text{para } t \in \mathbb{R},$$

donde s' denota la transpuesta de s .

- H5: *Condición de dependencia local*. La densidad de probabilidad conjunta f_{X_i, X_j} de (X_i, X_j) existe, es acotada y, para todo $u, v \in \mathbb{R}^d$, verifica para alguna constante $C > 0$, que $|f_{X_i, X_j}(u, v) - f_i(u)f_j(v)| < C$, para todo $u, v, i \neq j$.
- H6: $\psi(n, m) \leq C \min(n, m)$ y $\hat{n} b_n^{d_0} \rho_n^{N_0} \log \hat{n} b_n^3 u_n^{\theta_3} \rightarrow \infty$ con el coeficiente de mezcla $\theta > N(s(d+4) + 2d)$, $s > 2$ y

$$\begin{aligned} \theta_0 &= \frac{s(\theta + N(d+2))}{\theta - N(s(d+4) + 2d)}; & \theta_1 &= \frac{s(\theta + Nd)}{\theta - N(s(d+4) + 2d)}; \\ \theta_2 &= \frac{s(N(d+2) - \theta)}{\theta - N(s(d+4) + 2d)}; & \theta_3 &= \frac{-2(N(s+d) + \theta)}{\theta - N(s(d+4) + 2d)}. \end{aligned}$$

H7: $\psi(n, m) \leq C(n + m + 1)^{\tilde{\beta}}$ y $\hat{\mathbf{n}} b_{\mathbf{n}}^{d'_0} \rho_{\mathbf{n}}^{N'_0} \log \hat{\mathbf{n}}^{\theta'_2} u_{\mathbf{n}}^{\theta'_3} \rightarrow \infty$ con el coeficiente de mezcla $\theta > N(s(d + 3 + 2\tilde{\beta}) + 2(d + 1))$, $s > 2$, $\tilde{\beta} > 1$ y

$$\begin{aligned}\theta'_0 &= \frac{s(\theta + N(d + 3))}{\theta - N(s(d + 3 + 2\tilde{\beta}) + 2(d + 1))}; & \theta'_1 &= \frac{s(\theta + N(d + 1))}{\theta - N(s(d + 3 + 2\tilde{\beta}) + 2(d + 1))}; \\ \theta'_2 &= \frac{s(N(d + 1) - \theta)}{\theta - N(s(d + 3 + 2\tilde{\beta}) + 2(d + 1))}; & \theta'_3 &= \frac{-2(N(s + d + 1) + \theta)}{\theta - N(s(d + 3 + 2\tilde{\beta}) + 2(d + 1))}.\end{aligned}$$

H8: (i) Las densidades $f_{\mathbf{i}}$ y $f_{X_{\mathbf{i}}, Y_{\mathbf{i}}}$ de $X_{\mathbf{i}}$ y $(X_{\mathbf{i}}, Y_{\mathbf{i}})$ son tales que

$$\sup_{x \in \mathbb{R}^d, \|(\mathbf{i} - \mathbf{i}_0)/\mathbf{n}\| \leq h} |f_{\mathbf{i}}(x) - f(x)| = o(1), \quad \sup_{x \in \mathbb{R}^d, \|(\mathbf{i} - \mathbf{i}_0)/\mathbf{n}\| \leq h} |\varphi_{\mathbf{i}}(x) - \varphi(x)| = o(1)$$

cuando $h \rightarrow 0$ con $\varphi_{\mathbf{i}}(x) = \int y f_{X_{\mathbf{i}}, Y_{\mathbf{i}}}(x, y) dy$. La densidad condicional $f_{X_{\mathbf{i}}, X_{\mathbf{j}}|Y_{\mathbf{i}}, Y_{\mathbf{j}}}$ de $(X_{\mathbf{i}}, X_{\mathbf{j}})$ dado $(Y_{\mathbf{i}}, Y_{\mathbf{j}})$ y la densidad condicional $f_{X_{\mathbf{i}}|Y_{\mathbf{j}}}$ de $X_{\mathbf{i}}$ dado $Y_{\mathbf{j}}$ existen y

$$f_{X_{\mathbf{i}}, X_{\mathbf{j}}|Y_{\mathbf{i}}, Y_{\mathbf{j}}}(y, t|x, u) \leq C \quad \text{y} \quad f_{X_{\mathbf{i}}|Y_{\mathbf{j}}}(y|u) \leq C$$

para todo $x, y, u, t, \mathbf{i}, \mathbf{j}$.

(ii) $\mathbb{E}(|Y_{\mathbf{i}}|^s) < \infty$, $\sup_u \int |v|^s f_{X_{\mathbf{i}}, Y_{\mathbf{i}}}(u, v) dv < C$ para algún $s > 2$, para todo $\mathbf{i}$.

Sea $g_2(u) = \text{Var}(Y_{\mathbf{i}}|X_{\mathbf{i}} = u)$ independiente de $\mathbf{i}$ y continua en un entorno de x :

$$\sup_{u, \|u - x\| \leq b} |g_2(u) - g_2(x)| = o(1) \quad \text{cuando } b \rightarrow 0.$$

Adicionalmente, $g_s(u) = \mathbb{E}(|Y_{\mathbf{i}} - r(x)|^s | X_{\mathbf{i}} = u)$ es independiente de $\mathbf{i}$ y continua en un entorno de x . $g(u, v, x) = \mathbb{E}((Y_{\mathbf{i}} - r(x))((Y_{\mathbf{j}} - r(x)) | X_{\mathbf{i}} = u, X_{\mathbf{j}} = v), \forall \mathbf{i} \neq \mathbf{j}$, no depende de $\mathbf{i}, \mathbf{j}$ y es continua en un entorno de (x, x) .

Entonces

$$\lim_{n \rightarrow \infty} \sup_{x \in D} |r_n(x) - r(x)| = 0 \quad \text{c.c.}$$

Con

$$r_{\mathbf{n}}(x) = \frac{\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} Y_{\mathbf{i}} K_1((x - X_{\mathbf{i}})/b_{\mathbf{n}}) K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|)}{\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} K_1((x - X_{\mathbf{i}})/b_{\mathbf{n}}) K_{2, \rho_{\mathbf{n}}}(\|\mathbf{i}_0 - \mathbf{i}\|)},$$

en $\mathcal{I}_{\mathbf{n}}$ una grilla regular en $\mathbb{Z}^d$.

