

MAPEO ESPACIAL PARA VARIABLES ECONÓMICAS: ALTERNATIVAS FLEXIBLES AL MÉTODO CLÁSICO DE LA GEOESTADÍSTICA

Rodrigo García Arancibia
IECAL-FCE-UNL & CONICET

Trabajo en conjunto con Pamela Llop y Mariel Lovatto (FIQ-UNL & CONICET)

Workshop del IEF-FCE-UNC, 2022

- Trabajo: *Non-Parametric Prediction for univariate spatial data: Methods and Applications* (submitted)

- Trabajo: *Non-Parametric Prediction for univariate spatial data: Methods and Applications* (submitted)
- Proyecto PICT Joven de la Agencia (Ministerio de Ciencia, Tecnología e Innovación, Argentina): *Métodos de Predicción Espacial Semiparamétrica*

- Trabajo: *Non-Parametric Prediction for univariate spatial data: Methods and Applications* (submitted)
- Proyecto PICT Joven de la Agencia (Ministerio de Ciencia, Tecnología e Innovación, Argentina): *Métodos de Predicción Espacial Semiparamétrica*
- Parte de los métodos y resultados forman de la Tesis de Maestría titulada *Predicción no paramétrica para datos espaciales univariados* presentada por Mariel Lovatto bajo la dirección de Pamela Llop y Rodrigo García Arancibia.

Problema

- Predicción de una variable de interés en determinado punto geográfico a partir de mediciones de dicha variable en otras locaciones espaciales.

Problema

- Predicción de una variable de interés en determinado punto geográfico a partir de mediciones de dicha variable en otras locaciones espaciales.
- Bajo el enfoque de la estadística espacial uno de los métodos más utilizados para tal fin es el clásico método de *kriging*.

Problema

- Predicción de una variable de interés en determinado punto geográfico a partir de mediciones de dicha variable en otras locaciones espaciales.
- Bajo el enfoque de la estadística espacial uno de los métodos más utilizados para tal fin es el clásico método de *kriging*.
- Puede que los supuestos sean algo restrictivos y mas aún, para datos socioeconómicos.

Problema

- Predicción de una variable de interés en determinado punto geográfico a partir de mediciones de dicha variable en otras locaciones espaciales.
- Bajo el enfoque de la estadística espacial uno de los métodos más utilizados para tal fin es el clásico método de *kriging*.
- Puede que los supuestos sean algo restrictivos y mas aún, para datos socioeconómicos.
- Por ello, en este trabajo se presenta una propuesta flexible siguiendo el espíritu del *kriging*, donde en este caso los pesos son estimados de forma no paramétrica.

Enfoques de la estadística espacial

- En el campo de la estadística espacial se distinguen generalmente dos enfoques, a saber, el de la **econometría espacial** y el de la **geoestadística**.

Enfoques de la estadística espacial

- En el campo de la estadística espacial se distinguen generalmente dos enfoques, a saber, el de la **econometría espacial** y el de la **geoestadística**.
- **Econometría Espacial** \rightsquigarrow desarrollo de modelos de regresión espacial para validar o contrastar hipótesis de la teoría económica (Anselin10)

Enfoques de la estadística espacial

- En el campo de la estadística espacial se distinguen generalmente dos enfoques, a saber, el de la **econometría espacial** y el de la **geoestadística**.
- **Econometría Espacial** ~> desarrollo de modelos de regresión espacial para validar o contrastar hipótesis de la teoría económica (Anselin10) ~> Extensión a modelos de regresión aplicados a datos espaciales donde la autocorrelación espacial y la heterogeneidad espacial están presentes y se incorporan al modelado (Haining14).

Enfoques de la estadística espacial

- En el campo de la estadística espacial se distinguen generalmente dos enfoques, a saber, el de la **econometría espacial** y el de la **geoestadística**.
- **Econometría Espacial** ~> desarrollo de modelos de regresión espacial para validar o contrastar hipótesis de la teoría económica (Anselin10) ~> Extensión a modelos de regresión aplicados a datos espaciales donde la autocorrelación espacial y la heterogeneidad espacial están presentes y se incorporan al modelado (Haining14).
- **Geoestadística**: Tradicionalmente más relacionada con estudios de fenómenos físicos, geológicos y ambientales. Desde hace algunos años se ha incorporado como una herramienta más de las ciencias sociales, tanto para la predicción y mapeo de datos regionales (presentados como conteos de áreas pequeñas), como para modelar el comportamiento espacial de una variable de respuesta de interés.

Enfoques de la estadística espacial

- En el campo de la estadística espacial se distinguen generalmente dos enfoques, a saber, el de la **econometría espacial** y el de la **geoestadística**.
- **Econometría Espacial** ~> desarrollo de modelos de regresión espacial para validar o contrastar hipótesis de la teoría económica (Anselin10) ~> Extensión a modelos de regresión aplicados a datos espaciales donde la autocorrelación espacial y la heterogeneidad espacial están presentes y se incorporan al modelado (Haining14).
- **Geoestadística**: Tradicionalmente más relacionada con estudios de fenómenos físicos, geológicos y ambientales. Desde hace algunos años se ha incorporado como una herramienta más de las ciencias sociales, tanto para la predicción y mapeo de datos regionales (presentados como conteos de áreas pequeñas), como para modelar el comportamiento espacial de una variable de respuesta de interés. ~> más en el paradigma de ML.
- Extensión de la geoestadística en economía ~> fines predictivos.

Kriging clásico para predicción univariada

Método cuyo predictor es un promedio pesado de observaciones muestreadas sobre diferentes puntos espaciales. Más precisamente, para una locación $\mathbf{s}_0 \in D \subseteq \mathbb{R}^d$, el **predictor kriging** es definido como

$$Z_\omega(\mathbf{s}_0) = \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) \doteq \boldsymbol{\omega}^T \mathbf{Z}, \quad (0.1)$$

donde los pesos $\boldsymbol{\omega} \doteq (\omega(\mathbf{s}_1, \mathbf{s}_0), \dots, \omega(\mathbf{s}_n, \mathbf{s}_0))^T \in \mathbb{R}^n$ son obtenidos de forma tal que el predictor sea **insesgado** y de **varianza mínima**.

Kriging clásico para predicción univariada

Método cuyo predictor es un promedio pesado de observaciones muestreadas sobre diferentes puntos espaciales. Más precisamente, para una locación $\mathbf{s}_0 \in D \subseteq \mathbb{R}^d$, el **predictor kriging** es definido como

$$Z_{\omega}(\mathbf{s}_0) = \sum_{i=1}^n \omega_i Z(\mathbf{s}_i) \doteq \boldsymbol{\omega}^T \mathbf{Z}, \quad (0.1)$$

donde los pesos $\boldsymbol{\omega} \doteq (\omega(\mathbf{s}_1, \mathbf{s}_0), \dots, \omega(\mathbf{s}_n, \mathbf{s}_0))^T \in \mathbb{R}^n$ son obtenidos de forma tal que el predictor sea **insesgado** y de **varianza mínima**. Para la estimación de tales pesos es necesario modelar la estructura de correlación del fenómeno a estudiar, para lo cual es necesaria la estimación del semivariograma:

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

Kriging Ordinario y Universal

Kriging Ordinario y Universal

Predictor sea insesgado

$$\implies \mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0$$

Kriging Ordinario y Universal

Predictor sea insesgado

$$\implies \mathbb{E}[Z_{\omega}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0$$

La varianza del error sea mínima

$$\implies \underset{\omega}{\operatorname{argmín}} \operatorname{var}[Z_{\omega}(\mathbf{s}_0) - Z(\mathbf{s}_0)] \quad \text{sujeto a} \quad \omega^T \mathbf{1} = 1$$

Kriging Ordinario y Universal

Predictor sea insesgado

$$\implies \mathbb{E}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0$$

La varianza del error sea mínima

$$\implies \underset{\omega}{\operatorname{argmín}} \operatorname{var}[Z_\omega(\mathbf{s}_0) - Z(\mathbf{s}_0)] \quad \text{sujeto a} \quad \omega^T \mathbf{1} = 1$$

KO

$$\begin{aligned} \Gamma\omega + \lambda\mathbf{1} &= \gamma_0 \\ \omega^T \mathbf{1} &= 1 \end{aligned}$$

KU

$$\begin{aligned} \Gamma_Y\omega + \mathbf{F}\lambda &= \gamma_{Y,0} \\ \omega^T \mathbf{F} &= \mathbf{f}_0^T, \end{aligned}$$

Kriging Ordinario y Universal

Predictor sea insesgado

$$\implies \mathbb{E}[Z_{\omega}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0$$

La varianza del error sea mínima

$$\implies \underset{\omega}{\operatorname{argmín}} \operatorname{var}[Z_{\omega}(\mathbf{s}_0) - Z(\mathbf{s}_0)] \quad \text{sujeto a} \quad \omega^T \mathbf{1} = 1$$

KO

$$\begin{aligned} \Gamma \omega + \lambda \mathbf{1} &= \gamma_0 \\ \omega^T \mathbf{1} &= 1 \end{aligned}$$

KU

$$\begin{aligned} \Gamma_Y \omega + \mathbf{F} \lambda &= \gamma_{Y,0} \\ \omega^T \mathbf{F} &= \mathbf{f}_0^T, \end{aligned}$$

Pesos KO

$$\omega = \Gamma^{-1} \left(\gamma_0 - \frac{\mathbf{1}^T \Gamma^{-1} \gamma_0}{\mathbf{1}^T \Gamma^{-1} \mathbf{1}} \mathbf{1} \right)$$

Pesos KU

$$\begin{aligned} \omega_{UK} &\doteq \Gamma_Y^{-1} (\gamma_{Y,0} - \\ &\mathbf{F} (\mathbf{F}^T \Gamma_Y^{-1} \mathbf{F})^{-1} (\mathbf{F}^T \Gamma_Y^{-1} \gamma_{Y,0} - \mathbf{f}_0)) \end{aligned}$$

Matriz de semivariograma

Pesos KO

$$\omega = \Gamma^{-1} \left(\gamma - \frac{\mathbf{1}^T \Gamma^{-1} \gamma - 1}{\mathbf{1}^T \Gamma^{-1} \mathbf{1}} \mathbf{1} \right)$$

Pesos KU

$\omega_{UK} \doteq$

$$\Gamma_Y^{-1} (\gamma_{Y,0} - \mathbf{F} (\mathbf{F}^T \Gamma_Y^{-1} \mathbf{F})^{-1} (\mathbf{F}^T \Gamma_Y^{-1} \gamma_{Y,0} - \mathbf{f}_0))$$

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E} (Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2$$

Matriz de semivariograma

Pesos KO

$$\omega = \Gamma^{-1} \left(\gamma - \frac{\mathbf{1}^T \Gamma^{-1} \gamma - 1}{\mathbf{1}^T \Gamma^{-1} \mathbf{1}} \mathbf{1} \right)$$

Pesos KU

$$\omega_{UK} \doteq \Gamma_Y^{-1} (\gamma_{Y,0} - \mathbf{F}(\mathbf{F}^T \Gamma_Y^{-1} \mathbf{F})^{-1} (\mathbf{F}^T \Gamma_Y^{-1} \gamma_{Y,0} - \mathbf{f}_0))$$

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E}(Z(\mathbf{s} + \mathbf{h}) - Z(\mathbf{s}))^2$$

$$\Gamma = \begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) \\ \vdots & \ddots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_n - \mathbf{s}_n) \end{pmatrix}$$

$$\gamma_0 = \begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_0) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_0) \\ \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_0) \end{pmatrix}$$

Kriging de un núcleo simple (K1S)

Los pesos ω_i consisten en un sólo núcleo que controla la distancia espacial entre los puntos muestreados y el sitio donde deseamos realizar la predicción, esto es

$$\omega_i^{K1S} \doteq \frac{K_1\left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h}\right)}{\sum_{j=1}^n K_1\left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h}\right)}, \quad (0.2)$$

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

Kriging de un núcleo simple (K1S)

Los pesos ω_i consisten en un sólo núcleo que controla la distancia espacial entre los puntos muestreados y el sitio donde deseamos realizar la predicción, esto es

$$\omega_i^{K1S} \doteq \frac{K_1\left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h}\right)}{\sum_{j=1}^n K_1\left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h}\right)}, \quad (0.2)$$

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

- Ventaja: su simpleza, al involucrar un solo núcleo univariado y con ello, un único parámetro de suavizado.
- Desventaja: 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 Índice de Moran (K2IM)

Además del núcleo que controla las distancias, un segundo núcleo que controla la autocorrelación local medida a través del **Índice de Moran Local**:

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

Kriging de dos núcleos con Índice de Moran (K2IM)

Además del núcleo que controla las distancias, un segundo núcleo que controla la autocorrelación local medida a través del **Índice de Moran Local**:

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

Los pesos en este caso quedan determinados por

$$\omega_i^{K2IM} \doteq \frac{K_1\left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h}\right) K_2\left(\frac{1}{e^{I_i}}\right)}{\sum_{j=1}^n K_1\left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h}\right) K_2\left(\frac{1}{e^{I_j}}\right)}. \quad (0.3)$$

Kriging de dos núcleos con Índice de Moran (K2IM)

Además del núcleo que controla las distancias, un segundo núcleo que controla la autocorrelación local medida a través del **Índice de Moran Local**:

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

Los pesos en este caso quedan determinados por

$$\omega_i^{K2IM} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h} \right) K_2 \left(\frac{1}{e^{I_i}} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h} \right) K_2 \left(\frac{1}{e^{I_j}} \right)}. \quad (0.3)$$

- Objetivo del segundo núcleo: identificar sitios donde la variable de interés se encuentre rodeada por valores similares.
- Un $I_i > 0$, indica que la variable de interés en el sitio \mathbf{s}_i está rodeada por valores similares a ella \rightsquigarrow la transformación $\frac{1}{e^{I_i}}$ toma un valor pequeño haciendo que en ese caso el peso tome un valor grande

Kriging de dos núcleos con medianas (K2ME)

Además del núcleo que controla las distancias, un segundo núcleo que incluye cierta medida de similitud entre vecindarios definida a partir de la mediana

Kriging de dos núcleos con medianas (K2ME)

Además del núcleo que controla las distancias, un segundo núcleo que incluye cierta medida de similitud entre vecindarios definida a partir de la mediana \rightsquigarrow Para cada sitio consideramos la mediana de los valores de la variable de interés en sitios vecinos y luego definimos una distancia basada en la diferencia de dichas medianas. Formalmente, sea \tilde{Z}_s el vector cuyos componentes están dados por $\{Z(\mathbf{s}_i), \mathbf{i} \in V_s\}$ donde V_s es el conjunto de índices \mathbf{i} tal que \mathbf{s}_i es un vecino de \mathbf{s} . Supondremos que el cardinal de V_s es p por lo que $\tilde{Z}_s \in \mathbb{R}^p$. Definimos,

$$d_m(\tilde{Z}_s, \tilde{Z}_{s_0}) \doteq Med \tilde{Z}_s - Med \tilde{Z}_{s_0}, \quad (0.4)$$

donde $Med \tilde{Z}_s$ es la mediana del vector \tilde{Z}_s .

K2ME cont.

De esta manera, los pesos quedan dados por

$$\omega_i^{K2ME} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h_\beta} \right) K_2 \left(\frac{d_m(\tilde{\mathbf{Z}}_{s_i}, \tilde{\mathbf{Z}}_{s_0})}{h_\alpha} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h_\beta} \right) K_2 \left(\frac{d_m(\tilde{\mathbf{Z}}_{s_j}, \tilde{\mathbf{Z}}_{s_0})}{h_\alpha} \right)}. \quad (0.5)$$

K2ME cont.

De esta manera, los pesos quedan dados por

$$\omega_i^{K2ME} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h_\beta} \right) K_2 \left(\frac{d_m(\tilde{Z}_{s_i}, \tilde{Z}_{s_0})}{h_\alpha} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h_\beta} \right) K_2 \left(\frac{d_m(\tilde{Z}_{s_j}, \tilde{Z}_{s_0})}{h_\alpha} \right)}. \quad (0.5)$$

- En este tipo de pesos, los cuales están contruidos por el producto de dos funciones núcleo, los pesos mayores están asociados a observaciones para las cuales ambas distancias son pequeñas.

K2ME cont.

De esta manera, los pesos quedan dados por

$$\omega_i^{K2ME} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h_\beta} \right) K_2 \left(\frac{d_m(\tilde{\mathbf{z}}_{s_i}, \tilde{\mathbf{z}}_{s_0})}{h_\alpha} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h_\beta} \right) K_2 \left(\frac{d_m(\tilde{\mathbf{z}}_{s_j}, \tilde{\mathbf{z}}_{s_0})}{h_\alpha} \right)}. \quad (0.5)$$

- En este tipo de pesos, los cuales están contruidos por el producto de dos funciones núcleo, los pesos mayores están asociados a observaciones para las cuales ambas distancias son pequeñas.
- Sin embargo, un peso pequeño no necesariamente está asociado a una observación para la cual ambas distancias son grandes \rightsquigarrow basta con que solo una de las distancias sea grande (o un parámetro de suavizado pequeño) para que una función núcleo se haga muy pequeña y anule el peso a la observación, sin considerar el valor del otro núcleo \rightsquigarrow aun cuando exista similitud entre los vecindarios, si los mismos están alejados geográficamente, el peso asignado a ese sitio será pequeño.

Kriging de un núcleo Multiplicativo (K1M)

Como el anterior pero considerando un solo núcleo. En esta dirección, introducimos los pesos de núcleo multiplicativo dados por

$$\omega_i^{K1M} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0) * d_m(\tilde{Z}_{s_i}, \tilde{Z}_{s_0})}{h} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0) * d_m(\tilde{Z}_{s_j}, \tilde{Z}_{s_0})}{h} \right)}, \quad (0.6)$$

Kriging de un núcleo Multiplicativo (K1M)

Como el anterior pero considerando un solo núcleo. En esta dirección, introducimos los pesos de núcleo multiplicativo dados por

$$\omega_i^{K1M} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0) * d_m(\tilde{Z}_{\mathbf{s}_i}, \tilde{Z}_{\mathbf{s}_0})}{h} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0) * d_m(\tilde{Z}_{\mathbf{s}_j}, \tilde{Z}_{\mathbf{s}_0})}{h} \right)}, \quad (0.6)$$

- Esta estrategia brinda una ventaja computacional al requerir la búsqueda un solo parámetro de suavizado en contraposición con el de dos núcleos.
- Este predictor tiene un mejor comportamiento en sitios donde la mediana es representativa de la distribución de $Z(\mathbf{s}_0)$ \rightsquigarrow 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.

Kriging de un núcleo Aditivo (K1A)

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

$$\omega_i^{K1A} \doteq \frac{K_1 \left(\frac{d(\mathbf{s}_i, \mathbf{s}_0)}{h_\beta} + \frac{d_m(\tilde{\mathbf{Z}}_{s_i}, \tilde{\mathbf{Z}}_{s_0})}{h_\alpha} \right)}{\sum_{j=1}^n K_1 \left(\frac{d(\mathbf{s}_j, \mathbf{s}_0)}{h_\beta} + \frac{d_m(\tilde{\mathbf{Z}}_{s_j}, \tilde{\mathbf{Z}}_{s_0})}{h_\alpha} \right)}. \quad (0.7)$$

En esta nueva versión es también la distancia más grande la que cuenta con la mayor contribución al peso, haciendo que la función núcleo devuelva valores pequeños y dejando sin influencia el comportamiento de la otra distancia.

Evaluación de los métodos

- Simulaciones (varios escenarios) (50 rep.)
- Predicción de Precios de Vivienda (Saltillo-Mexico)
- Predicción de Precios de Alojamientos (Singapur)

Evaluación de los métodos

- Simulaciones (varios escenarios) (50 rep.)
- Predicción de Precios de Vivienda (Saltillo-Mexico)
- Predicción de Precios de Alojamientos (Singapur)

Estimación de parámetros

- En los Kirging no paramétricos \rightsquigarrow Para el parámetro de suavizado y elección de vecinos más cercanos hacemos k-fold cross validation sobre la muestra de entrenamiento.
- En los Kriging paramétricos (parámetros del variograma teórico) \rightsquigarrow Ajustamo usando el variograma empírico a partir de la muestra de entrenamiento usando modelo esférico, modelo exponencial y sinc (dependiendo de cómo se generan los datos), con mínimos cuadrados pesados

Escenarios de las simulaciones

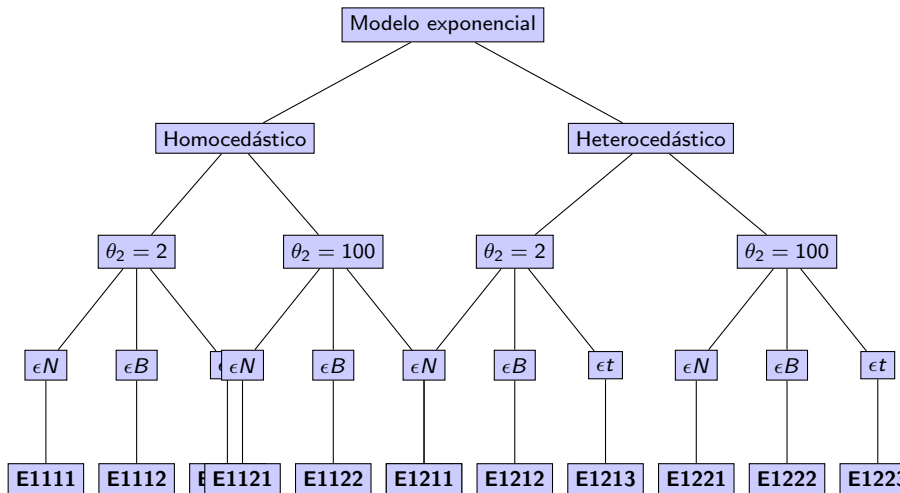


Figura: Combinación de parámetros considerados en cada escenario partiendo del modelo de covarianza exponencial.

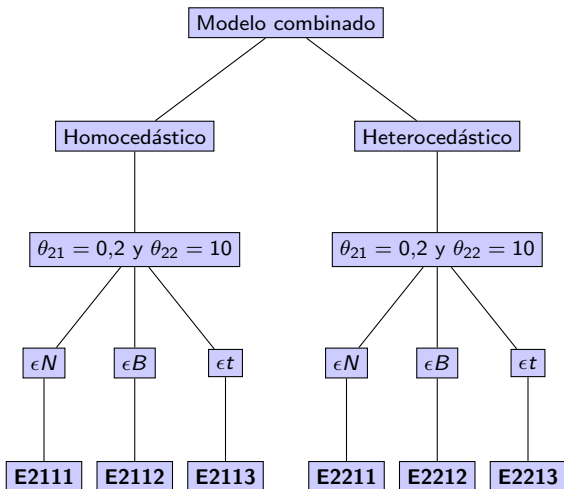


Figura: Combinación de parámetros considerados en cada escenario partiendo del modelo de covarianza combinado.

Resultados de las simus 1

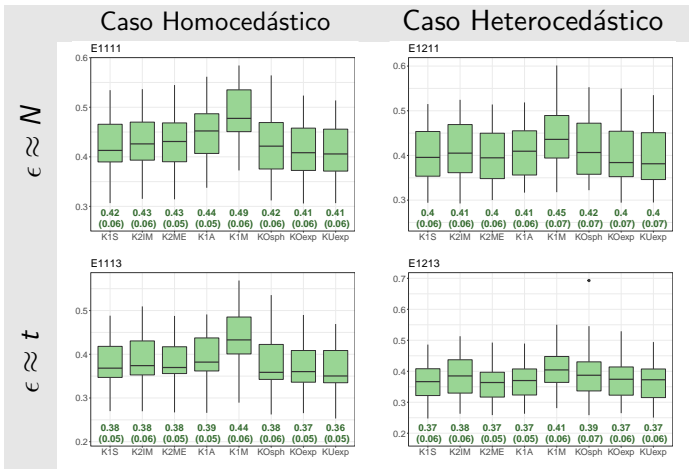


Figura: Para procesos con modelos de covarianza exponencial y rango $\theta_2 = 2$.

Resultados simus 2

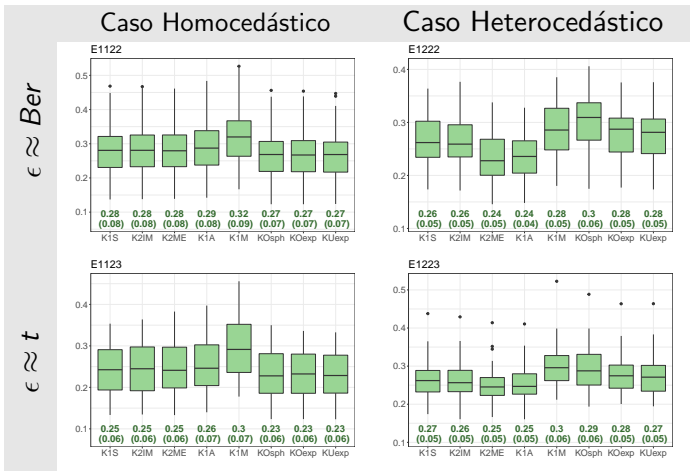


Figura: Modelos de covarianza exponencial y rango $\theta_2 = 100$.

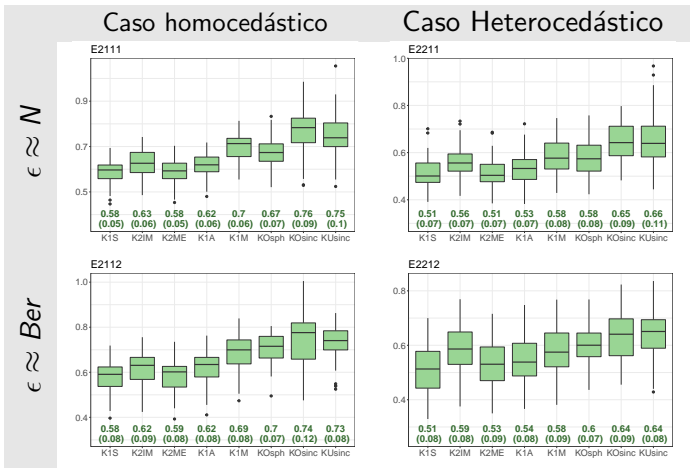


Figura: Distribución del MAE, sobre 50 repeticiones, para procesos con modelos de covarianza combinada y rangos $\theta_{21} = 0,2$ y $\theta_{22} = 10$.

Predicción Precios de Viviendas (Saltillo)

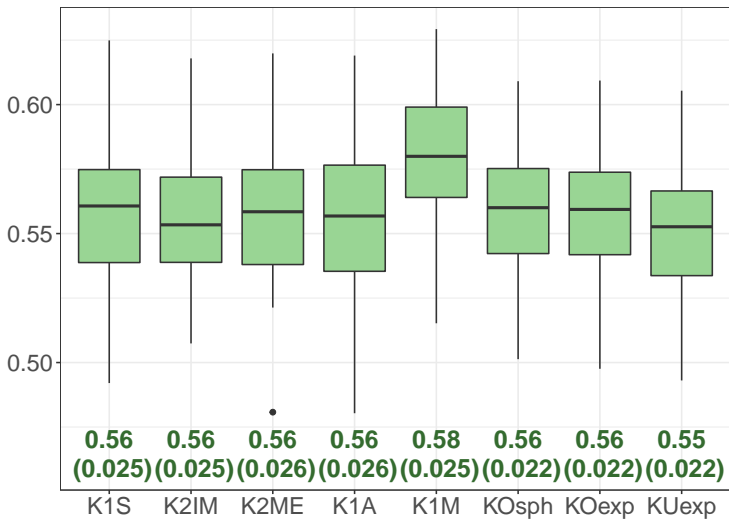


Figura: Distribución del MAE, sobre 50 repeticiones, para los datos Saltillo.

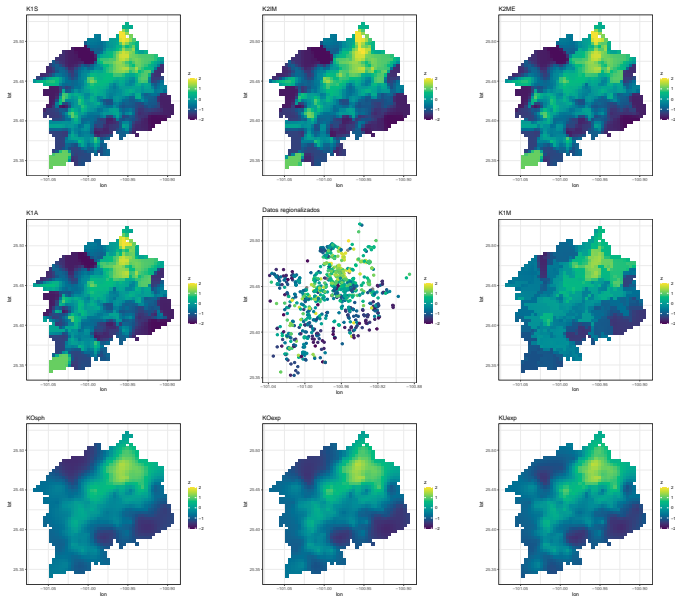


Figura: Datos Saltillo: Mapas predictivos para cada método.

Predicción Precios de Hospedajes (Singapur)

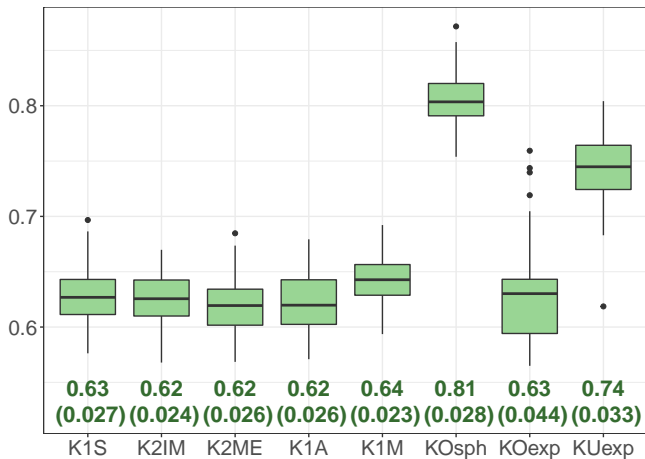


Figura: Distribución del MAE, sobre 50 repeticiones, para los datos de Singapur.

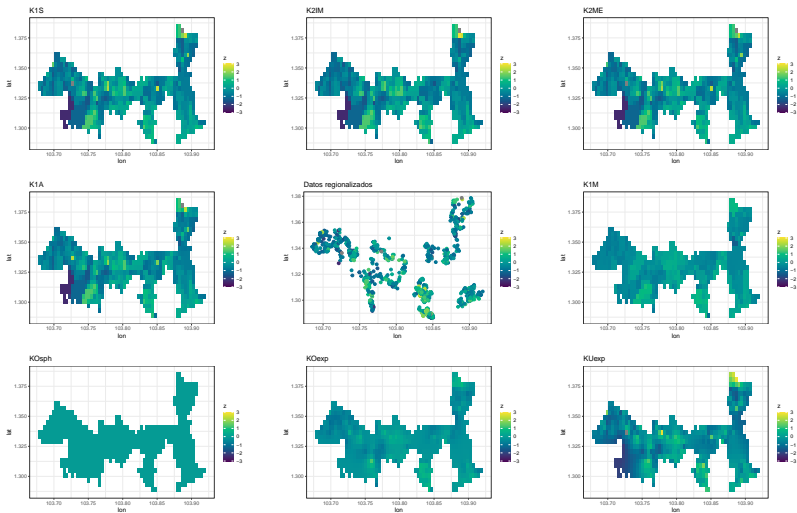


Figura: Datos de Singapur: Mapas predictivos para cada método.

Extensiones y trabajo actual

- Propiedades asintóticas.
- Mejoras para bajar el costo computacional.
- Co-kriging no paramétrico.
- Nuevas aplicaciones sociales y Económicas.

GRACIAS!