



Geostatistical new functions and radial basis functions in R program: Package geospt

Carlos E. Melo

Cadastral and Geodesy Engineering
Faculty of Engineering
District University
Francisco José de Caldas

Oscar O. Melo

Department of Statistics
Faculty of Sciences
National University of Colombia

Sandra E. Melo

Department of Agronomy
Faculty of Agricultural Sciences
National University of Colombia

Abstract

Proponemos un conjunto de funciones diseñadas en el programa R, las cuales permiten realizar un análisis geoestadístico junto con la ayuda de otros paquetes previamente diseñados en R tales como; geoR, gstat y sgeostat, entre otros. Las contribuciones expuestas aquí son: una función para la construcción del variograma de media recortada, una función para la construcción del pocketplot (útil para el análisis de estacionariedad local), una serie de funciones de base radial espaciales (multicuadrática, multicuadrática inversa, spline con tensión, spline capa delgada, exponencial, Gaussiana, completamente regularizada spline) para predecir, optimizar y realizar validación cruzada leave-one-out, una función para diseño de red y una función que genera un resumen de estadísticas obtenidas de la validación cruzada, para evaluar el funcionamiento de los métodos de interpolación (geoestadísticos y determinísticos) construídos en función de los errores. Los fundamentos teóricos se presentan brevemente junto con la implementación mediante algunos ejemplos prácticos.

Keywords: análisis geoestadístico, variograma media recortada, pocketplot, y funciones de base radial.

1. Introduction

Para realizar un análisis geoestadístico es necesario considerar una serie de pasos: un primer paso consiste en analizar la calidad y la cantidad de datos requeridos, es decir el muestreo espacial. El segundo paso es el análisis exploratorio, el cual se basa en el uso de técnicas estadísticas convencionales y en el análisis estructural de los datos, con el objetivo de identificar la presencia de anisotropía o isotropía y la tendencia. En el modelado del variograma, se

evalúa la relación espacial entre los valores de la variable regionalizada, y se ajusta un modelo de variograma al variograma experimental. Una vez el modelo de variograma se encuentra, los valores de predicción se pueden generar usando la interpolación Kriging para la construcción del mapa de predicción de la variable explicada. Sin embargo, hay métodos deterministas, donde el modelo de interpolación no requiere de un modelo de variograma, como es el caso de las RBFs. Después de esto, para elegir el mejor método de interpolación se utiliza la validación cruzada. El último paso consiste en la generación de los mapas de predicciones de la variable regionalizada y de las desviaciones estándar, junto con su interpretación y análisis.

Actualmente, la realización de estos procedimientos es viable gracias a los modernos programas informáticos existentes. Sin embargo, no se puede decir que exista un sólo programa informático que tenga implementadas todas las herramientas geoestadísticas, esto junto con la carencia de funciones en el programa R en cuanto a las funciones de base radial espaciales y espacio-temporales, y con respecto a el pocket plot, entre otras, motiva la realización de la librería expuesta aquí, la cual fue útil en el desarrollo de esta investigación.

En la librería (Melo, Santacruz, and Melo 2015) proponemos una serie de funciones que están diseñadas en el programa R. Estas permiten un análisis geoestadístico más completo junto con la ayuda de paquetes previamente diseñados en R, tales como: **geoR** (Jr and Diggle 2015), **gstat** (Pebesma 2004) y **sgeostat** (original by James J. Majure Iowa State University and port + extensions by Albrecht Gebhardt 2013), entre otros. De esta manera, estas contribuciones son: una función para la construcción del variograma experimental de la media recortada, una función para generar un diseño de red de grillas regulares asociadas a muestreos condicionales, una función para la construcción del pocketplot para datos grillados (útil para el análisis de estacionariedad local), y funciones de base radial (multicuadrática, multicuadrática inversa, spline con tensión, completamente regularizada spline y spline capa delgada) para optimizar, predecir y realizar validación cruzada en el espacio, una función para producir un gráfico que muestra el comportamiento de los parámetros **eta** y **rho** asociados con la función de base radial, y una función que genera una tabla con el resumen de las estadísticas de la validación cruzada para evaluar la exactitud de los métodos de interpolación (geoestadísticos y determinísticos) con base en los errores de predicción. Se describen brevemente algunas de las funciones, y luego se ilustra su funcionamiento con varios ejercicios. El paquete está implementado en el programa (Team 2015) y se encuentra disponible en el Comprehensive R Archive Network (CRAN) en <https://cran.r-project.org/web/packages/geospt>.

Hay por lo menos tres desarrollos para interpoladores que conducen a la misma forma funcional para la interpolación, el spline capa delgada, las funciones de base radial y el método conocido en geoestadística como kriging, todos interrelacionados al ser su función kernel definida positiva. Este artículo se centra en el diseño R funciones para realizar procedimientos geoestadísticos.

La organización del artículo es la siguiente. En la sección 2 se presenta un resumen de los fundamentos teóricos que soportan las funciones implementadas, la sección 3 está dedicada a describir el uso de las funciones implementadas en la librería, la sección 4 considera una aplicación con datos trabajados en Hengl (2007), la sección 5 concluye con un breve resumen de características ya disponibles y otras mejoras que más adelante se implementarán.

2. Métodos geoestadísticos y funciones de base radial

2.1. Variograma media recortada

El variograma experimental $2\hat{\gamma}(\mathbf{h})$ es útil en la interpolación kriging, ya que este interpolador lo considera en su estructura. Un estimador natural basado en el método de momentos, debido a Matheron [Cressie \(1993\)](#), está dado por

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{N(\mathbf{h})} [Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 \quad (1)$$

donde $N(\mathbf{h})$ denota todos aquellos pares $(\mathbf{s}_i, \mathbf{s}_j)$ para los que $\mathbf{s}_i - \mathbf{s}_j = \mathbf{h}$ y $|N(\mathbf{h})|$ denota el cardinal de $N(\mathbf{h})$, $Z(\mathbf{s}_i)$ es el valor muestral en la localización \mathbf{s}_i , y $Z(\mathbf{s}_j)$ es el valor muestral en la localización \mathbf{s}_j . Dicho estimador es generalmente sesgado en presencia de información atípica, afectando el estimador. Un estimador robusto a esa situación es el de la media recortada [Cressie and Hawkins \(1980\)](#), y está dado por

$$2\hat{\gamma}(\mathbf{h}) = \frac{\left[\text{med} \left(|Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{\frac{1}{2}} \right) \right]^4}{0.457 + 0.494/N(\mathbf{h})} \quad (2)$$

Para ajustar el modelo del variograma hay algunos métodos tales como mínimos cuadrados ordinarios (MCO), mínimos cuadrados ponderados (MCP) y máxima verosimilitud restringida (MVR).

2.2. Pocket plot

El Pocket Plot (llamado así debido a su uso en la detección de bolsillos de no estacionariedad) es una técnica necesaria para identificar un área localizada atípica con respecto al modelo de estacionariedad, es construida para aprovechar la naturaleza espacial de los datos a través de las coordenadas de filas y columnas (estes "x" y nortes "y" respectivamente). Para la ilustración de este ejemplo, ver la siguiente Figura 1

En geoestadística se pretende estimar las relaciones espaciales entre los datos de los puntos (modelamiento del variograma). Luego este estimado es usado para el desarrollo del método kriging y para estimar la variabilidad del predictor. Aunque el estimador de ([Cressie and Hawkins 1980](#)), ofrece una estimación robusta para el variograma, hay aun una fracción de las diferencias $(Z_i - Z_j)$, que resulta ser inapropiada en la estimación del variograma de Cressie. Las ubicaciones sobre la grilla que exhiben diferentes medidas del resto se deben identificar. Estos bolsillos de no estacionariedad, una vez descubiertos, pueden ser removidos de la estimación del variograma, pero naturalmente eventualmente deben ser modelados e incorporados en las apreciaciones finales del recurso analizado. El Pocket Plot "Gráfico de Bolsillo", es una simple idea que se ilustrará sobre las diferencias norte-sur de los datos de coal-ash¹ concentrados sobre la fila j de la grilla, para alguna otra fila, k por ejemplo, hay un cierto número N_{jk} , de diferencias de datos definidas, cuyas localizaciones están a una distancia $h = |j - k|$ en la dirección norte-sur. Sea \bar{Y}_{jk} la media de estas *diferencias*^{1/2}, promediadas sobre los N_{jk} términos, y se define:

$$\bar{\bar{Y}}_h = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} |Z_i - Z_j|^{1/2} \quad (3)$$

¹Este registro de datos del porcentaje de coal ash encontrado en muestras mineras originalmente reportadas por ([Gomez and Hazen 1970](#)) y posteriormente utilizado ([Cressie 1993](#)). Los datos se pueden descargar de la librería `gstat` o `sp` del programa R

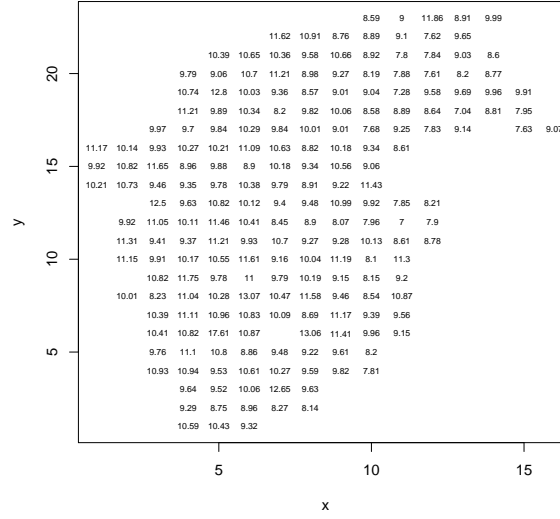


Figure 1: Ubicación espacial de una muestra de cenizas de carbón (coal-ash), las unidades están en % en ubicaciones reorientadas (Cressie, 1993)

$\bar{\bar{Y}}_h$ es una media ponderada de los \bar{Y}_{jks} tales que $|j - k| = h$. Luego se define

$$P_{jk} = \bar{Y}_{jk} - \bar{\bar{Y}}_h \quad (4)$$

($P_{jk} : k = 1, 2, \dots$), es la contribución del residual de la fila j , al estimador del variograma en la diferencia de rezagos. Idealmente, estos puntos serán repartidos a ambos lados del cero, pero si hay algo inusual en la fila j , entonces se dará una singular contribución en todos los rezagos y típicamente mostrará una dispersión de puntos por encima del nivel cero. Ahora la fila j varía y dispersa los puntos, lo que constituye el diagrama de bolsillo, ilustrado en la Figura 2, donde la parte central de la dispersión se sustituye por la caja de un diagrama de caja (Velleman and Hoaglin 1981, chap. 3)

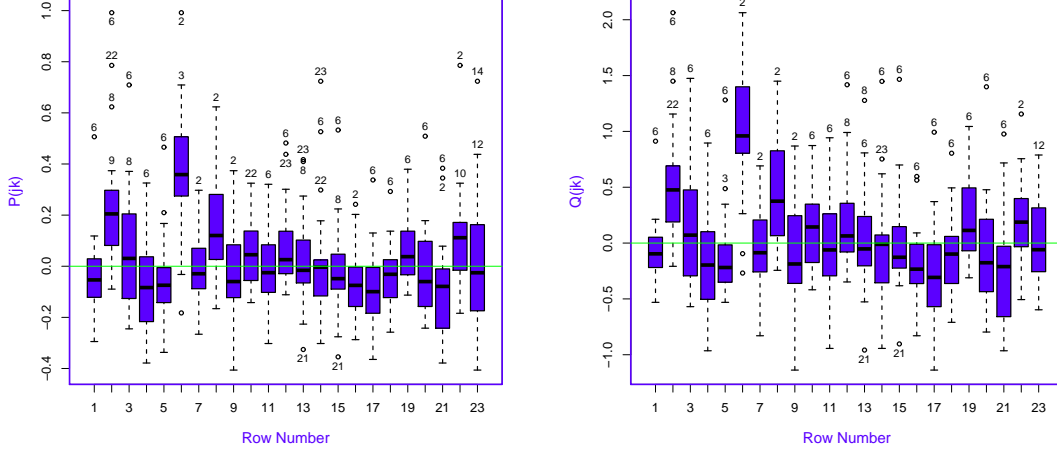
Una modificación adicional del pocket plot sería graficar valores normalizados de P_{jk} , el gráfico puede ser obtenido a partir de:

$$Q_{jk} = N_{jk}^{1/2} \left\{ \left(\frac{\bar{Y}_{jk}}{\bar{\bar{Y}}_h} \right) - 1 \right\} \quad (5)$$

A partir de los resultados de Cressie (1985), $var(P_{jk}) = (2\gamma(h))^{(1/2)} / N_{jk}$ la cual justifica el cálculo de Q_{jk} . Este cambio sólo afectará la diseminación de los puntos y no el panorama general de estar por encima del nivel cero Cressie (1993).

2.3. Radial basis function

En interpolación espacial, hay métodos que no requieren información del modelo de dependencia espacial tales como el variograma o el covariograma, estos se denominan determinísticos y son del interés en esta sección.



(a) Probabilidad: Unidades sobre el eje vertical
están en $(\%coalash)^{1/2}$

(b) Varianzas estandarizadas.

Figure 2: POCKET-PLOT en dirección sur-norte.

Dadas las localizaciones $\mathbf{s}_1, \dots, \mathbf{s}_n$ en un espacio Euclideo n -dimensional y los datos observados $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$, la forma general del interpolador de las funciones de base radial es

$$Z^*(\mathbf{s}) = \sum_{i=1}^n \omega_i \phi(\mathbf{s} - \mathbf{s}_i) + \sum_{k=0}^p \nu_k f_k(\mathbf{s}) \quad (6)$$

donde f_k ; $k = 0, \dots, p$, son las funciones linealmente independientes como en el estimador kriging, $Z(\mathbf{s})$ es la función a ser interpolada, \mathbf{s}_i , $i = 1, \dots, n$ son las localizaciones y ϕ es la función kernel que satisface apropiadamente la condición de definida positiva. En el caso donde el interpolador es exacto, e.d., $Z^*(\mathbf{s}) = Z(\mathbf{s})$ para $i = 1, \dots, n$, Micchelli ha mostrado que los coeficientes son determinados, dada la condición en g y en $f_k(\mathbf{s})$. Además la prueba de conjetura de Franke, (Micchelli 1986) demuestra, las condiciones suficientes para garantizar la no singularidad del método cuando se utilizan otras funciones de base radial. El predictor asociado con (6) es una combinación lineal de las funciones de base

$$\widehat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \omega_i(\mathbf{s}) \phi(\mathbf{s}_i - \mathbf{s}_0) + \sum_{k=0}^p \nu_k f_k(\mathbf{s}) \quad (7)$$

donde $\phi(\mathbf{s}_i - \mathbf{s}_0)$ es una función de base radial obtenida a partir de la distancia Euclidea entre la localización de predicción \mathbf{s}_0 y cada una de las localizaciones \mathbf{s}_i . Los parámetros ν y ω pueden ser estimados por mínimos cuadrados penalizados, minimizando la siguiente expresión

$$\sum_{i=1}^n [Z(\mathbf{s}_i) - g(\mathbf{s}_i)]^2 + \rho \int_{\mathbb{R}^2} J_m(g(\mathbf{s})) d\mathbf{s} \quad (8)$$

donde $J_m(g(\mathbf{s}))$ es una medida de la rugosidad de la función spline g (definida en términos

de las m -ésimas derivadas de g), $g(\mathbf{s}_i)$ es una función de valor-real, dada por

$$g(\mathbf{s}_i) = \sum_{l=0}^k \nu_l f_l(\mathbf{s}_i) + \sum_{j=1}^n \omega_j \phi(\mathbf{s}_i - \mathbf{s}_j), \quad i = 1, \dots, n$$

y $\rho > 0$ actúa como un parámetro de suavizamiento. Los parámetros ν_k , están asociados a la tendencia, la cual usualmente se expresa en términos de las coordenadas "x" y "y". Los parámetros se encuentran al resolver el siguiente sistema de ecuaciones

$$\begin{pmatrix} \Phi + \rho I & \mathbf{f} \\ \mathbf{f}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \mathbf{Z} \\ 0 \end{pmatrix} \quad (9)$$

Si no hay tendencia \mathbf{f} viene a ser un vector de unos y $\boldsymbol{\nu}$ un parámetro de sesgo. Un estimador alternativo esta dado por

$$Z^*(s) = \sum_{i=1}^n \lambda_i(s) Z(s_i) \quad (10)$$

Luego el predictor como una combinación lineal de $Z(\mathbf{s}_i)$'s, esta dado por: $\widehat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i(\mathbf{s}) Z(\mathbf{s}_i)$, al establecer la condición

$$\sum_{i=1}^n \lambda_i f_l(\mathbf{s}_i) = \boldsymbol{\lambda}' \mathbf{f} = f_l(\mathbf{s}_0), \quad l = 0, \dots, k$$

donde $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$, $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ y $\mathbf{f} = (f_l(\mathbf{s}_1), \dots, f_l(\mathbf{s}_n))'$.

y resolver por minimos cuadrados ordinarios se llega al siguiente sistema matricial,

$$\begin{pmatrix} \Phi + \rho I & \mathbf{f} \\ \mathbf{f}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \phi_0 \\ \mathbf{f}(\mathbf{s}_0) \end{pmatrix} \quad (11)$$

Resolviendo el sistema, los coeficientes para $\boldsymbol{\lambda}$ y $\boldsymbol{\alpha}$ están dados por

$$\begin{aligned} \widehat{\boldsymbol{\lambda}}' &= \left\{ \phi_0 + \mathbf{f}' (\Phi + \rho I)^{-1} \mathbf{f} \right\}^{-1} \left[\mathbf{f}(\mathbf{s}_0) - \mathbf{f}' (\Phi + \rho I)^{-1} \phi_0 \right] (\Phi + \rho I)^{-1} \\ \widehat{\boldsymbol{\alpha}} &= - \left[\mathbf{f}' (\Phi + \rho I)^{-1} \mathbf{f} \right]^{-1} \left[\mathbf{f}(\mathbf{s}_0) - \mathbf{f}' (\Phi + \rho I)^{-1} \phi_0 \right] \end{aligned} \quad (12)$$

donde Φ es una matrix con el i, j -ésimo elemento $\phi(\mathbf{s} - \mathbf{s}_i)$ para el i, j -ésimo par de datos, $\mathbf{1}$ es un vector columna de unos, $\boldsymbol{\alpha}$ es el vector de multiplicadores de Lagrange de tamaño $k \times 1$, y ϕ_0 corresponde a la RBF evaluada entre los vecinos y el puntos donde se desea predecir, e.d. $\phi(\mathbf{s} - \mathbf{s}_0)$. Este predictor tiene la ventaja que muestra los pesos para cada dato, que dependen del número de vecinos y el parámetro de suavizado principalmente.

Los interpoladores (6) y (10) representan dos enfoques diferentes para el mismo problema y en algunos casos espaciales son formalmente equivalentes. Tal como se señala en (Matheron 1981), a través de un análisis Bayesiano (Kimeldorf and Wahba 1970), a través de demostraciones de conexión entre kriging y splines (Cressie 1989), (Hutchinson and Gessler 1994) y (Kent and Mardia 1994).

Hay un enlace entre los métodos espaciales splines y kriging, el cual fue llamado equivalentemente "cercano" porque el TPS (una clase de spline) corresponde a una covarianza generalizada específica, mientras que el estimador kriging y el interpolador RBF sólo requieren el uso de un kernel con propiedades adecuadas como la de definida positiva. En general, esto permite adaptar la función kernel a un conjunto de datos particular (Cressie 1989; Myers 1992). La mayor diferencia es que el usuario establece el parámetro de suavizamiento en los splines, mientras en el caso de kriging, el suavizamiento se determina de forma objetiva.

Definición 1: Una función $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ es llamada radial, si existe una función invariante $\Phi(s) = [0, \infty)$ tal que:

$$\Phi(s) = \phi(r), \text{ donde } r = \|s\|$$

y $\|\cdot\|$ es la norma Euclidiana \mathbb{R}^d . Esto significa que el valor de la función Φ en el punto $s \in \mathbb{R}^d$ únicamente depende de la norma de s .

Multicuadrática (MQ)

(Hardy 1990) llamo multicuadrático al método porque considero que la característica principal es la de ser una superposición de superficies cuádricas.

Definición 2: Dado un conjunto de n puntos distintos $\{x_i\}_{i=1}^n \in \mathbb{R}^d$ y sus correspondientes valores escalares $\{f_i\}_{i=1}^n \in \mathbb{R}$, el interpolador multicuadrático de los datos tiene la siguiente forma:

$$p(s_i) = \sum_{j=1}^n \mathbf{b}_j \sqrt{\eta_j^2 + \delta_i^2}, \quad j = 1, \dots, n$$

donde los coeficientes \mathbf{b}_j se determinan mediante la imposición de las condiciones de interpolación $p(s_i) = f_i$, para $i = 1, \dots, n$, y η_j es un parámetro de suavizado. De aquí se obtiene el siguiente sistema de ecuaciones lineales y simétricas

$$\Phi \mathbf{b} = \mathbf{f}$$

donde $\mathbf{b} = (\mathbf{b}_1, \dots, \mathbf{b}_n)'$ y las entradas de Φ están dadas por $\phi_{ij} = \sqrt{\eta^2 + \delta_{ij}^2}$.

Multicuadrática Inversa (IM)

Una variación de la función multicuadrática fue introducida por Hardy and Gopfert (1975), y esta dada por

$$\phi(\delta) = 1/\sqrt{\eta^2 + \delta^2}$$

donde $\eta \neq 0$ es un parámetro de suavizado de libre escogencia. (Franke 1982) encuentra que esta función de base radial puede proporcionar excelentes aproximaciones, incluso cuando el número de centros (vecinos más cercanos) es pequeño.

Spline con Tensión (ST)

Esta función esta dada por la expresión

$$\phi(\delta) = \ln(\eta \cdot \delta/2) + K_0(\eta \cdot \delta) + C_E$$

donde $K_0(x)$ es la función modificada de Bessel ([Abramowitz and Stegun 1965](#), pág. 374) y $C_E = -\int_0^\infty (\ln(x)/e^x)dx = 0.5772161$ es la constante de Euler ([Abramowitz and Stegun 1965](#), pág. 255).

Spline Capa Delgada (TPS)

Este spline fue introducido en el diseño geométrico por ([Duchon 1976](#)). El nombre, spline capa delgada, se refiere a una analogía física que implica la flexión de una hoja delgada de metal. Más tarde ([Thiébaux and Pedder 1987](#)) describió la TPS como un spline cúbico de 2 dimensiones (superficie). En el caso de un espacio Euclidiano con $d = 2$, la TPS tendrá la siguiente forma:

$$\phi(\delta) = \begin{cases} (\eta \cdot \delta)^2 \log(\eta \cdot \delta) & \text{si } \delta \neq 0 \text{ y } \eta > 0 \\ 0 & \text{si } \delta = 0 \end{cases}$$

Completamente regularizada spline (CRS)

Una variante de la TPS que usa la función base spline regularizada, se denomina CRS y viene dada por

$$\phi(\delta) = -\sum_{k=1}^{\infty} \frac{(-1)^k (\eta \cdot \delta)^{2k}}{k!k} = \ln(\eta \cdot \delta/2)^2 + E_1(\eta \cdot \delta/2)^2 + C_E$$

donde $E_1(\cdot)$ es la función integral exponencial ([Abramowitz and Stegun 1965](#), pág. 227) y C_E es la constante de Euler definida anteriormente.

Exponencial (EXPON)

La función de base radial útil en la interpolación espacial es la exponencial que permite evitar los puntos de inflexión y contiene splines cúbicos como un caso especial [Späth \(1969\)](#). La función esta viene dada por

$$\phi(\delta) = e^{-\eta\delta}$$

donde $\eta \neq 0$ es el parámetro de suavizado.

Esta función en interpolación espacial se caracteriza por no generar predicciones fuera del rango muestral.

Gaussiana (GAU)

Una de las funciones de base radial más popular, junto con la TPS, es la Gaussiana. ([Schagen 1979](#)) fue el primero en usar la Gaussiana como función de base radial. Esta función esta dada por

$$\phi(\delta) = e^{-\eta\delta^2}$$

donde $\eta \neq 0$ es el parámetro de suavizado de libre elección.

El parámetro de suavizamiento óptimo η , el cual es un parámetro de libre elección, se encuentra al minimizar la raíz del cuadrado medio del error de predicción (RMSPE) haciendo

uso de la validación cruzada. Algunas descripciones adicionales de RBFs y sus relaciones con los splines y kriging se pueden encontrar en (Bishop 1995, p. 164), (Chilès and Delfiner 1999, pag. 272) y (Cressie 1993, pag. 180).

Los splines han demostrado ser muy adecuados para la interpolación de alturas densamente muestreadas y variables climáticas (Hutchinson 1995; Mitáš and Mitášová 1999). Sin embargo, su mayor crítica es la incapacidad para incorporar información auxiliar para modelar la parte determinista de la variación (Hengl 2007). Las RBFs se recomiendan utilizar para calcular superficies suavizadas de un gran número de datos. Las funciones producen buenos resultados para variaciones superficiales suaves tales como la elevación (Johnston, Ver, Krivoruchko, and Lucas 2001).

2.4. Resumen validación cruzada

La validación cruzada leave-one-out (LOOCV, por sus siglas en inglés) consiste en excluir la observación de uno de los n puntos muestrales (por lo general asociados a un vecindario), y con los $n - 1$ valores restantes y la función de base radial con sus parámetros η y ρ , predecir vía splines el valor de la variable en estudio en la ubicación del punto que se excluyó. Si la función de base radial presenta un buen funcionamiento, entonces la diferencia entre el valor observado y el valor predicho debe ser pequeña y se podrá producir el mapa. Este procedimiento se realiza en forma secuencial con cada uno de los puntos muestrales y así se obtiene un conjunto de n errores de predicción. Entonces el conjunto de predicciones se puede resumir por alguna medida de precisión de la predicción. La idea se remonta por lo menos a Mosteller and Wallace (1963); Stone (1974).

La función `criterio.cv` de validación cruzada presentada aquí funciona también para el caso de interpolación vía kriging. Por consiguiente, se considerará también la desviación estándar del kriging. Sea $\hat{Z}_{[i]}(\mathbf{s}_i)$ el valor predicho a partir de la validación cruzada, y sea $\hat{\sigma}_{[i]}(\mathbf{s}_i)$ la predicción para la desviación estándar en la localización \mathbf{s}_i , con estos estadísticos se construyen la media de los errores de predicción, la raíz media del cuadrado de los errores de predicción ($RMSPE$), error estándar promedio del kriging de los errores de predicción ($ASEPE$), la media estandarizada de los errores de predicción ($MSPE$), la raíz media estandarizada del cuadrado de los errores predichos ($RMS SPE$) y el coeficiente de determinación (R^2). A continuación se presentan dichos estadísticos, los cuales son útiles para evaluar la capacidad predictiva del método (RBF o kriging)

$$\begin{aligned}
MPE &= \frac{\sum_{i=1}^n (\hat{Z}_{[i]}(s_i) - Z(s_i))}{n} \\
RMSP E &= \sqrt{\frac{\sum_{i=1}^n (\hat{Z}_{[i]}(s_i) - Z(s_i))^2}{n}} \\
ASEPE &= \frac{\sum_{i=1}^n \hat{\sigma}_{[i]}(s_i)}{n} \\
MSPE &= \frac{\sum_{i=1}^n \left((\hat{Z}_{[i]}(s_i) - Z(s_i)) / \hat{\sigma}_{[i]}(s_i) \right)^2}{n} \\
RMSSPE &= \sqrt{\frac{\sum_{i=1}^n \left((\hat{Z}_{[i]}(s_i) - Z(s_i)) / \hat{\sigma}_{[i]}(s_i) \right)^2}{n}} \\
R^2 &= 1 - \frac{\sum_{i=1}^n (\hat{Z}_{[i]}(s_i) - Z(s_i))^2}{\sum_{i=1}^n (Z(s_i) - \bar{Z}(s_i))^2}
\end{aligned} \tag{13}$$

Donde $\hat{Z}_{[i]}(s_i)$ es el valor de la predicción y $\hat{\sigma}_{[i]}(s_i)$ es la desviación estándar, obtenidos a partir de la validación cruzada. Y $Z(s_i)$ es el valor muestreado en la localización s_i .

Una variación de la metodología previa consiste en dividir la muestra en dos submuestras; la primera submuestra es usada para modelar el variograma, y la otra submuestra es utilizada para validar el método kriging. Después de eso, las medidas de validación se pueden construir a partir de los valores observados y pronosticados (Bivand, Pebesma, and Rubio 2008). Si todo va bien, el $RMSP E$ debería ser lo más pequeño posible (cercano a cero) y R^2 debe estar cerca de 1.

3. Implementation in R

En esta sección se enfatiza en el uso de las funciones de la librería asociados con los conceptos teóricos definidos en la sección previa.

3.1. Funciones geoestadísticas

variograma media recortada

Para el caso de la media recortada se programo la función 2, modificando la suma de la formula de Cressie-Hawkins por la media recortada, en esta propuesta el usuario puede escoger el porcentaje de recorte. En el caso en que el porcentaje sea del 50%, este estimador coincidirá con el de la mediana, el cual es más robusto frente a la presencia de datos atípicos, mientras que si el porcentaje de recorte es del 0%, el estimador coincidirá con el estimador robusto de Cressie-Hawkins. En Bárdossy (2001) se compara el estimador clásico, robusto y de la media recortada (con un 10% de recorte), se considera un atípico, ver abajo. Se encuentra que el

estimador de la media recortada arroja mejores resultados ante la presencia de datos atípicos y por ende es más robusto, resultados similares se muestran por medio de simulaciones en Roustant, Dupuy, and Helbert (2007).

La función `est.variograms()` esta construida a partir de la función `est.variogram()` del paquete `sgeostat` en <https://cran.r-project.org/web/packages/sgeostat>, se implementa la media recortada en su funcionamiento. Como ejemplo se considera la base de datos `maas` del paquete `sgeostat`, especificando un recorte del 10% a continuación:

```
R> library(sgeostat)
R> data(maas)
R> maas.point <- point(maas)
R> maas.pair <- pair(maas.point, num.lags=24, maxdist=2000)
R> maas.v <- est.variograms(maas.point,maas.pair,'zinc',trim=0.1)
R> maas.v
```

El resultado es el siguiente:

	lags	bins	classic	robust	med	trimmed.mean	n
1	1	41.66667	101947.2	65465.76	36286.13	57015.22	31
2	2	125.00000	113158.9	61238.92	33444.66	51991.43	184
3	3	208.33333	143501.3	79790.82	53728.38	67770.61	279
4	4	291.66667	177257.6	101478.44	63406.79	86754.46	336
5	5	375.00000	239373.8	144476.65	103685.85	125286.53	367
6	6	458.33333	233764.5	145387.50	115946.06	125355.24	404
7	7	541.66667	273382.4	194285.17	186095.48	177289.00	421
8	8	625.00000	280300.4	197139.93	215218.63	180371.19	441
9	9	708.33333	308830.8	227925.27	273564.52	207709.69	455
10	10	791.66667	297263.4	225228.13	240608.52	210802.15	447
11	11	875.00000	337402.5	250439.56	276672.91	230168.09	461
12	12	958.33333	321287.9	226290.79	246422.02	199083.61	433
13	13	1041.66667	342465.0	252177.03	262795.80	229030.66	417
14	14	1125.00000	371965.3	289594.79	303591.84	271317.58	387
15	15	1208.33333	309236.5	232539.63	234756.15	212280.02	386
16	16	1291.66667	315844.0	239704.08	238300.05	217875.01	360
17	17	1375.00000	347594.5	239448.38	246261.11	210848.17	343
18	18	1458.33333	300932.6	226781.23	226889.51	203460.52	354
19	19	1541.66667	290834.7	210952.98	183415.61	190246.32	330
20	20	1625.00000	260444.7	197217.81	163738.82	174456.98	327
21	21	1708.33333	315371.1	228165.97	206878.84	205701.77	319
22	22	1791.66667	270525.7	198176.63	163732.14	181498.03	323
23	23	1875.00000	255374.6	174233.92	147363.74	155691.27	288
24	24	1958.33333	275440.4	193038.79	168454.22	171184.29	277

pocket.plot() function

Para este caso, consideraremos la base de datos `coalash` mencionada en la sección 2.2. La función requiere el nombre del `data.frame`, el tipo de gráfico asociado con la probabilidad o

la varianza estandarizada del pocket plot en las direcciones sur-norte o este-oeste; pocketplot de probabilidades por fila, es decir, horizontal "sur-norte" "PPR", pocketplot de probabilidades por columnas, es decir, vertical "este-oeste" "PPC", pocketplot de varianza por filas, es decir horizontal "sur-norte" "PVR" y pocketplot de varianzas por columnas, es decir vertical "este-oeste" "PVC", las coordenadas "X" y "Y", el nombre de la variable a analizar "Z", y la identificación de los atípicos (automática "F" o personal "T"). En la Figura 2 claramente las filas 2, 6, y 8 son atípicas, esto sirve como verificación de que estas filas son potencialmente problemáticas. El siguiente código en R, describe la situación de un análisis de estacionariedad local en probabilidades del % ceniza de carbón en dirección sur-norte:

```
R > library(gstat)
R > data(coalash)
R > pocket.plot(coalash, "PPR", coalash$x, coalash$y, coalash$coalash)
```

El resultado obtenido se muestra en la Figura 2a y el asociado con varianzas estándar se muestra en la Figura 2b.

cross validation summary

Para generar el resumen de estadísticas de la validación cruzada, proponemos la función `criterio.cv()`, Esta genera un data frame con los estadísticos presentados en la expresión 13, obtenidos por validación cruzada leave-one-out. Para que funcione correctamente, ingresamos un `data.frame` con las coordenadas de los datos, columnas de predicción y varianza de la predicción de las muestras, los valores observados, los residuales, el zscore (residual dividido por el error estandar del kriging) y grupo. En el caso de usar la función `rbf.tcv`, la varianza de predicción y el zscore tendran NA's. A continuación se muestra un corto ejemplo del funcionamiento de `criterio.cv()` con la base de datos *meuse* de la librería *gstat*:

```
R > library(gstat)
R > data(meuse)
R > coordinates(meuse) <- ~x+y
R > m <- vgm(.59, "Sph", 874, .04)

R > # leave-one-out cross validation:
R > out <- krige.cv(log(zinc)~1, meuse, m, nmax = 40)
R > criterio.cv(out)
```

El resultado obtenido se muestra a continuación

	MPE	ASEPE	RMSPE	MSPE	RMSSPE	MAPPE	CCPE	R2	pseudoR2
1	0.006674145	0.4188814	0.3873933	0.01150903	0.924489	0.04821387	0.8428837	0.7101429	0.7104529

3.2. Uso funciones RBF

El funcionamiento presentado en esta sección con el paquete de funciones RBF, considera una base de datos empírica de la precipitación denominada `preci` de la librería *geospt*. La función `rbf()` es construida a partir de las expresiones (6) y (7) de la subsección 2.3. Esta expresión requiere; el factor de suavizamiento η y de robustez ρ (se recomienda encontrar los

valores optimos con la función `graph.rbf`), la formula $z \sim 1$ (sin tendencia) o $z \sim x + y$ (con tendencia), las coordenadas planas de la muestra `coordinates` utilizadas en la predicción, las coordenadas del punto a predecir, o data frame de nuevos puntos a predecir `newdata`, el número de vecinos `n.neigh`, este ultimo si se llega a desear un determinado tamaño de vecindario, también es necesario especificar la RBF dada `func` (ver Tabla 1). El código de uso para esta función es el siguiente:

```
library(geospt)
R > data(preci)
R > coordinates(preci) <- ~x+y
# prediction case: a grid of points
R > puntos<-expand.grid(x=seq(min(preci$x),max(preci$x),0.05),
                        y=seq(min(preci$y),max(preci$y),0.05))
R > coordinates(puntos) <- ~x+y
R > pred.rbf <- rbf(prec~x+y, preci, eta=0.1460814, rho=0, newdata=puntos,
                  n.neigh=10, func="TPS")
R > coordinates(pred.rbf) = c("x", "y")
R > gridded(pred.rbf) <- TRUE

# show prediction map
R > spplot(pred.rbf["var1.pred"], cuts=40, col.regions=bpy.colors(100),
R >      main = "", key.space=list(space="right", cex=0.8))
```

Previamente se optimizo sólo η con la función `graph.rbf`, obteniendo un valor de $\eta = 0.1460814$ con la RBF "TPS" y el resultado se muestra a continuación:

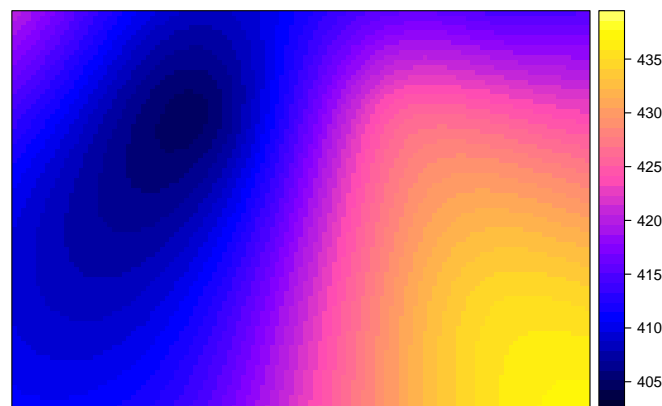


Figure 3: Superficie de interpolación TPS de la precipitación

La función `rbf.cv()` requiere de las siguientes entradas; el factor de suavizamiento `eta`, el parámetro de robustez `ρ`, la formula descripta previamente, las coordenadas de la muestra `coordinates`, el tamaño del vecindario `n.neigh`, y la RBF dada `func`. El ejemplo se muestra a continuación:

```
library(geospt)
R > data(preci)
R > coordinates(preci)<~x+y
R > rbf.cv(prec~1, preci, eta=0.2589, rho=0, n.neigh=9, func="M")
```

Con lo cual se obtiene un $RMSPE = 7.691789$.

La función `rbf.phi()` requiere; una distancia entre un par de puntos s_1 y s_2 , un factor de suavizamiento `eta` y una RBF `func`. El siguiente ejemplo produce un valor igual a 42.88812:

```
R > data(preci)
R > d1 <- dist(rbind(preci[1,],preci[2,]))
R > RBF.phi(distance=d1, eta=0.5, func="TPS")
```

La función `rbf.tcv()` requiere; una formula, unos parámetros `eta` y `rho`, un tamaño de vecindario `n.neigh`, y una RBF `func`, ya explicadas anteriormente. El resultado de esta función es necesario para la construcción del resumen de estadísticos de la validación cruzada leave-one-out (LOOCV). A continuación se presenta un ejemplo:

```
R > data(preci)
R > coordinates(preci)<~x+y
R > rbf.tcv(prec~x+y, preci, eta=0.1460814, rho=0, n.neigh=9, func="TPS")
```

El resultado obtenido es

	var1.pred	var1.var	observed	residual	zscore	fold	x	y
1	420.9332	NA	420	-0.93315414	NA	1	1	1
2	410.6660	NA	410	-0.66604423	NA	2	2	1
3	401.7670	NA	405	3.23295070	NA	3	3	3
4	413.9099	NA	415	1.09013829	NA	4	4	3
5	429.5202	NA	430	0.47981416	NA	5	5	5
6	424.9259	NA	425	0.07408364	NA	6	6	5
7	415.2948	NA	415	-0.29484136	NA	7	7	6
8	437.2620	NA	435	-2.26200498	NA	8	8	6
9	425.1257	NA	425	-0.12570136	NA	9	9	6
10	429.7477	NA	430	0.25225005	NA	10	10	7

Como se puede observar se generan valores NA en las columnas `var1.var` y `zscore`, esto se debe a que en las RBF no se tienen varianzas para las predicciones de los errores, lo cual no ocurre en los krigeados.

La función `graph.rbf()` permite ver el comportamiento de los parámetros de suavizamiento y robustez η y ρ respectivamente, a la vez que encuentra los valores óptimos en caso que el usuario lo pida. Al encontrar los valores óptimos de dichos parámetros y al definir un vecindario de un determinado tamaño en una RBF específica podemos encontrar una buena superficie de interpolación. En esta función se debe especificar; la formula, las coordenadas de la muestra `coordinates`, el tamaño del vecindario `n.neigh`, la RBF dada `func` descriptas previamente y los operadores lógicos `eta.opt` y `rho.opt` que indicarán si se optimizan o no

(TRUE o FALSE), el limite máximo para los parámetros η y ρ , `eta.dmax` y `rho.dmax` respectivamente, los valores iniciales para buscar los valores optimos de los parámetros definidos en el vector `x0`, el número de iteraciones deseado en la busqueda de los valores optimos de los parámetros dado por la opción `iter`, y el operador lógico `P.T` que permitirá visualizar la tabla con los valores asociados al gráfico junto con los valores optimos obtenidos. A continuacion se muestra el gráfico obtenido con 5 RBF en un determinado intervalo para el parámetro $0 \leq \eta \leq 1.6$.

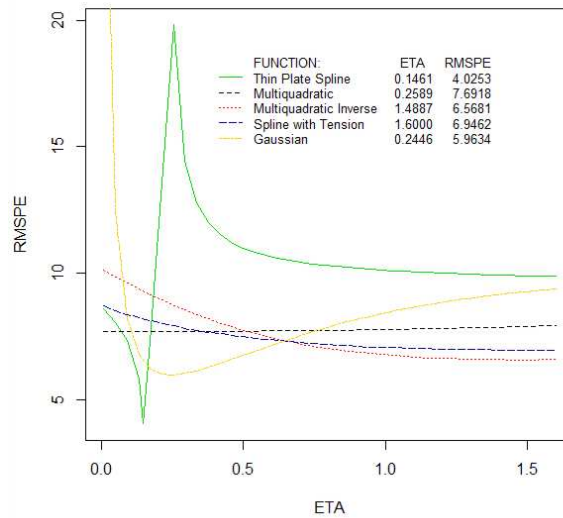


Figure 4: Optimización de η , en funciones de base radial

Como se puede ver en la Figura 4, el *RMSPE* más bajo es generado con el spline capa delgada. El código sólo para la función Gaussiana es:

```
R > data(preci)
R > coordinates(preci) ~ x+y
R > graph.rbf(preci ~ 1, prec, eta.opt=TRUE, rho.opt=FALSE, n.neigh=9, func=
  "GAU", np=40, eta.dmax=1.6, P.T=TRUE)
```

La Tabla 1 provee una breve descripción de las funciones implementadas en la librería **geospt**.

4. Aplicación

A continuación se presenta una aplicación en donde se ponen a prueba las funciones diseñadas para la interpolación espacial con las RBFs.

4.1. Base de datos

La temperatura media terrestre diaria en Croacia fue estudiada por Hengl (2009). La información es tomada de <http://spatial-analyst.net/book/HRclim2008> y fue proporcionada

Función	Descripción
<code>criterio.cv()</code>	Genera un data frame de un resumen de valores estadísticos asociados con la validación cruzada: <i>MPE</i> , <i>ASEPE</i> , <i>RMSPE</i> , <i>MSPE</i> , <i>RMSSPE</i> , <i>MAPPE</i> , <i>CCPE</i> , R^2 y <i>pseudoR2</i> .
<code>network.design</code>	Genera un promedio con los errores estándar del kriging de los errores de predicción (<i>ASEPE</i>) asociado a un diseño de red condicional.
<code>est.variograms()</code>	Calcula el variograma experimental: clásico, robusto, mediana y media recortada. (Cressie 1993; Bárdossy 2001)
<code>pocketplot()</code>	Gráfica el pocketplot de probabilidad o de varianzas estandarizadas en las direcciones norte-sur o este-oeste, (ver Cressie (1993))
<code>rbf()</code>	Genera predicciones individuales a partir de las siguientes funciones de base radial: completamente regularizada spline (<i>CRS</i>), spline con tensión (<i>ST</i>), Gaussiana (<i>GAU</i>), exponencial (<i>EXPON</i>), trigonométrica (<i>TRI</i>) spline capa delgada (<i>TPS</i>), multicuadrática inversa (<i>IM</i>) y multicuadrática (<i>M</i>)
<code>rbf.cv()</code>	Extrae un valor <i>RMSPE</i> , resultado de la validación cruzada leave-one-out
<code>rbf.cv1()</code>	Retorna un valor <i>RMSPE</i> , resultado de LOOCV (permite optimizar los parámetros <i>eta</i> y <i>eta</i> simultáneamente).
<code>RBF.phi()</code>	Produce un valor numérico obtenido a partir de la función de base radial, generado con una determinada distancia entre dos puntos, el valor del parámetro de suavizamiento <i>eta</i> , y una función: "CRS", "ST", "GAU", "EXPON", "TRI", "TPS", "IM" or "M"
<code>rbf.tcv()</code>	Genera una tabla las coordenadas de los datos, las predicciones, los valores observados, los residuales, las varianzas de la predicciones, los zscore (residual dividido por el error estándar) de la variable analizada, los cuales son resultado de LOOCV
<code>graph.rbf()</code>	Produce un gráfico que describe el comportamiento de los parámetros <i>eta</i> y <i>rho</i> , asociados con la función de base radial.

Table 1: Funciones de la librería **geospt**

por Melita Perčec Tadić, de la Organización Meteorológica y de Servicios Hidrológicos Croatas (Hengl 2009). Croacia es un país relativamente pequeño, pero cuenta con varias regiones de clima diferente que son el resultado de su posición específica en el mar Adriático y de la topografía muy diversa que va desde las llanuras en el este, a través de una zona central montañosa que separa el territorio continental de la parte marítima del país. La región de estudio se caracteriza por una amplia gama de características topográficas y climáticas, lo que permite evaluar correctamente la metodología propuesta con respecto a la tradicional, ya que las temperaturas promedio de la tierra en tal región se ven fuertemente influenciadas por la topografía.

La temperatura media fue medida desde enero hasta diciembre de 2008. En la mayoría de las estaciones meteorológicas, la temperatura se mide tres veces al día, a las 7 am, 1 pm y 9 pm, y la media de la temperatura diaria (ΔT en un día) se calcula como un promedio ponderado, de acuerdo a la siguiente expresión

$$\Delta T = \frac{T_{(7am)} + T_{(1pm)} + 2 \cdot T_{(9pm)}}{4}$$

Luego la temperatura media mensual se obtiene de la media diaria mencionada anteriormente.

Teniendo en cuenta que se dispone de una composición de imágenes (imágenes MODIS de 1 km de resolución, de 8 días, dispuestas al público) de la temperatura media diaria, es decir, de 3 a 4 registros mensuales. Las mediciones de temperatura se recogen automáticamente en 159 estaciones meteorológicas. Como cuatro estaciones meteorológicas no tenían registros disponibles para enero y dado que en algunas estaciones habian datos faltantes (o perdidos), entonces dichas estaciones se retiraron del análisis. Por lo tanto, se consideran sólo 146 estaciones y se calculo con los valores observados removiendo los datos faltantes (o perdidos) la temperatura media mensual. En esta aplicación se analiza la temperatura media mensual terrestre en Croacia sólo para el mes de enero de 2008 a partir de 146 estaciones meteorológicas.

Las coordenadas geográficas (latitud y longitud) fueron transformadas a un sistema de coordenadas cartesianas (x, y) . La ubicación de las 146 estaciones meteorológicas se muestra en la Figura 6a.

4.2. Mapa de predicción

Inicialmente se realiza la optimización de los parámetros de suavizamiento η y ρ a partir de los datos previamente descritos. Para esto, se trabaja con la función `graph.rbf()`, esta función se evalúa en las funciones de base radial mencionadas en 2.3, encontrando uno de los mejores ajustes en la función M , con parámetros $\eta = 1e - 05$ y $\rho = 0$ ($RMSPE = 1.23022$). Esta función trabaja internamente con las funciones `rbf.cv()` y `rbf.cv1()`, las cuales emplean `optimize()` (para un parámetro) y `bobyqa()` (para los dos parámetros) respectivamente. En la optimización no se utilizó la función `optim()`, descrita en Mittelhammer, Judge, and Miller (2000) ya que demandaba más tiempo que `optimize()`. Para optimizar los dos parámetros es necesario que `eta.opt=TRUE` y `rho.opt=TRUE` así se obtendrá un gráfico tipo lattice (para el caso Figura 5),

La función `optimize()` del programa R descrita en Brent (1973), busca el valor óptimo del parámetro η o ρ generando un valor de $RMSPE$ a partir de la función `rbf.cv()` (para el caso `bobyqa()` converge rápidamente en la iteración 19), en un intervalo para dichos parámetros establecidos por el usuario, por ejemplo para funciones multicuadráticas los óptimos para η y ρ suelen encontrarse cerca a 0. Por lo cual, un buen intervalo sería entre 0 y 1 tal como lo muestra la Figura 5.

Los datos y el shapefile son precargados, seguidamente se genera una grilla de 70000 puntos dentro de la región analizada, con el fin de generar predicciones de la temperatura media terrestre. Esta grilla es obtenida usando la función `spsample()` del paquete `sp`. Las predicciones son generadas con la función `rbf()`, esta función requiere el valor de los parámetros "`eta`" y ρ , la formula que define la tendencia o modelo: $z \sim 1$ para casos sin tendencia y $z \sim x + y$ para tendencia lineal, las coordenadas de los puntos muestreados `coordinates`, las coordenadas de los nuevos puntos `newdata`, el número de vecinos `n.neigh`, y el tipo de función de base radial `func`. Estas predicciones son luego convertidas a un objeto de clases `SpatialPixelsDataFrame` y `sp`, con la instrucción `coordinates()`, del paquete `sp`, y finalmente con la función `spplot()` se obtiene el mapa de las predicciones de la variable analizada, el cual se muestra en la Figura 6b.

```
R> library(geosptdb)
R> data(croatia)
R> data(croatiadb)
```

```

R> croatia.jan <- croatiadb[croatiadb$t==1,c(1:2,4)]
R> coordinates(croatia.jan) <- ~x+y
R> rbf.cv(MTEMP~1, croatia.jan, eta=1e-05, rho=0, n.neigh=10, func="M")
1.23022

R> graph.rbf(MTEMP~1, croatia.jan, eta.opt=T, rho.opt=T, n.neigh=10, func="M",
eta.dmax=2, rho.dmax=2, iter=80)

# prediction case a grid of points
R> pts <- spsample(croatia, n=70000, type="regular")
R> pred.rbf <- rbf(MTEMP~1, croatia.jan, eta=1e-05, rho=0, newdata= pts,
n.neigh=10, func="M")
R> coordinates(pred.rbf) = c("x", "y")
R> gridded(pred.rbf) <- TRUE
R> spplot(pred.rbf["var1.pred"], cuts=40, scales = list(draw =T), col.regions
=bpy.colors(100), key.space=list(space="right", cex=0.8))

```

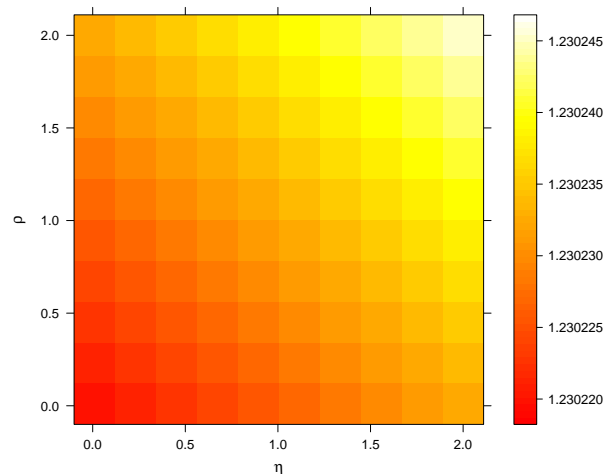
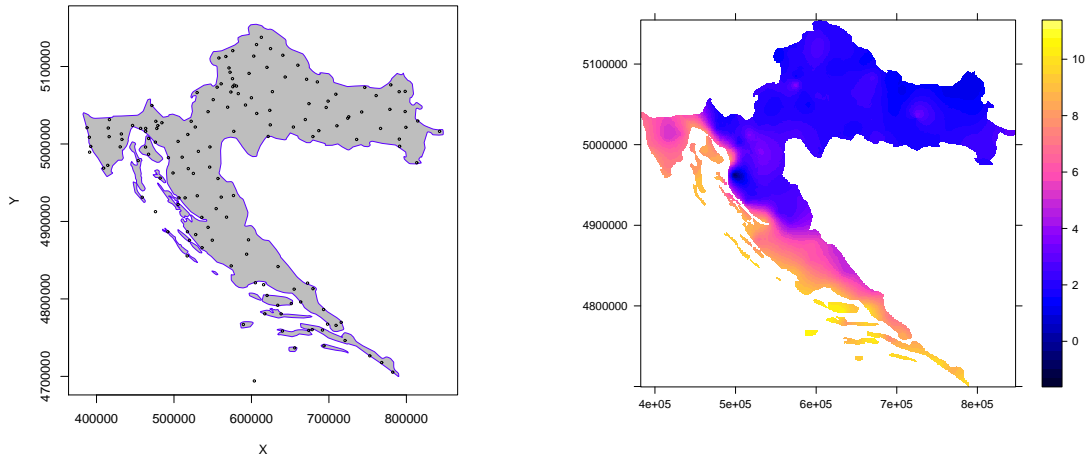


Figure 5: Lattice de η y ρ para función multicuadrática

5. Conclusiones

Las funciones contenidas en la librería **geospt** estan programadas en código R y en general producen resultados en un tiempo corto, sin embargo algunas pueden programarse en C con el fin de obtener resultados mas rapidos.

La última versión 1.0-1 de **geospt**, permite realizar optimización para los parámetros η y ρ en el caso del uso de las funciones de base radial para la obtención de superficies de interpolación,



(a) Localización de las estaciones meteorológicas (b) Predicción del promedio de la temperatura mensual
Enero 2008 (función multicuadrática)

Figure 6: Mapa de Croacia

lo cual implica una alta demanda en tiempo computacional. Por lo cual, se recomienda el uso de procesadores modernos y de alta capacidad o limitar el número de iteraciones con el fin de obtener resultados en el menor tiempo posible. La instrucción `graph.rbf` para cuando `eta.opt` y `rho.opt` son iguales a `TRUE` permite obtener un grafico Lattice con las diferentes combinaciones entre los valores de los parámetros en un rango predefinido por el usuario para dichos parámetros.

Una gran ventaja de las funciones programadas en cuanto a la interpolación espacial con las RBFs, es que estas permiten considerar la tendencia, lo cual en algunas aplicaciones como es el caso del manejo de variables ambientales permite obtener mejores resultados con relación a cuando se omite la tendencia en la interpolación a pesar de la no estacionariedad del proceso observado.

6. Acknowledgments

Core Spatial Data Research (Faculty of Engineering, District University Francisco José de Caldas) and Applied Statistics in Experimental Research, Industry and Biotechnology (National University of Colombia).

References

- Abramowitz M, Stegun IA (1965). *Handbook of Mathematical Functions*. Dover, New York.
- Bárdossy A (2001). *Introduction to Geostatistics*. Institut für Wasserbau der Universität Stuttgart.

- Bishop CM (1995). *Neural Networks for Pattern Recognition*. Oxford Press, Oxford.
- Bivand R, Pebesma E, Rubio V (2008). *Applied Spatial Data Analysis with R*. Springer, New York.
- Brent R (1973). *Algorithms for minimization without derivatives*. Prentice-Hall, Englewood Cliffs.
- Chilès JP, Delfiner P (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New York.
- Cressie N (1985). “Fitting variogram models by weighted least squares.” *Journal of the International Association for Mathematical Geology*, **17**, 563–586.
- Cressie N (1989). “Geostatistics.” *The American Statistician*, **43**, 197–202.
- Cressie N (1993). *Statistics for Spatial Data*. Revised Edition. John Wiley & Sons Inc., New York.
- Cressie N, Hawkins DM (1980). “Robust estimation of the variogram.” *Mathematical Geology*, **12**, 115–125.
- Duchon J (1976). “Interpolation des Fonctions de Deux Variables Suivant le Principe de la Flexion des Plaques Minces.” *Rairo Analyse Numerique*, **10**, 5–12.
- Franke R (1982). “Smooth Interpolation of Scattered Data by Local Thin Plate Splines.” *Computer and Mathematics with Applications*, **8**, 273–281.
- Gomez M, Hazen K (1970). “Evaluating sulfur and ash distribution in coal seams by statistical response surface regression analysis.” *Technical report*, U.S. Bureau of Mines Report RI 7377.
- Hardy R, Gopfert W (1975). “Least squares prediction of gravity anomalies, geoidal undulations, and detections of the vertical with multiquadric harmonic functions.” *Geophysical Research Letters*, **2**, 423–426.
- Hardy RL (1990). “Theory and applications of the multiquadric-biharmonic method. 20 Years of discovery 1968-1988.” *Computers & Mathematics with Applications*, **19**, 163–208.
- Hengl T (2007). “A Practical Guide to Geostatistical Mapping of Environmental Variables.” *Technical report*, JRC Scientific and Technical Research series, Office for Official Publications of the European Communities Luxembourg, EUR 22904 EN.
- Hengl T (2009). *A Practical Guide to Geostatistical Mapping*. 2nd edition. University of Amsterdam, Amsterdam.
- Hutchinson MF (1995). “Interpolating mean rainfall using thin plate smoothing splines.” *International Journal of Geographical Information Systems*, **9**, 385–403.
- Hutchinson MF, Gessler PE (1994). “Splines Ü more than just a smooth interpolator.” *Geoderma*, **62**, 45–67.

- Johnston K, Ver J, Krivoruchko K, Lucas N (2001). *Using ArcGIS Geostatistical Analysis*. ESRI.
- Jr PJR, Diggle PJ (2015). *geoR: Analysis of Geostatistical Data*. R package version 1.7-5.1, URL <http://CRAN.R-project.org/package=geoR>.
- Kent JT, Mardia KV (1994). “The link between kriging and thin-plate splines.” *Probability, Statistics and Optimization*, Wiley, pp. 325–339.
- Kimeldorf G, Wahba G (1970). “A correspondence between Bayesian estimation of stochastic processes and smoothing by splines.” *Annals of Mathematical Statistics*, **41**, 495–502.
- Matheron G (1981). “Splines and kriging: Their formal equivalence. In Down-to-Earth Statistics: Solutions Looking for Geological Problems.” *D. F. Merriam, ed. Syracuse University Geological Contributions*, pp. 77–95.
- Melo C, Santacruz A, Melo O (2015). *geospt: An R package for spatial statistics*. R package version 1.0-1, URL <https://CRAN.R-project.org/package=geospt>.
- Michelli C (1986). “Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Functions.” *Constructive Approximation*, **2**, 11–22.
- Mitáš L, Mitášová H (1999). *Geographical Information Systems: Principles, Techniques, Management and Applications*, volume 1, chapter Spatial interpolation, pp. 481–492. Wiley.
- Mittelhammer R, Judge G, Miller D (2000). *Econometric Foundations*. New York.
- Mosteller F, Wallace DL (1963). “Inference in an Authorship Problem.” *Journal of the American Statistical Association*, **58**(302), 275–309. doi:10.1080/01621459.1963.10500849.
- Myers D (1992). “Kriging, Cokriging, Radial Basic Functions and The Role of Positive Definiteness.” *Computers Mathematical Application*, **24**, 139–148.
- original by James J Majure Iowa State University S, port + extensions by Albrecht Gebhardt R (2013). *sgeostat: An Object-oriented Framework for Geostatistical Modeling in S+*. R package version 1.0-25, URL <http://CRAN.R-project.org/package=sgeostat>.
- Pebesma EJ (2004). “Multivariable geostatistics in S: the gstat package.” *Computers & Geosciences*, **30**, 683–691.
- Roustant O, Dupuy D, Helbert C (2007). “Robust Estimation of the Variogram in Computer Experiments.” *Technical report*, Ecole des Mines, Département 3MI, 158 Cours Fauriel, 42023 Saint-Etienne, France.
- Schagen IP (1979). “Interpolation in two dimensions: a new technique.” *Journal of the Institute of Mathematics and its Applications*, **23**, 53–59.
- Späh H (1969). “Exponential spline interpolation.” *Computing*, **4**, 225–233.
- Stone M (1974). “Cross-Validatory Choice and Assessment of Statistical Predictions.” *Journal of the Royal Statistical Society. Series B (Methodological)*, **36**(2), 111–147. ISSN 00359246.

Team RDC (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>.

Thiébaux H, Pedder M (1987). *Spatial Objective Analysis: With Applications in Atmospheric Science*. Academic Press. London.

Velleman PF, Hoaglin DC (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury, Boston, MA.

Affiliation:

Carlos E. Melo

Faculty of Engineering, Cadastral and Geodesy Engineering

District University Francisco José de Caldas

11001 Bogotá D.C., Colombia

E-mail: cmelo@udistrital.edu.co

Telephone: +57 /1/3239300

Oscar O. Melo

Department of Statistics, Faculty of Sciences

National University of Colombia

11001 Bogotá D.C., Colombia

E-mail: oomelom@unal.edu.co

Telephone: +57 /1/3165000/ext 13169

Sandra E. Melo

Department of Agronomy, Faculty of Agricultural Sciences

National University of Colombia

11001 Bogotá D.C., Colombia

E-mail: semelo@unal.edu.co

Telephone: +57 /1/3165000/ext 19120