Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

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

Carlos E. Melo
Cadastral and Geodesy Engineering
Faculty of Engineering
Universidad Distrital
Francisco José de Caldas

Sandra E. Melo
Department of Agronomy
Faculty of Agricultural Sciences
Universidad Nacional
de Colombia

Oscar O. Melo
Department of Statistics
Faculty of Sciences
Universidad Nacional
de Colombia

Abstract

We propose a set of functions designed on the software R, which permit carrying out a geostatistical analysis with the support of other packages previously designed on R such as geoR, gstat and sgeostat. The package geospt supplies tools to construct the trimmed mean variogram and to make the pocket plot that is useful for the local stationarity analysis. A series of radial basis spatial functions aimed to predict, to optimize and to realize leave-one-out cross-validation are presented. These spatial functions are: multiquadratic, inverse multiquadratic, tension splines, thin plate splines, exponential, Gaussian and completely regularized spline. Furthermore, we propose a function for net designing and a function that generates a summary of the resulting statistics from cross-validation to evaluate how works the interpolation methods (geostatistical and deterministic), which are constructed in function of the errors. The fundamental theory is briefly presented together with its implementation through some practical examples.

Keywords: geostatistical analysis, trimmed mean variogram, pocket plot, and radial basis functions.

1. Introduction

Implementing a geostatistical analysis requires consideration of a sequence of steps. The first step, as in almost any study based on data, is to examine the quality and quantity of the required data, i.e. the spatial sampling. The second step is to perform an exploratory analysis, which is based on conventional statistical techniques and the structural analysis of the data, aimed at identifying directional influences and global trends. In the variogram modeling, the spatial relationship between the values of the regionalized variable is evaluated,

and a variogram model is fitted to the experimental variogram. Once the variogram model is found, the forecast map can be generated using kriging interpolation for the construction of the dependent variable predictions. Nonetheless, there are some deterministic methods, such as radial basis functions (RBFs), where the interpolation model does not require a variogram model. In order to choose the best interpolation method, the cross-validation is computed. The last step consists of the generation, interpretation and analysis of the prediction maps of the regionalized variable and the standard deviation.

Nowadays, the realization of these procedures is feasible thanks to the modern software available. However, it is impossible to affirm that there is one single software that includes all the geostatistical tools; this fact, along with the lack of tools in the software R of spatial and spatio-temporal RBFs, and the use of the pocket plot, motivate the realization of the library hereby exposed, which was useful during the development of this research.

In the library **geospt** (Melo, Santacruz, and Melo 2012), we propose a set of functions designed with the software R. These functions allow a more complete geostatistical analysis with the support of packages previously designed on R like **geoR** (Ribeiro and Diggle 2001; Ribeiro Jr, Diggle, Schlather, Bivand, and Ripley 2020), **gstat** (Pebesma 2004; Gräler, Pebesma, and Heuvelink 2016) and **sgeostat** (original by James J. Majure Iowa State University and port + extensions by Albrecht Gebhardt 2016), among others. In this regard, our contributions are: a function for the trimmed mean experimental variogram construction, a function to generate the design of the regular grid net associated with conditional samples, a function for the pocket plot construction for gridded data that is useful for local stationarity analysis, a function to generate a chart which shows the parameter's behavior of `eta` and `rho` associated to the radial basis function, a function that generates a table with the summary of the cross validation statistics to evaluate the exactitude of the interpolation (geostatistical and deterministic) methods based on the prediction errors, and radial basis functions (multiquadratic, inverse multiquadratic, tension spline, completely regularized spline, and thin plate splines) to optimize, to predict and to perform the cross validation in the space. Some of the RBFs are briefly described, and then they are illustrated by several exercises. The package is implemented in the software R (R Core Team 2020), which is available on the Comprehensive R Archive Network (CRAN) in `https://cran.r-project.org/web/packages/geospt`.

There are at least two developments for interpolators that lead to the same functional form in interpolation: RBFs and the method known in geostatistics as kriging. All of them are interrelated by their positive definite kernel function. This paper focuses on the R design of functions for performing geostatistical processes.

The plan of the paper is the following. Section 2 presents a summary of the fundamental theory that supports the implemented functions. Section 3 is dedicated to describe the usage of the functions implemented in the library. Section 4 considers an application of the processed data in Hengl (2007). Finally, Section 5 concludes with a brief summary of the available features and other improvements that will be implemented in another version.

2. Geostatistical methods and radial basis functions

2.1. Trimmed mean variogram

The experimental variogram, $2\widehat{\gamma}(\boldsymbol{h})$, is useful in the kriging interpolation because this inter-

polator considers it in its structure. A natural estimator based on the momentum method, due to Matheron (Cressie 1993), is given by

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N_{\boldsymbol{h}}|} \sum_{N(\boldsymbol{h})} \left[Z\left(\boldsymbol{s}_i\right) - Z\left(\boldsymbol{s}_j\right)\right]^2 \tag{1}$$

where $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$ and $N_{\boldsymbol{h}}$ represents the number of distinct elements of $N(\boldsymbol{h})$, and $Z\left(\boldsymbol{s}_i\right)$ and $Z\left(\boldsymbol{s}_j\right)$ are the observed values at the locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, respectively. This estimator is generally biased in the presence of atypical data, affecting the estimator. In this situation, a robust estimator for gaussian independent called trimmed mean variogram is (Cressie and Hawkins 1980)

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{\left[\frac{1}{|N_{\boldsymbol{h}}|} \sum\limits_{N(\boldsymbol{h})} \left|Z\left(\boldsymbol{s}_i\right) - Z\left(\boldsymbol{s}_j\right)\right|^{\frac{1}{2}}\right]^4}{0.457 + 0.494/|N(\boldsymbol{h})|} \tag{2}$$

To fit the variogram model, there are some methods such as ordinary least squares (OLS), weighted least squares (WLS), and restricted maximum likelihood (REML).

2.2. Pocket plot

The pocket plot whose name comes from the detection of non-stationarity pockets is a technique necessary to identify an atypical localized area with respect to the assumption of stationarity. It is built to take advantage on the spatial nature of the data through the coordinates of rows and columns (east "x" and north "y", respectively). To illustrate this example, please check the Figure 1.

In geostatistics, it is pretended to estimate the spatial relations between point data (variogram model). Thereupon, the estimated relations are used to compute the kriging method and to estimate the predictor variability. Although the Cressie and Hawkins (1980) estimator offers a robust approach to variogram estimation, there is a fragment of the differences, $Z(\boldsymbol{s}_i + t\boldsymbol{e}) - Z(\boldsymbol{s}_i)$, which results inappropriate in the Cressie's variogram estimation. The locations on the grid with different measurements must be identified. These non-stationarity pockets, once they have been revealed, can be removed from the estimation of the variogram. However, they eventually should be modeled and incorporated for the final assessments of the analyzed resource. The pocket plot is a simple idea which illustrates the north-south differences of the coal-ash data[1] concentrated in the j row from the grid. For any other row, k say, there is a specific number (N_{jk}) of data differences defined, located at a distance $t = \mid j - k \mid$ on the north-south direction. Let $\overline{Y}_{jk}$ be the mean value of these $\mid$ differences $\mid^{1/2}$, averaged over the N_{jk} terms, and let $\overline{\overline{Y}}_t$ be a weighted mean of the $\overline{Y}_{jk}$s, which is given by

$$\overline{\overline{Y}}_t = \frac{1}{N_{\boldsymbol{h}}} \sum_{N(\boldsymbol{h})} \mid Z(\boldsymbol{s}_i + t\boldsymbol{e}) - Z(\boldsymbol{s}_i) \mid^{1/2} .$$

where $\boldsymbol{e}$ is the unit vector in the north-south direction. Then define

$$P_{jk} = \overline{Y}_{jk} - \overline{\overline{Y}}_t, \quad k = 1, 2, \ldots,$$

[1] This information concerning the registered coal-ash percentage found in mining samples originally reported by Gomez and Hazen (1970) and later used by Cressie (1993). Data can be downloaded from the library **gstat** or **sp** from the software R

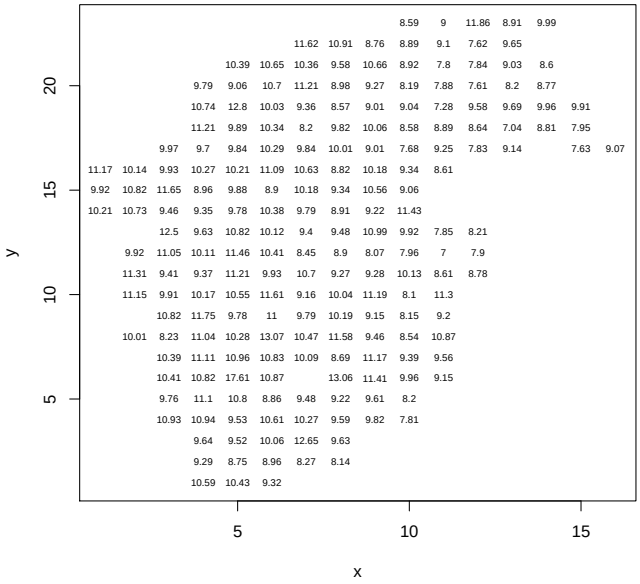


Figure 1: Core measurements (in % coal ash) at reoriented locations (Cressie 1993)

the residual contribution of row j to the variogram estimator at differing lags. Ideally, these points will be spread on both sides of the zero level. However, if there is something unusual about row j, an unusual contribution will be given at all lags and usually, there will be a point dispersion over the zero level. In such a way, the row j varies and spreads the points, which constitutes the pocket plot, as seen in Figure 2, where the central part of the scatter is substituted by the box of a box plot (Velleman and Hoaglin 1981, Chap. 3).

An additional modification of the pocket plot would consist to plot normalized values of P_{jk}, the plot could be obtained from

$$Q_{jk}{=}N_{jk}^{1/2}\left\{\left(\frac{\overline{Y}_{jk}}{\overline{\overline{Y}}_t}\right)-1\right\}. \tag{3}$$

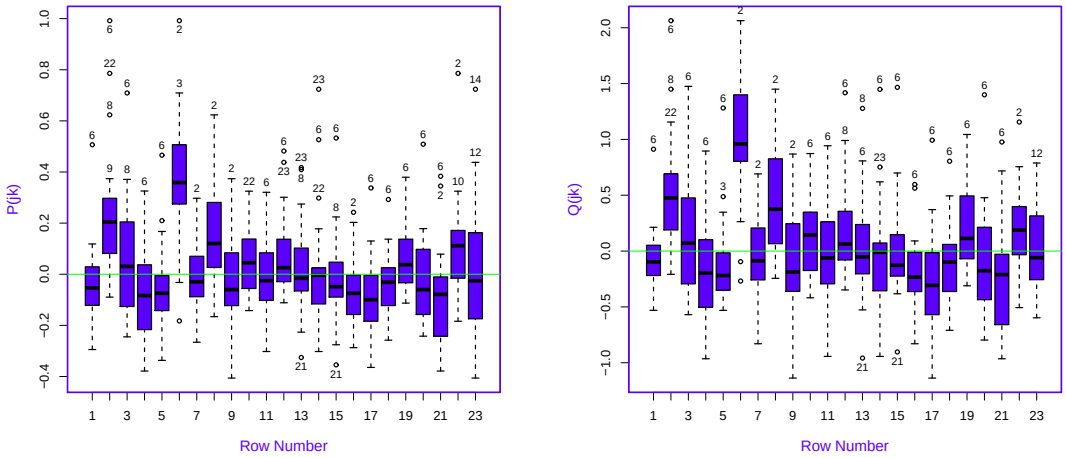
From the results obtained in Cressie (1985), $\text{var}(P_{jk}) = (2\gamma(\boldsymbol{h}))^{1/2}/N_{jk}$, which justifies computing Q_{jk}. This change will only affect the spread of the points and not the overall picture of being above the zero level (Cressie 1993).

2.3. Radial basis function

In spatial interpolation, there exist methods that do not require information from a spatial dependence model such as the variogram or covariogram, which are called deterministic and are the ones of interest in this section. The model can be expressed as

$$Z(\boldsymbol{s}_i) = g(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i), \qquad i = 1, \ldots, n \tag{4}$$

where $\varepsilon(\boldsymbol{s}_i)$ is the unobserved noise term which may or may not be included depending on the application, and it is assumed to be normal mutually independent with mean 0 and



(a) Probability: Units on the vertical axis are in (% coal ash)$^{1/2}$

(b) Standardized variances.

Figure 2: Pocket plot in the north-south direction.

variance τ^2. Given the location $\boldsymbol{s}_i, ..., \boldsymbol{s}_n$ in an Euclidean n-dimensional space, each $g(\boldsymbol{s}_i)$ is a real-valued function given by

$$\begin{aligned} g(\boldsymbol{s}_i) &= \sum_{i'=1}^{n} \omega_{i'} \phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}) + \sum_{k=0}^{p} \nu_k f_k(\boldsymbol{s}_i) \\ &= \boldsymbol{\phi}^t(\boldsymbol{s}_i)\boldsymbol{\omega}_s + \boldsymbol{f}^t(\boldsymbol{s}_i)\boldsymbol{\nu}_s \end{aligned} \tag{5}$$

where $Z^*(\boldsymbol{s}_i)$ is the function to be interpolated, $\boldsymbol{\phi}(\boldsymbol{s}_i) = (\phi(\boldsymbol{s}_i - \boldsymbol{s}_1), \ldots, \phi(\boldsymbol{s}_i - \boldsymbol{s}_n))^t$ with $\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'})$ a basis function, $\boldsymbol{\omega}_s = (\omega_1, \ldots, \omega_n)^t$ with each ω_i an unknown weight, $\boldsymbol{f}(\boldsymbol{s}_i) = (f_0(\boldsymbol{s}_i), f_1(\boldsymbol{s}_i), \ldots, f_p(\boldsymbol{s}_i))^t$ with each $f_k(\boldsymbol{s}_i)$ is a linearly independent real-valued function in the location $\boldsymbol{s}_i$ $(i = 1, ..., n)$, and $\boldsymbol{\nu}_s = (\nu_0, \nu_1, \ldots, \nu_p)^t$ with each ν_k a trend model coefficient. Writing (5) in matrix form,

$$\boldsymbol{g}_s = \boldsymbol{\Phi}_s \boldsymbol{\omega}_s + \boldsymbol{F}_s \boldsymbol{\nu}_s \tag{6}$$

where $\boldsymbol{g}_s = (g(\boldsymbol{s}_1), \ldots, g(\boldsymbol{s}_n))^t$, $\boldsymbol{\Phi}_s =$ is a $n \times n$ matrix with elements $\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'})$, $i, i' = 1, \ldots, n$, $\boldsymbol{F}_s = (\boldsymbol{f}(\boldsymbol{s}_1), \ldots, \boldsymbol{f}(\boldsymbol{s}_n))^t$ is a $n \times (p+1)$ matrix.

The parameters $\boldsymbol{\omega}_s$ and $\boldsymbol{\nu}_s$ can be estimated by penalized least squares, minimizing the following expression

$$\sum_{i=1}^{n} [Z(\boldsymbol{s}_i) - g(\boldsymbol{s}_i)]^2 + \rho \int_{\mathbb{R}^2} J_m(\kappa(\boldsymbol{s}))d\boldsymbol{s} \tag{7}$$

where $\rho > 0$ acts as a smoothing parameter and $J_m(\kappa(\boldsymbol{s}))$ is a measure of roughness of the spline function κ which is defined in terms of the mth derivatives of κ. If $\rho = 0$, RBFs such as exponential (EXPON), Gaussian (GAU), multiquadratic (M) and inverse multiquadratic (IM) could be used.

Writing $m = 2$ and replacing (6) in (7), we obtain

$$\begin{aligned} L(\boldsymbol{\omega}_s, \boldsymbol{\nu}_s) =& (\boldsymbol{Z}_s - \boldsymbol{\Phi}_s\boldsymbol{\omega}_s - \boldsymbol{F}_s\boldsymbol{\nu}_s)^t(\boldsymbol{Z}_s - \boldsymbol{\Phi}_s\boldsymbol{\omega}_s - \boldsymbol{F}_s\boldsymbol{\nu}_s) + \rho \int_{\mathbb{R}^2} \left[\kappa''(\boldsymbol{s})\right]^2 d\boldsymbol{s} \\ =& (\boldsymbol{Z}_s - \boldsymbol{\Phi}_s\boldsymbol{\omega}_s - \boldsymbol{F}_s\boldsymbol{\nu}_s)^t(\boldsymbol{Z}_s - \boldsymbol{\Phi}_s\boldsymbol{\omega}_s - \boldsymbol{F}_s\boldsymbol{\nu}_s) + \rho\boldsymbol{\omega}_s^t\boldsymbol{\Phi}_s\boldsymbol{\omega}_s \end{aligned}$$

where $\boldsymbol{Z}_s = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))^t$, $\int_{\mathbb{R}^2} \left[\kappa''(\boldsymbol{s})\right]^2 d\boldsymbol{s} = \int_{\mathbb{R}^2} \kappa''(\boldsymbol{s}) \left[\kappa''(\boldsymbol{s})\right]^t d\boldsymbol{s} = \boldsymbol{\omega}_s^t\boldsymbol{\Phi}_s\boldsymbol{\omega}_s$ with $\kappa''(\boldsymbol{s}) = \sum_{i'=1}^{n} \omega_{i'}\phi''(\boldsymbol{s} - \boldsymbol{s}_{i'}) = \left[\boldsymbol{\phi}''(\boldsymbol{s})\right]^t \boldsymbol{\omega}_s$, $\boldsymbol{\Phi}_s = \int_{\mathbb{R}^2} \boldsymbol{\phi}''(\boldsymbol{s}) \left[\boldsymbol{\phi}''(\boldsymbol{s})\right]^t d\boldsymbol{s}$ and $\boldsymbol{\phi}''(\boldsymbol{s}) = (\phi''(\boldsymbol{s} - \boldsymbol{s}_1), \ldots, \phi''(\boldsymbol{s} - \boldsymbol{s}_n))^t$.

After differentiating with respect to the vectors $\boldsymbol{\omega}_s$ and $\boldsymbol{\nu}_s$, we find that $(\boldsymbol{\omega}_s, \boldsymbol{\nu}_s)$ are the solution of the following system of linear equations

$$\begin{pmatrix} \boldsymbol{\Phi}_s + \rho\boldsymbol{I} & \boldsymbol{F}_s \\ \boldsymbol{F}_s^t & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega}_s \\ \boldsymbol{\nu}_s \end{pmatrix} = \begin{pmatrix} \boldsymbol{Z}_s \\ \boldsymbol{0} \end{pmatrix} \tag{8}$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix and $\rho\boldsymbol{I}$ can be interpreted as a white noise added to the variances at the data locations (Lloyd 2010).

The proof for the Franke conjecture (Michelli 1986) demonstrates the sufficient conditions for guarantee the non-singularity of the method when other radial basis functions are used. Then, $\boldsymbol{\Phi}_s + \rho\boldsymbol{I}$ is a non-singular matrix. Therefore, the weights $\boldsymbol{\omega}_s$ and the parameter vector $\boldsymbol{\nu}_s$ are estimated, respectively, by

$$\begin{aligned} \widehat{\boldsymbol{\omega}}_s =& (\boldsymbol{\Phi}_s + \rho\boldsymbol{I})^{-1} \left\{ \boldsymbol{I} - \boldsymbol{F}_s \left[\boldsymbol{F}_s^t(\boldsymbol{\Phi}_s + \rho\boldsymbol{I})^{-1}\boldsymbol{F}_s \right]^{-1} \boldsymbol{F}_s^t(\boldsymbol{\Phi}_s + \rho\boldsymbol{I})^{-1} \right\} \boldsymbol{Z}_s \\ \widehat{\boldsymbol{\nu}}_s =& \left[\boldsymbol{F}_s^t(\boldsymbol{\Phi}_s + \rho\boldsymbol{I})^{-1}\boldsymbol{F}_s \right]^{-1} \boldsymbol{F}_s^t(\boldsymbol{\Phi}_s + \rho\boldsymbol{I})^{-1}\boldsymbol{Z}_s \end{aligned}$$

Having estimated the parameters $\boldsymbol{\omega}_s$ and $\boldsymbol{\nu}_s$, we are now ready to perform spatial prediction at a new location, $\boldsymbol{s}_0$. The aim is to predict the value of $Z(\boldsymbol{s}_0)$ based on a set of observations $\boldsymbol{Z}_s^*$. For this, the RBF predictor is given by

$$\widehat{Z}(\boldsymbol{s}_0) = \hat{g}(\boldsymbol{s}_0) = \sum_{i=1}^{n_h} \lambda_i Z(\boldsymbol{s}_i) = \boldsymbol{\lambda}^t\boldsymbol{Z}_s^* \tag{9}$$

subject to

$$\sum_{r=1}^{n_h} \lambda_r f_l(\boldsymbol{s}_r) = \boldsymbol{\lambda}^t\boldsymbol{f}_l = f_l(\boldsymbol{s}_0), \qquad l = 0, \ldots, p$$

where n_h is the size of the neighborhood, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{n_h})^t$, $\boldsymbol{f}_l = (f_l(\boldsymbol{s}_1), \ldots, f_l(\boldsymbol{s}_{n_h}))^t$ and $\boldsymbol{Z}_s^* = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_{n_h}))^t$.

The expected error is equal to zero, i.e., $\mathsf{E}\left(\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)\right) = 0$ and the mean square error of prediction by kriging, σ_K^2, using the radial basis function is given by

$$\begin{aligned} \sigma_K^2(\boldsymbol{s}_0) =& \mathsf{E}\left\{ \left[\hat{Z}(\boldsymbol{s}_0) - Z(\boldsymbol{s}_0)\right]^2 \right\} \cong -\sum_{i=1}^{n_h}\sum_{i'=1}^{n_h} \lambda_i\lambda_{i'}\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}) + 2\sum_{i=1}^{n_h} \lambda_i\phi(\boldsymbol{s}_i - \boldsymbol{s}_0) \\ =& -\boldsymbol{\lambda}^t\boldsymbol{\Phi}_{s_0}\boldsymbol{\lambda} + 2\boldsymbol{\lambda}^t\boldsymbol{\phi}_0 \end{aligned}$$

where $\boldsymbol{\Phi}_{s_0}$ is a $n_h \times n_h$ matrix with elements $\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'})$, and $\boldsymbol{\phi}_0 = (\phi(\boldsymbol{s}_1 - \boldsymbol{s}_0), \ldots, \phi(\boldsymbol{s}_n - \boldsymbol{s}_0))'$ corresponds to the radial function vector evaluated between the neighbors and the point where we aim to predict, i.e. $\phi(\boldsymbol{s}_r - \boldsymbol{s}_0)$, $r = 1, \ldots, n_h$.

The weights are determined minimizing the following penalized expression

$$l(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \boldsymbol{\lambda}^t \left(\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0\right) \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^t \boldsymbol{\phi}_0 + 2\boldsymbol{\alpha}^t \left(\boldsymbol{F}_{s_0}^t \boldsymbol{\lambda} - \boldsymbol{f}_s(\boldsymbol{s}_0)\right)$$

where $\boldsymbol{I}_0$ is the $n_h \times n_h$ identity matrix, $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_p)^t$ is the vector of $p+1$ Lagrange multipliers associated with the unbiasedness constraint, $\boldsymbol{F}_{s_0}$ is similar to (6) using only n_h observations, and $\boldsymbol{f}_s(\boldsymbol{s}_0) = (f_0(\boldsymbol{s}_0), \ldots, f_p(\boldsymbol{s}_0))^t$.

After differentiating with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$, equating the result to zero and performing some algebraic procedures, the following matrix system is found

$$\begin{pmatrix} \boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0 & \boldsymbol{F}_{s_0} \\ \boldsymbol{F}_{s_0}^t & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi}_0 \\ \boldsymbol{f}_s(\boldsymbol{s}_0) \end{pmatrix} \tag{10}$$

Note that $\boldsymbol{\Phi}_{s_0}$ is a positive define matrix then $\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0$ is a nonsingular matrix. Thus to solve the system, the coefficients for $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ are given by

$$\begin{aligned} \widehat{\boldsymbol{\lambda}}^t = & \left\{ \boldsymbol{\phi}_0 + \boldsymbol{F}_{s_0} \left[\boldsymbol{F}_{s_0}^t (\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0)^{-1} \boldsymbol{F}_{s_0}\right]^{-1} \left[\boldsymbol{f}_s(\boldsymbol{s}_0) - \boldsymbol{F}_{s_0}^t (\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0)^{-1} \boldsymbol{\phi}_0\right] \right\}^t \\ & (\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0)^{-1} \\ \widehat{\boldsymbol{\alpha}} = & - \left[\boldsymbol{F}_{s_0}^t (\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0)^{-1} \boldsymbol{F}_{s_0}\right]^{-1} \left[\boldsymbol{x}_s(\boldsymbol{s}_0) - \boldsymbol{F}_{s_0}^t (\boldsymbol{\Phi}_{s_0} + \rho \boldsymbol{I}_0)^{-1} \boldsymbol{\phi}_0\right] \end{aligned} \tag{11}$$

Once we know $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$, an approximate estimation for the mean square error of prediction can be written as

$$\widehat{\sigma}_K^2(\boldsymbol{s}_0) \cong \sum_{r=1}^{n_h} \widehat{\lambda}_r \phi(\boldsymbol{s}_r - \boldsymbol{s}_0) + \rho \sum_{r=1}^{n_h} \widehat{\lambda}_r^2 + \sum_{l=0}^{p} \widehat{\alpha}_l f_l(\boldsymbol{s}_0)$$

The interpolators (5) and (9) represent two different focuses for the same problem and in some spatial cases are formally equivalent. This was exposed in Matheron (1981) through a Bayesian analysis (Kimeldorf and Wahba 1970). Then, there is a connection between kriging and splines demonstrations (Cressie 1989; Hutchinson and Gessler 1994; Kent and Mardia 1994). There is a link between the spatial methods kriging and splines, which was named equivalently "near" because of the thin plate spline (TPS, a class of spline) corresponds to a specific generalized covariance, whereas the kriging estimator or the RBF interpolator only requires the use of a kernel with appropriate positive definiteness properties. In general, this allows adapting the kernel function to a set of particular data (Cressie 1989; Myers 1992). Then, for both splines and kriging, the smoothing is determined objectively.

The splines have demonstrated to be so adequate for interpolating densely sampled heights and climatic variables (Hutchinson 1995; Mitáš and Mitášová 1999). Nevertheless, its main criticism comes from the incapacity to incorporate auxiliary information for modeling the deterministic part of the variation (Hengl 2007). The RBFs are recommended to be used for computing smoothed surfaces with a great deal of data. The functions produce proper results for superficial soft variations like elevation (Johnston, Ver, Krivoruchko, and Lucas 2001). A recent study is presented in (Melo, Melo, and Mateu 2018), where trending RBFs are used in an application of calcium concentration, there the library **geospt** was utilized.

Several RBFs that are considered in this paper are presented below:

Multiquadratic

Hardy (1990) called multiquadratic this method because he considered that the main feature is the superposition of quadratic surfaces. The multiquadratic radial basis function is given by

$$\phi(\delta) = \sqrt{\eta^2 + \delta^2}, \quad \eta \neq 0.$$

where η is a smoothed parameter of free choice.

Inverse multiquadratic

A variation of the multiquadratic function was introduced by Hardy and Gopfert (1975), and it is the inverse multiquadratic given by

$$\phi(\delta) = 1/\sqrt{\eta^2 + \delta^2}, \quad \eta \neq 0$$

Franke (1982) found that this radial basis function provides excellent approximations, even when the number of centers (nearest neighbors) is small.

Spline with tension

The spline with tension (ST) function is given by the expression

$$\phi(\delta) = \begin{cases} \ln(\eta \cdot \delta/2) + K_0(\eta \cdot \delta) + C_E & \text{if} \quad \delta \neq 0 \\ 0 & \text{if} \quad \delta = 0 \end{cases}$$

where $K_0(x)$ is the modified Bessel function (Abramowitz and Stegun 1965, p. 374) and $C_E = -\int_0^\infty (\ln(x)/e^x)dx = 0.5772161$ is the Euler's constant (Abramowitz and Stegun 1965, p. 255).

Thin plate spline

This spline was introduced in the geometrical design by Duchon (1976). Its name, thin plate spline (TPS), refers to a physical analogy which implies the flexion of a thin metal sheet. Later, Thiébaux and Pedder (1987) described the TPS as a cubic spline of two dimensions (surface). In the case of an Euclidean space with $d = 2$, the TPS will have the following form:

$$\phi(\delta) = \begin{cases} (\eta \cdot \delta)^2 \ln(\eta \cdot \delta) & \text{if} \quad \delta > 0, \eta > 0 \\ 0 & \text{if} \quad \delta = 0 \end{cases}$$

Completely regularized spline

A variant of the TPS that uses the regularized spline basis function is known as completely regularized spline (CRS), which is described by

$$\phi(\delta) = \begin{cases} \ln(\eta \cdot \delta/2)^2 + E_1(\eta \cdot \delta/2)^2 + C_E & \text{if} \quad \delta \neq 0 \\ 0 & \text{if} \quad \delta = 0 \end{cases}$$

where $E_1(\cdot)$ is the exponential integral function (Abramowitz and Stegun 1965, p. 227) and C_E is the Euler's constant previously defined.

Exponential

The exponential radial basis function useful for the spatial interpolation is the exponential one, which allows avoiding inflection points and contains cubic splines as a special case (Späh 1969). This function is expressed by

$$\phi(\delta) = e^{-\eta\delta}, \qquad \eta \neq 0.$$

This function, in spatial interpolation, is characterized by not generating predictions out of the sampling range.

Gaussian

One of the most popular radial basis functions, along with the TPS function, is the Gaussian. Schagen (1979) was the first one who used the Gaussian function as a radial basis function. This function is given by

$$\phi(\delta) = e^{-\eta\delta^2}, \qquad \eta \neq 0.$$

For all radial basis functions, the optimal smoothing parameter η, which is a free choice parameter, can be found by minimizing the root of the mean square prediction error (RMSPE) for the cross validation. Some additional descriptions of RBFs and their relations with splines and kriging can be found in Bishop (1995, p. 164), Chilès and Delfiner (1999, p. 272) and Cressie (1993, p. 180).

2.4. Cross validation summary

The cross validation idea was initially worked by Mosteller and Wallace (1963) and Stone (1974). The leave-one-out cross-validation (LOOCV) consists of removing one observation from the n sample points (generally associated to a neighborhood), and then, with the remaining $n-1$ values and a selected RBF with its parameters η and ρ, we predict the variable value of study at the point location that was removed. This procedure is done sequentially with each sample point, and thus, a set of n prediction errors are obtained. If the RBF model describes well the spatial structure, then the difference between the observed and predicted values should be small. This procedure is justified because the RBF interpolation methods are accurate, i.e., the predicted and observed values match at the sampled points. Then, the set of predictions can be summarized by a measure of the prediction precision.

The function, `criterio.cv`, of cross validation hereby presented works for the case of kriging interpolation too. Hence, it will be considered the kriging standard deviation. Let $\hat{Z}_{[i]}(\boldsymbol{s}_i)$ be the predicted value of the cross validation and $\hat{\sigma}_{[i]}(\boldsymbol{s}_i)$ the prediction for the standard deviation in the position $\boldsymbol{s}_i$. With these statistics, it is possible to construct the mean prediction error (MPE), the root-mean-square prediction error ($RMSPE$), the average standard prediction error ($ASPE$), mean standardized prediction error ($MSPE$), the root-mean-squared standardized prediction error ($RMSSPE$), the determination coefficient (R^2), the absolute percentage prediction error ($MAPPE$), the correlation coefficient of the prediction error ($CCPE$), and the pseudo determination correlation coefficient of the prediction error ($pseudoR2$). These statistics are useful to evaluate the predictive capacity of the method (RBF or kriging), and they are presented below.

$$
\begin{aligned}
MPE =& \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Z}_{[i]}(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i)\right) \\
RMSPE =& \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Z}_{[i]}(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i)\right)^2} \\
ASPE =& \frac{1}{n}\sum_{i=1}^{n}\hat{\sigma}_{[i]}(\boldsymbol{s}_i) \\
MSPE =& \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\hat{Z}_{[i]}(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i)\right)}{\hat{\sigma}_{[i]}(\boldsymbol{s}_i)} \\
RMSSPE =& \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left(\hat{Z}_{[i]}(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i)\right)}{\hat{\sigma}_{[i]}(\boldsymbol{s}_i)}\right]^2} \\
R^2 =& 1 - \frac{\sum_{i=1}^{n}\left(\hat{Z}_{[i]}\left(\boldsymbol{s}_i\right) - Z\left(\boldsymbol{s}_i\right)\right)^2}{\sum_{i=1}^{n}\left(Z\left(\boldsymbol{s}_i\right) - \bar{Z}\left(\boldsymbol{s}_i\right)\right)^2} \\
MAPPE =& \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{Z}_{[i]}(\boldsymbol{s}_i) - Z(\boldsymbol{s}_i)}{Z(\boldsymbol{s}_i)}\right| \\
CCPE =& \frac{\sum_{i=1}^{n}\left(\hat{Z}_{[i]}\left(\boldsymbol{s}_i\right) - \bar{\hat{Z}}_{[i]}\left(\boldsymbol{s}_i\right)\right)\left(Z\left(\boldsymbol{s}_i\right) - \bar{Z}\left(\boldsymbol{s}_i\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Z}_{[i]}\left(\boldsymbol{s}_i\right) - \bar{\hat{Z}}_{[i]}\left(\boldsymbol{s}_i\right)\right)^2}\sqrt{\sum_{i=1}^{n}\left(Z\left(\boldsymbol{s}_i\right) - \bar{Z}\left(\boldsymbol{s}_i\right)\right)^2}} \\
pseudoR2 =& CCPE^2
\end{aligned}
\tag{12}
$$

where $\hat{Z}_{[i]}\left(\boldsymbol{s}_i\right)$ is the prediction value and $\hat{\sigma}_{[i]}(\boldsymbol{s}_i)$ is the standard deviation, which are obtained from the cross validation, and $Z\left(\boldsymbol{s}_i\right)$ as was defined above.

A variation of the previous methodology consists on fractioning the sample in two sub-samples; the first sub-sample is used to model the variogram and the other one is used to validate the kriging method. After that, the validation measures can be constructed from the observed and predicted values (Bivand, Pebesma, and Rubio 2008). If everything goes right, the $RMSPE$ should be as small as possible (closer to zero) and the R^2 should be close to 1.

3. Implementation in R

In this section, there is a special emphasis on the use of the functions from the library associated with the theoretical concepts defined in the previous section.

3.1. Geostatistical functions

Trimmed mean variogram

For the case of the trimmed mean, the function given in equation (2) was programmed, modifying the sum of the Cressie-Hawkins formula by the trimmed mean. In this purpose, the user can choose the trim percentage. In case that the percentage is equal 50%, this estimator will coincide with the median, which is more robust in the presence of atypical data, while a trim percentage of 0% will result in an estimator that coincides with the Cressie-Hawkins robust estimator. Bárdossy (2001) compared the classic, robust and mean trimmed (with 10% trim) estimators considered an atypical data. Then, using this trimmed mean, we found that with this estimator, better results are obtained in presence of atypical data and, hence, it is more robust. Similar results were shown by simulations in Roustant, Dupuy, and Helbert (2007).

The function `est.variograms()` is built from `est.variogram()` of the package **sgeostat** in https://cran.r-project.org/web/packages/sgeostat implementing the trimmed mean on its calculations. For example, `maas` dataset from the package **sgeostat** is considered, specifying a trim of 10% below:

```
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
```

The result is the following:

```
   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

In this case, we will consider the coalash dataset mentioned in section 2.2. The function requires the `data.frame`, the sort of graph associated to the probability or the standardized variance pocket plot on directions north-south or east-west; probability pocket plot per row, i.e. horizontal "south-north" `"PPR"`, pocket plot of probabilities by columns, i.e. vertical "east-west" `"PPC"`, variance pocket plot by rows, i.e. horizontal "south-north" `"PVR"` and the variance pocket plot by columns, i.e. vertical "east-west" `"PVC"`, the coordinates "X" and "Y", the name of the variable to be analyzed `"Z"`, and the identification of atypical data (automatic "F" or personal "T"). On the Figure 2 clearly, the rows 2, 6 and 8 are atypical. That fact is useful as a verification that these rows are potentially problematic. The following code in R describes the situation of an analysis of local stationarity on probabilities of % coal-ash in the direction south-north:

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

The obtained result is shown in Figure 2(a) and the associated one with standardized variances is shown in Figure 2(b).

cross validation summary

For generating the summary of cross validation statistics, we propose the function `criterio.cv()`, which generates a data frame with the statistics presented in the expression (12), obtained by leave-one-out cross validation. In order that this works well, we introduced a `data.frame` with: the coordinates of the data, prediction columns and prediction variances of the samples, the observed values, the residual values, the zscore (residual value over the standard error from kriging) and group. In case of using the function `rbf.tcv`, the prediction variance and the zscore will have NA values. Following, a brief example of the `criterio.cv()` functioning is shown based on the *meuse* database of the library `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)
```

The obtained result is shown below

```
         MPE      ASPE     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. RBF functions usage

In this section, we present how RBFs package work for which we consider an experimental precipitation data set called `preci` of the library **geospt**. The function `rbf()` is conducted from the expressions (5) and (9) presented in the subsection 2.3. Those expressions require: the smoothing (η) and the robustness (ρ) factors which are recommended to find the optimal values through the `graph.rbf` function, the formula $z \sim 1$ (without trend) or $z \sim x + y$ (with trend), the flat coordinates of the sample used for the prediction "`coordinates`", the coordinates of the point to be predicted or the data frame of new points to be predicted "`newdata`" and the number of neighbors "`n.neigh`" if a determinate neighborhood size is desired. Also, it is necessary to specify the RBF given "`func`" (see table 1). The code used for this function is the following:

```
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) = ~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))
```

Previously η was optimized using the function `graph.rbf`, obtaining a value $\eta = 0.1460814$ with the RBF "TPS". The result is shown in Figure 3.

The function `rbf.cv()` requires the following inputs: the smoothing factor (`eta`), the robustness parameter (`rho`), the formula described before, the coordinates of the sample (`coordinates`), the size of neighborhood (`n.neigh`), and the RBF given (`func`). The example is shown below:

```
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")
```

Running all this, it is obtained $RMSPE = 7.691789$.

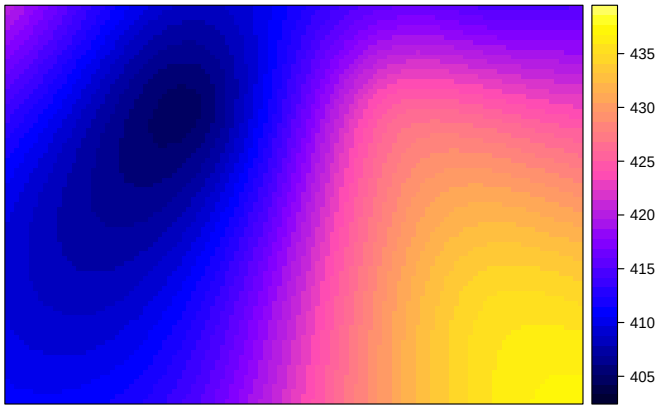


Figure 3: TPS interpolation surface of the precipitation

The function `RBF.phi()` requires: a distance between a couple of points $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$, a smoothing factor `eta` and a RBF `func`. The following example results in the value 42.88812:

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

The function `rbf.tcv()` requires: a formula, the parameters `eta` and `rho`, a neighborhood size (`n.neigh`), and a RBF (`func`), which explained above. The result of this function is necessary to construct the statistics summary of the leave-one-out cross validation (LOOCV). An example is presented below:

```
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")
```

The result obtained is

```
   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
```

As seen in this table, NA values are generated in the columns, *var1.var* and *zscore*, which is due to the lack of variances for the error predictions. It does not occur with the kriging computed data.

The `graph.rbf()` function allows observing the behavior of the smoothing and robustness parameters, η and ρ respectively, and finding the optimal value in case the user requires it. When the optimal values for these parameters are found and a neighborhood of a determined size in a specific RBF is defined, we can find an appropriate interpolation surface. For this function it is necessary to specify: the formula, the coordinates of the sample (`coordinates`), the neighborhood size (`n.neigh`), the RBF given (`func`) previously described, the logic operators `eta.opt` and `rho.opt` which indicates if the optimization will be performed or not (TRUE or FALSE), the maximum limit for the parameters η and ρ (`eta.dmax` and `rho.dmax`, respectively), the initial values to look for optimal values of the definite parameters in the vector `x0`, the number of desired iterations for finding the optimal parameters given by the option `iter`, and the logic operator `P.T` which allows the visualization of table that contains the values associated to the graph with the optimal values obtained. The Figure 4 shows the graph obtained with five RBFs in $0 \leq \eta \leq 1.6$.

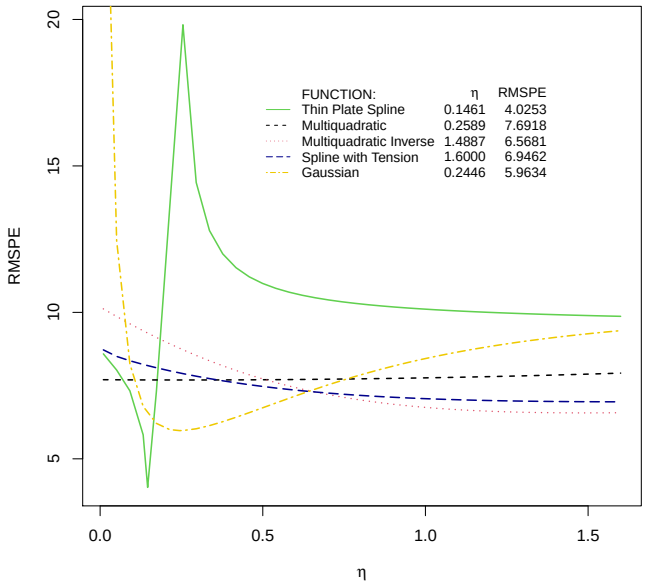


Figure 4: Optimization of η, in radial basis functions

According to the Figure 4, the lowest $RMSPE$ is generated with a TPS. The code only for the Gaussian function is:

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

Table 1 provides a brief description of the functions implemented in the library **geospt**.

Function	Description
`criterio.cv()`	It generates a object of class data.frame with a summary of statistics associated values using cross validation: MPE, $ASPE$, $RMSPE$, $MSPE$, $RMSSPE$, $MAPPE$, $CCPE$, R^2 and $pseudoR2$.
`network.design`	It generates an average with the standard kriging of the prediction errors (ASPE) associated to a conditional net design.
`est.variograms()`	It calculates the experimental variogram: classic, robust, median and trimmed mean (Cressie 1993; Bárdossy 2001).
`pocketplot()`	It graphs the pocket plot of probability or standardized variances on directions north-south or east-west (see Cressie (1993)).
`rbf()`	It generates individual predictions from the following radial basis functions: completely regularized spline (CRS), spline with tension (ST), Gaussian (GAU), exponential ($EXPON$), trigonometric (TRI), thin plate spline (TPS), inverse multiquadratic (IM) and multiquadratic (M).
`rbf.cv()`	It extracts a $RMSPE$ value, resulting from the leave-one-out cross validation.
`rbf.cv1()`	It returns a $RMSPE$ value, resulting from the LOOCV which allows the simultaneous optimization of the parameters `eta` and `rho`.
`RBF.phi()`	It produces: a numeric value obtained from the radial basis function which is generated with a determined distance between two points, the value of the smoothing parameter `eta`, and a function "CRS", "ST", "GAU", "EXPON", "TRI", "TPS", "IM" or "M".
`rbf.tcv()`	It generates a table with: the coordinates of the data, the predictions, the observed values, the residual values, the prediction variances, and the zscore values (residual value over standard error) of the analyzed variable which are the result of the LOOCV.
`graph.rbf()`	It produces a graph that describes the behavior of the parameters `eta` and `rho`, which are associated to the RBF.

Table 1: Functions of the library **geospt**

4. Application

In this section, an application is presented where the functions designed for spatial interpolation with RBFs are tested.

4.1. Data set

The average daily temperature in Croatia was measured during January 2008. This is measured three times per day in most of the weather stations, at 7 am, 1 pm and 9 pm. The mean daily temperature (ΔT in a day) is calculated as a weighted average, according to the following expression:

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

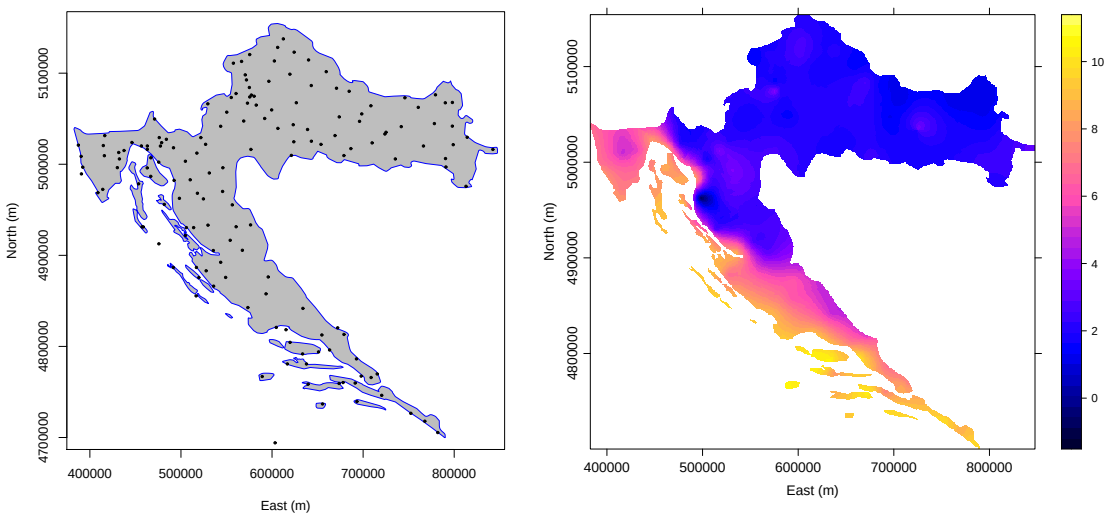
The information was taken from `http://spatial-analyst.net/book/HRclim2008` and was

provided by Melita Perčec Tadić, from the Croatian Meteorological and Hydrological Service (DHMZ) (Hengl 2009).

Afterward, the monthly mean temperature is obtained from the daily mean previously mentioned, taking into account that a composition of images is available (MODIS images with 1 km resolution every 8 days, with access to all public) of the daily mean temperature, i.e. from 3 to 4 monthly registers. The temperature measurements are automatically collected by 159 meteorological stations. Given that four of the stations had not available registers for January and some of the stations hat missing (or lost) data, these stations were omitted from the analysis. Consequently, only 146 stations were considered, and the computing was performed with observed information of monthly mean temperature, removing missing or lost data. In this application, it is analyzed the average daily temperature in Croatia for January in 2008 from 146 meteorological stations.

Croatia is a relatively small country, but it has several regions of different weather as a result of its specific position on the Adriatic Sea and fairly diverse topography ranging from plains on the east, through a central mountainous region separating the continental from the maritime part of the country. The study region is characterized by a wide range of topographic and climatic features. This allows to properly assess the proposed methodology with respect to the traditional one, given that the mean earth's temperatures in this region are strongly influenced by the topography. Temperature measurements are automatically collected at 154 meteorological stations.

The geographical coordinates (latitude and longitude) were transformed to a system of Cartesian coordinates (x, y). The datum used was WGS84 zone 33, and the transformation method used is known as Bursa Wolf. The location of 146 stations is shown in the Figure 5(a).



(a) Locations of meteorological stations

(b) Prediction of average monthly temperature in January 2008 (multiquadratic function)

Figure 5: Croatia Map

4.2. Prediction map

Initially, the optimization of the two parameters of smoothing η and ρ is performed. For this aim, it is used the function `graph.rbf()`, which is evaluated in the RBFs mentioned in 2.3. Here, we found that one of the better fits is using the function M with parameters $\eta = 1e{-}05$ and $\rho = 0$ ($RMSPE = 1.23022$). This function works internally with the functions `rbf.cv()` and `rbf.cv1()`, which use `optimize()` (for one parameter) and `bobyqa()` (for both parameters), respectively. In the optimization process, the function `optim()` described by Mittelhammer, Judge, and Miller (2000) was not used because it needs of more processing time than `optimize()`. For optimizing both parameters, it is necessary that `eta.opt=TRUE` and `rho.opt=TRUE`, thus a lattice graph will be obtained (see Figure 6 for the study case).

The function `optimize()` of the software R, described in Brent (1973), looks for the optimal parameter η or ρ by generating a value of $RMSPE$ from the function `rbf.cv()` in an interval for such parameters established by the user. For the study case, `bobyqa()` converges rapidly in the iteration number 19. For instance, in multiquadratic function, the optimal values for η and ρ are usually found close to 0. Then, an appropriate interval would be between 0 and 1 as shown by Figure 6.

The dataset and the shapefile are preload from **geosptdb** (Melo and Melo 2015), and subsequently, a 70000 points grid is generated within the analyzed region in order to generate predictions of the mean terrestrial temperature. This grid is obtained by using the function `spsample()` from the package **sp**. The predictions are generated from the function `rbf()`. This function requires: the values of the parameters `"eta"` and `"rho"`, the formula that defines the trend or model ($z \sim 1$ for cases without trend and $z \sim x + y$ for linear trend), the coordinates of the sampled points (”`coordinates`”), the new locations coordinates (”`newdata`”), the number of neighbors (”`n.neigh`”), and the RBF (”`func`”). These predictions are eventually turned to a class object codeSpatialPixelsDataFrame and `sp`, with the instruction `coordinates()`, from the package **sp**. Finally, with the function `spplot()`, the predictions map is obtained for the analyzed variable, which is shown in the Figure 5(b). The code to develop all this is the follow:

```
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) = ~x+y
R> gridded(pred.rbf) <- TRUE
```

```
R> spplot(pred.rbf["var1.pred"],cuts=40,scales=list(y=list(rot=90),draw =T),
          col.regions=bpy.colors(100),key.space=list(space="right",cex=0.8),
          xlab="East (m)",ylab="North (m)")
```

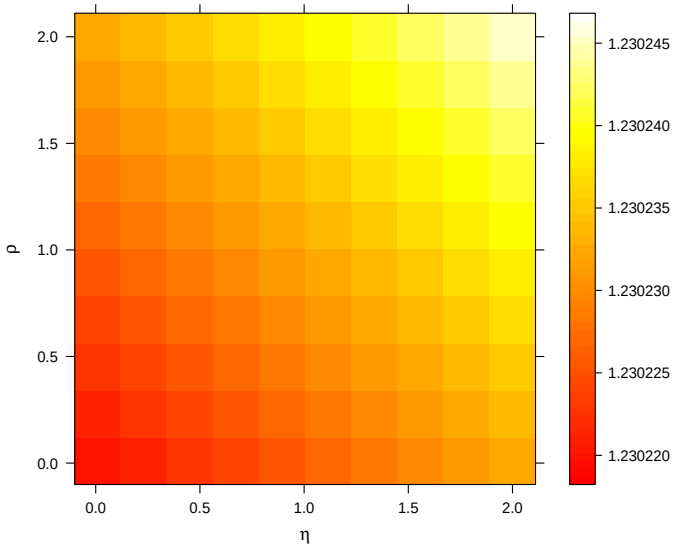


Figure 6: Lattice of η y ρ for multicuadratic function

5. Conclusions

The last version of **geospt** allows performing the optimization for the parameters η and ρ in the case of the RBFs to obtain the interpolation surfaces, which implies a greater demand of computational time. For this reason, we recommend using modern and high capability processors or limiting of iterations number in order to obtain results in a minor time. The instruction `graph.rbf` for the case where `eta.opt` and `rho.opt` are equal to *TRUE* allows to obtain a Lattice chart with the different combinations between the parameters values for a range predefined by the user. One of the greatest advantages of the programmed functions, regarding spatial interpolation with RBFs, is that they allow considering the trend. This fact, for some applications such as the interpolation of environmental variables, may lead to obtain better results than when the trend modeling is omitted in spite of the non-stationarity of the observed process.

6. Acknowledgments

Work partially funded and supported by: Core Spatial Data Research (Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Grant COL0013969), and Applied Statistics in Experimental Research, Industry and Biotechnology (Universidad Nacional de Colombia, Grant COL0004469).

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." *Joumal of the Intemational 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 Appllications*, **8**, 273–281.

Gomez M, Hazen K (1970). "Evaluating sulfur and ash distribution in coal seems by statistical response surface regression analysis." *Technical report*, U.S. Bureau of Mines Report RI 7377.

Gräler B, Pebesma E, Heuvelink G (2016). "Spatio-Temporal Interpolation using gstat." *The R Journal*, **8**, 204–218. URL https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

Hardy R, Gopfert W (1975). "Least squares prediction of gravity anomalies, geodial 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 Comunities 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 Ű more than just a smooth interpolator." *Geoderma*, **62**, 45–67.

Johnston K, Ver J, Krivoruchko K, Lucas N (2001). *Using ArcGIS Geostatistical Analysis*. ESRI.

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.

Lloyd CD (2010). *Local Models for Spatial Analysis*. Second edition. Taylor & Francis Group, Boca Raton-Florida.

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, Melo O (2015). *geosptdb: Spatio-Temporal; Inverse Distance Weighting and Radial Basis Functions with Distance-Based Regression*. R package version 0.5-0, URL geosptdb.r-forge.r-project.org/.

Melo C, Santacruz A, Melo O (2012). *geospt: An R package for spatial statistics*. R package version 1.0-0, URL geospt.r-forge.r-project.org/.

Melo CE, Melo OO, Mateu J (2018). "A distance-based model for spatial prediction using radial basis functions." *AStA Advances in Statistical Analysis*, **102**(2), 263–288. doi:10.1007/s10182-017-0305-4. URL https://ideas.repec.org/a/spr/alstar/v102y2018i2d10.1007_s10182-017-0305-4.html.

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 (2016). *sgeostat: An Object-Oriented Framework for Geostatistical Modeling in S+*. R package version 1.0-27, URL https://CRAN.R-project.org/package=sgeostat.

Pebesma EJ (2004). "Multivariable geostatistics in S: the gstat package." *Computers & Geosciences*, **30**, 683–691.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ribeiro PJ, Diggle PJ (2001). "geoR: a package for geostatistical analysis." *R-NEWS*, **1**(2), 14–18. ISSN 1609-3631, URL http://CRAN.R-project.org/doc/Rnews/.

Ribeiro Jr PJ, Diggle PJ, Schlather M, Bivand R, Ripley B (2020). *geoR: Analysis of Geostatistical Data*. R package version 1.8-1, URL https://CRAN.R-project.org/package=geoR.

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.

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
Universidad Distrital Francisco José de Caldas
11001 Bogotá D. C., Colombia
E-mail: cmelo@udistrital.edu.co
Telephone: +57 /1/3239300

Sandra E. Melo
Department of Agronomy, Faculty of Agricultural Sciences
Universidad Nacional de Colombia
11001 Bogotá D. C., Colombia
E-mail: semelom@unal.edu.co
Telephone: +57 /1/3165000/ext 19021

Oscar O. Melo
Department of Statistics, Faculty of Sciences
Universidad Nacional de Colombia
11001 Bogotá D. C., Colombia
E-mail: oomelom@unal.edu.co
Telephone: +57 /1/3165000/ext 13206

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd