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
Ingeniería Catastral y Geodesia
Facultad de Ingeniería
Universidad Distrital
Francisco José de Caldas

Oscar O. Melo
Departamento de Estadística
Facultad de Ciencias
Universidad Nacional
de Colombia

Sandra E. Melo
Departamento de Agronomia
Facultad de Ciencias Agrarias
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. That is geoR, gstat and sgeostat. The package geospt supplies tools for constructing the trimmed mean variogram, to making the pocket plot (useful for the local stationarity analysis), a series of radial basis spatial functions (multiquadratic, inverse multiquadratic, tension splines, thin plate splines, exponential, Gaussian, completely regularized spline) aimed to predict, optimize and realize leave-one-out cross-validation, a function for net designing and a function which generates a summary of the resulting statistics from cross-validation, to evaluate how works the interpolation (geostatistical and deterministic) methods, constructed in function of the errors. The fundamental theory is briefly presented coupled with its implementation through some practical examples.

Keywords: geostatistical analysis, trimmed mean variogram, pocketplot, 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 the 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 prediction

values can be generated using kriging interpolation for the construction of the predictions map of the dependent variable. Nonetheless, there are some deterministic methods, such as RBFs, where the interpolation model does not require a variogram model. Afterward, 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.

Currently, 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 spatiotemporal radial basis functions, 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 (Melo, Santacruz, and Melo 2015) 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** (Jr and Diggle 2015), **gstat** (Pebesma 2004) and **sgeostat** (original by James J. Majure Iowa State University and port + extensions by Albrecht Gebhardt 2013), among others. In this regard, our contributions are: a function for the trimmed mean experimental variogram construction, a function for generating the design of the regular grid net associated with conditional samples, a function for the pocket plot construction for gridded data (useful for local stationarity analysis) and radial basis functions (multiquadratic, inverse multiquadratic, tension spline, completely regularized spline and thin plate splines) to optimize, predict and perform the cross validation in the space, a function to generate a chart which shows the parameter's behavior of `eta` and `rho` associated to the radial basis function, and a function that generates a table with the summary of the cross validation statistics for evaluating the exactitude of the interpolation (geostatistical and deterministic) methods based on the prediction errors. Some of the functions are briefly described, and then its functioning is illustrated with some exercises. The package is implemented in the software R (Team 2015) which is available on the Comprehensive R Archive Network (CRAN) in `https://cran.r-project.org/web/packages/geospt`.

There are at least, three developments for interpolators that lead to the same functional form in interpolation: thin plate spline, radial basis function and the method known in geostatistics as kriging; all of them interrelated by their positive definite kernel function. This article focuses on the R design of functions for performing geostatistical processes.

The paper's organization is the following: In the section 2 a summary of the fundamental theory that supports the implemented functions is presented, the section 3 is dedicated to describe the usage of the functions implemented on the library, the section 4 considers an application of the processed data in Hengl (2007), and the 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, since this interpolator 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 $N(\boldsymbol{h})$ represents all of the pairs $(\boldsymbol{s}_i, \boldsymbol{s}_j)$ where $\boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}$ and $|N(\boldsymbol{h})|$ represents the cardinal number of $N(\boldsymbol{h})$, $Z\left(\boldsymbol{s}_i\right)$ is the sampling value in the position $\boldsymbol{s}_i$, and $Z\left(\boldsymbol{s}_j\right)$ is the sampling value in the position $\boldsymbol{s}_j$. This estimator is generally biased in the presence of atypical data, affecting the estimator. A robust estimator in this situation is the trimmed mean one Cressie and Hawkins (1980) and is given by

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

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

2.2. Pocket plot

The pocket plot (whose name comes from the detection of non-stationarity pockets) is a necessary technique to identify local anomalies with respect to the assumption of stationarity. It is constructed to take advantage on the spatial nature of the data through the rows and columns coordinates (East "x" and North "y" respectively). To illustrate this example, please check the following Figure 1

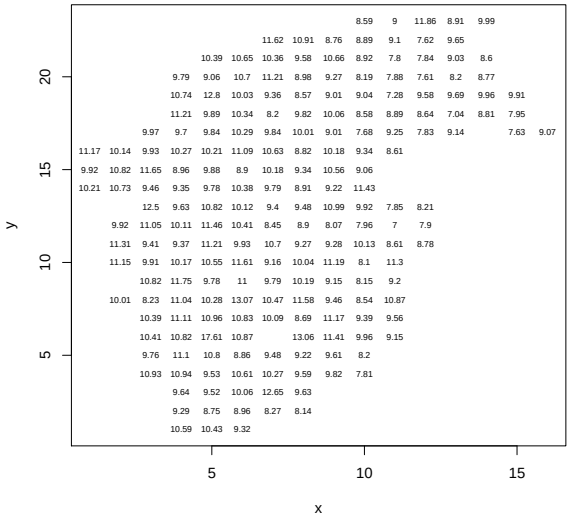


Figure 1: Core measurements (in % coal ash) at reoriented locations Cressie (1993)

In geostatistics, it is pretended to estimate the spatial relations between point data (variogram modeling). Thereupon, the estimated relations are used to compute the kriging method and

to estimate the predictor variability. Even though the Cressie and Hawkins (1980) estimator offers a robust estimation for the variogram, there is a fragment of the differences $(Z_i - Z_j)$, 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 but, naturally, 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, for instance k, there is a specific number N_{jk} of defined data differences, located at a distance $h =\mid j-k \mid$ 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 it is defined as:

$$\overline{\overline{Y}}_h = \frac{1}{\mid N(\boldsymbol{h}) \mid} \sum_{N(\boldsymbol{h})} \mid Z_i - Z_j \mid^{1/2} \tag{3}$$

$\overline{\overline{Y}}_h$ is a weighted mean of the $\overline{Y}_{jk}$s such that $\mid j-k \mid= h$. Then define

$$P_{jk} = \overline{Y}_{jk} - \overline{\overline{Y}}_h \tag{4}$$

$(P_{jk} : k = 1, 2, \ldots)$, is the residual contribution of the row j, to the variogram estimator on the biases differences. Ideally, these points are spread at both sides of the zero value. However, if there is something unusual in the row j, a singular contribution will be given for all the biases 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 chart, as seen on the Figure 2, where the central part of the dispersion is substituted by the box of a box diagram (Velleman and Hoaglin 1981, chap. 3)

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

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

From the results of Cressie (1985), $var(P_{jk}) = (2\gamma(h))^{(1/2)} / N_{jk}$ which justify the calculation of Q_{jk}. This change will only affect the spread of the points and not the panorama general to be above zero levels Cressie (1993).

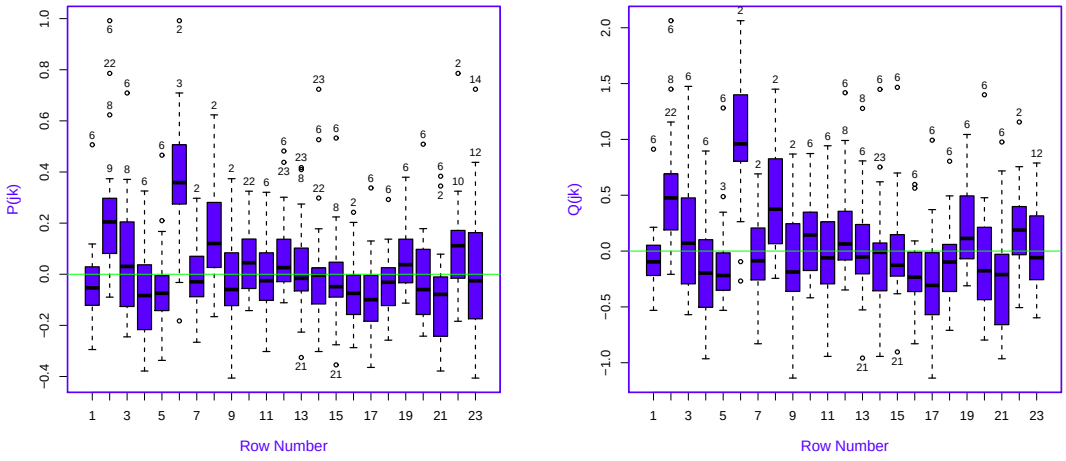
2.3. Radial basis function

For the spatial interpolation, there are some methods that do not require information from the spatial dependence model, such as the variogram or covariogram, which are known as deterministic and constitute the key point of this section.

Given the positions $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ in an Euclidean n-dimensional space and the observed data $Z(\boldsymbol{s}_1), ...Z(\boldsymbol{s}_n)$, the general form of the radial basis functions interpolator is

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

[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"



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

(b) Standardized variances.

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

where f_k; $k = 0, ..., p$, are the linearly independent functions as in the kriging estimator, $Z(\boldsymbol{s})$ is the function to be interpolated, $\boldsymbol{s}_i$, $i = 1, ..., n$ are the locations and ϕ is the kernel function which properly satisfies the positive defined condition. In the case where the interpolator is exact, i.e. $Z^*(\boldsymbol{s}) = Z(\boldsymbol{s})$ for $i = 1, ..., n$, 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. The predictor associated to (6) is a linear combination of the basis functions.

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

where $\phi(\boldsymbol{s}_i - \boldsymbol{s}_0)$ is a radial basis function obtained from the Euclidean distance between the position of the prediction $\boldsymbol{s}_0$ and every single location $\boldsymbol{s}_i$. The parameters $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$ can be estimated by penalized least squares regression, minimizing the following expression

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

where $J_m(g(\boldsymbol{s}))$ is a rugosity measure of the spline function g (defined in terms of the mth derivatives of g), $g(\boldsymbol{s}_i)$ is a real value function given by

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

and $\rho > 0$ acts as a smoothing parameter. The parameters ν_k, are associated to the trend, which usually is expressed in terms of coordinates "x" and "y". The parameters are found

after solving the following equations system

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

If there is no trend, $\mathbf{f}$ is a vector of ones and $\boldsymbol{\nu}$ is a bias parameter. An alternative estimator is given by

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

Then, the predictor, as a linear combination of $Z(\boldsymbol{s}_i)$'s, is given by: $\widehat{Z}(\boldsymbol{s}_0) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{s}) Z(\boldsymbol{s}_i)$, when the following condition is established

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

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$, $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))'$ and $\boldsymbol{f} = (f_l(\boldsymbol{s}_1), \ldots, f_l(\boldsymbol{s}_n))'$.
After solving by ordinary least squares, the following matrix system is found,

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

Solving the system, the coefficients for $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ are given by

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

where Φ is a matrix with the i, jth element $\phi(\boldsymbol{s} - \boldsymbol{s}_i)$ for the i, jth pair of data, $\mathbf{1}$ is a column-vector of ones, $\boldsymbol{\alpha}$ is the vector of Lagrange multipliers with a size $k \times 1$ and ϕ_0 corresponds to the RBF evaluated between the neighbors and the points where a prediction is desired, i.e. $\phi(\boldsymbol{s} - \boldsymbol{s}_0)$. This predictor has the advantage that it shows the weights for every single data, which mainly depend on the number of neighbors and the smoothing parameter.

The interpolators (6) and (10) represent two different focuses for the same problem and in some spatial cases are formally equivalent. Just as it is exposed in (Matheron 1981), through a Bayesian analysis(Kimeldorf and Wahba 1970), the connection demonstrations between kriging and splines (Cressie 1989), (Hutchinson and Gessler 1994) and (Kent and Mardia 1994).

There is a link between the spatial methods kriging and splines, which was named equivalently "near" because of the TPS (a class of spline) that corresponds to a specific generalized covariance, whilst the kriging estimator and the RBF interpolator just require the use of a kernel with suitable properties like the positive definite. Generally speaking, this allows adapting the kernel function to a set of particular data (Cressie 1989; Myers 1992). The greatest difference is that the user sets the smoothing parameter for the splines, while it is objectively established in the kriging case.

Definition 1: a function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is called radial if there is an invariable function $\Phi(\boldsymbol{s}) = [0, \infty)$ such that:

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

and$\parallel \cdot \parallel$ is the Euclidian norm $\mathbb{R}^d$. This means that the value of the function Φ in the point $\boldsymbol{s} \in \mathbb{R}^d$ just depends on the norm of $\boldsymbol{s}$.

Multiquadratic (MQ)

(Hardy 1990) called multiquadratic this method because he considered that the main feature is the superposition of quadratic surfaces.
Definition 2: Given a set of n different points $\{\boldsymbol{s}_i\}_{i=1}^n \in \mathbb{R}^d$ and their corresponding scalar values $\{f_i\}_{i=1}^n \in \mathbb{R}$, the multiquadratic interpolator of the data has the following form:

$$p(\boldsymbol{s}_i) = \sum^{n} \mathtt{b}_j \sqrt{\eta^2 + \delta_{ij}^2}, \qquad j = 1, \ldots, n$$

Where the coefficients $\mathtt{b}_j$ are determined through the imposition of the interpolation conditions $p(\boldsymbol{s}_i) = f_i$, for the $i = 1, \ldots, n$, δ_{ij} is the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, and η is a smoothing parameter. Hence the following system of linear and symmetric equations

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

donde $\mathtt{b} = (\mathtt{b}_1, \ldots, \mathtt{b}_n)'$ and the inputs of Φ are given by $\phi_{ij} = \sqrt{\eta^2 + \delta_{ij}^2}$.

Inverse Multiquadratic (IM)

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

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

where $\eta \neq 0$ is a smoothed parameter of free choice. (Franke 1982) found that this radial basis function provides excellent aproximations, even when the number of centers (nearest neighbors) is small.

Tension Spline (TS)

This function is given by the expression

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

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

Thin Plate Spline (TPS)

This spline was introduced in the geometrical design by (Duchon 1976). Its name, thin plate spline, 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:

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

Completely regularized spline (CRS)

A variant of the TPS that uses the regularized spline basis function is known as CRS and is described by

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

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 (EXPON)

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

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

where $\eta \neq 0$ is the smoothing parameter.

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

Gaussian (GAU)

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

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

where $\eta \neq 0$ is the free choice smoothing parameter.

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).

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).

2.4. Cross validation summary

The leave-one-out cross-validation (LOOCV) consists of excluding the observation of one of the n sampling points (generally associated with a neighborhood), and, with the remaining $n-1$ and the radial basis function with its parameters η and ρ, predict via splines the value of the study variable in the location of the excluded point. If the radial basis function presents an appropriate behavior, the difference between the observed and the predicted values should be minor and the map could be produced. This procedure is performed in a sequential way, with each of the sampling points and thus a set of n prediction errors is obtained. Then, the set of predictions can be summarized by a measure of the prediction precision. The idea dates back to, at least, Mosteller and Wallace (1963); Stone (1974).

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 too. 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 prediction error mean, the root of the mean square prediction error ($RMSPE$), the kriging prediction errors standard mean ($ASEPE$), the mean standardized prediction errors ($RMSSPE$) and the determination coefficient (R^2). Following, these statistics are presented, which are useful to evaluate the predictive capacity of the method (either RBF or kriging).

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

Where $\hat{Z}_{[i]}\left(s_i\right)$ is the prediction value and $\hat{\sigma}_{[i]}(s_i)$ the standard deviation, obtained from the

cross validation, and $Z(s_i)$ is the localization sampled value s_i.

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 forecasted 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 an especial 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 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. In Bárdossy (2001) are compared the classic, robust and mean trimmed (with 10% trim) estimators, an atypical is considered, see below. It is found that the estimator with the trimmed mean calculates better results in the presence of atypical data and, hence, is more robust. Similar results are 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, the trimmed mean is implemented on its calculations. For example, it is considered the database `maas` from the package **sgeostat**, 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 database 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, which 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 in 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)
```

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

cross validation summary

For generating the summary of cross validation statistics, we propose the function `criterio.cv()`, this generates a data frame with the statistics presented in the expression 13, obtained by leave-one-out cross validation. In order that it works properly, 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 divided

on the standard 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     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. RBF functions usage

The presented functioning in this section, with the RBF functions package, considers an experimental precipitation database `preci` of the library **geospt**. The function `rbf()` is conducted from the expressions (6) and (7) of the subsection 2.3. This expression requires the smoothing η and the robustness ρ factors (it is recommended to find the optimal values through the function `graph.rbf`), 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`", the number of neighbors "`n.neigh`", if a determinate neighborhood size is desired, and it is necessary to specify the RBF given "`func`" (see the table 1). The using code 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) = 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))
```

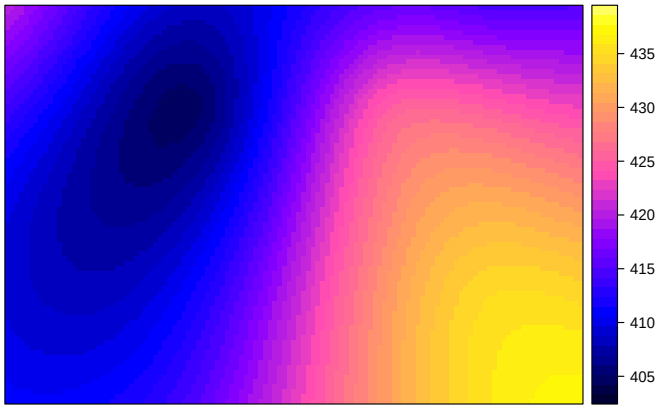


Figure 3: Interpolation TPS surface of the precipitation

Previously, just η was optimized with the function `graph.rbf`, obtaining a value $\eta = 0.1460814$ with the RBF "TPS" and the result is shown below:

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")
```

With all of this, it is obtained $RMSPE = 7.691789$.

The function `RBF.phi()` requires; a distance between a couple of points $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$, a smoothing factor `eta` and an 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 an RBF `func`, explained above. The result of this function is necessary for constructing 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* due to the lack of variances for the error predictions, which does not occur with the kriging computed data.

The function `graph.rbf()` allows observing the behavior of the smoothing and robustness parameters, η and ρ respectively, and finds 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 and 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 the table that contains the values associated to the graph with the optimal values obtained. Following, the graph obtained with 5 RBF in a determined interval for the parameter $0 \leq \eta \leq 1.6$ is shown.

As can be seen in the Figure 4, the lowest $RMSPE$ is generated with a thin plate spline. 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)
```

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

4. Application

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

4.1. Database

The daily mean terrestrial temperature in Croatia was studied by Hengl (2009). The information was taken from http://spatial-analyst.net/book/HRclim2008 and was provided

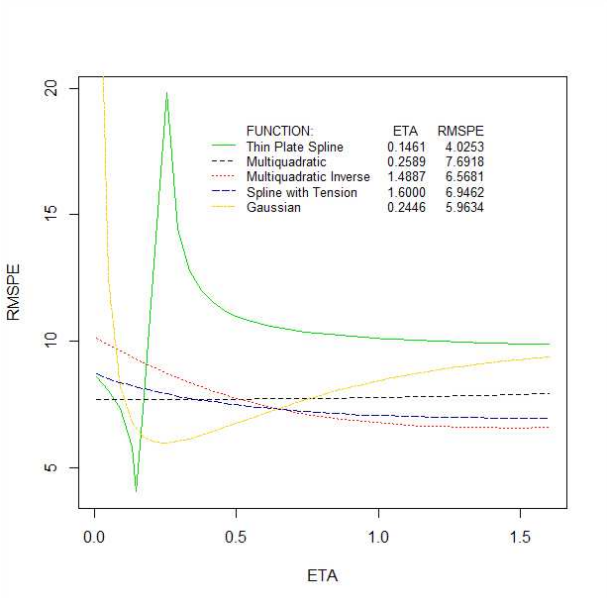


Figure 4: Optimization of η, in radial basis functions

by Melita Perčec Tadić, from the Croatian Meteorological and Hydrological Service (DHMZ) (Hengl 2009). Croatia is a relatively small country, but it counts with several regions of different weather as a result of its specific position near the Adriatic Sea and the diverse topography from plains on the east, through the central mountainous zone that separates the continental territory from the maritime part of the country. The study region is characterized by a wide range of topographic and climatic features. This allows to assess correctly the proposed methodology and compare it with the traditional one, given that the mean Earth's temperatures in such a region are strongly influenced by topography.

The mean temperature was observed from January to December in 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:

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

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, just 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 monthly mean terrestrial temperature in Croatia only for January in 2008 from 146 meteorological stations.

The geographical coordinates (latitude and longitude) were transformed to a system of Carte-

Function	Description
`criterio.cv()`	Generates a object of class data.frame with a summary of statistics associated values with the cross validation: MPE, $ASEPE$, $RMSPE$, $MSPE$, $RMSSPE$, $MAPPE$, $CCPE$, R^2 y $pseudoR2$.
`network.design`	Generates an average with the standard kriging errors of the prediction errors (ASEPE) associated to a conditional net design.
`est.variograms()`	Calculates the experimental variogram: classic, robust, median and trimmed mean. (Cressie 1993; Bárdossy 2001)
`pocketplot()`	Graphs the pocket plot of probability or standardized variances on directions north-south or east-west, (see Cressie (1993))
`rbf()`	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()`	Extracts a $RMSPE$ value, resulting from the leave-one-out cross validation
`rbf.cv1()`	Returns a $RMSPE$ value, resulting from the LOOCV (allows the simultaneous optimization of the parameters `eta` and `rho`).
`RBF.phi()`	Produces a numeric value obtained from the radial basis function, 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()`	Generates a table with the coordinates of the data, the predictions, the observed values, the residual values, the prediction variances, the zscore values (the residual value divided by the standard error) of the analyzed variable, which are the result of the LOOCV
`graph.rbf()`	Produces a graph that describes the behavior of the parameters `eta` and `rho`, associated to the radial basis function.

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

sian 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 6a.

4.2. Prediction Map

Initially, it is performed the optimization of the parameters of smoothing η and ρ from the data described before. For this aim, it is used the function `graph.rbf()`, which is evaluated in the radial basis functions mentioned in 2.3, finding one of the better adjustments in 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()`, that uses `optimize()` (for one parameter) and `bobyqa()` (for both parameters) respectively. In the optimization, it was not used the function `optim()`, described in Mittelhammer, Judge, and Miller (2000) because it has need of more time than `optimize()`. For optimizing both parameters, it is necessary that `eta.opt=TRUE` and `rho.opt=TRUE`, thus will be obtained a lattice graph (for the case Figure5).

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()` (for the case,

`bobyqa()` converges rapidly in the iteration number 19), in an interval for such parameters established by the user. For instance, in multiquadratic functions the optimal values for η and ρ are usually found close to 0. Provided that, an appropriate interval would be between 0 and 1 as shown by the Figure 5.

The data and the shapefile are pre-charged; subsequently, a 70000 points grid is generated within the analyzed region, aimed 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 coordinates of the new points "`newdata`", the number of neighbors "`n.neigh`", and the type of radial basis function "`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()` it is obtained the predictions map of the analyzed variable, which is shown in the Figure 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))
```

5. Conclusions

The functions contained within the library **geospt** are programmed on R code and generally produce results in a short time. Nonetheless, some of them can be programmed on `C` in order to obtain faster results.

The last version 1.0-1 of **geospt** allows performing the optimization for the parameters η and ρ, in the case of the use of the radial basis functions to obtain the interpolation surfaces, which implies a greater demand of computational time. For this reason, it is recommended the usage of modern and high capability processors or the limitation of the number of iterations in order

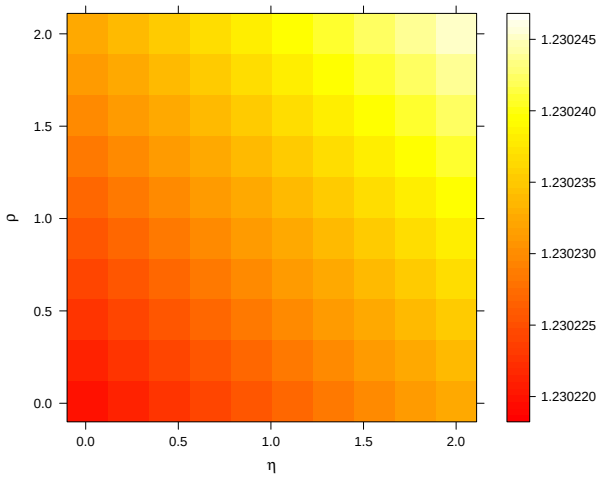
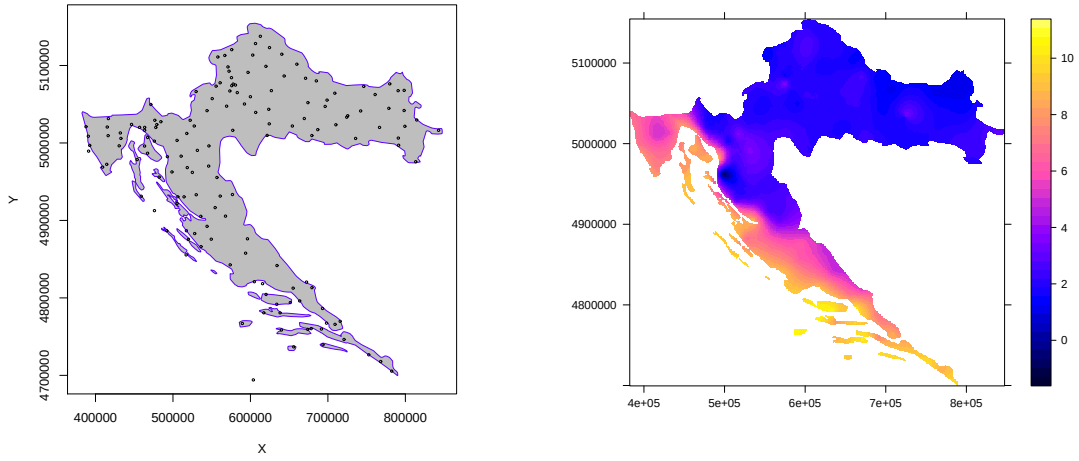


Figure 5: Lattice of η y ρ for multicuadratic function



(a) Locations of meteorological stations

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

Figure 6: Croatia Map

to obtain results in the minor possible time. The instruction `graph.rbf` for the case where `eta.opt` and `rho.opt` are equal to *TRUE*, allows obtaining a Lattice chart with the different combinations between the values of the parameters for a range predefined by the user for such parameters.

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 obtaining better results than when the trend modeling is omitted, in spite of the non-stationarity of the observed process.

6. Acknowledgments

Core Spatial Data Research (Faculty of Engineering, Universidad Distrital Francisco José de Caldas) and Applied Statistics in Experimental Research, Industry and Biotechnology (Universidad Nacional de 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." *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.

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.

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

Oscar O. Melo
Departamento de Estadística, Facultad de Ciencias
Universidad Nacional Colombia
11001 Bogotá D.C., Colombia
E-mail: oomelom@unal.edu.co
Telephone: +57 /1/3165000/ext 13169

Sandra E. Melo
Departamento de Agronomia, Facultad de Ciencias Agrarias
Universidad Nacional Colombia
11001 Bogotá D.C., Colombia
E-mail: semelo@unal.edu.co
Telephone: +57 /1/3165000/ext 19120

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