Environmetrics
The official journal of The International Environmetrics Society
An Association of The International Statistical Institute

Interpolation using distance-based space-time radial basis functions

Journal:	*Environmetrics*
Manuscript ID	env-19-0034
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	15-Mar-2019
Complete List of Authors:	Melo, Carlos; Universidad Distrital Francisco José de Caldas, Faculty of Engineering Melo, Oscar; Universidad Nacional de Colombia, Statistics Mateu, Jorge; University Jaume I, Mathematics
Keywords:	Distance-based methods, Land surface temperature, Principal Coordinates, Radial basis functions, Smoothing parameter, Spatio-temporal prediction
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
space time RBF 13-03-19.tex svjour3.cls rbfca.bib	

Environmetrics manuscript No.
(will be inserted by the editor)

Interpolation using distance-based space-time radial basis functions

Carlos E. Melo · Oscar O. Melo · Jorge Mateu

Abstract We propose a new method of spatio-temporal interpolation using distance-based space-time radial basis functions, which combines radial basis functions and principal coordinates. We consider space-time interpolation using radial basis functions in a space-time metric model, with a trend built from mixed variables transformed to principal coordinates. In order to evaluate the interpolation accuracy of our method, a simulation study for space-time Gaussian random fields is carried out for five distinct radial basis functions under a variety of practical scenarios, incorporating principal coordinates. We analyze Earth's monthly average temperature in Croatia, incorporating the trend from the principal coordinates obtained as a function of the space-time locations, distance from the sea, elevation, and season. Our technique is recommended since it does not require the estimation of a space-time variogram, which usually requires too much computational time. In addition, the

Carlos E. Melo

Faculty of Engineering, Universidad Distrital Francisco José de Caldas

Tel.: +057-1-2059671

E-mail: cmelo@udistrital.edu.co

Oscar O. Melo

Department of Statistics, Faculty of Sciences, Universidad Nacional de Colombia

Tel.: +057-1-2059671

E-mail: oomelom@unal.edu.co

Jorge Mateu

Department of Mathematics, Universitat Jaume I, Campus Riu Sec, E-12071, Castellón, Spain

Fax: +34-964-728429

E-mail: mateu@mat.uji.es

combination of radial basis functions and a distance-based method provides more flexibility in the estimation of the smoothing parameters.

Keywords Distance-based methods · Land surface temperature · Principal Coordinates · Radial basis functions · Smoothing parameter · Spatio-temporal prediction

1 Introduction

In the last years, there has been tremendous growth in models and techniques for analyzing spatio-temporal data. In Rouhani & Hall (1989), Rouhani & Myers (1990), Dimitrakopoulos & Luo (1994), Kyriakidis & Journel (1999), Cressie & Huang (1999), Christakos (2000), De Cesare et al. (2001), Gneiting & Schlather (2002), Myers et al. (2002), Mateu et al. (2003), Kolovos et al. (2004), Banerjee et al. (2004), Sahu & Mardia (2005), Chen et al. (2006), Gneiting et al. (2007), Bellier et al. (2007), Planque et al. (2007), Ye (2008), Ye et al. (2011), Kondoh et al. (2011), among others, we can find a summary of the main techniques for modeling space-time data, along with numerous practical applications for a variety of natural phenomena. For instance, measurements of air pollution (De Cesare et al. 1997, Şen et al. 2006), precipitation (Yavuz & Erdoğan 2012) and ground temperature (Perčec-Tadić 2010, Hengl et al. 2012) are often observed daily over a number of places in many countries worldwide. In this way, for statistical and geographical analysis, the availability of satellite images and designed software motivates the embodiment of the present research.

Global methods, such as trend surface analysis, use all available data for prediction; instead local methods, including inverse distance weighting, usually use only some subset of the data for prediction. One benefit of local methods is that computational time is reduced by using only some subset of the data for prediction purposes. So, radial basis functions (RBFs) such as multiquadratic (MQ) or completely regularized spline (CRS) are useful in the construction of digital elevation models, as shown in Mitášová & Hofierka (1993). One variation of the multiquadratic function is called the inverse multiquadratic (IMQ), introduced by Hardy & Gopfert (1975). Späh (1969) described a method that allows to avoid inflexion points and contains cubic splines as a special case, using piecewise cubic and exponential (EXP) spline interpolation. Later, the thin plate spline (TPS) was introduced in a

geometric design by Duchon (1976), and the Gaussian (GAU) approximation used by Schagen (1979) was presented as a popular variant to TPS. Finally, Mitáš & Mitášová (1988), Mitášová & Hofierka (1993) and Mitášová & Mitáš (1993) developed the formulation of the spline with tension (ST), and implemented a segmentation algorithm with a flexible size of overlapping neighborhood.

The link between splines and kriging was called "near" equivalent, because the TPS corresponds to a specific generalized covariance, whereas the kriging estimator or the RBF interpolator only require the use of a kernel with appropriate positive definiteness properties. In general, this allows adapting the kernel function to a particular data set (Cressie 1989, Myers 1992). Then, for both splines and kriging, the smoothing is determined objectively.

In this paper, we consider the problem of choosing a model based on distances that incorporates information that influences the response variable. In particular, in spatio-temporal data analysis, we often have to deal with variables of different nature that are associated with the response variable: categorical and binary variables such as type of soil or rock, and continuous variables (e.g. the spatial locations or additional covariates). Our aim here is to present a unified approach that uses RBFs in such spatio-temporal contexts where the explanatory variables are mixed. Therefore, this paper proposes a new method using distances between individuals such as the Gower's distance (Gower 1968), although some other Euclidean distances may be used. Our distance-based spatio-temporal radial basis functions (DBSTRBFs) method is applied in space-time metric models to predict the trend and to estimate the covariance structure when the explanatory variables are mixed.

The distance-based (DB) spatio-temporal method we propose is based on methods developed by Cuadras & Arenas (1990) and Cuadras et al. (1996), who presented some results of a DB model for prediction with mixed variables. This strategy is an excellent alternative because it takes full advantage of the information obtained due to the relationship between observations, which can be established through the use of spectral decomposition using any Euclidean distance. Consequently, this approach allows for any number of principal coordinates at the sample locations to be included in the model to improve the overall predictions.

In our case, the principal coordinates are obtained using a DB method from the covariates associated with the response variable, and the spatio-temporal locations defined by polynomials of degree 1, 2 or 3. The selection of the principal coordinates is performed using the higher t-test values and a significant fall in the lack of predictability (see the beginning of Section 2), i.e. the principal coordinates that are more associated with the response variable. In this way, to evaluate the interpolation accuracy of our method, unconditional simulations were performed for five distinct radial basis functions under a variety of practical scenarios, and the results show that RBFs using a DB method have the advantage of working with mixed variables in the trend and do not require the estimation of a space-time variogram, which usually requires too much computational time.

This paper is organized as follows. Section 2 develops the methodological proposal introducing the local linear DB trend. In that section, RBFs are constructed from the DB trend, some RBFs are described, and an approximation from the spline interpolation to the kriging model for the prediction is shown. Section 3 presents an intensive simulation study based on several space-time spline models. Section 4 develops an application that illustrates the proposed methodology. We analyze the Earth's monthly average temperature in Croatia, where the trend is incorporated from the principal coordinates obtained as a function of the space-time locations, distance from the sea, elevation, and season. The paper ends with some conclusions.

2 Local linear distance-based trend

Let $\{Z(\boldsymbol{s},t), \boldsymbol{s} \in \boldsymbol{D}, t \in Q\}$ be a random spatio-temporal process. Here $\boldsymbol{s}$ varies on a given set $\boldsymbol{D} \subseteq \mathbb{R}^d$ (usually, $d = 2$), and $Q \subseteq \mathbb{Z}_+$ or $\mathbb{R}_+$, so that the models developed are suitable for both a discrete-time and a continuous-time setting. Without loss of generality, we take $Q \subseteq \mathbb{R}_+$. Suppose that this process is observed on a set of spatio-temporal locations $\{(\boldsymbol{s}_1,t_1),\ldots,(\boldsymbol{s}_n,t_1),\ldots,(\boldsymbol{s}_1,t_T),\ldots,(\boldsymbol{s}_n,t_T)\}$ obtaining a set of values $\{Z(\boldsymbol{s}_1,t_1),\ldots,Z(\boldsymbol{s}_n,t_1),\ldots,Z(\boldsymbol{s}_1,t_T),\ldots,Z(\boldsymbol{s}_n,t_T)\}$.

Assume the stochastic process follows a random function model, which can be written as

$$Z(\boldsymbol{s}_i,t_j) = \mu(\boldsymbol{s}_i,t_j) + \varepsilon(\boldsymbol{s}_i,t_j), \qquad i = 1,...,n, j = 1,...,T \tag{1}$$

where $\boldsymbol{Z}(\boldsymbol{s}_i, t_j)$ is the regionalized variable taken at the location $(\boldsymbol{s}_i, t_j)$, $\mu(\boldsymbol{s}_i, t_j)$ is a deterministic function associated with the trend at location $(\boldsymbol{s}_i, t_j)$, and $\varepsilon(\boldsymbol{s}_i, t_j)$ is a stationary stochastic component at location $(\boldsymbol{s}_i, t_j)$ with zero mean and variogram $2\gamma(\cdot, \cdot)$. The trend vector is formed by categorical, continuous and binary variables, and it is modeled as

$$\mu(\boldsymbol{s}_i, t_j) = \theta_0 + \boldsymbol{v}'(\boldsymbol{s}_i, t_j)\boldsymbol{\theta}, \qquad i = 1, ..., n, j = 1, ..., T \tag{2}$$

where θ_0 is a parameter related to the unknown intercept, $\boldsymbol{v}(\boldsymbol{s}_i, t_j) = (v_1(\boldsymbol{s}_i, t_j), \ldots, v_p(\boldsymbol{s}_i, t_j))'$ is a vector containing explanatory variables associated to the spatio-temporal location $(\boldsymbol{s}_i, t_j)$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$ is a vector of unknown parameters.

In matrix form, model (1) can be expressed as

$$\boldsymbol{Z}_{st} = \mathbf{1}\theta_0 + \boldsymbol{V}\boldsymbol{\theta} + \boldsymbol{e}_{st} \tag{3}$$

where $\mathbf{1}$ is a vector of dimension $nT \times 1$ which is associated with the intercept, $\boldsymbol{Z}_{st} = (Z(\boldsymbol{s}_1, t_1), \ldots, Z(\boldsymbol{s}_n, t_1), \ldots, Z(\boldsymbol{s}_1, t_T), \ldots, Z(\boldsymbol{s}_n, t_T))'$, $\boldsymbol{V} = (\boldsymbol{V}_1, ..., \boldsymbol{V}_p)$ is a design matrix of dimension $nT \times p$ with p explanatory variables which are given by $\boldsymbol{V}_k = (v_k(\boldsymbol{s}_1, t_1), \ldots, v_k(\boldsymbol{s}_n, t_1), \ldots, v_k(\boldsymbol{s}_1, t_T), \ldots, v_k(\boldsymbol{s}_n, t_T))'$, $k = 1, ..., p$, and $e_{st} = (e(\boldsymbol{s}_1, t_1), \ldots, e(\boldsymbol{s}_n, t_1), \ldots, e(\boldsymbol{s}_1, t_T), \ldots, e(\boldsymbol{s}_n, t_T))'$. For the sake of simplicity, we assume that $\boldsymbol{V} = \boldsymbol{H}\boldsymbol{V}^*$ where $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{nT}\mathbf{1}\mathbf{1}'$ is the centering matrix with $\boldsymbol{I}$ the identity matrix of size $nT \times nT$, and $\boldsymbol{V}^*$ is the matrix of original explanatory variables that may involve continuous, categorical and binary variables, or even a mixture of them. Note that it is not necessary to centralize the covariates in the proposed model; this is performed with the purpose to build the distances, but depending on the distance used we do not always need to centralize the covariates.

Now, the idea is to transform the explanatory variables using the method based on distances. To do this we must define some similarity (or Euclidean distance) measures that depend on the explanatory variables characteristics. Distance-based regression was introduced by Cuadras (1989) and has been developed in Boj et al. (2007), Boj et al. (2010), Cuadras & Arenas (1990) and Cuadras et al. (1996). This DB method is an excellent alternative because it takes full advantage of the information obtained due to the relationship between observations, and can be established through the use of the spectral decomposition using any Euclidean distance. We use the DB method in the spatio-temporal model

to estimate both the trend and the covariance structure, which are fundamental in interpolation and monitoring problems.

Let $\Omega = \{\vartheta_1, \ldots, \vartheta_{nT}\}$ be a set consisting of n individuals in T times. Let $\delta_{rr'} = \delta(\vartheta_r, \vartheta_{r'}) = \delta(\vartheta_{r'}, \vartheta_r) \geq \delta(\vartheta_r, \vartheta_r) = 0$ be a distance (or dissimilarity) function defined on Ω. Suppose that the distance matrix with dimension $nT \times nT$, $\Delta = (\delta_{rr'})$ is Euclidean.

If the vector $\boldsymbol{v}(\boldsymbol{s}_i, t_j)$ given in (2) is formed by binary, categorical and continuous variables, then the similarity according to Gower (1968) can be defined for mixed variables as

$$m_{rr'} = \frac{\sum_{h=1}^{p_1} \left(1 - \frac{|v_h(\boldsymbol{a}_r) - v_h(\boldsymbol{a}_{r'})|}{G_h}\right) + c_{1rr'} + \varpi_{rr'}}{p_1 + (p_2 - c_{2rr'}) + p_3}, \qquad r, r' = 1, ..., nT \tag{4}$$

where G_h is the range of the h-th continuous variable, p_1 is the number of continuous variables, $c_{1rr'} = c_1(\boldsymbol{a}_r, \boldsymbol{a}_{r'})$ and $c_{2rr'} = c_2(\boldsymbol{a}_r, \boldsymbol{a}_{r'})$ are the number of positive and negative matches, respectively, for p_2 binary variables, and $\varpi_{rr'} = \varpi(\boldsymbol{a}_r, \boldsymbol{a}_{r'})$ is the number of matches for p_3 multistate variables, with $\boldsymbol{a}_r$ a vector of spatio-temporal locations. In the case that the explanatory variables are binary or categorical, the similarity (4) reduces to the Sokal-Michener similarity coefficient

$$m_{rr'} = \frac{c_{1rr'} + c_{4rr'}}{c_{1rr'} + c_{2rr'} + c_{3rr'} + c_{4rr'}}$$

where $c_{1rr'}$, $c_{2rr'}$, $c_{3rr'}$, $c_{4rr'}$ are the frequencies of (1,1), (1,0), (0,1) and (0,0), respectively. Note that for l categories of one categorical variable, $l - 1$ dummy variables are created of ones and zeros corresponding to the presence or absence of one category (e.g., with 1 for the presence of A category in that variable and zero for the non-presence of that category, and we did the same procedure for $B, C, ...$ categories). Therefore, the Sokal-Michener similarity coefficient can be applied because we have $l - 1$ binary variables.

Through the transformation

$$\delta_{rr'} = \sqrt{1 - m_{rr'}} \tag{5}$$

it is possible to obtain Euclidean distances. If all explanatory variables in (2) are continuous, the square distance would be defined as

$$\begin{aligned} \delta^2_{rr'} =& (\boldsymbol{v}(\boldsymbol{a}_r) - \boldsymbol{v}(\boldsymbol{a}_{r'}))'(\boldsymbol{v}(\boldsymbol{a}_r) - \boldsymbol{v}(\boldsymbol{a}_{r'})) \\ =& \boldsymbol{v}'(\boldsymbol{a}_r)\boldsymbol{v}(\boldsymbol{a}_r) + \boldsymbol{v}'(\boldsymbol{a}_{r'})\boldsymbol{v}(\boldsymbol{a}_{r'}) - 2\boldsymbol{v}'(\boldsymbol{a}_r)\boldsymbol{v}(\boldsymbol{a}_{r'}) \end{aligned} \tag{6}$$

or alternatively, the absolute distance $\delta^2_{rr'} = \sum_{h=1}^{p} \left| v_h(\boldsymbol{a}_r) - v_h(\boldsymbol{a}_{r'}) \right|$. Thus, in the case of information based only on the spatial locations at time t, (w_x, w_y, t), the distances are directly given by

$$\delta_{rr'} = \sqrt{(w_{x_r} - w_{x_{r'}})^2 + (w_{y_r} - w_{y_{r'}})^2 + (t_r - t_{r'})^2}$$

Expressions for the Gower similarity as given in equation (4) will be useful to the extent of having information associated with mixed variables, not only for the sampling sites at time t but also for the non-sampled, which limits its use in unsampled areas.

After selecting a distance, let $\boldsymbol{A}_{nT\times nT} = (a_{rr'})$ be the matrix with entries $a_{rr'} = -\delta^2_{rr'}/2$ and $\boldsymbol{B} = \boldsymbol{HAH}$ where $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{nT}\mathbf{1}\mathbf{1}'$ is the above centering matrix. It is known that $\boldsymbol{B}$ is a semi-definite positive matrix (Mardia et al. 2002) of rank k, so the matrix $\boldsymbol{X}$ of principal coordinates is obtained from the spectral decomposition as

$$\boldsymbol{B} = \boldsymbol{HAH} = \boldsymbol{U\Lambda U}' = \boldsymbol{XX}' \tag{7}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{A}$, $\boldsymbol{X} = \boldsymbol{U\Lambda}^{1/2}$ is a $nT \times nT$ matrix of rank $k \leq nT - 1$, and $\boldsymbol{U}$ contains the standardized coordinates.

Thus, the trend given in (3) becomes

$$\boldsymbol{Z}_{st} = \beta_0\mathbf{1} + \boldsymbol{X\beta} + \boldsymbol{\xi}_{st} \tag{8}$$

where $\boldsymbol{Z}_{st}$ is given in (3), β_0 and $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)'$ are the unknown parameters, $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_k)$ is the principal coordinates matrix, and $\boldsymbol{\xi}_{st} = (\xi(\boldsymbol{s}_1, t_1), ..., \xi(\boldsymbol{s}_n, t_T))'$ with each $\xi(\boldsymbol{s}_i, t_j)$ a stationary stochastic component with zero mean and variogram $2\gamma(\cdot, \cdot)$. Note that $\mathbf{1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ are eigenvectors of $\boldsymbol{B}$ with eigenvalues $0, \lambda_1, \ldots, \lambda_k$, respectively, and $\boldsymbol{X}_i'\boldsymbol{X}_i = \lambda_i$, $\boldsymbol{X}_i'\boldsymbol{X}_j = 0$ $(i \neq j)$, and $\boldsymbol{X}_i'\mathbf{1} = 0$, $i, j = 1, \ldots, k$.

To avoid the problem of having a coefficient of determination $R^2 \simeq 1$ when the rank of $\boldsymbol{X}$ is $k = nT - 1$, it is necessary to consider only the most correlated eigenvectors of $\boldsymbol{B}$ given by $(\mathbf{1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$, with the regionalized variable $\boldsymbol{Z}_{st}$, i.e., the most significantly correlated covariates with $\boldsymbol{Z}_{st}$. In order to decide whether a predictor variable, i.e., a column of $\boldsymbol{X}$, has to be included or deleted, the variables can be arranged in descending order of absolute correlations with $\boldsymbol{Z}_{st}$,

$$r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_1) > \cdots > r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_k) > \cdots > r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_{nT-1}) \tag{9}$$

where $r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_i) = \frac{\boldsymbol{Z}'_{st}\boldsymbol{X}_i\boldsymbol{X}'_i\boldsymbol{Z}_{st}}{\lambda_i \sum_{i'=1}^{nT}(Z(\boldsymbol{a}_{i'})-\bar{Z})^2}$, $i = 1, \ldots, nT-1$, with $\bar{Z} = \frac{\sum_{i=1}^{nT} Z(\boldsymbol{a}_i)}{nT}$.

Moreover, a principal coordinate $\boldsymbol{X}_i$ should be deleted if the null hypothesis $\beta_i = 0$ is not rejected. A test can be built as (Cuadras et al. 1996)

$$u_i = \frac{\hat{\beta}_i}{\|\boldsymbol{Z}_{st} - \hat{\beta}_0\mathbf{1} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2}\sqrt{\lambda_i(nT-k-1)}, \qquad i = 1, \ldots, nT-1 \tag{10}$$

where $\hat{\beta}_0 = \bar{Z}$, $\hat{\boldsymbol{\beta}} = \boldsymbol{\Lambda}^{-1}\boldsymbol{X}'\boldsymbol{Z}_{st}$ and $\hat{\beta}_i$ is the i-th component of $\hat{\boldsymbol{\beta}}$. u_i follows a t-student distribution with $(nT-k-1)$ degrees of freedom.

Another possibility is through the explained variability given by the predictor variables, which are established by the largest eigenvalues $\lambda_1 > \cdots > \lambda_k > \lambda_{k+1} > \cdots > \lambda_{nT-1}$ choosing the k first principal coordinates. However, a dimension with small eigenvalue can be correlated with the response variable. Then, this dimension is correlated with "noise" rather than the main variability of the data (Cuadras 1993).

Another good alternative for the selection of principal coordinates is performed in a similar way to the selection of the number of variables in multivariate regression, using the statistic called $C_p - Mallows$. That is, we draw a graph representing the points $(i, 1-c(i))$ $i = 1, \ldots, k, k+1, \ldots, nT-1$, and then the point with a significant fall in the lack of predictability, given by $1-c(i)$, is determined. The predictability $c(i)$ is given by

$$c(0) = 0, \qquad c(i) = \frac{\sum_{j=1}^{i} r_j^2\lambda_j}{\sum_{j=1}^{nT-1} r_j^2\lambda_j} \tag{11}$$

where $r_j^2 = r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_j)$, and λ_j is the j-th eigenvalue associated with $\boldsymbol{X}_j$, $j = 1, 2, \ldots, k, k+1, \ldots, nT-1$ (see Cuadras et al. (1996) for more details).

Finally, in any of the four methods presented above, the $\boldsymbol{X}_{k+1}, \ldots, \boldsymbol{X}_{nT-1}$ principal coordinates must be removed.

2.1 Relation with the classical spatio-temporal geostatistical model

Model (8) depends on the chosen distance $\delta_{rr'}$. Therefore, when the predictor variables are continuous and the Euclidean distance is used, the DB spatio-temporal geostatistical model is compatible with the

classical spatio-temporal geostatistical model. This equivalence also holds for qualitative and mixture (continuous and categorical) explanatory variables when an appropriate dissimilarity measure is used.

If all the explanatory variables in (3), $\boldsymbol{V}^*$, are continuous, the squared distance is given by (6). Thus, the distance matrix $\boldsymbol{D} = (\delta_{rr'})$ is obtained, and

$$\boldsymbol{A} = -\frac{1}{2}\left[\text{diag}(\boldsymbol{V}^*\boldsymbol{V}^{*\prime})\mathbf{1}' + \mathbf{1}\text{diag}(\boldsymbol{V}^*\boldsymbol{V}^{*\prime})' - 2\boldsymbol{V}^*\boldsymbol{V}^{*\prime}\right]$$

where $\text{diag}(\boldsymbol{V}^*\boldsymbol{V}^{*\prime})$ is a vector which contains the diagonal terms of the matrix $\boldsymbol{V}^*\boldsymbol{V}^{*\prime}$. Therefore,

$$\boldsymbol{B} = \boldsymbol{H}\boldsymbol{A}\boldsymbol{H} = \boldsymbol{H}\boldsymbol{V}^*\boldsymbol{V}^{*\prime}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{V}' = \boldsymbol{X}\boldsymbol{X}'$$

because $\boldsymbol{H}[\text{diag}(\boldsymbol{V}^*\boldsymbol{V}^{*\prime})\mathbf{1}']\boldsymbol{H} = \mathbf{0}$ and $\boldsymbol{H}[\mathbf{1}(\text{diag}(\boldsymbol{V}^*)\boldsymbol{V}^{*\prime})']\boldsymbol{H} = \mathbf{0}$. Then, the spatio-temporal geostatistical model based on distances introduced in (8) is a centered spatio-temporal geostatistical model (3), i.e., it produces the same predictions at p dimensions than the model given in (3).

However, it would not be necessary to consider an Euclidean p-dimensional distance as given in Equation (6). Let E_l $(k \leq l \leq nT-1)$ be the space spanned by the columns of $\boldsymbol{X}$, where $\boldsymbol{X}$ is a metric scaling solutions obtained from a distance applied to the same data. Then, by taking $k > p$, i.e., the most suitable columns of $\boldsymbol{X}$, the DB spatio-temporal geostatistical model improves the classical spatio-temporal geostatistical model when $(\boldsymbol{Z}_{st} - \hat{\theta}_0\mathbf{1}) \in E_l$. Note that this is always true for $l = nT-1$ with $l > p$.

To illustrate the comparison between the DB spatio-temporal geostatistical model and the classical spatio-temporal geostatistical model, let us focus on the R_k^2 coefficient that can be written as

$$R_k^2 = \sum_{j=1}^{k} r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_j)$$

Note that $R_k^2 < R_{k+1}^2$ because the principal coordinates $\boldsymbol{X}_j$'s $(j = 1, ..., k)$ are uncorrelated with the principal coordinate $\boldsymbol{X}_{k+1}$. Moreover, R_k^2 is maximized for a given k provided that the first k ordered principal coordinates are selected.

On the other hand, we can write

$$R^2(\boldsymbol{Z}_{st}, \boldsymbol{V}^*) \geq \sum_{j=1}^{p} r^2(\boldsymbol{Z}_{st}, \boldsymbol{V}_j^*)$$

where $R^2(\boldsymbol{Z}_{st}, \boldsymbol{V}^*)$ is the coefficient of determination between $\boldsymbol{Z}_{st}$ and $\boldsymbol{V}^*$ in the classical spatio-temporal geostatistical model, and $r^2(\boldsymbol{Z}_{st}, \boldsymbol{V}_j^*)$ is the correlation coefficient between $\boldsymbol{Z}_{st}$ and the j-th explanatory variable ($j = 1, ..., p$), $\boldsymbol{V}_j^*$.

Without loss of generality, suppose $p = 1$ and note that applying the DB spatio-temporal geostatistical model, R_l^2 can be written as

$$R_l^2 = \sum_{j=1}^{l} r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_j), \quad 1 \leq k \leq l \leq nT - 1$$

because the principal coordinates $(\boldsymbol{X}_1, ..., \boldsymbol{X}_{nT-1})$ are uncorrelated. Also, we can represent as $R^2(\boldsymbol{Z}_{st}, \boldsymbol{V}_1^*)$, the coefficient of determination between $\boldsymbol{Z}_{st}$ and the explanatory variable $\boldsymbol{V}_1^*$ using the classical spatio-temporal geostatistical model. Therefore, if we take $l = k > 1$ (e.g., $k = 2$) then we find that

$$R_l^2 = \sum_{j=1}^{l} r^2(\boldsymbol{Z}_{st}, \boldsymbol{X}_j) \geq R^2(\boldsymbol{Z}_{st}, \boldsymbol{V}_1^*)$$

So, the DB geostatistical model improves the classical geostatistical model. However, this inequality is not too relevant if $R^2(\boldsymbol{Z}_{st}, \boldsymbol{V}_1^*)$ is close to 1, and so, the classical geostatistical model is sufficiently good, and thus the DB spatio-temporal geostatistical model improvement is not necessary. This result holds when the explanatory variables are qualitative or mixed. This approach aims at improving the predictions by allowing for inclusion of any number of principal coordinates at the sample locations. Note that universal kriging traditionally considers only information of a matrix $\boldsymbol{V}$ of size $nT \times p$. Therefore, when we want to improve predictions regardless of the effect that some covariates have on the spatial response variable, the methodology proposed in this paper can be used.

2.2 Radial basis functions built from the distance-based spatio-temporal drift

In spatio-temporal interpolation, there exist methods that do not require information from a spatio-temporal dependence model such as the variogram or covariogram, these are called deterministic and are the ones of interest in this section. Model (1), using a DB format, can be expressed in general form as

$$Z(\boldsymbol{s}_i, t_j) = g(\boldsymbol{s}_i, t_j) + \varepsilon(\boldsymbol{s}_i, t_j), \qquad i = 1, ..., n,\ j = 1, ..., T \tag{12}$$

where $\varepsilon(\boldsymbol{s}_i, t_j)$ is the unobserved noise term which may or may not be included depending on the application, and it is assumed to be normal. In addition, $g(\boldsymbol{s}_i, t_j)$ is a real-valued function given by

$$\begin{aligned} g(\boldsymbol{s}_i, t_j) =& \sum_{l=0}^{k} \nu_l x_l(\boldsymbol{s}_i, t_j) + \kappa(\boldsymbol{s}_i, t_j), \quad i = 1, ..., n,\ j = 1, ..., T \\ =& \sum_{l=0}^{k} \nu_l x_l(\boldsymbol{s}_i, t_j) + \sum_{i'=1}^{n} \sum_{j'=1}^{T} \omega_{i'j'} \phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'}) \\ =& \boldsymbol{x}'(\boldsymbol{s}_i, t_j)\boldsymbol{\nu}_{st} + \boldsymbol{\phi}'(\boldsymbol{s}_i, t_j)\boldsymbol{\omega}_{st} \end{aligned} \tag{13}$$

where $\kappa(\boldsymbol{s}_i, t_j) = \sum_{i'=1}^{n} \sum_{j'=1}^{T} \omega_{i'j'} \phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'})$, $\boldsymbol{x}'(\boldsymbol{s}_i, t_j) = (1, x_1(\boldsymbol{s}_i, t_j), ..., x_k(\boldsymbol{s}_i, t_j))'$ with each $x_l(\boldsymbol{s}_i, t_j)$ a linearly independent real-valued function in the locations of $(\boldsymbol{s}_i, t_j)$, $\boldsymbol{\nu}_{st} = (\nu_0, \nu_1, ..., \nu_k)'$ with each ν_l a trend model coefficient, $\boldsymbol{\phi}(\boldsymbol{s}_i, t_j) = (\phi(\boldsymbol{s}_i - \boldsymbol{s}_1, t_j - t_1), ..., \phi(\boldsymbol{s}_i - \boldsymbol{s}_n, t_j - t_T))'$ with $\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'})$ a basis function, i.e., a scalar function of the Euclidean distance between $(\boldsymbol{s}_i, t_j)$ and $(\boldsymbol{s}_{i'}, t_{j'})$, and $\boldsymbol{\omega}_{st} = (\omega_{11}, ..., \omega_{nT})'$ with $\omega_{i'j'}$ an unknown weight.

Writing (13) in matrix form,

$$\boldsymbol{g}_{st} = \boldsymbol{X}_{st}\boldsymbol{\nu}_{st} + \boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st} \tag{14}$$

where $\boldsymbol{X}_{st} = (\mathbf{1}, \boldsymbol{X}) = (\mathbf{1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_k)$ is a $nT \times (k+1)$ matrix with elements $\mathbf{1}$ and $\boldsymbol{X}_l = (x_l(\boldsymbol{s}_1, t_1), ..., x_l(\boldsymbol{s}_n, t_T))'$ $(l = 1, \ldots, k)$, and $\boldsymbol{\Phi}_{st}$ is a $nT \times nT$ matrix with elements $\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_i - t_{j'})$, $i, i' = 1, ..., n$ and $j, j' = 1, ..., T$.

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

$$\sum_{i=1}^{n} \sum_{j=1}^{T} [Z(\boldsymbol{s}_i, t_j) - g(\boldsymbol{s}_i, t_j))]^2 + \rho \int_{\mathbb{R}^2 \times \mathbb{R}} J_m(\kappa(\boldsymbol{s}, t)) d(\boldsymbol{s}, t) \tag{15}$$

where $J_m(\kappa(\boldsymbol{s}, t))$ is a measure of roughness of the spline function κ (defined in terms of m-th degree derivatives of κ) when $\rho > 0$, which acts as a smoothing parameter. If $\rho = 0$, we can use RBFs such as EXP, GAU, MQ and IMQ (see Table 1).

Writing $m = 2$ and replacing (14) in (15), we obtain

$$\begin{aligned} L(\nu_{st}, \boldsymbol{\omega}_{st}) =& (\boldsymbol{Z}_{st} - \boldsymbol{X}_{st}\boldsymbol{\nu}_{st} - \boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st})'(\boldsymbol{Z}_{st} - \boldsymbol{X}_{st}\boldsymbol{\nu}_{st} - \boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st}) + \rho \int_{\mathbb{R}^2 \times \mathbb{R}} [\kappa''(\boldsymbol{s}, t)]^2 d(\boldsymbol{s}, t) \\ =& (\boldsymbol{Z}_{st} - \boldsymbol{X}_{st}\boldsymbol{\nu}_{st} - \boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st})'(\boldsymbol{Z}_{st} - \boldsymbol{X}_{st}\boldsymbol{\nu}_{st} - \boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st}) + \rho \boldsymbol{\omega}'_{st}\boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st} \end{aligned}$$

where $\int_{\mathbb{R}^2\times\mathbb{R}} [\kappa''(\boldsymbol{s},t)]^2 d(\boldsymbol{s},t) = \int_{\mathbb{R}^2\times\mathbb{R}} \kappa''(\boldsymbol{s},t)[\kappa''(\boldsymbol{s},t)]' d(\boldsymbol{s},t) = \boldsymbol{\omega}'_{st}\boldsymbol{\Phi}_{st}\boldsymbol{\omega}_{st}$ with $\kappa''(\boldsymbol{s},t) = \sum_{i=1}^{n}\sum_{j=1}^{T} \omega_{ij}\phi''(\boldsymbol{s}-\boldsymbol{s}_i, t-t_j) = [\boldsymbol{\phi}''(\boldsymbol{s},t)]'\boldsymbol{\omega}_{st}$, $\boldsymbol{\Phi}_{st} = \int_{\mathbb{R}^2\times\mathbb{R}} \boldsymbol{\phi}''(\boldsymbol{s},t)[\boldsymbol{\phi}''(\boldsymbol{s},t)]' d(\boldsymbol{s},t)$ and $\boldsymbol{\phi}''(\boldsymbol{s},t) = (\phi''(\boldsymbol{s}-\boldsymbol{s}_1, t-t_1), ..., \phi''(\boldsymbol{s}-\boldsymbol{s}_n, t-t_T))'$.

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

$$\begin{pmatrix} \boldsymbol{\Phi}_{st} + \rho\boldsymbol{I} & \boldsymbol{X}_{st} \\ \boldsymbol{X}'_{st} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega}_{st} \\ \boldsymbol{\nu}_{st} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Z}_{st} \\ \boldsymbol{0} \end{pmatrix} \tag{16}$$

where $\boldsymbol{I}$ is the $nT \times nT$ identity matrix and ρ can be interpreted as white noise added to the variances at the data locations. If there were no trend, $\boldsymbol{X}_{st}$ becomes a vector of ones, and $\boldsymbol{\nu}_{st}$ a bias parameter.

We note that positive definiteness plays a central role with respect to the uniqueness of the coefficients in the interpolator $\hat{g}(\boldsymbol{s}_i, t_j)$. In order to generalize the interpolators, it is necessary to consider more general forms of positive definiteness.

Definition 1 Let $\mathbf{1}, \boldsymbol{X}_1, ..., \boldsymbol{X}_k$ be linearly independent real-valued variables defined on $\mathbb{R}^d$, and $\boldsymbol{\Phi}_{st}$ a real symmetric matrix. Then, $\boldsymbol{\Phi}_{st}$ is positive definite with respect to $\mathbf{1}, \boldsymbol{X}_1, ..., \boldsymbol{X}_k$ if and only if for all sets of points $(\boldsymbol{s}_1, t_1), ..., (\boldsymbol{s}_n, t_T)$ in $\mathbb{R}^d \times \mathbb{R}$ we have that $\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{T}\sum_{j'=1}^{T} q_{ij}q_{i'j'}\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'}) \geq 0$ for all q_{ij} $(i, i' = 1, ..., n,\ j, j' = 1, ..., T)$, where q_{ij} is a scalar (not all zero) and such that $\sum_{i=1}^{n}\sum_{j=1}^{T} x_l(\boldsymbol{s}_i, t_j)q_{ij} = 0$ for all $l = 1, ..., k$.

By Definition 1, $\boldsymbol{\Phi}_{st}$ is a positive definite matrix, then $\boldsymbol{\Phi}_{st} + \rho\boldsymbol{I}$ is a non-singular matrix. Therefore, the weights ω_{st} and the parameter vector υ_{st} are estimated, respectively, by

$$\begin{aligned} \widehat{\boldsymbol{\omega}}_{st} =& (\boldsymbol{\Phi}_{st} + \rho\boldsymbol{I})^{-1}\left\{\boldsymbol{I} - \boldsymbol{X}_{st}\left[\boldsymbol{X}'_{st}(\boldsymbol{\Phi}_{st} + \rho\boldsymbol{I})^{-1}\boldsymbol{X}_{st}\right]^{-1}\boldsymbol{X}'_{st}(\boldsymbol{\Phi}_{st} + \rho\boldsymbol{I})^{-1}\right\}\boldsymbol{Z}_{st} \\ \widehat{\boldsymbol{\nu}}_{st} =& \left[\boldsymbol{X}'_{st}(\boldsymbol{\Phi}_{st} + \rho\boldsymbol{I})^{-1}\boldsymbol{X}_{st}\right]^{-1}\boldsymbol{X}'_{st}(\boldsymbol{\Phi}_{st} + \rho\boldsymbol{I})^{-1}\boldsymbol{Z}_{st} \end{aligned}$$

Now, as well as being a means to characterize spatio-temporal structure, the variogram is used in kriging to assign weights to observations to predict the value of some property at locations $(\boldsymbol{s}_i, t_j)$ for which data are not available, or where there is a desire to predict in a different support (Lloyd 2011). To use the variogram in kriging, a mathematical model must be fitted, such that the coefficients may then be used in the system of kriging equations. In this paper, we work with the spatio-temporal

metric model. This space-time model has a stationary covariance showing geometric anisotropy on $\mathbb{R}^d \times \mathbb{R}_+$; therefore, a metric in the space-time domain that can be directly used with isotropic models can be defined as

$$C_{st}(\delta_s, \delta_t) = C(\delta_{st}^2) = C(q_1^2\delta_s^2 + q_2^2\delta_t^2) \tag{17}$$

where $\delta_{st}^2 = q_1^2\delta_s^2 + q_2^2\delta_t^2$, $q_1, q_2 \in \mathbb{R}$ are constants that define the spatio-temporal metric, and δ_s and δ_t are the usual Euclidean distances in space and time, respectively. This model assumes the same structure of dependence in the space and time, and only allows changes in the range of both covariance functions. Several applications of this model can be found in Armstrong et al. (1993) and Snepvangers et al. (2003).

Several RBFs that are considered in this paper using the DB approach are presented in Table 1. The optimal smoothing parameter η, which is a parameter of free choice, is found by minimizing the root-mean-square prediction error (RMSPE) using cross-validation. Further descriptions of RBFs and their relationships to splines and kriging can be found in Cressie (1993, p. 180), Bishop (1995, p. 164), and Chilès & Delfiner (1999, p. 272).

2.3 Distance-based spatio-temporal prediction using RBFs

Having estimated the parameters $\boldsymbol{\nu}_{st}$ and $\boldsymbol{\omega}_{st}$, we are now ready to perform spatio-temporal prediction at a new location, $(\boldsymbol{s}_0, t_0)$, from the closest observations and where a set of mixed explanatory variables have been observed. Therefore, assume that on the set of mixed explanatory variables, a new individual $(nT+1)$ is observed, with known $\boldsymbol{v}(\boldsymbol{s}_0, t_0) = (v_1(\boldsymbol{s}_0, t_0), \ldots, v_p(\boldsymbol{s}_0, t_0))'$. Then, the distances between the new individual and each of the individuals involved in model (1) are to be calculated, i.e., $\delta_{0(ij)} = \delta\left(v(\boldsymbol{s}_0, t_0), v(\boldsymbol{s}_i, t_j)\right)$, $i = 1, ..., n$, $j = 1, ..., T$. From these distances, a prediction can be given using a result proposed by Cuadras & Arenas (1990) and Gower (1971), which relates the vector $\boldsymbol{\delta}_0 = \left(\delta_{0(11)}^2, \ldots, \delta_{0(nT)}^2\right)'$ of squared distances with the vector $\boldsymbol{x}\left(\boldsymbol{s}_0, t_0\right) = \left(x_1\left(\boldsymbol{s}_0, t_0\right), \ldots, x_k\left(\boldsymbol{s}_0, t_0\right)\right)'$ of principal coordinates associated to the new individual. This relation is given by

$$\delta_{0(ij)}^2 = \left(\boldsymbol{x}\left(\boldsymbol{s}_0, t_0\right) - \boldsymbol{x}\left(\boldsymbol{s}_i, t_j\right)\right)'\left(\boldsymbol{x}\left(\boldsymbol{s}_0, t_0\right) - \boldsymbol{x}\left(\boldsymbol{s}_i, t_j\right)\right)$$

with $i = 1, ..., n$, $j = 1, ..., T$. Then, we have that

$$\boldsymbol{x}\left(\boldsymbol{s}_0, t_0\right) = \frac{1}{2}\boldsymbol{\Lambda}^{-1}\boldsymbol{X}'_{st}(\boldsymbol{b} - \boldsymbol{\delta}_0)$$

where $\boldsymbol{b} = (b_{11}, ..., b_{(nT)(nT)})'$ is a vector formed by the diagonal elements of $\boldsymbol{B}$, and $b_{rr} = \boldsymbol{x}'(\boldsymbol{a}_r)\boldsymbol{x}(\boldsymbol{a}_i)$ with $r = 1, ..., nT$ with $\boldsymbol{a}_r$ defined above.

The aim is to predict the value of $Z\left(\boldsymbol{s}_0, t_0\right)$ based on a set of observations $\boldsymbol{Z}^*_{st}$. For this, the RBF predictor is given by

$$\widehat{Z}(\boldsymbol{s}_0, t_0) = \widehat{g}(\boldsymbol{s}_0, t_0) = \sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\varphi_{ij}Z(\boldsymbol{s}_i, t_j) = \boldsymbol{\varphi}'_{st}\boldsymbol{Z}^*_{st} \tag{18}$$

subject to

$$\sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\varphi_{ij}x_l(\boldsymbol{s}_i, t_j) = \boldsymbol{\varphi}'_{st}\boldsymbol{X}_l = x_l(\boldsymbol{s}_0, t_0), \quad l = 0, \ldots, k$$

where n_h and t_h are the sizes of neighborhoods associated to the space and time, respectively, $\boldsymbol{\varphi}_{st} = (\varphi_{11}, \ldots, \varphi_{n_h t_h})'$, $\boldsymbol{Z}^*_{st} = (Z(\boldsymbol{s}_1, t_1), \ldots, Z(\boldsymbol{s}_{n_h}, t_{t_h}))'$, and $\boldsymbol{X}_l = (x_l(\boldsymbol{s}_1, t_1), \ldots, x_l(\boldsymbol{s}_{n_h}, t_{t_h}))$.

Note that the expected error is equal to zero, i.e.,

$$\mathbb{E}\left(\hat{Z}(\boldsymbol{s}_0, t_0) - Z(\boldsymbol{s}_0, t_0)\right) = 0$$

and the mean square error of prediction by kriging, σ^2_K, using the approximation with a distance-based space-time radial basis function (DBSTRBF) is given by

$$\begin{aligned}
\sigma^2_K(\boldsymbol{s}_0, t_0) =& \mathbb{E}\left\{\left[\hat{Z}(\boldsymbol{s}_0, t_0) - Z(\boldsymbol{s}_0, t_0)\right]^2\right\} \\
\cong& \sum_{i=1}^{n_h}\sum_{i'=1}^{n_h}\sum_{j=1}^{t_h}\sum_{j'=1}^{t_h}\varphi_{ij}\varphi_{i'j'}\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'}) - 2\sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\varphi_{ij}\phi(\boldsymbol{s}_i - \boldsymbol{s}_0, t_j - t_0) \\
\cong& \boldsymbol{\varphi}'_{st}\boldsymbol{\Phi}_0\boldsymbol{\varphi}_{st} - 2\boldsymbol{\varphi}'_{st}\boldsymbol{\phi}_0
\end{aligned}$$

where $\boldsymbol{\Phi}_0$ is a $n_h t_h \times n_h t_h$ matrix with elements $\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'})$, and $\boldsymbol{\phi}_0 = (\phi(\boldsymbol{s}_1 - \boldsymbol{s}_0, t_1 - t_0), \ldots, \phi(\boldsymbol{s}_{n_h} - \boldsymbol{s}_0, t_{t_h} - t_0))'$ corresponds to the radial function vector evaluated between the neighbors and the point where we aim to predict, i.e. $\phi(\boldsymbol{s}_i - \boldsymbol{s}_0, t_j - t_0)$, $i = 1, \ldots, n_h$, $j = 1, ..., t_h$.

The weights are determined minimizing the following penalized expression

$$\begin{aligned}
l(\boldsymbol{\varphi}_{st}, \boldsymbol{\alpha}) =& \sum_{i=1}^{n_h}\sum_{i'=1}^{n_h}\sum_{j=1}^{t_h}\sum_{j'=1}^{t_h}\varphi_{ij}\varphi_{i'j'}\phi(\boldsymbol{s}_i - \boldsymbol{s}_{i'}, t_j - t_{j'}) - 2\sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\varphi_{ij}\phi(\boldsymbol{s}_i - \boldsymbol{s}_0, t_j - t_0) \\
&+ \rho\int_{\mathbb{R}^2\times\mathbb{R}} J_m(\kappa(\boldsymbol{s}, t))d(\boldsymbol{s}, t) + 2\sum_{l=0}^{k}\alpha_l\left(\sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\varphi_{ij}x_l(\boldsymbol{s}_i, t_j) - x_l(\boldsymbol{s}_0, t_0)\right)
\end{aligned}$$

or equivalently, the previous expression becomes

$$l(\boldsymbol{\varphi}_{st}, \boldsymbol{\alpha}) = \boldsymbol{\varphi}'_{st}\left(\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0\right)\boldsymbol{\varphi}_{st} - 2\boldsymbol{\varphi}'_{st}\boldsymbol{\phi}_0 + 2\boldsymbol{\alpha}'\left(\boldsymbol{X}'_0\boldsymbol{\varphi}_{st} - \boldsymbol{x}(\boldsymbol{s}_0, t_0)\right)$$

where $\boldsymbol{I}_0$ is the $n_h t_h \times n_h t_h$ identity matrix, $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_k)'$ is the vector of $(k+1)$ Lagrange multipliers associated with the unbiasedness constraint, $\boldsymbol{X}_0$ is similar to (14) using only $n_h t_h$ observations, and $\boldsymbol{x}(\boldsymbol{s}_0, t_0) = (1, x_1(\boldsymbol{s}_0, t_0), ..., x_k(\boldsymbol{s}_0, t_0))'$.

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

$$\begin{pmatrix} \boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0 & \boldsymbol{X}_0 \\ \boldsymbol{X}'_0 & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\varphi}_{st} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi}_0 \\ \boldsymbol{x}\left(\boldsymbol{s}_0, t_0\right) \end{pmatrix} \tag{19}$$

By Definition 1, $\boldsymbol{\Phi}_0$ is positive definite, then $\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0$ is invertible, and the system can be solved with the coefficients $\boldsymbol{\varphi}_{st}$ and $\boldsymbol{\alpha}$ given by

$$\begin{aligned} \widehat{\boldsymbol{\varphi}}'_{st} &= \left\{\boldsymbol{\phi}_0 + \boldsymbol{X}_0\left[\boldsymbol{X}'_0(\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0)^{-1}\boldsymbol{X}_0\right]^{-1}\left[\boldsymbol{x}(\boldsymbol{s}_0, t_0) - \boldsymbol{X}'_0(\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0)^{-1}\boldsymbol{\phi}_0\right]\right\}'(\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0)^{-1} \\ \widehat{\boldsymbol{\alpha}} &= -\left[\boldsymbol{X}'_0(\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0)^{-1}\boldsymbol{X}_0\right]^{-1}\left[\boldsymbol{x}(\boldsymbol{s}_0, t_0) - \boldsymbol{X}'_0(\boldsymbol{\Phi}_0 + \rho \boldsymbol{I}_0)^{-1}\boldsymbol{\phi}_0\right] \end{aligned} \tag{20}$$

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

$$\widehat{\sigma}^2_K(\boldsymbol{s}_0, t_0) \cong \sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\widehat{\varphi}_{ij}\phi(\boldsymbol{s}_i - \boldsymbol{s}_0, t_j - t_0) + \rho\sum_{i=1}^{n_h}\sum_{j=1}^{t_h}\widehat{\varphi}^2_{ij} + \sum_{l=0}^{k}\widehat{\alpha}_l x_l(\boldsymbol{s}_0, t_0)$$

If we want to evaluate or compare adjustments between DBSTRBFs, it is recommended to use cross-validation (leave-one-out). This criterium is now presented.

2.4 Evaluation of interpolation accuracy

We consider the root-mean-square prediction error (RMSPE) to evaluate the DBSTRBF interpolation accuracy. The RMSE is obtained by leave-one-out using cross-validation (LOOCV), and it can be used to compare the performance of several interpolation methods. The method that yields the smallest value of RMSE is the best. LOOCV consists in removing one observation from the nT sample points (usually related to a neighborhood), and then, with the remaining $(nT-1)$ values and a selected

STRBF, 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 nT prediction errors are obtained. This procedure is justified because the DBSTRBF interpolation methods are accurate, i.e., the predicted and observed values match in the sampled points. In this way, the LOOCV procedure gives an idea of how good the forecasts are, and provides information about which model gives more accurate predictions. The expression for the RMSPE is given by

$$\text{RMSPE} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{T}\left(\hat{Z}_{[ij]}\left(\boldsymbol{s}_i, t_j\right) - Z\left(\boldsymbol{s}_i, t_j\right)\right)^2}{nT}} \tag{21}$$

where $\hat{Z}_{[ij]}\left(\boldsymbol{s}_i, t_j\right)$ is the predicted value obtained from cross-validation, and $Z\left(\boldsymbol{s}_i, t_j\right)$ is the sampled value at location $(\boldsymbol{s}_i, t_j)$.

A variation of the previous methodology is to divide the sample into two sub-samples; the first subsample is used to select the η and ρ smoothing parameters, and the second subsample is used to validate the DBSTRBF method. After that, validation measures can be constructed from the observed and predicted values (Bivand et al. 2008). If all goes well, the RMSPE should be as small as possible (close to zero).

Finally, our procedure can be summarized as follows:

1. Obtain the principal coordinates from the spectral decomposition of the matrix of similarities (or distances) calculated from the explanatory variables.
2. Select the most correlated or significant principal coordinates with the regionalized variable, $\boldsymbol{Z}_{st}$. In this step, we recommend using the criterium given in (9) to make a first selection in order to leave out principal coordinates which are poorly correlated with the regionalized variable. Then, using the criteria (10) or (11), we choose the most significant principal coordinates using DB regression.
3. Optimize the parameters η and ρ of the DB spatio-temporal interpolator using radial basis functions by means of LOOCV from expression (21), and using (20) in different neighborhoods of a predetermined size. The sizes of the neighborhoods, n_h and t_h, can also be chosen within the optimization process.

4. Make predictions at sampled and unsampled points to generate the prediction map using the DBSTRBFs method, i.e., by means of $\hat{Z}(s_0, t_0) = \hat{\boldsymbol{\varphi}}'_{st}\boldsymbol{Z}^*_{st}$.

3 Simulated experiments

This section describes a simulation study conducted to evaluate the efficiency of the spatio-temporal proposed method under different conditions associated with the smoothing parameters and RBFs using the DB method. In particular, we aim to study the effects of: (i) noise level, (ii) design density, (iii) degree of spatio-temporal variation, and (iv) noise variance function. These configurations and scenarios are presented in Table 2.

This study considers simulations in three dimensions (w_x, w_y, t), plus a Binomial random variable with $V_1 \sim Bi(nT, 0.4)$, sample sizes $nT = 150, 250$ associated to $n = 25$ points in space in each case and $T = 6$ and $T = 10$ points in time, respectively, neighborhood sizes $n_h t_h = 8, 16$, smoothing parameters $\eta = 0.01, 0.1$, and $r = 1, 3$ in the variance factor. Assume we have a nominal variable associated with three particular regions in the unit square, as shown in Figure 1. Since there are three regions, only two dummy variables (D_2 and D_3) are considered to avoid problems of singularity. In addition, $\varepsilon(\boldsymbol{s}_i, t_j)$ follows a stationary isotropic Gaussian random field for a space-time process with mean zero and covariance function given by

$$C(\boldsymbol{h}, u) = \sigma^2 \left[\left(a_1^2 \|\boldsymbol{h}\|^2 + a_2^2 |u|^2 \right)^{1/2} / a_3 \right], \qquad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

where $\sigma^2 = 1$, $a_1 = a_2 = a_3 = 1$ are constants that define the spatio-temporal metric, $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_{i'}$ and $u = t_j - t_{j'}$. We can find applications of this covariance model in Armstrong et al. (1993) or Dimitrakopoulos & Luo (1994), among others. For trend parameters, we assume the following values $\beta_0 = 10$, $\beta_1 = -4$, $\beta_2 = 2$, $\beta_3 = -4$ and $\beta_4 = 3$, with $w_{x_{ij}}$ and $w_{y_{ij}}$ associated to the spatial locations, and t_{ij} associated to the time, where ij is the ij-th simulated observation. In Table 3, the simulated scenarios are presented. The proposed method is tested with five RBFs (MQ, TPS, CRS, ST, and EXP) considering the metric model given in (17) with $q_1 = q_2 = 1$. A total of 80 scenarios were simulated, and for each of them, the process was repeated 100 times.

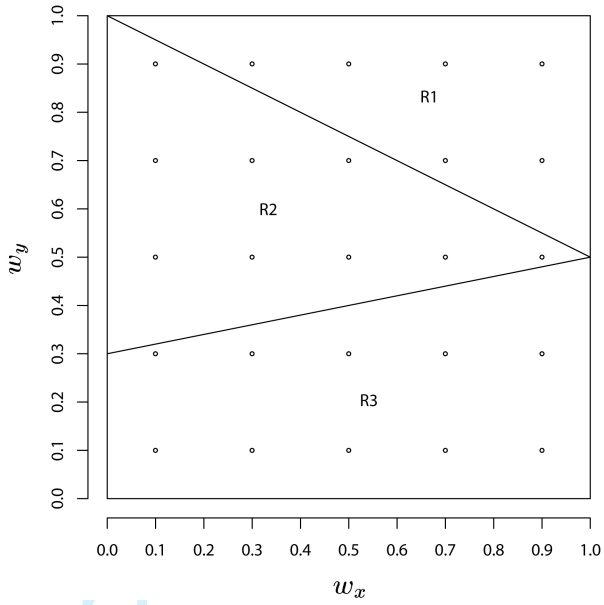

Fig. 1 Spatial location of sampling points and associated regions defining the nominal variable

For each simulated data set, we assessed the quality of the fit with the RMSPE, which was obtained by LOOCV method. The results are shown in Tables 4 and 5. We initially considered using a positive parameter ρ, but the values of RMSPE showed no significant differences with those obtained when $\rho = 0$, in particular when the multiquadratic and exponential functions were used.

Tables 4 and 5 show the average of RMSPE values from 100 simulations per case and for the 80 cases described in Table 3. The DBSTRBF method worked well for large neighborhoods, indicating a gain (a decrease) of 70.6% of the RMSPE values when $n_h t_h = 32$ with respect to $n_h t_h = 8$. When $\eta = 0.1$, there was a reduction of 2.5% compared with $\eta = 0.01$ in the average of RMSPE values. Taking into account parameter r, there was a general slight increase (loss) of 1.43% in the RMSPE values when $r = 3$ compared with $r = 1$. Just on the contrary, when $nT = 150$ the RMSPE values were 15% larger than those obtained when $nT = 250$.

Regarding the scenarios associated with generic forms, the lowest RMSPE values corresponded to the noise level and factor variance cases, with RMSPE average values of 1.86 and 1.90, respectively. For block design density and spatio-temporal variance cases, the RMSPE average values were 3.38 and 2.16, respectively. In terms of the DBSTRBF method, we found: i) for the *noise level* case, the DBSTRBF method that produced the lowest RMSPE average was the CRS, while TPS showed the

highest; ii) for the *density block design* case, the DBSTRBF method with smaller RMSPE average was the MQ, while CRS showed again the highest; iii) in terms of *spatio-temporal variance*, the lowest RMSPE average value was observed with MQ, while TPS showed once more the highest RMSPE average values; and iv) the *variance factor* case with better results in terms of RMSPE averages was the CRS, and again, the DBSTRBF method built with the TPS function showed the worst results.

Since the larger RMSPE average values were in those cases with sample size $nT = 150$, these cases are shown in separate boxplots to the cases with sample size $nT = 250$, see Figures 2 and 3. In these graphs, we found that the RMSPE average values are smaller for $n_h t_h = 32$ than for $n_h t_h = 8$. Additionally, according to Figure 2, we note the following: i) in terms of *noise level*, we found less variability when $r = 1$ and $n_h t_h = 8$ except for CRS-RBF, where the variability is too small when $r = 1$ and $n_h t_h = 32$; ii) in case of *design function*, the DBSTRBF method that showed the highest variability was the TPS, while MQ and EXP radial basis functions showed less variability, especially for $r = 1$ and $n_h t_h = 32$; iii) under the *spatio-temporal variation* scenarios, when $r = 1$ and $n_h t_h = 8$ a general large variability was generated under any particular case, except for EXP when $\eta = 0.1$; and iv) for the *variance factor*, variability was similar in all the functions tested. In general, the largest RMSPE average values were found for the neighborhood size $n_h t_h = 8$.

According to Figure 3, we note that: i) when considering the *noise level*, there was a larger variability with $r = 3$, increasing the RMSPE average value with $n_h t_h = 8$; ii) in case of *design function*, TPS, CRS and ST radial basis functions showed larger variability with $r = 3$ and $n_h t_h = 8$ regardless of η, while for MQ and EXP radial basis functions, the largest variabilities were shown in $r = 3$ and $n_h t_h = 8$ in both RBFs but with $\eta = 0.1$ and $\eta = 0.01$, respectively; iii) in case of *spatio-temporal variation*, the largest variability was for $r = 3$ and $n_h t_h = 8$, especially for ST and TPS radial basis functions, and the smallest variability was for the CRS radial basis function. In general, the RMSPE average value was smaller when $n_h t_h = 32$.

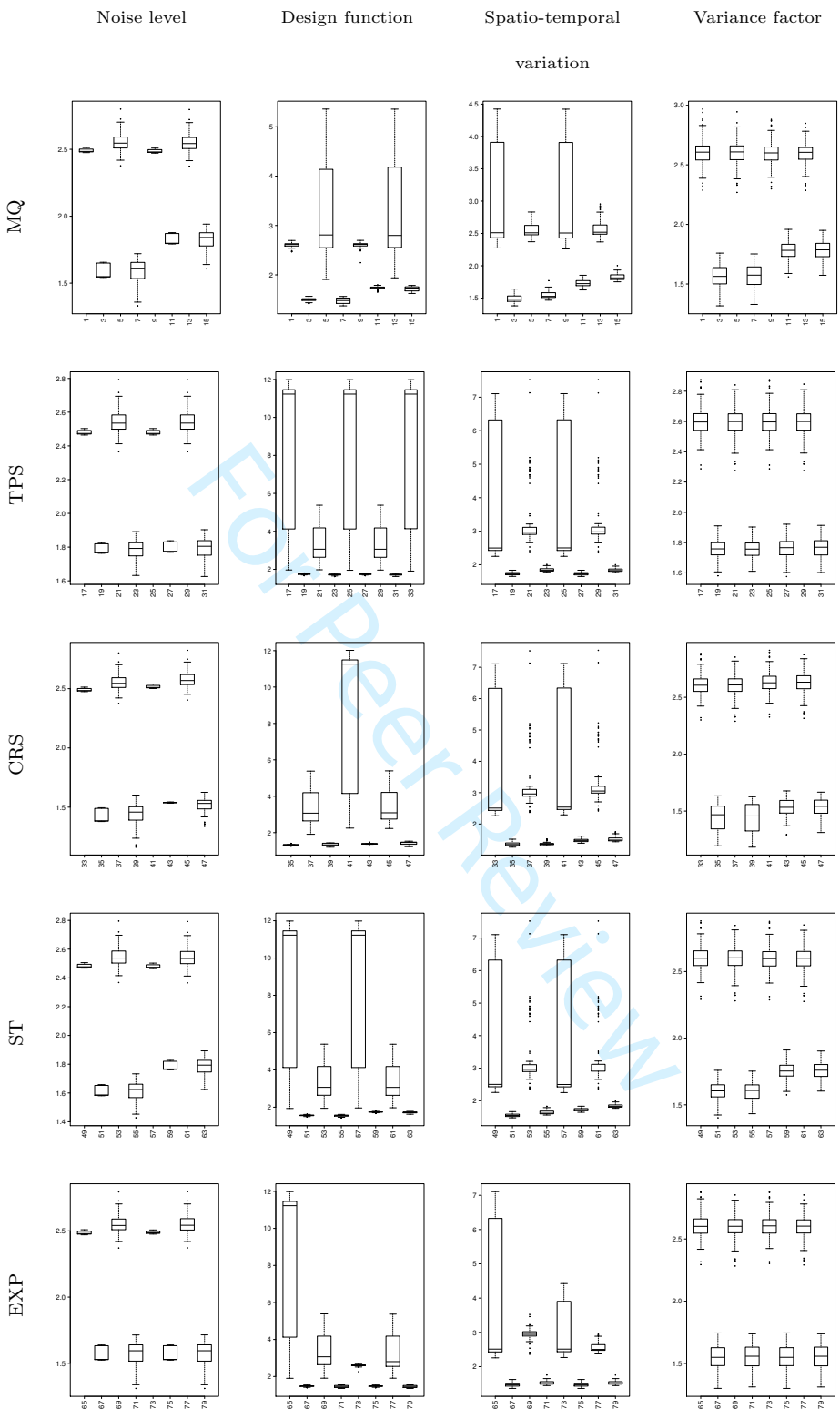

Fig. 2 RMSPE average values for the simulated scenarios with 6 points in time

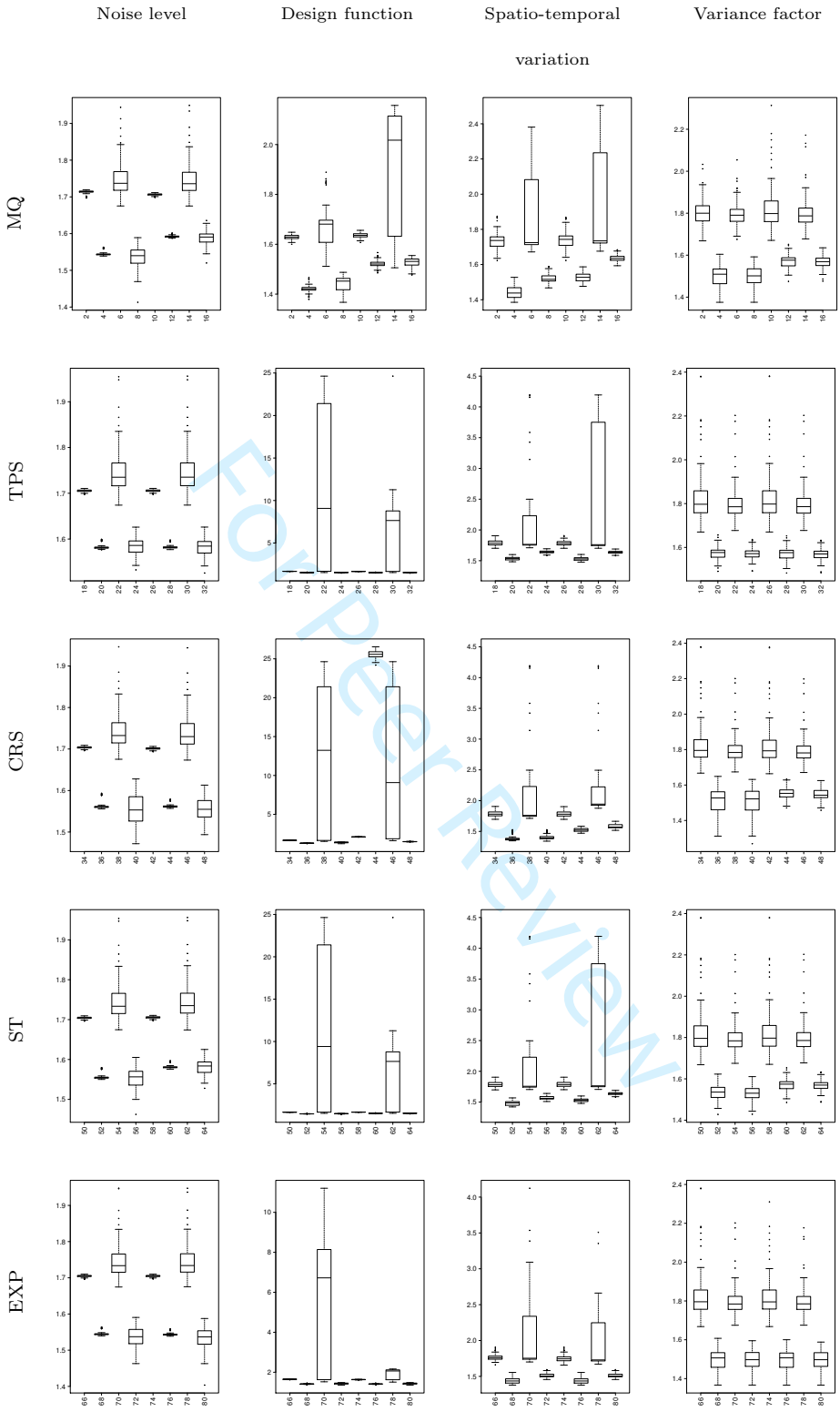

Fig. 3 RMSPE average values for the simulated scenarios with 10 points in time

4 Application

We analyzed the monthly average temperature in Croatia in 155 meteorological stations, the monthly average temperature in Croatia was measured from January to December of 2008. This information is taken from http://spatial-analyst.net/book/HRclim2008, and it was provided by Melita Perčec Tadić, Croatian Meteorological Organization and Hydrological Services (Hengl 2009). Croatia is a relatively small country, but it has several different climates, which are a result of its specific position on the Adriatic Sea and fairly diverse topography ranging from plains on the east, through a central mountainous separating the continental from the maritime part of the country. The study region is characterized by a wide range of topographical and climatological features, which allow to properly assess the proposed methodology with respect to the traditional one, because the earth's average temperatures in this region are strongly influenced by the topography. On most meteorological stations, the temperature is measured three times a day, at 7 am, 1 pm and 9 pm (Hengl et al. 2012). The mean day-time temperature (ΔT in one day) is calculated as a weighted average (Hiebl et al. 2009) as follows:

$$\Delta T = \frac{T_{(7am)} + T_{(1pm)} + 2 \cdot T_{(9pm)}}{4} \tag{22}$$

Then the monthly average temperature is obtained from the daily average mentioned above, considering the daily average every 8 days, i.e., 3 to 4 monthly records. Temperature measurements are automatically collected at the 159 meteorological stations. As four meteorological stations had not available records for some months, then these stations were removed from the analysis. So, only 155 stations were considered, and the monthly average temperature was calculated with the observed values removing missing data. The spatial distribution of the stations is not optimal (Zaninovic et al. 2008, Perčec-Tadić 2010), there is a certain sub-sampling at higher altitudes and in areas with lower population density (for practical reasons, the areas of higher population density have been given a priority). Consequently, the accuracy of the mapping will be lower at higher altitudes and highlands (Hengl 2009).

The w_x and w_y locations were obtained from a transformation of the geographical locations given in latitude and longitude. The location of the 155 meteorological stations is shown in Figure 4 (spatial

locations). The principal coordinates were calculated from the spectral decomposition generated by: the spatial locations (w_x, w_y), the month, Digital Elevation Model (DEM, in meters), topographically weighted distance from the coast line (DSEA, in km), Topographic Wetness Index (TWI), and the season. Both spatial locations and time were standardized to give equal weight to all spatio-temporal dimensions. Furthermore, a DB spatio-temporal regression is performed with the principal coordinates and the Earth's monthly average temperature. In this process, we found that each one of the first 10 principal coordinates had a higher statistical significance (at the 5% level) with the Earth's average temperature. The DB space-time regression explained 96.1% of the variability in monthly temperatures. So, these 10 principal coordinates were considered in the trend, and finally, LOOCV was performed selecting the DBSTRBF method that showed the lowest RMSPE.

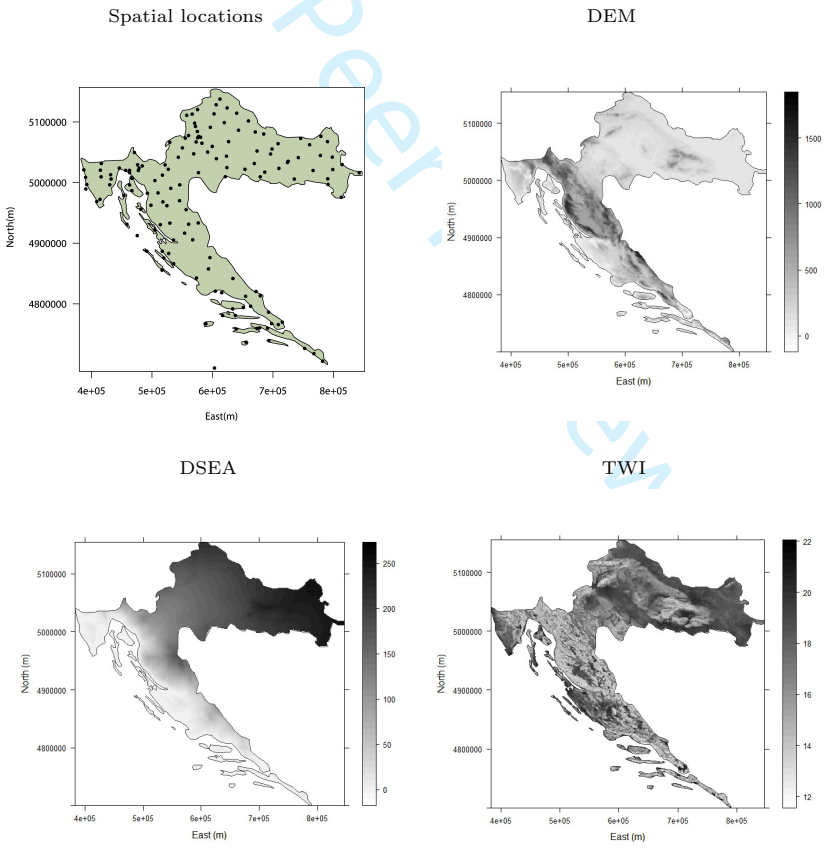

Fig. 4 Spatial locations of meteorological stations in Croatia and static topographic predictors: Digital Elevation Model (DEM, in meters), topographically weighted distance from the coast line (DSEA, in km) and Topographic Wetness Index (TWI)

Additionally, the LOOCV method using the metric model given in (17) with parameters $q_1 = q_2 = 1$ is selected to check the performance of the six models herein. Table 6 shows the RMSPE values for the interpolation results by six DBSTRBFs, and the best radial functions are the CRS and ST which show the smallest RMSPE values. The corresponding prediction maps are shown in Figure 5.

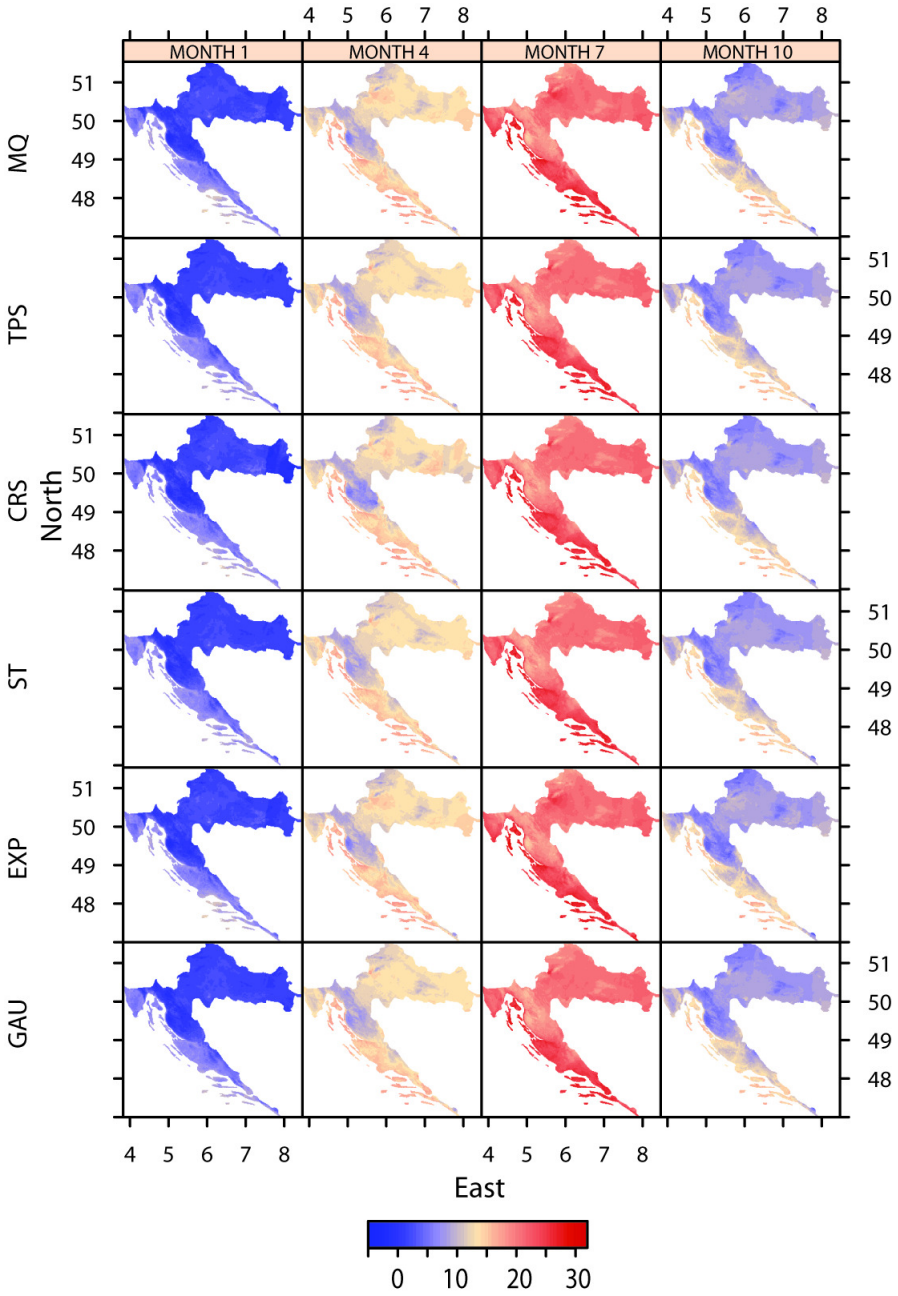

Fig. 5 Prediction maps for Earth's monthly average temperature in Croatia under the DBSTRBF method in January, April, July and October (units of the locations East and North in 100000 of meters)

5 Conclusions

The DBSTRBF method has the advantage of working with mixed variables in the trend and does not require the estimation of a spatio-temporal variogram, which usually requires too much computational time. This can be relevant in many disciplines of geosciences and environmental areas where the researches can have mixed regionalized covariates. Also, the proposed method provides flexibility in the fitting of the smoothing parameters when RBFs using DB method are employed, and it shows, both under simulations and application, some advantages in reducing the error depending from which DBSTRBF is used. We suggest that the cross-validation calculated for splines may be a reliable measure of overall prediction error.

In a variety of studies, detection of variability among areas is a quite difficult task, and so the proposed DBSTRBF method is expected to be useful as it does not despise the existing information. The DBSTRBF method could produce better estimates of the regionalized variable in the time if the number of principal coordinates is increased; however, in this procedure, we should not include principal coordinates with eigenvalues close to zero because these coordinates could cause problems of no invertibility in the matrix associated with the trend. It is also possible to include more significative principal coordinates in the trend and variogram models. A detailed simulation study in this way should be performed. On the other hand, computational complexity is an issue: the RBF methods used in this paper take sometime in processing to generate predictions. This indicates that there is still opportunity to improve the data processing in the proposed method.

Acknowledgements Work partially funded by grant MTM2016-78917-R from the Spanish Ministry of Science and Education, Core Spatial Data Research (Faculty of Engineering, Universidad Distrital Francisco José de Caldas) by grant COL0013969, and Applied Statistics in Experimental Research, Industry and Biotechnology (Universidad Nacional de Colombia) by grant COL0004469.

References

Armstrong, M., Chetboun, G. & Hubert, P. (1993), Kriging the rainfall in lesotho, *in* S. A., ed., 'Geostatistics Tróia '92', Vol. 5, Quantitative Geology and Geostatistics, Springer, Dordrecht, pp. 661–

672.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton-Florida.

Bellier, E., Planque, B. & Petitgas, P. (2007), 'Historical fluctuations in spawning location of anchovy (engraulis encrasicolus) and sardine (sardina pilchardus) in the bay of biscay during 1967-1973 and 2000-2004', *Fisheries Oceanography* **16**(1), 1–15.

Bishop, C. M. (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.

Boj, E., Claramunt, M. M. & Fortiana, J. (2007), 'Selection of predictors in distance-based regression', *Communications in Statistics - Theory and Methods* **36**(1), 87–98.

Boj, E., Delicado, P. & Fortiana, J. (2010), 'Distance-based local linear regression for functional predictors', *Computational Statistics and Data Analysis* **54**(2), 429–437.

Chen, L., Fuentes, M. & Davis, J. M. (2006), Spatial temporal statistical modelling and prediction of environmental processes, *in* O. U. Press, ed., 'Hierarchical Modelling for the Environmental Sciences', pp. 121–144.

Chilès, J. P. & Delfiner, P. (1999), *Geostatistics: modeling spatial uncertainty*, John Wiley & Sons, New York.

Christakos, G. (2000), *Modern Spatiotemporal Geostatistics*, Oxford University Press, Oxford.

Cressie, N. (1989), 'Geostatistics', *The American Statistician* **43**(4), 197–202.

Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley & Sons Inc, New York.

Cressie, N. & Huang, H. C. (1999), 'Classes of nonseparable, spatio-temporal stationary covariance functions', *Journal of the American Statistical Association* **94**(448), 1330–1340.

Cuadras, C. M. (1989), Distance analysis in discrimination and classification using both continuous and categorical variables, *in* 'Dodge, Y. (Ed.), Statistical Data Analysis and Inference. North-Holland Publishing Co.', Amsterdam, pp. 459–473.

Cuadras, C. M. (1993), 'Interpreting an inequality in multiple regression', *The American Statistician* **47**(4), 256–258.

Cuadras, C. M. & Arenas, C. (1990), 'A distance based regression model for prediction with mixed data', *Communications in Statistics - Theory and Methods* **19**(6), 2261–2279.

Cuadras, C. M., Arenas, C. & Fortiana, J. (1996), 'Some computational aspects of a distance-based model for prediction', *Communications in Statistics - Simulation and Computation* **25**(3), 593–609.

De Cesare, L., Myers, D. E. & Posa, D. (1997), Spatial temporal modeling of so2 in milan district, *in* 'Baaffi, E., and Schofield, N., eds., Geostatistics Wollongong '96', Vol. 2, Kluwer Academic Publishers, Dordrecht, pp. 1031–1042.

De Cesare, L., Myers, D. E. & Posa, D. (2001), 'Estimating and modeling space-time correlation structures', *Statistics and Probability Letters* **51**(1), 9–14.

Dimitrakopoulos, R. & Luo, X. (1994), Spatiotemporal modelling: Covariances and ordinary kriging systems, *in* 'Dimitrakopoulos R. (eds) Geostatistics for the Next Century. Quantitative Geology and Geostatistics', Vol. 6, Springer, Dordrecht.

Duchon, J. (1976), 'Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces', *Rairo Analyse Numerique* **10**(12), 5–12.

Gneiting, T., Genton, M. G. & Guttorp, P. (2007), Geostatistical space-time models, stationarity, separability and full symmetry, *in* 'Finkenstadt, B., Held, L. and Isham, V. (eds.), Statistical Methods for Spatio-Temporal Systems', Chapman & Hall/CRC, Boca Raton, pp. 151–175.

Gneiting, T. & Schlather, M. (2002), Space-time covariance models, *in* 'El-Shaarawi, A. H. and Piegorsch, W. W. (eds.), Encyclopedia of Environmetrics', Vol. 4, John Wiley & Sons, Chichester, pp. 2041–2045.

Gower, J. C. (1968), 'Adding a point to vector diagrams in multivariate analysis', *Biometrika* **55**(3), 582–585.

Gower, J. C. (1971), 'A general coefficient of similarity and some of its properties', *Biometrics* **27**(4), 857–871.

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**(10), 423–426.

Hengl, T. (2009), *A Practical Guide to Geostatistical Mapping*, 2nd edn, University of Amsterdam, Amsterdam.

Hengl, T., Heuvelink, G. B. M., Perčec-Tadić, M. & Pebesma, E. J. (2012), 'Spatio-temporal prediction of daily temperatures using time-series of modis lst images', *Theoretical and Applied Climatology* **107**(1-2), 265–277.

Hiebl, J., Auer, I., Böhm, R., Schöner, W., Maugeri, M., Lentini, G., Spinoni, J., Brunetti, M., Nanni, T., Perčec-Tadić, M., Bihari, Z., Dolinar, M. & Müller-Westermeier, G. (2009), 'A high-resolution 1961-1990 monthly temperature climatology for the greater alpine region', *Meteorologische Zeitschrift* **18**(5), 507–530.

Kolovos, A., Christakos, G., Hristopulos, D. T. & Serre, M. L. (2004), 'Methods for generating non-separable spatiotemporal covariance models with potential environmental applications', *Advances in Water Resources* **27**(8), 815–830.

Kondoh, H., Koizumi, T. & Ikeda, K. (2011), 'A geostatistical approach to spatial density distributions of sika deer (*Cervus nippon*)', *Journal of Forest Research* **18**(1), 93–100.

Kyriakidis, P. C. & Journel, A. G. (1999), 'Geostatistical space-time models: a review', *Mathematical Geology* **31**(6), 651–684.

Lloyd, C. D. (2011), *Local Models for Spatial Analysis*, second edition edn, Taylor & Francis Group, Boca Raton.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (2002), *Multivariate Analysis*, Academic Press, London.

Mateu, J., Montes, F. & Fuentes, M. (2003), 'Recent advances in space-time statistics with applications to atmospheric data: An overview', *Journal of Geophysical Research* **108**(D24), 1–6.

Mitáš, L. & Mitášová, H. (1988), 'General variational approach to the interpolation problem', *Computers and Mathematics with Applications* **16**(12), 983–992.

Mitášová, H. & Hofierka, J. (1993), 'Interpolation by regularized spline with tension: Ii. aplication to terrain modeling and surface geometry analysis', *Mathematical Geology* **25**(6), 657–669.

Mitášová, H. & Mitáš, L. (1993), 'Interpolation by regularized spline with tension: I. theory and implementation', *Mathematical Geology* **25**(6), 641–655.

Myers, D. (1992), 'Kriging, cokriging, radial basic functions and the role of positive definiteness', *Computers Mathematics Applied* **24**(12), 139–148.

Myers, D. E., De Iaco, S., Posa, D. & De Cesare, L. (2002), 'Space-time radial basis functions', *Computers and Mathematics with Applications* **43**(3-5), 539–549.

Perčec-Tadić, M. (2010), 'Gridded croatian climatology for 1961-1990', *Theoretical and Applied Climatology* **102**(1-2), 87–103.

Planque, B., Bellier, E. & Lazure, P. (2007), 'Modelling potential spawning habitat of sardine (sardina pilchardus) and anchovy (engraulis encrasicolus) in the bay of biscay', *Fisheries Oceanography* **16**(1), 16–30.

Rouhani, S. & Hall, T. L. (1989), Space-time kriging of groundwater data, *in* 'Armstrong M. (eds) Geostatistics. Quantitative Geology and Geostatistics', Vol. 4, Springer, Dordrecht, pp. 639–651.

Rouhani, S. & Myers, D. E. (1990), 'Problems in space-time kriging of hydrogeological data', *Mathematical Geology* **22**(5), 611–623.

Sahu, S. K. & Mardia, K. V. (2005), Recent trends in modeling spatio-temporal data, *in* 'Proceedings of the special meeting on Statistics and Environment', Messina.

Schagen, I. P. (1979), 'Interpolation in two dimensions - a new technique', *Journal of the Institute of Mathematics and its Applications* **23**(1), 53–59.

Şen, Z., Altunkaynak, A. & Özger, M. (2006), 'Space-time interpolation by combining air pollution and meteorologic variables', *Pure and Applied Geophysics* **163**(7), 1435–1451.

Snepvangers, J. J. J. C., Heuvelink, G. B. M. & Huisman, J. A. (2003), 'Soil water content interpolation using spatio-temporal kriging with external drift', *Geoderma* **112**(3-4), 253–271.

Späh, H. (1969), 'Exponential spline interpolation', *Computing* **4**(3), 225–233.

Yavuz, H. & Erdoğan, S. (2012), 'Spatial Analysis of Monthly and Annual Precipitation Trends in Turkey', *Water Resources Management* **26**(3), 609–621.

Ye, J. (2008), Geostatistical methods for spatio-temporal analysis of fMRI data, PhD thesis, Department of Statistics, University of Georgia, Athens.

Ye, J., Lazar, N. & Li, Y. (2011), 'Sparse geostatistical analysis in clustering fmri time series', *Journal of Neuroscience Methods* **199**(2), 336–345.

Zaninovic, K., Gajic-Capka, M. & Perçec-Tadić, M. (2008), Meteorological and hydrological service republic of croatia, Technical report, Zagreb.

Table 1 Functional forms of some RBFs

RBF	Functional form
EXP	$\phi(\delta_{st}) = e^{-\eta\delta_{st}}, \quad \eta \neq 0$
GAU	$\phi(\delta_{st}) = e^{-\eta\delta_{st}^2}, \quad \eta \neq 0$
MQ	$\phi(\delta_{st}) = \sqrt{\eta^2 + \delta_{st}^2}, \quad \eta \neq 0$
IMQ	$\phi(\delta_{st}) = 1/\sqrt{\eta^2 + \delta_{st}^2}, \quad \eta \neq 0$
TPS	$\phi(\delta_{st}) = \begin{cases} (\eta \cdot \delta_{st})^2 \log(\eta \cdot \delta_{st}) & \text{if} \quad \delta_{st} \neq 0, \eta > 0 \\ 0 & \text{if} \quad \delta_{st} = 0 \end{cases}$
CRS	$\phi(\delta_{st}) = \begin{cases} \ln\left(\eta\frac{\delta_{st}}{2}\right)^2 + E_1\left(\eta\frac{\delta_{st}}{2}\right)^2 + C_E & \text{if} \quad \delta_{st} \neq 0 \\ 0 & \text{if} \quad \delta_{st} = 0 \end{cases}$ where $\eta > 0$, ln is the natural logarithm, $E_1(x)$ is the exponential integral function, and C_E is the Euler constant.
ST	$\phi(\delta_{st}) = \begin{cases} \ln(\eta \cdot \delta_{st}/2) + K_0(\eta \cdot \delta_{st}) + C_E & \text{if} \quad \delta_{st} \neq 0 \\ 0 & \text{if} \quad \delta_{st} = 0 \end{cases}$ where $K_0(x)$ is the modified Bessel function and C_E is the Euler constant.

Table 2 Scenarios considered in the spatio-temporal simulated experiments

Factor	Generic form
Noise level	$z_r(\boldsymbol{s}_i,t_j)=\beta_0+\beta_1V_{ij}+\beta_2D_{(ij)2}+\beta_3D_{(ij)3}+\beta_4t_{ij}+f(w_{x_{ij}})+f(w_{y_{ij}})+f(t_{ij})$ $+f(w_{x_{ij}})f(w_{y_{ij}})+f(w_{x_{ij}})f(t_{ij})+f(w_{y_{ij}})f(t_{ij})+\sigma_r\varepsilon(\boldsymbol{s}_i,t_j)$ where $\sigma_r=0.02+0.04(r-1)^2$
Design density	$z_r(\boldsymbol{s}_i,t_j)=\beta_0+\beta_1V_{ij}+\beta_2D_{(ij)2}+\beta_3D_{(ij)3}+\beta_4t_{ij}+f(X_{r(ij)})+f(Y_{r(ij)})+f(T_{r(ij)})$ $+f(X_{r(ij)})f(Y_{r(ij)})+f(X_{r(ij)})f(T_{r(ij)})+f(Y_{r(ij)})f(T_{r(ij)})+\sigma\varepsilon(\boldsymbol{s}_i,t_j)$ where $\sigma=0.1$, $X_{r(ij)}=F_r^{-1}(X_{ij})$, $Y_{r(ij)}=F_r^{-1}(Y_{ij})$, $T_{r(ij)}=F_r^{-1}(T_{ij})$ with $T_{ij}=t_{ij}/t_{max}$, $t_{max}=6,10$
Spatio-temporal variation	$z_r(\boldsymbol{s}_i,t_j)=\beta_0+\beta_1V_{ij}+\beta_2D_{(ij)2}+\beta_3D_{(ij)3}+\beta_4t_{ij}+f(w_{x_{ij}})+f(w_{y_{ij}})+f(t_{ij})$ $+f(w_{x_{ij}})f(w_{y_{ij}})+f(w_{x_{ij}})f(t_{ij})+f(w_{y_{ij}})f(t_{ij})+\sigma\varepsilon(\boldsymbol{s}_i,t_j)$ where $\sigma=0.2$, $f_r(l_{ij})=\sqrt{l_{ij}(1-l_{ij})}\sin\left[\frac{2\pi\left\{1+2^{(9-4r)/5}\right\}}{l_{ij}+2^{(9-4r)/5}}\right]$
Variance function	$z_r(s_i,t_j)=\beta_0+\beta_1V_{ij}+\beta_2D_{(ij)2}+\beta_3D_{(ij)3}+\beta_4t_{ij}+f(w_{x_{ij}})+f(w_{y_{ij}})$ $+f(w_{x_{ij}})f(w_{y_{ij}})+\sqrt{\varsigma_1+\varsigma_2+\varsigma_3+\varsigma_1\varsigma_2+\varsigma_1\varsigma_3+\varsigma_2\varsigma_3}\varepsilon(\boldsymbol{s}_i,t_j)$ where $\varsigma_1=\upsilon_r(w_{x_{ij}})$, $\varsigma_2=\upsilon_r(w_{y_{ij}})$, $\varsigma_3=\upsilon_r(t_{ij})$, with $\upsilon_r(l_{ij})=[0.15\{1+0.4(2r-7)(l_{ij}-0.5)\}]^2$
Assumptions and other choices	
$V_{ij}\sim Binomial(nT,0.4)$, $\varepsilon(\boldsymbol{s}_i,t_j)\overset{iid}{\sim}N(0,0.1)$; $nT=150$ ($n=25$ points in space and $T=6$ points in time) and $nT=250$ ($n=25$ points in space and $T=10$ points in time); $X_{ij},Y_{ij},T_{ij}\overset{iid}{\sim}Uniform\,[0,1]$; F_r is the $Beta\left(\frac{r+4}{5},\frac{11-r}{5}\right)$;$f(l_{ij})=1.5\phi\left(\frac{l_{ij}-0.5}{0.15}\right)-\phi\left(\frac{l_{ij}-0.8}{0.04}\right)$; $\phi(u)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-u^2}{2}\right)$ $r=1,3$, $l_{ij}=w_{x_{ij}},w_{y_{ij}},t_{ij}$, $i=1,...,n$, $j=1,...,T$	

Table 3 Simulated scenarios. The natural numbers in the last five columns (from 1 to 80) represent the number of scenario

Model parameters			nT	Radial basis function				
η	r	$n_h t_h$		MQ	TPS	CRS	ST	EXP
0.01	1	8	150	1	17	33	49	65
			250	2	18	34	50	66
		32	150	3	19	35	51	67
			250	4	20	36	52	68
	3	8	150	5	21	37	53	69
			250	6	22	38	54	70
		32	150	7	23	39	55	71
			250	8	24	40	56	72
0.1	1	8	150	9	25	41	57	73
			250	10	26	42	58	74
		32	150	11	27	43	59	75
			250	12	28	44	60	76
	3	8	150	13	29	45	61	77
			250	14	30	46	62	78
		32	150	15	31	47	63	79
			250	16	32	48	64	80

Table 4 RMSPE under the DBSTRBFs method for the scenarios presented in Table 3 (noise level and block design density cases)

Parameter			nT	Noise level					Block design density				
η	r	$n_h t_h$		MQ	TPS	CRS	ST	EXP	MQ	TPS	CRS	ST	EXP
0.01	1	8	150	2.49	2.48	2.49	2.48	2.48	2.61	8.06	8.05	8.05	8.05
			250	1.71	1.71	1.70	1.70	1.70	1.63	1.67	1.66	1.66	1.65
		32	150	1.58	1.78	1.41	1.60	1.56	1.50	1.75	1.33	1.56	1.48
			250	1.54	1.58	1.56	1.55	1.54	1.42	1.53	1.29	1.46	1.41
	3	8	150	2.55	2.54	2.55	2.55	2.55	3.43	3.56	3.56	3.56	3.55
			250	1.75	1.74	1.74	1.74	1.74	1.67	11.77	11.85	11.77	5.03
		32	150	1.59	1.78	1.44	1.61	1.57	1.48	1.72	1.35	1.54	1.46
			250	1.54	1.58	1.55	1.55	1.54	1.44	1.53	1.38	1.48	1.44
0.1	1	8	150	2.49	2.48	2.52	2.48	2.49	2.60	8.05	8.13	8.06	2.60
			250	1.71	1.71	1.70	1.71	1.70	1.64	1.66	2.10	1.66	1.63
		32	150	1.82	1.79	1.54	1.78	1.56	1.74	1.74	1.39	1.74	1.48
			250	1.59	1.58	1.56	1.58	1.54	1.52	1.52	25.56	1.52	1.41
	3	8	150	2.55	2.54	2.58	2.54	2.55	3.45	3.56	3.64	3.56	3.45
			250	1.75	1.74	1.74	1.74	1.74	1.92	5.69	11.84	5.69	1.92
		32	150	1.82	1.79	1.52	1.78	1.57	1.72	1.72	1.44	1.72	1.46
			250	1.59	1.58	1.56	1.58	1.53	1.53	1.53	1.49	1.53	1.44

Table 5 RMSPE under the DBSTRBFs method for the scenarios presented in Table 3 (spatio-temporal variation and variance factor cases)

Parameter			nT	Spatio-temporal variation					Variance function				
η	r	$n_h t_h$		MQ	TPS	CRS	ST	EXP	MQ	TPS	CRS	ST	EXP
0.01	1	8	150	3.00	3.77	3.78	3.77	3.78	2.61	2.60	2.61	2.60	2.61
			250	1.74	1.79	1.78	1.79	1.76	1.81	1.84	1.83	1.83	1.83
		32	150	1.49	1.74	1.36	1.55	1.48	1.56	1.76	1.45	1.60	1.55
			250	1.45	1.54	1.39	1.48	1.44	1.50	1.57	1.51	1.53	1.50
	3	8	150	2.55	3.35	3.34	3.34	2.94	2.60	2.59	2.60	2.59	2.60
			250	1.88	3.05	3.04	3.04	2.02	1.80	1.81	1.81	1.81	1.81
		32	150	1.55	1.85	1.37	1.64	1.53	1.56	1.76	1.44	1.60	1.55
			250	1.52	1.64	1.41	1.57	1.51	1.50	1.57	1.51	1.53	1.50
0.1	1	8	150	3.00	3.77	3.81	3.77	3.01	2.60	2.60	2.64	2.60	2.61
			250	1.74	1.78	1.78	1.79	1.75	1.83	1.84	1.83	1.84	1.83
		32	150	1.73	1.73	1.48	1.73	1.48	1.78	1.77	1.53	1.76	1.55
			250	1.53	1.53	1.52	1.53	1.44	1.57	1.57	1.55	1.57	1.50
	3	8	150	2.57	3.36	3.42	3.35	2.56	2.59	2.59	2.62	2.59	2.59
			250	1.94	3.04	3.17	3.05	1.95	1.81	1.81	1.80	1.81	1.80
		32	150	1.83	1.84	1.52	1.84	1.52	1.78	1.77	1.53	1.76	1.55
			250	1.64	1.64	1.58	1.64	1.51	1.57	1.57	1.55	1.57	1.49

Table 6 Comparison of several DBSTRBF methods for the average temperatures in Croatia using LOOCV

	MQ	TPS	CRS	ST	EXP	GAU
η	0.001	0.001	0.200	0.001	0.001	0.010
ρ	0.000	0.001	0.000	0.100	0.000	0.000
RMSPE	2.333	2.284	2.242	2.251	2.311	2.274