


Article

Use of Correlated Data for Nonparametric Prediction of a Spatial Target Variable

Pilar García-Soidán ^{1,*}  and Tomás R. Cotos-Yáñez ² 

¹ Department of Statistics and Operations Research, University of Vigo, Campus A Xunqueira, 36005 Pontevedra, Spain

² Department of Statistics and Operations Research, University of Vigo, Campus As Lagoas, 32004 Ourense, Spain; cotos@uvigo.es

* Correspondence: pgarcia@uvigo.es

Received: 24 September 2020; Accepted: 18 November 2020; Published: 20 November 2020



Abstract: The kriging methodology can be applied to predict the value of a spatial variable at an unsampled location, from the available spatial data. Furthermore, additional information from secondary variables, correlated with the target one, can be included in the resulting predictor by using the cokriging techniques. The latter procedures require a previous specification of the multivariate dependence structure, difficult to characterize in practice in an appropriate way. To simplify this task, the current work introduces a nonparametric kernel approach for prediction, which satisfies good properties, such as asymptotic unbiasedness or the convergence to zero of the mean squared prediction error. The selection of the bandwidth parameters involved is also addressed, as well as the estimation of the remaining unknown terms in the kernel predictor. The performance of the new methodology is illustrated through numerical studies with simulated data, carried out in different scenarios. In addition, the proposed nonparametric approach is applied to predict the concentrations of a pollutant that represents a risk to human health, the cadmium, in the floodplain of the Meuse river (Netherlands), by incorporating the lead level as an auxiliary variable.

Keywords: bandwidth parameter; cokriging; covariogram; kernel method; prediction

1. Introduction

A major challenge for the World Health Organization (WHO) is to protect human health. With this aim, the WHO promotes a variety of strategies and plans, such as those addressed to provide scientific information about the harmful effects of certain chemicals on the population and recommendations for their management. Cadmium (Cd) is one of the elements included in the group of pollutants under control, whose toxicity can provoke renal or respiratory dysfunctions to humans and affect the skeleton system, as well as it is linked to cancer [1].

The presence of Cd in nature is scarce, mainly associated with zinc and lead. The wide industrial use of this metal releases Cd to the air, which may cause pollution in air, water and soil [2], even on areas far from the sources of emission, by the atmosphere or the river transport. This contaminant is absorbed by certain plants and aquatic organisms, so it can easily enter the food chain. Indeed, some research has been developed to predict the Cd concentrations in crops by taking into account the Cd content in soil, as well as other factors [3]. Thus, periodic assessments to check the soil levels of this variable would be advisable, since they would allow the authorities to detect those zones at risk and to adopt the necessary strategies to alleviate it. This task can be accomplished by collecting data of the target pollutant at some spatial locations in the corresponding domain and then applying appropriate tools to predict its values at a number of unsampled sites, so as to map the concentrations achieved in the whole region.

In geostatistics, prediction of one variable at an unsampled location is usually obtained by applying the kriging methodology on the data collected for the target variable [4,5]. Furthermore, when secondary variables have also been measured at a number of spatial locations and they present a significant correlation with the main variable, the whole data set can be used to improve the results. This is the key idea of the cokriging procedures [6], whose application may vary depending on whether or not the variables share the set of sampled sites, leading to the isotopic or heterotopic cases, respectively.

To solve a cokriging equation system, estimation of the multivariate dependence structure is needed, which is a complex task, since it demands the use of valid functions that must satisfy some relation constraints between each pair of variables. The linear coregionalization model (LCM) was one of the first proposals suggested to simplify the characterization of the multivariate dependence [7]. This approach is based on the idea that the required functions can be derived as a linear combination of a number of theoretical valid models [8]. Then, an appropriate selection of parameters and models is needed, which can be obtained in an iterative way, as developed in [9]. However, the LCM mechanism becomes limited for a large number of secondary variables, thus prompting alternative methods, such as the one presented in [10], derived from Bochner's theorem. Other approaches have been designed to provide more flexible models, based on considering moving averages [11] or multivariate Matérn functions [12]. The work presented in [13] offers a widespread review of procedures for multivariate dependence estimation, including the aforementioned proposals and their extension to the nonstationary setting.

The use of an auxiliary variable can be advisable, for instance, when it has been observed at more sites than the target one and both variables exhibit a strong dependence. Even, sometimes, the measurement of the secondary information is less time consuming and/or expensive than the main one. Thus, the incorporation of the resulting data can be advantageous and it provides a larger sample than the one derived from the principal variable, on which the cokriging approaches can be applied. This methodology has been widely used to solve a variety of problems, specially to assess the contamination level by any pollutant, whose concentration can be influenced by the presence of other variables (chemical elements, atmospheric variables, etc.). This way of proceeding has proved to result in an efficiency gain to determine the air pollution levels [14,15], the groundwater quality parameters [16,17] or the soil concentrations of different contaminants [18,19], among other examples. In addition, some research has been developed to extend the cokriging techniques for its application to functional random fields, taken as main or secondary variables [20,21], although the current work is restricted to the scalar case.

Next, a scheme of the universal cokriging approach is presented. Let $\{Z(s) = (Z_1(s), \dots, Z_p(s)) / s \in D \subset R^d\}$ be a p -variate random process, where s represents a spatial location in the observation region D . The more general case (heterotopy) is considered here, so that data from the different variables are not necessarily collected at the same sampling sites, leading us to assume that Z_i is observed on the set $S_i = \{s_{i1}, \dots, s_{in_i}\}$ of n_i locations, for $n_i > 0$ and $i = 1, 2, \dots, p$. Let $\{(s_{ij_i}, Z_i(s_{ij_i})) / 1 \leq j_i \leq n_i, 1 \leq i \leq p\}$ denote the whole set of data. Since only one incomplete realization of the multivariate process is available, it is necessary to assume some kind of stationarity on Z for developing inference on this process.

Suppose that Z_i can be modeled as

$$Z_i(s) = \mu_i(s) + Y_i(s), \quad (1)$$

for $i = 1, 2, \dots, p$, where $\mu_i(\cdot)$ is the deterministic trend and $Y_i(\cdot)$ is a second-order stationary random process with zero mean. Then, it follows that:

- (i) $E[Z_i(s)] = \mu_i(s)$, for all $s \in D$ and $i = 1, 2, \dots, p$.
- (ii) $\text{Cov}[Z_i(s), Z_{i'}(s')] = C_{ii'}(s - s')$, for all $s, s' \in D$ and $i, i' = 1, 2, \dots, p$.

Functions $C_{ii'}$ are referred to as cross-covariograms or cross-covariances.

From (ii), one has that

$$\text{Cov}[Z_i(s), Z_{i'}(s)] = C_{ii'}(0), \quad \text{Var}[Z_i(s)] = C_{ii}(0) = \sigma_i^2,$$

for all s and $i, i' = 1, 2, \dots, p$.

If we assumed instead intrinsic stationarity from the random processes Y_i , condition (ii) would be replaced by:

$$(ii') \quad \text{Cov}[Z_i(s) - Z_i(s'), Z_{i'}(s) - Z_{i'}(s')] = 2\gamma_{ii'}(s - s'), \text{ for all } s, s' \in D \text{ and } i, i' = 1, 2, \dots, p.$$

The cross-semivariograms $\gamma_{ii'}$ are symmetric functions, although this property does not necessarily follow for the cross-covariograms, when $i \neq i'$. Under second-order stationarity, the relationship between $\gamma_{ii'}$ and $C_{ii'}$ is given by

$$\gamma_{ii'}(t) = 2C_{ii'}(0) - C_{ii'}(t) - C_{ii'}(-t), \tag{2}$$

for $t \in \mathbb{R}^d$.

Without loss of generality, we take Z_1 as the main variable, while the remaining ones are considered as secondary variables. Then, a kriging predictor for Z_1 at an unobserved location s can be constructed as follows

$$\hat{Z}_1(s) = \sum_{i,j_i} \lambda_{ij_i} Z_i(s_{ij_i}), \tag{3}$$

where the parameters λ_{ij_i} would be obtained by minimizing

$$E \left[(\hat{Z}_1(s) - Z_1(s))^2 \right],$$

subject to the unbiasedness condition, namely,

$$\mu_1(s) = E \left[\hat{Z}_1(s) \right] = \sum_{i,j_i} \lambda_{ij_i} E \left[Z_i(s_{ij_i}) \right]. \tag{4}$$

For equality (4) to hold, it is sufficient to require that $\sum_{1,j_1} \lambda_{1j_1} = 1$ and $\sum_{i,j_i} \lambda_{ij_i} = 0$, for all $i \neq 1$. This would lead to the following objective function

$$E \left[(\hat{Z}_1(s) - Z_1(s))^2 \right] - l_1 \left(\sum_{j_1} \lambda_{1j_1} - 1 \right) - \sum_i l_i \sum_{j_i} \lambda_{ij_i}, \tag{5}$$

with l_i denoting the Lagrange multipliers.

By deriving function (5) with respect to each λ_{ij_i} and l_i , for each i and j_i , as well as equaling the result to zero, an equation system is obtained, whose solution provides the values of the unknown parameters λ_{ij_i} . They would be expressed in terms of the cross-covariograms $C_{ii'}$ or cross-semivariograms $\gamma_{ii'}$, difficult to be characterized in practice [6]. Indeed, the number of functions to be estimated amounts to $\frac{p^2+p}{2}$ and they must satisfy the positive definiteness property or the conditionally negative definiteness property, respectively, to guarantee the existence of a solution of the cokriging system. Besides, the different functions cannot be derived in an independent way, due to the underlying dependence among them.

Furthermore, the accuracy of the cokriging methodology is conditioned by different factors. A remarkable feature is related to the selection of models needed to estimate the dependence structure and the limitations of the cross-validation techniques to check the adequateness of the resulting functions [22]. Also, a low correlation level between the target and the secondary variables may discourage from applying cokriging on them [23] over univariate kriging, due to the increase in the uncertainty when the dimension augments. On the other hand, the quality of the data may

have an effect on the results, which is particularly negative for samples of small sizes or unevenly distributed [24].

To avoid the aforementioned problems, the current research provides a nonparametric alternative for prediction in this setting, derived by extending the univariate approach without covariables that was proposed in [25]. The current work is organized as follows. Section 2 introduces the main hypotheses to be required. The proposed nonparametric predictor is described in Section 3, as well as its properties, whose proofs are outlined in Appendix A and B. Section 4 presents the results of the numerical studies with simulated data, carried out to illustrate the performance of the new methodology. Section 5 shows the application of our proposal to map the cadmium concentrations on the floodplain soil of the Meuse river (Netherlands), by incorporating data of the lead content, due to the correlation between both variables.

2. Hypotheses

To derive our proposal, the following conditions are required:

Hypothesis 1 (H1). $\{Z(s) = (Z_1(s), \dots, Z_p(s)) / s \in D \subset \mathbb{R}^d\}$ is a p -variate random process, with $E[Z_i(s)^2] < +\infty$, for all $s \in D$ and $i = 1, \dots, p$.

Hypothesis 2 (H2). D is a bounded region, with positive d -dimensional volume.

Hypothesis 3 (H3). $Z_i(\cdot)$ can be modeled as given in (1), for all $i = 1, \dots, p$.

Hypothesis 4 (H4). μ_1 is twice continuously differentiable in a neighborhood of s , for all $s \in D$.

Hypothesis 5 (H5). C_{1i} is twice continuously differentiable in a neighborhood of 0, for all $i = 1, \dots, p$.

Hypothesis 6 (H6). A random design is assumed for the spatial locations, which follow a density function f on D that is bounded and twice continuously differentiable in a neighborhood of s , for all $s \in D$.

Hypothesis 7 (H7). K_d is a d -variate, compactly supported and symmetric kernel function.

Hypothesis 8 (H8). L is a univariate, compactly supported and symmetric kernel function.

3. Nonparametric Predictor

In the univariate setting, the nonparametric prediction of variable Z_1 at an unsampled location s can be addressed by only considering the data collected for the target variable, together with the application of the methodology developed in [25], yielding

$$\hat{Z}_1(s) = \frac{\sum_j K_d\left(\frac{s-s_{1j}}{h}\right) Z_1(s_{1j})}{\sum_j K_d\left(\frac{s-s_{1j}}{h}\right)},$$

where h represents a bandwidth parameter.

Based on this idea, a new nonparametric predictor can be designed so as to incorporate the whole observed values, provided either by the main variable (Z_1) or the correlated secondary ones, through a weighted average, in the following manner

$$\hat{Z}_1(s) = \sum_{i,j_i} p_{ij_i}(s) \left(\mu_1(s_{ij_i}) + \frac{\alpha_{1i}\sigma_1}{\sigma_i} (Z_i(s_{ij_i}) - \mu_i(s_{ij_i})) \right), \tag{6}$$

with

$$p_{ij_i}(s) = \frac{w_{ij_i}(s)}{\sum_{i,j_i} w_{ij_i}(s)} \text{ and } w_{ij_i}(s) = L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) K_d \left(\frac{s - s_{ij_i}}{h} \right),$$

where h and h_i represent bandwidth parameters and $\alpha_{i'}$ equals -1, 0 or 1, depending on whether $C_{i'}$ (0) is negative, zero or positive, respectively. Each weight $p_{ij_i}(s)$ has been defined to account for the correlation between $Z_1(s)$ and $Z_i(s_{ij_i})$. To avoid the border effects in predictor (6), boundary kernels could be used instead of symmetric ones.

Predictor (6) satisfies good properties, under appropriate hypotheses, which include conditions (H1)–(H8), as well as the convergence to 0 of h_i and h , as n_i and $n = \sum_i n_i$ diverge to $+\infty$, respectively, with n_i denoting the number of observations collected for Z_i , for $i = 1, \dots, p$. On the one hand, the nonparametric predictor is asymptotically unbiased, so that

$$B(s) = \text{Bias}(s) = E[\hat{Z}_1(s) - Z_1(s)] \rightarrow 0. \tag{7}$$

On the other, its mean squared prediction error (MSPE) converges to zero, namely,

$$M(s) = \text{MSPE}(s) = E \left[(\hat{Z}_1(s) - Z_1(s))^2 \right] \rightarrow 0. \tag{8}$$

The proofs of these properties are outlined in Appendix A and B. Also, a proposal for approximation of $M(s)$ in practice is included in Appendix B, given by (A13) and defined in terms of the cross-covariances C_{1i} . It is important to notice that both properties, (7) and (8), are checked under the assumption of a random design for the spatial sampling locations, although they also follow for predictor (6) when deterministic designs are considered instead.

Next, we deal with the selection of the bandwidth parameters needed for implementation of predictor (6). Starting with h_i , write a_i for some sequence of positive values converging to 1, as n_i tends to $+\infty$, with $a_i < 1$. Suppose that the correlation between $Z_1(s)$ and $Z_i(s)$ is considered significant if $\text{Corr}[Z_1(s), Z_i(s)]^2 > a_i$. Seeking to meet this requirement for the data used to compute (6), h_i can be taken as

$$h_i = b^{-1} \sigma_1^2 \sigma_i^2 (1 - a_i), \tag{9}$$

where b stands for the support of L .

Regarding the bandwidth h , global or local selectors can be considered. An option of the former group may be derived from the traditional cross-validation method [26], adapted to this setting. To proceed in this way, for each j_1 and a given h , we would leave the observation $Z_1(s_{1j_1})$ out and use the remaining data to predict Z_1 at s_{1j_1} , through (6). Denote by $\hat{Z}_{1h}^{-(j_1)}(s_{1j_1})$ the resulting value and repeat this scheme for different choices of $h \in H$, where H represents an appropriate set of positive values. A global bandwidth h can be obtained as

$$h_{glo} = \text{argmin}_{h \in H} \left\{ \frac{1}{n_1} \sum_{j_1} \left(Z_1(s_{1j_1}) - \hat{Z}_{1h}^{-(j_1)}(s_{1j_1}) \right)^2 \right\}.$$

More generally, the set of sampling locations S_1 can be divided into two disjoint subsets, S_{1t} and S_{1v} , to be employed for training and validation, respectively. By omitting the data $Z_1(s_{1j_1})$, with $s_{1j_1} \in S_{1v}$, we can predict them from the values of the process at the training set S_{1t} and some h , through (6). Write $\hat{Z}_{1h}^{-(v)}(s_{1j_1})$ for the predictions attained and obtain a global selector h as follows

$$h_{glo} = \text{argmin}_{h \in H} \left\{ \frac{1}{|S_{1v}|} \sum_{j_1/s_{1j_1} \in S_{1v}} \left(Z_1(s_{1j_1}) - \hat{Z}_{1h}^{-(v)}(s_{1j_1}) \right)^2 \right\}, \tag{10}$$

with $|A|$ denoting the cardinal of the set A .

If a local bandwidth $h = h(s)$ is preferred instead, a simple idea to derive it can be based on applying the balloon approach [27]. With this aim, bear in mind that the kernel density K_d is compactly supported, so that $K_d(y) = 0$, for all $y \in R^d$ such that $\|y\| \geq c_d$ and some $c_d > 0$. Then, h can be taken to equal the m -th percentile (for instance, m equal to 10% or 20%) of the values $c_d^{-1} \|s - s_{ij_i}\|$, for each $s \in D$. More complex alternatives for acquiring local bandwidths can be derived from the bootstrap techniques, as the one introduced in [25] for univariate prediction without covariables.

In addition to the bandwidth parameters, implementation of predictor (6) also requires estimation of other unknown terms, dependent on the trend and the covariogram functions. These issues can be addressed in a variety of ways, according to the conditions satisfied by the underlying random processes. Furthermore, parametric or nonparametric approaches can be selected to derive the necessary estimates.

Starting with the trend function $\mu_i(\cdot)$, the sample mean could be applied, under the assumption of constant trend,

$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j_i} Z_i(s_{ij_i}).$$

However, when the constant condition fails for the trend, this function can be approximated through a parametric model (e.g. a polynomial function), although misspecification of the selected model is an important problem that may arise with the parametric procedures and it could affect the accuracy of the results derived from it. Nonparametric alternatives for trend estimation include, for instance, the Nadaraya-Watson approach

$$\tilde{\mu}_i(s) = \frac{\sum_{j_i} K_d\left(\frac{s-s_{ij_i}}{g_i}\right) Z_i(s_{ij_i})}{\sum_{j_i} K_d\left(\frac{s-s_{ij_i}}{g_i}\right)},$$

where g_i is a bandwidth parameter, or the local linear trend [28].

The conditions required in the current work lead us to assume constant variance from each process Z_i . Thus, a natural estimator of σ_i^2 is provided by the sample variance

$$\hat{\sigma}_i^2 = \frac{1}{n_i} \sum_{j_i} (Z_i(s_{ij_i}) - \tilde{\mu}_i(s_{ij_i}))^2.$$

In addition, we could apply a kernel-based approach for approximation of the variance, as the one given by

$$\tilde{\sigma}_i^2 = \frac{\sum_{j_1, j_2} K_d\left(\frac{s_{j_1}-s_{j_2}}{g_i}\right) (Z_i(s_{j_1}) - \tilde{\mu}_i(s_{j_1})) (Z_i(s_{j_2}) - \tilde{\mu}_i(s_{j_2}))}{\sum_{j_1, j_2} K_d\left(\frac{s_{j_1}-s_{j_2}}{g_i}\right)},$$

with g_i representing a bandwidth parameter and $\tilde{\mu}_i$ denoting an estimator of the trend μ_i .

Estimation of function C_{1_i} can be addressed through a parametric fit, which would be also strongly dependent on an adequate selection of the parametric model. A nonparametric mechanism for the latter issue is provided in [13], based on the method of moments, which can be written as follows

$$\hat{C}_{1_i}(t) = \frac{1}{|N_{1_i}(t)|} \sum_{j_1, j_i \in N_{1_i}(t)} (Z_1(s_{1_{j_1}}) - \tilde{\mu}_1(s_{1_{j_1}})) (Z_i(s_{ij_i}) - \tilde{\mu}_i(s_{ij_i})),$$

for $t \in R^d$, with $N_{1_i}(t) = \{(j_1, j_i) / s_{1_{j_1}} - s_{ij_i} \approx t\}$ and $\tilde{\mu}_i$ denoting an estimator of the trend function μ_i .

A smoothed version of the cross-covariance C_{1i} can be derived from the kernel method, proposed in [29], as

$$\hat{C}_{1i}(t) = \frac{\sum_{j_1, j_i} K_d \left(\frac{s_{1j_1} - s_{ij_i} - t}{g_i} \right) (Z_1(s_{1j_1}) - \tilde{\mu}_1(s_{1j_1})) (Z_i(s_{ij_i}) - \tilde{\mu}_i(s_{ij_i}))}{\sum_{j_1, j_i} K_d \left(\frac{s_{1j_1} - s_{ij_i} - t}{g_i} \right)},$$

for some bandwidth parameter g_i and estimator $\tilde{\mu}_i$ of μ_i .

Remark 1. The new prediction methodology must perform adequately for Gaussian data, similarly to what might be expected for the cokriging predictor (3). However, the latter one demands an appropriate characterization of the whole cross-covariograms $C_{ii'}$ (or cross-semivariograms), which should additionally satisfy a number of restrictions so as to be valid for prediction, due to the underlying dependence among them, unlike what is needed to develop our proposal. Indeed, predictor (6) and its approximated error (A13) only involve terms dependent on the functions C_{1i} , whose estimators are not required to accomplish the aforementioned conditions, thus simplifying the practical implementation of the nonparametric approach.

Remark 2. For construction of a prediction map, implementation of (6) is not so computationally demanding for each site as the cokriging predictor, since the terms $K_d \left(\frac{s - s_{ij_i}}{h} \right)$ are the unique ones that involve the target location.

Remark 3. Under isotopy, the new approach for prediction can be adapted in a simple way. With this aim, suppose that Z_i is observed on the set $S_i = S = \{s_j\}_{j=1}^n$ of n locations, for all $i = 1, \dots, p$. The kernel-type predictor of Z_1 at location s would then be given by

$$\hat{Z}_1(s) = \sum_j p_{1j}(s) \left(\mu_1(s_j) + \sum_i p_{2i} \left(\frac{\alpha_{1i} \sigma_1}{\sigma_i} (Z_i(s_j) - \mu_i(s_j)) \right) \right),$$

where

$$p_{1j}(s) = \frac{K_d \left(\frac{s - s_j}{h} \right)}{\sum_j K_d \left(\frac{s - s_j}{h} \right)} \text{ and } p_{2i} = \frac{L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right)}{\sum_i L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right)},$$

with h and h_i denoting the bandwidth parameters, for $i = 1, \dots, p$.

4. Numerical Studies

This section presents the results of the numerical studies with simulated data, developed to illustrate the behavior of the proposed nonparametric predictor (np). For the sake of comparison, the cokriging (cokr) and the kriging (kr) approaches were also implemented.

The samples used in the analyses were generated from bivariate random processes, modeled as given in (1), for $D = [0, 1] \times [0, 1]$ and $p = 2$. To derive correlated variables Z_1 and Z_2 , we took Y_1 and Y_2 to follow a linear model of coregionalization [9], so that $Y_i(s) = M_i^t (V_1(s), V_2(s))^t$, where V_i represent univariate and independent stationary processes, with zero mean and unit variance, and vectors $M_i = (m_{1i}, m_{2i})^t$, for some real values m_{ij} and $i, j = 1, 2$. Equivalently, the bivariate random process was obtained as

$$Z(s) = \begin{pmatrix} Z_1(s) \\ Z_2(s) \end{pmatrix} = \begin{pmatrix} \mu_1(s) \\ \mu_2(s) \end{pmatrix} + M \begin{pmatrix} V_1(s) \\ V_2(s) \end{pmatrix}, \tag{11}$$

with function μ_i denoting the trend of Z_i and $M = (m_{ij})$ being a 2×2 matrix, whose i -th column equals M_i , for $i = 1, 2$.

Various semivariogram models γ_i were considered to characterize the dependence structure of V_i , with different values for the nugget effect τ_i^2 and the range (or the asymptotic range) ϕ_i , which are

specified in each analysis. Under this configuration, one has the following cross-covariograms for process Z :

$$\begin{aligned} C_{11}(t) &= m_{11}^2 (1 - \gamma_1(t)) + m_{21}^2 (1 - \gamma_2(t)), \\ C_{22}(t) &= m_{21}^2 (1 - \gamma_1(t)) + m_{22}^2 (1 - \gamma_2(t)), \\ C_{12}(t) &= m_{11}m_{12} (1 - \gamma_1(t)) + m_{21}m_{22} (1 - \gamma_2(t)). \end{aligned}$$

The cross-semivariograms are given by:

$$\begin{aligned} \gamma_{11}(t) &= m_{11}^2 \gamma_1(t) + m_{21}^2 \gamma_2(t), \\ \gamma_{22}(t) &= m_{21}^2 \gamma_1(t) + m_{22}^2 \gamma_2(t), \\ \gamma_{12}(t) &= m_{11}m_{12} \gamma_1(t) + m_{21}m_{22} \gamma_2(t). \end{aligned} \tag{12}$$

The correlation function between Z_1 and Z_2 amounts to

$$\rho_{12}(t) = \text{Corr} [Z_1(s), Z_2(s + t)] = \frac{m_{11}m_{12} (1 - \gamma_1(t)) + m_{21}m_{22} (1 - \gamma_2(t))}{\sqrt{(m_{11}^2 + m_{21}^2) (m_{12}^2 + m_{22}^2)}}$$

and, therefore:

$$\rho_{12}(0) = \frac{m_{11}m_{12} + m_{21}m_{22}}{\sqrt{(m_{11}^2 + m_{21}^2) (m_{12}^2 + m_{22}^2)}}.$$

We focused on prediction with heterotopic data. Thus, the sampling locations for Z_1 and Z_2 were respectively selected as random sites s_{1j_1} drawn from the uniform distribution on D , for $j_1 = 1, \dots, n_1$, and equispaced sites s_{2j_2} within D , for $j_2 = 1, \dots, n_2$. In addition, we took a fixed regular grid of 10×10 locations s_{0k} in D , with $k = 1, \dots, 100$, for the prediction points. Then, for each model, 500 samples were generated from Z_1 at locations s_{1j_1} and s_{0k} , as well as from Z_2 at sites s_{2j_2} , for different sizes n_1 and n_2 . To assess the results derived for each prediction method, the corresponding averaged mean squared prediction error (AMSPE) was approximated as follows:

$$\text{AMSPE} = 100^{-1} \sum_{k=1}^{100} \text{MSPE} (s_{0k}) = (100 \cdot 500)^{-1} \sum_{k=1}^{100} \sum_{m=1}^{500} \left(\hat{Z}_1^{(m)} (s_{0k}) - Z_1^{(m)} (s_{0k}) \right)^2,$$

where $Z_1^{(m)} (s_{0k})$ denotes the value assigned to $Z_1 (s_{0k})$ in the m -sample and $\hat{Z}_1^{(m)} (s_{0k})$ stands for the prediction of $Z_1 (s_{0k})$ obtained through the m -sample, for $m = 1, \dots, 500$.

Implementation of the three predictors was developed by using different packages of the R library. The kriging and the cokriging predictors were computed with the `gstat` [30] and the `RandomFields` [31] packages. The `npsp` package [32] was employed to obtain predictor (6), combined with the `locpol` package [33], to yield the nonparametric (local-linear) trend estimates. Regarding the bandwidths needed to derive (6), h_i was chosen as given in (9) and selector h was attained through (10), by considering S_{1t} as the set of sampling locations for Z_1 and S_{1v} as the set of prediction sites.

For the first analysis, constant trends μ_1 and μ_2 were taken, as well as Gaussian random fields V_1 and V_2 , with zero mean, unit variance and isotropic semivariograms following a pure nugget model and a spherical model, respectively, where $\tau_1^2 = 1$, $\tau_2^2 = 0$ and $\phi_2 = 0.5$. We checked the performance of the nonparametric predictor for distinct choices of vectors M_1 and M_2 , given by $M_1 = (\sqrt{0.1}, \sqrt{0.2})^t$ and $M_2 = (\sqrt{0.9}, \sqrt{0.8})^t$, $M_1 = (0, \sqrt{0.2})^t$ and $M_2 = (\sqrt{0.9}, \sqrt{0.25})^t$ and $M_1 = (0, \sqrt{0.2})^t$ and $M_2 = (\sqrt{0.9}, \sqrt{0.1})^t$, so that $\rho_{12}(0)$ approximately equals 0.99, 0.74 and 0.57, respectively. For implementation of the kriging and cokriging, constant trends were assumed and the estimation of the dependence structure was addressed parametrically in both cases, through the theoretical models. Table 1 summarizes the values achieved for the AMSPE with different sample sizes,

where the values of $\rho_{12}(0)$ have been included in the table to allow the identification of the models. However, it is important to notice that the relationship among the values of $\rho_{12}(0)$ for the different models does not necessarily hold for the cross-correlation at other lags.

Table 1. AMSPE values obtained for several sample sizes n_1 and n_2 and correlations ρ_{12} , for data drawn from random fields Z_1 and Z_2 with constant trends, derived from Gaussian processes V_1 and V_2 with nugget and spherical semivariograms, respectively. For implementation of the kriging and cokriging approaches, the dependence structures of V_1 and V_2 were estimated through the theoretical models.

(n_1, n_2)	$\rho_{12}(0) = 0.99$			$\rho_{12}(0) = 0.74$			$\rho_{12}(0) = 0.57$		
	cokr	np	kr	cokr	np	kr	cokr	np	kr
(100,100)	0.26	0.31	0.38	0.19	0.23	0.28	0.21	0.25	0.29
(400,400)	0.17	0.19	0.27	0.10	0.12	0.20	0.10	0.14	0.18
(900,900)	0.13	0.15	0.24	0.07	0.09	0.15	0.07	0.11	0.15
(100,625)	0.18	0.20	0.38	0.14	0.17	0.29	0.17	0.21	0.30
(625,100)	0.17	0.20	0.26	0.08	0.10	0.17	0.08	0.09	0.14

The kriging technique does not incorporate the information provided by the secondary variable, giving rise to the largest errors, as shown in Table 1. The best performance is exhibited by the cokriging predictor, for all sample sizes and correlation functions. An intermediate behavior is observed for the nonparametric approach, with values close to the cokriging predictions, specially for greater sizes. Another remarkable feature is the reduction in the errors achieved for the same size n_1 , with the cokriging method and the new predictor, under an increment of the secondary sample size. Anyway, the results obtained with the proposed methodology are appealing, if we bear in mind the advantages given to the kriging and cokriging procedures. Indeed, the latter methods are optimal under gaussianity and their application was accomplished by a correct specification of the trend and the semivariogram functions, parametrically estimated through the theoretical models.

A deeper analysis was carried out for the bivariate data generated with vectors $M_1 = (\sqrt{0.1}, \sqrt{0.2})^t$ and $M_2 = (\sqrt{0.9}, \sqrt{0.8})^t$ that provide the highest cross-correlation at lag 0, namely, $\rho_{12}(0) = 0.99$. Table 2 and Figure 1 display the results achieved for the AMSPE, with several sizes n_1 and n_2 , where a similar pattern to the one previously described is shown. Indeed, those methods that make use of the auxiliary data (cokriging and nonparametric predictors) yield the smallest errors and they decrease for larger sizes n_1 and n_2 , with a slight outperformance of the cokriging approach over the nonparametric one. This fact may again be due to the favorable scenario designed for the cokriging technique, although the difference between both procedures tends to diminish as n_1 and n_2 increase.

Table 2. AMSPE values obtained for several sample sizes n_1 and n_2 and correlation $\rho_{12}(0) = 0.99$, for data drawn from random fields Z_1 and Z_2 with constant trends, derived from Gaussian processes V_1 and V_2 with nugget and spherical semivariograms, respectively. For implementation of the kriging and cokriging approaches, the dependence structures of V_1 and V_2 were estimated through the theoretical models.

n_1	$n_2 = 100$			$n_2 = 225$			$n_2 = 400$			$n_2 = 625$			$n_2 = 900$		
	cokr	np	kr	cokr	np	kr	cokr	np	kr	cokr	np	kr	cokr	np	kr
100	0.26	0.31	0.39	0.22	0.25	0.37	0.20	0.22	0.37	0.18	0.21	0.38	0.17	0.18	0.39
225	0.21	0.26	0.30	0.19	0.22	0.29	0.18	0.21	0.30	0.16	0.19	0.30	0.16	0.18	0.32
400	0.19	0.21	0.28	0.18	0.19	0.28	0.17	0.19	0.27	0.15	0.18	0.27	0.15	0.17	0.28
625	0.17	0.20	0.26	0.17	0.18	0.25	0.16	0.17	0.25	0.15	0.17	0.26	0.14	0.16	0.25
900	0.16	0.18	0.24	0.16	0.17	0.24	0.15	0.16	0.25	0.14	0.16	0.23	0.13	0.15	0.24

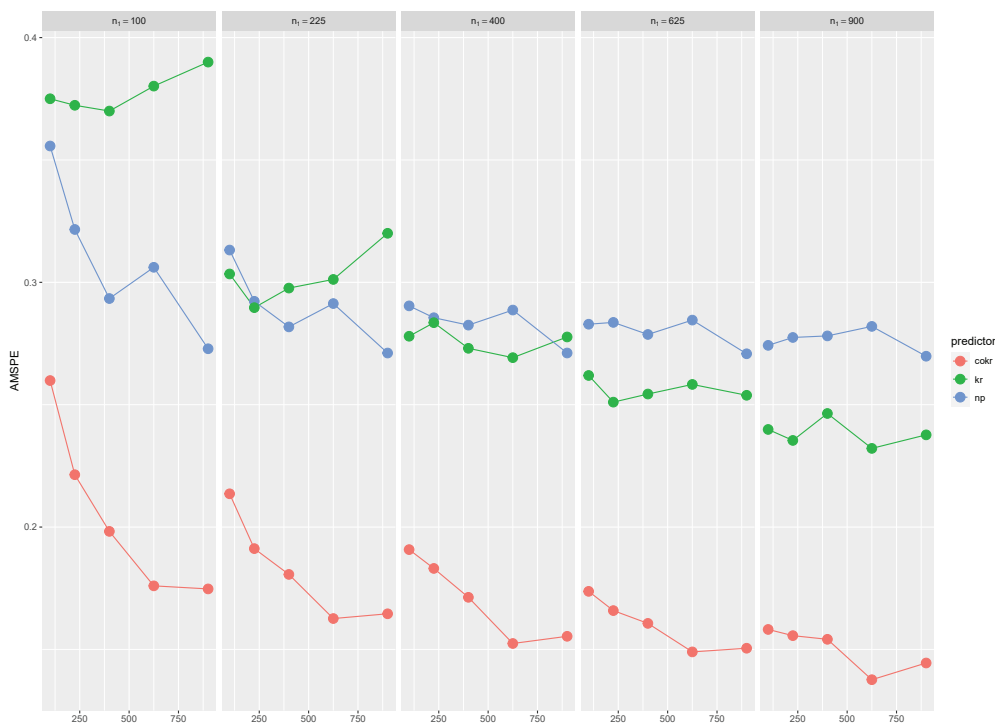


Figure 1. Plots of the AMSPE values obtained for several sample sizes n_1 and n_2 and correlation $\rho_{12}(0) = 0.99$, for data drawn from random fields Z_1 and Z_2 with constant trends, derived from Gaussian processes V_1 and V_2 with nugget and spherical semivariograms, respectively. For implementation of the kriging and cokriging approaches, the dependence structures of V_1 and V_2 were estimated through the theoretical models.

Thus, the first study confirms some benefits of using the cokriging technique, when Gaussian processes are considered, with a correct specification of the estimated trend and spatial dependence. Then, a further step in our research is addressed to analyze the development of the new methodology for prediction, under deviations from the aforementioned features.

We started by checking the effect of misspecification in the parametric estimation of the semivariogram functions. For the latter purpose, in the second study, bivariate data were drawn from independent Gaussian processes V_1 and V_2 , with isotropic exponential (Exp) and Matérn (Mat) semivariograms, respectively. The Matérn model involves an additional parameter ν and the particular choice of $\nu = 0.5$ provides the exponential model. To generate the samples in our analyses, we considered constant trends μ_1 and μ_2 , as well as the semivariogram parameters $\tau_i^2 = 0$ and $\phi_i = 0.5$, for $i = 1, 2$, together with $\nu = 2$ for the Matérn function. The cross-semivariograms $\gamma_{ii'}$ were obtained as given in (12), through several combinations of semivariogram models selected for V_1 and V_2 , which include the theoretical ones (Exp-Mat), as well as misspecified alternatives obtained by considering the spherical (Sph) family for V_1 or the Gaussian (Gau) family for V_2 . Figure 2 depicts the form of the cross-semivariograms employed in the current study, so that the functions derived under a correct specification of the theoretical model are compared with their counterparts, achieved with a partially misspecified model (Figure 2(a)) and a totally misspecified one (Figure 2(b)). These plots show that the wrong selection of the parametric families leads to inaccurate estimates of the semivariograms at small distances, with the largest deviations observed for the totally misspecified scenario, as expected, for which the resulting nuggets clearly differ from the corresponding values attained through the theoretical model. These variations of the semivariograms estimates at lags close to zero should have a noticeable effect on prediction, which was checked in the analysis described below.

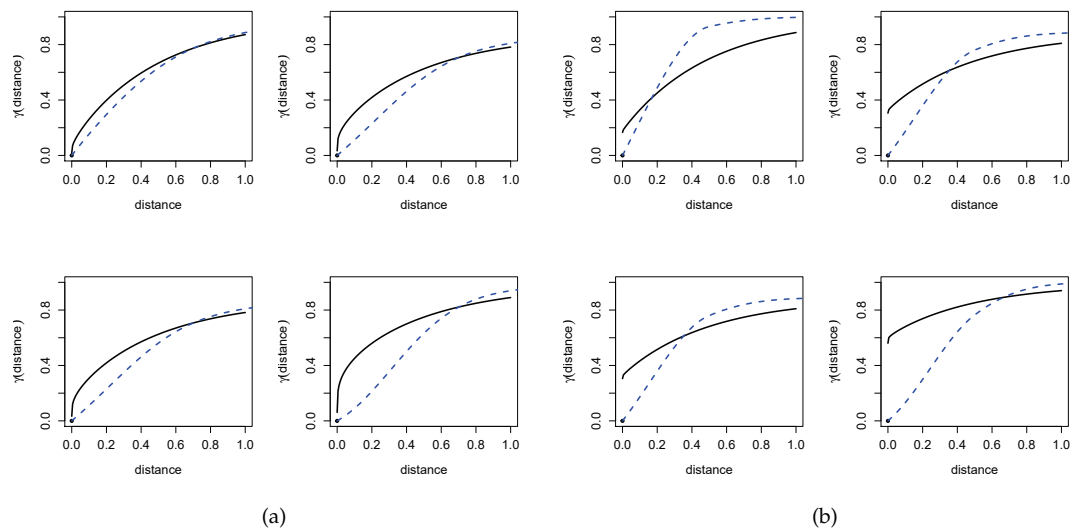


Figure 2. (a) Plots of γ_{ii}^h under a correct characterization of γ_1 and γ_2 (Exp-Mat, black solid line) and a misspecified combination (Exp-Gau, blue dashed line). (b) Plots of γ_{ii}^h under a correct characterization of γ_1 and γ_2 (Exp-Mat, black solid line) and a misspecified combination (Sph-Gau, blue dashed line).

Table 3 presents the results of the second numerical study, addressed to compare the performance of the three prediction approaches, under different specifications of the dependence structure. As expected, the use of the theoretical models (Exp-Mat) gives rise to the best results, with the smallest errors attained through the cokriging and the nonparametric methods, showing some distance to the corresponding values provided by the kriging predictor. On the other hand, the misspecification does not have an effect on the new methodology for prediction, unlike what happens to the other procedures. Indeed, the wrong selection of one of the semivariogram models (Exp-Gau) leads to an augment in the errors derived through the kriging and cokriging techniques. This trend is even more evident when using a totally misspecified model (Sph-Gau), which results in an additional increment of the AMSPE values obtained.

Table 3. AMSPE values obtained for several sample sizes n_1 and n_2 , for data drawn from random fields Z_1 and Z_2 with constant trends, derived from Gaussian processes V_1 and V_2 with exponential and Matérn semivariograms, respectively. For implementation of the kriging and cokriging approaches, the dependence structures of V_1 and V_2 were estimated through the theoretical models (Exp-Mat), together with misspecified alternatives either for the second process (Exp-Gau) or for both processes (Sph-Gau).

(n_1, n_2)	Exp-Mat			Exp-Gau			Sph-Gau		
	cokr	np	kr	cokr	np	kr	cokr	np	kr
(100,100)	0.16	0.20	0.27	0.33	0.21	0.40	0.43	0.29	0.48
(225,225)	0.10	0.11	0.20	0.25	0.18	0.33	0.35	0.24	0.40
(400,400)	0.08	0.08	0.18	0.21	0.16	0.29	0.31	0.22	0.37
(625,625)	0.06	0.08	0.16	0.18	0.14	0.26	0.28	0.21	0.34
(900,900)	0.06	0.07	0.15	0.17	0.14	0.24	0.28	0.20	0.34

In the previous studies, the trends μ_1 and μ_2 were taken as constant values, so the third analysis is focused on exploring the performance of the three methods considered for prediction under non-constant trend. With this aim, data were generated from model (11), with the same characteristics as in the second study, except that now linear trends were selected, given by $\mu_i(s) = 1 + (-1)^{i+1}(s_1 + s_2)$, for $s = (s_1, s_2)$ and $i = 1, 2$. Table 4 shows the errors achieved with the

proposed nonparametric predictor and the cokriging approach. In the former case, the trend was nonparametrically estimated, whereas for the cokriging predictor, the approximation of the trend was developed through a constant and a linear models.

Table 4. AMSPE values obtained for several sample sizes n_1 and n_2 , for data drawn from random fields Z_1 and Z_2 with linear trends, derived from Gaussian processes V_1 and V_2 with exponential and Matérn semivariograms, respectively. For implementation of the cokriging approach, the dependence structures of V_1 and V_2 were estimated through the theoretical models and the trend estimation was addressed either by considering a constant model (cm) or a linear model (lm).

(n_1, n_2)	np	cokr-lm	cokr-cm
(100,100)	0.20	0.19	0.23
(225,225)	0.11	0.10	0.15
(400,400)	0.08	0.08	0.11
(625,625)	0.07	0.06	0.09
(900,900)	0.06	0.05	0.09

The results in Table 4 reflect that the misspecification of the trend models yields the largest errors, even though the cokriging approach was implemented with the right models for the semivariograms. On the other hand, under a correct characterization of the trend, the cokriging and the nonparametric predictors have a similar performance, with a slight superiority of the former one. In this respect, recall the advantages given to the cokriging prediction in this scenario, since this technique is optimal under gaussianity from the random processes, together with the fact that the dependence structures and trends involved in this approach were correctly specified.

In the final numerical study, our aim was to analyze the effects that the deviations from gaussianity might have on the results achieved with the three predictors considered in the current research. Thus, to simulate a non-Gaussian bivariate process, modeled as in (11), we drew non-Gaussian processes V_1 and V_2 , by taking them as $V_i(s) = W_i(s)^2$, for some Gaussian processes W_i , with zero mean and unit variance, for $i = 1, 2$. This way of proceeding led to non-Gaussian but stationary random fields V_i , with mean 1, variance 2 and semivariogram $\gamma_i(t) = 2\gamma_{W_i}(t)(2 - \gamma_{W_i}(t))$, where γ_{W_i} stands for the semivariogram of W_i . In particular, for data generation in the current analysis, W_1 and W_2 were selected to have the same characteristics as V_i in the second study, so their semivariograms followed the isotropic exponential and Matérn models, respectively, with parameters $\nu = 2$, $\tau_i^2 = 0$ and $\phi_i = 0.5$, for $i = 1, 2$. The results are summarized in Table 5.

Table 5. AMSPE values obtained for several sample sizes n_1 and n_2 , for data drawn form random fields Z_1 and Z_2 with constant trends, derived from $V_1 = W_1^2$ and $V_2 = W_2^2$, where W_1 and W_2 are zero mean Gaussian processes, with exponential and Matérn semivariograms, respectively. For implementation of the cokriging approach, the dependence structures of V_1 and V_2 were estimated through the exponential and Matérn models, respectively.

(n_1, n_2)	cokr	np	kr
(100,100)	0.34	0.09	0.43
(225,225)	0.24	0.08	0.35
(400,400)	0.23	0.08	0.31
(625,625)	0.20	0.06	0.28
(900,900)	0.12	0.04	0.21

The errors displayed in Table 5 give account of the best performance of the nonparametric predictor over the kriging and cokriging techniques, when the underlying distribution of the random processes V_1 and V_2 departs from gaussianity. Indeed, several problems arise in this scenario with the application of the kriging and cokriging methods. On the one hand, these approaches are no longer

optimal for non-Gaussian processes and, on the other, their behavior is highly dependent on the use of proper specifications of the trend and the semivariogram functions.

5. Assessment of Cd Concentrations in the Floodplain of the Meuse River

Now, we describe the application of the proposed methodology to assess the contamination by Cd in the geographical area associated to the Meuse data set. These data, presented in [34], contain measurements of different pollutants (zinc, cadmium, lead, copper or organic matter, among others), taken on alluvial soils in the floodplain of the Meuse river, close to Stein (Netherlands). A map of the zone is displayed in Figure 3, designed with the RgoogleMaps package [35] of the R library, which shows the 155 sampling locations considered for data collection.

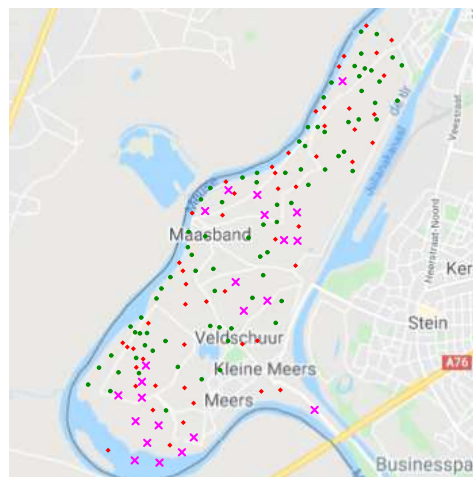


Figure 3. Area in the floodplain of the Meuse river considered for measurement of different pollutants at 155 sampling locations. The 132 points marked on the map display those sites taken for training, by using the values of cadmium and lead respectively collected at the 52 red points and 80 green points. The 23 crosses represent those locations used for validation.

The aforementioned sample has been studied with different aims and methods, which include the application of the cokriging techniques. An example can be found in [36], where cokriging was employed with lead (Pb) as target variable and organic matter or zinc as separate auxiliary variables. In the current research, we focused our attention on Cd, since it represents a risk to human health, with the aim of mapping the concentrations of this pollutant on the whole area of interest. This task was accomplished by using the collected Cd and Pb values, due to the association between both elements, whose correlation coefficient with the available data amounted to 0.7989.

A preliminary analysis of the Cd and Pb data depicted asymmetry, thus they were log-transformed to continue the study, where we applied the universal cokriging and the proposed nonparametric tool. The former approach was implemented by fitting a linear model of coregionalization with isotropic exponential semivariograms.

For the choice of bandwidth h in (6), as given in (10), we randomly divided the set of sampling sites into two disjoint subsets, one for training (with 132 sites) and the other for validation (with 23 sites), marked on Figure 3 as points and crosses, respectively. To proceed in this way, we first selected the spatial locations for the training set from the whole set of sampling sites and the remaining locations formed the validation set. In addition, aiming to solely use the value of one variable at each site, the training set was again split into two separate groups, so that we just considered the values of Cd collected at one of them (with 52 sites) and those of Pb at the other (with 80 sites), represented by red and green points on Figure 3, respectively. The resulting training data set was used to compute predictor (6) with different bandwidths and then we selected the one that best approximated the

log-values of the Cd concentrations at the validation set. The bandwidth h_{glo} achieved was employed to proceed with the new methodology.

Finally, the predictions derived with the data in the log-scale and both procedures were back-transformed to the original scale. The results attained for the Cd concentrations (in mg/kg soil), on the geographical area associated to the Meuse data set, are represented in Figure 4. To measure the accuracy of each mechanism, the AMSPE values were computed at the validation set and they amounted to 2.2737 and 0.8189 for the cokriging and the nonparametric approaches, respectively, thus giving advantage to the latter one.

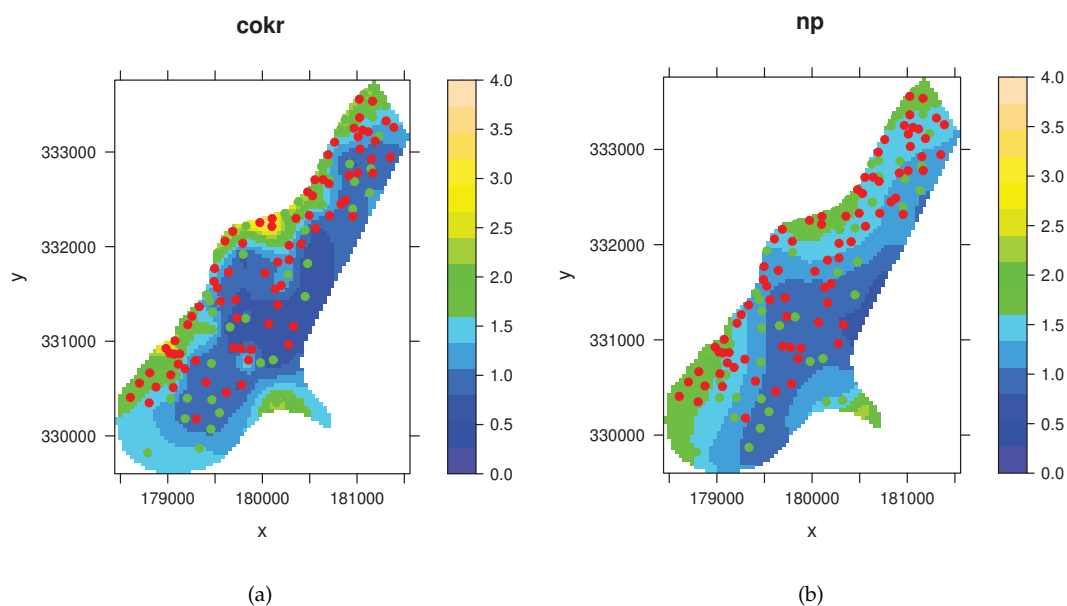


Figure 4. (a): Prediction map obtained for the Cd concentrations (in mg/kg soil), with Pb as auxiliary variable, through the universal cokriging. (b): Prediction map obtained for the Cd concentrations (in mg/kg soil), with Pb as auxiliary variable, through the nonparametric predictor.

The maps displayed in Figure 4 give evidence of the differences between the two prediction methods considered in the current study, although they also exhibit some common patterns. On the one hand, the area with the smallest levels of Cd, identified with the dark blue color, is located in each case on the central part of the region, towards the eastern border, being of larger size the one achieved with the universal cokriging. Then, as we move away from each of the aforementioned areas, the predicted Cd content increases and the highest values are mainly concentrated on the western border of each domain and its vicinity. Nevertheless, the resulting hot spots range in different magnitudes for the two prediction approaches, so that the universal cokriging provides the largest Cd accumulations, even surpassing the green color scale and reaching values above 3 mg/kg soil, unlike what happens to the non-parametric method.

In summary, both ways of proceeding allow the identification of those zones that pose a threat to the population health for their pollutant content. However, the smaller error (AMSPE) achieved by the nonparametric procedure and its good theoretical properties present this predictor as a better candidate for assessment of the points at risk and the consequent decision-making.

6. Conclusions

The current work deals with the construction of a prediction map that shows the concentrations of a pollutant, the cadmium, in the whole region of interest. This issue may be addressed by using the available data of the target variable (Cd) and also incorporating information of a secondary correlated one (Pb), collected at various spatial sites. In geostatistics, problems of this kind are typically solved

through the kriging or the cokriging techniques, depending on whether data of one or more variables are employed, respectively. These methods are optimal for Gaussian processes, although their practical implementation requires an appropriate characterization of the underlying dependence structure. This is not an easy task to be accomplished, specially when cokriging is applied, due to the number of functions to be approximated and the properties that the resulting estimates must satisfy. Thus, our research introduces an alternative kernel-type approach for spatial prediction with covariables, which was designed to overcome the previous drawbacks. Indeed, the new methodology just involves p cross-covariances (or cross-semivariograms), for a p -variate random process, instead of the $\frac{p^2+p}{2}$ functions of more complex estimation that are needed for cokriging. To characterize the aforementioned p functions, as well as other unknown terms in our proposal, such as those dependent on the trends, we recommend proceeding through nonparametric procedures, to avoid the misspecification problem present in the parametric estimation. The choice of bandwidths, required to implement the kernel predictor, is also addressed in the current work. In particular, we propose a cross-validation method to derive a global bandwidth h , based on considering random training and validation sets. However, further research should be developed to analyze alternative mechanisms for the selection of sites to be included in each set and, even, to provide additional approaches for bandwidth selection, which could be derived from the application of the bootstrap methodology. The numerical studies carried out in the current study give account of the sensitivity of the kriging and the cokriging techniques to the proper specification of the trend and the cross-semivariograms, unlike what happens to the nonparametric predictor. In addition, deviations from gaussianity in the data are less noticeable when prediction is developed with the new methodology than with the other approaches. Then, the use of the nonparametric proposal is recommended over the cokriging to obtain more accurate predictions of the Cd concentrations, with Pb as auxiliary variable, so as to detect those areas with highest accumulation of Cd, which is crucial for health protection.

Author Contributions: Conceptualization, P.G.-S.; methodology, P.G.-S.; software, T.R.C.-Y.; validation, P.G.-S. and T.R.C.-Y.; formal analysis, P.G.-S. and T.R.C.-Y.; investigation, P.G.-S. and T.R.C.-Y.; resources, P.G.-S. and T.R.C.-Y.; data curation, T.R.C.-Y.; writing—original draft preparation, P.G.-S. and T.R.C.-Y.; writing—review and editing, P.G.-S. and T.R.C.-Y.; visualization, P.G.-S. and T.R.C.-Y.; supervision, P.G.-S.; project administration, P.G.-S.; funding acquisition, P.G.-S. and T.R.C.-Y. All authors have read and agreed to the published version of the manuscript.

Funding: The first author’s research was partially funded by the Spanish National Research and Development Program project TEC2015–65353–R, by the ERDF and by the Xunta de Galicia (Spain), under project GRC 2015/018 and under agreement for funding AtlantTIC (Atlantic Research Center for Information and Communication Technologies). The second author’s work has been partially supported by grant MTM2017–89422–P from the Spanish Ministry of Science and Innovation and by the Xunta de Galicia (Spain), under project GRC ED431C 2016/040.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Property (7)

In this appendix, we check that relation (7) holds or, equivalently, that $\hat{Z}_1(s)$ is asymptotically unbiased. With this aim, bear in mind that

$$\begin{aligned}
 B(s) &= E [\hat{Z}_1(s) - Z_1(s)] = E [E [\hat{Z}_1(s) - Z_1(s) / s_{ij_i}, \forall i, j_i]] = & (A1) \\
 &= E \left[\frac{\sum_{i,j_i} w_{ij_i}(s) E [A_{1ij_i}(s) / s_{ij_i}, \forall i, j_i]}{\sum_{i,j_i} w_{ij_i}(s)} \right] = E \left[\frac{A_2(s)}{A_3(s)} \right],
 \end{aligned}$$

with

$$\begin{aligned}
 w_{ij_i}(s) &= L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) K_d \left(\frac{s - s_{ij_i}}{h} \right), \\
 A_{1ij_i}(s) &= \mu_1(s_{ij_i}) + \frac{\alpha_{1i} \sigma_1}{\sigma_i} (Z_i(s_{ij_i}) - \mu_i(s_{ij_i})) - Z_1(s), \\
 A_2(s) &= \sum_{i,j_i} w_{ij_i}(s) (\mu_1(s_{ij_i}) - \mu_1(s)), \\
 A_3(s) &= \sum_{i,j_i} w_{ij_i}(s).
 \end{aligned} \tag{A2}$$

It follows that

$$\begin{aligned}
 A_2(s) &= \sum_i n_i L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} K_d \left(\frac{s - u}{h} \right) (\mu_1(u) - \mu_1(s)) f(u) du \quad a.s. \\
 A_3(s) &= \sum_i n_i L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} K_d \left(\frac{s - u}{h} \right) f(u) du \quad a.s.
 \end{aligned}$$

By making the change of variable $y = \frac{s-u}{h}$ in the previous integrals, one has that

$$\begin{aligned}
 A_2(s) &= \sum_i n_i h^d L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) \int_{supp(K_d)} K_d(y) (\mu_1(s - hy) - \mu_1(s)) f(s - hy) dy \quad a.s. \\
 A_3(s) &= \sum_i n_i h^d L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) \int_{supp(K_d)} K_d(y) f(s - hy) dy \quad a.s.
 \end{aligned}$$

with $supp(K_d)$ denoting the support of K_d . Then, we can form the Taylor expansions for $\mu_1(s - hy)$ and $f(s - hy)$ about s , to obtain that

$$A_2(s) = \sum_i n_i L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) O(h^{d+2}) = O\left(h^{d+2} \sum_i n_i\right) \quad a.s. \tag{A3}$$

$$A_3(s) = \sum_i n_i L \left(\frac{C_{1i}(0)^2 - \sigma_1^2 \sigma_i^2}{h_i} \right) (h^d f(s) + O(h^{d+2})) = O\left(h^d \sum_i n_i\right) \quad a.s. \tag{A4}$$

Relation (A1) and the orders achieved in (A3) and (A4), as well as the application of the Slutsky's theorem, lead us to $B(s) = O(h^2)$. Therefore, predictor (6) is asymptotically unbiased.

Appendix B. Proof of Property (8)

This appendix outlines the proof of relation (8), which establishes the convergence to zero of the mean squared prediction error.

To start, one has that

$$\begin{aligned}
 M(s) &= E \left[(\hat{Z}_1(s) - Z_1(s))^2 \right] = E \left[E \left[(\hat{Z}_1(s) - Z_1(s))^2 / s_{ij_i}, \forall i, j_i \right] \right] \\
 &= E \left[\frac{\sum_{i_1, j_{i_1}} \sum_{i_2, j_{i_2}} E \left[A_{4i_1 j_{i_1} i_2 j_{i_2}}(s) / s_{i_k j_{i_k}}, \forall i_k, j_{i_k}, k \right]}{A_3(s)^2} \right] \\
 &= E \left[\frac{\sum_{i_1, j_{i_1}} \sum_{i_2, j_{i_2}} A_{5i_1 j_{i_1} i_2 j_{i_2}}(s)}{A_3(s)^2} \right] + E \left[\frac{A_2(s)^2}{A_3(s)^2} \right] = E \left[\frac{\sum_{i_1, j_{i_1}} \sum_{i_2, j_{i_2}} A_{5i_1 j_{i_1} i_2 j_{i_2}}(s)}{A_3(s)^2} \right] + O(h^4), \tag{A5}
 \end{aligned}$$

by using in the latter equality the orders established in (A3) and (A4), together with

$$\begin{aligned}
 A_{4i_1 j_{i_1} i_2 j_{i_2}}(s) &= \prod_k w_{i_k j_{i_k}}(s) A_{1i_k j_{i_k}}(s), \\
 A_{5i_1 j_{i_1} i_2 j_{i_2}}(s) &= w_{i_1 j_{i_1}}(s) w_{i_2 j_{i_2}}(s) \left(\frac{\alpha_{1i_1} \alpha_{1i_2} \sigma_1^2}{\sigma_{i_1} \sigma_{i_2}} C_{i_1 i_2}(s_{i_1 j_{i_1}} - s_{i_2 j_{i_2}}) \right. \\
 &\quad \left. - \frac{\alpha_{1i_1} \sigma_1}{\sigma_{i_1}} C_{i_1 1}(s_{i_1 j_{i_1}} - s) - \frac{\alpha_{1i_2} \sigma_1}{\sigma_{i_2}} C_{i_2 1}(s - s_{i_2 j_{i_2}}) + C_{11}(0) \right).
 \end{aligned}$$

Later, we will prove that

$$A_{5i_1 j_{i_1} i_2 j_{i_2}}(s) = O(h^{2d} h_{i_1} + h^{2d} h_{i_2} + h^{2d+2}) \quad a.s. \tag{A6}$$

Thus, by considering (A4)–(A6), as well as applying Slutsky’s theorem, we obtain that the mean squared prediction error of $\hat{Z}_1(s)$ tends to zero, since $M(s) = o(1)$.

Then, it only remains to state the validity of (A6). This relation follows straightforwardly if $L \left(\frac{C_{i_1 i_1}(0)^2 - \sigma_1^2 \sigma_{i_1}^2}{h_{i_1}} \right) = 0$ or $L \left(\frac{C_{i_2 i_2}(0)^2 - \sigma_1^2 \sigma_{i_2}^2}{h_{i_2}} \right) = 0$, which yields $A_{5i_1 j_{i_1} i_2 j_{i_2}}(s) = 0$. On the other hand, when $L \left(\frac{C_{i_k i_k}(0)^2 - \sigma_1^2 \sigma_{i_k}^2}{h_{i_k}} \right)$ differs from 0, for $k = 1, 2$, we will see that the order established in (A6) holds, due to the properties indicated below, whose demonstration will be provided later,

$$\alpha_{1i_k} C_{1i_k}(0) = \sigma_1 \sigma_{i_k} + O(h_{i_k}), \quad \text{for } k = 1, 2, \tag{A7}$$

$$C_{i_1 i_2}(t) = \frac{\alpha_{1i_2} \sigma_{i_2} C_{i_1 1}(t)}{\sigma_1} + O(h_{i_2}), \quad \text{for all } t \in \mathbb{R}^d. \tag{A8}$$

Indeed, the use of (A7) and (A8) leads us to

$$\begin{aligned}
 A_{5i_1j_{i_1}i_2j_{i_2}}(s) &= w_{i_1j_{i_1}}(s)w_{i_2j_{i_2}}(s) \left(\frac{\alpha_{1i_1}\sigma_1}{\sigma_{i_1}}C_{i_11}(s_{i_1j_{i_1}} - s_{i_2j_{i_2}}) \right. \\
 &\quad \left. - \frac{\alpha_{1i_1}\sigma_1}{\sigma_{i_1}}C_{i_11}(s_{i_1j_{i_1}} - s) - \frac{\alpha_{1i_2}\sigma_1}{\sigma_{i_2}}C_{1i_2}(s - s_{i_1j_{i_1}}) + \sigma_1^2 \right) + O(h_{i_2}) \\
 &= L \left(\frac{C_{1i_1}(0)^2 - \sigma_1^2\sigma_{i_1}^2}{h_{i_1}} \right) L \left(\frac{C_{1i_2}(0)^2 - \sigma_1^2\sigma_{i_2}^2}{h_{i_2}} \right) h^{2d} f(s)^2 \\
 &\quad \cdot \left(\frac{\alpha_{1i_1}\sigma_1}{\sigma_{i_1}}C_{i_11}(0) - \frac{\alpha_{1i_1}\sigma_1}{\sigma_{i_1}}C_{i_11}(0) - \frac{\alpha_{1i_2}\sigma_1}{\sigma_{i_2}}C_{1i_2}(0) + \sigma_1^2 \right) + O(h^{2d}h_{i_2} + h^{2d+2}) \\
 &= L \left(\frac{C_{1i_1}(0)^2 - \sigma_1^2\sigma_{i_1}^2}{h_{i_1}} \right) L \left(\frac{C_{1i_2}(0)^2 - \sigma_1^2\sigma_{i_2}^2}{h_{i_2}} \right) h^{2d} f(s)^2 \\
 &\quad \cdot \left(\frac{\alpha_{1i_1}^2\sigma_1^2\sigma_{i_1}}{\sigma_{i_1}} - \frac{\alpha_{1i_1}^2\sigma_1^2\sigma_{i_1}}{\sigma_{i_1}} - \frac{\alpha_{1i_2}^2\sigma_1^2\sigma_{i_2}}{\sigma_{i_2}} + \sigma_1^2 \right) + O(h^{2d}h_{i_1} + h^{2d}h_{i_2} + h^{2d+2}) \\
 &= O(h^{2d}h_{i_1} + h^{2d}h_{i_2} + h^{2d+2}) \quad a.s.
 \end{aligned}
 \tag{A9}$$

which yields relation (A6).

Now, to complete this proof, we check that properties (A7) and (A8) follow, when $L \left(\frac{C_{1i_k}(0)^2 - \sigma_1^2\sigma_{i_k}^2}{h_{i_k}} \right) \neq 0$, for $k = 1, 2$. With this aim, bear in mind that L is compactly supported and that the previous terms differ from zero, to obtain that $\|C_{1i_k}(0)^2 - \sigma_1^2\sigma_{i_k}^2\| \leq h_{i_k} \text{supp}(L)$, for $k = 1, 2$, with $\text{supp}(L)$ denoting the support of function L . The latter inequality directly leads to the relation established in (A7).

On the other hand, for $t \in R^d$, one has

$$\begin{aligned}
 C_{i_1i_2}(t) &= \text{Cov} [Z_{i_1}(s), Z_{i_2}(s - t)] = \text{Cov} \left[Z_{i_1}(s), Z_{i_2}(s - t) - \frac{C_{1i_2}(0)}{\sigma_1^2}Z_1(s - t) + \frac{C_{1i_2}(0)}{\sigma_1^2}Z_1(s - t) \right] \\
 &= \frac{C_{1i_2}(0)}{\sigma_1^2} \text{Cov} [Z_{i_1}(s), Z_1(s - t)] + \text{Cov} \left[Z_{i_1}(s), Z_{i_2}(s - t) - \frac{C_{1i_2}(0)}{\sigma_1^2}Z_1(s - t) \right] \\
 &= \text{(I)} + \text{(II)}.
 \end{aligned}
 \tag{A10}$$

We first deal with (I), to achieve that

$$\text{(I)} = \frac{C_{1i_2}(0)C_{i_11}(t)}{\sigma_1^2} = \frac{\alpha_{1i_2}\sigma_{i_2}C_{i_11}(t)}{\sigma_1} + O(h_{i_2}), \tag{A11}$$

where relation (A7) has been used to obtain the last term.

Secondly, the application of Hölder’s inequality on (II) yields

$$\begin{aligned}
 |\text{(II)}| &\leq (\text{Var} [Z_{i_1}(s)])^{1/2} \left(\text{Var} \left[Z_{i_2}(s - t) - \frac{C_{1i_2}(0)}{\sigma_1^2}Z_1(s - t) \right] \right)^{1/2} \\
 &= \sigma_{i_1} \left(\text{Var} [Z_{i_2}(s - t)] + \frac{C_{1i_2}(0)^2}{\sigma_1^4} \text{Var} [Z_1(s - t)] - \frac{2C_{1i_2}(0)}{\sigma_1^2} \text{Cov} [Z_{i_2}(s - t), Z_1(s - t)] \right)^{1/2} \\
 &= \sigma_{i_1} \left(\sigma_{i_2}^2 + \frac{C_{1i_2}(0)^2\sigma_1^2}{\sigma_1^4} - \frac{2C_{1i_2}(0)^2}{\sigma_1^2} \right)^{1/2} = \sigma_{i_1} \left(\sigma_{i_2}^2 - \frac{C_{1i_2}(0)^2}{\sigma_1^2} \right)^{1/2} \\
 &= \sigma_{i_1} \left(\sigma_{i_2}^2 - \frac{\sigma_1^2\sigma_{i_2}^2}{\sigma_1^2} + O(h_{i_2}) \right)^{1/2} = O(h_{i_2}),
 \end{aligned}
 \tag{A12}$$

where the penultimate term was derived from (A7). Then, combination of (A10)–(A12) allows us to conclude the proof of property (A8).

The final comment in this appendix is related to the practical implementation of the mean squared prediction error $M(s)$, for a given data set. By taking into account (A5), as well as (A2) and (A9), we propose approximating $M(s)$ through

$$\hat{M}(s) = \frac{1}{\left(\sum_{i,j_i} w_{ij_i}(s)\right)^2} \sum_{i_1,j_{i_1}} \sum_{i_2,j_{i_2}} w_{i_1j_{i_1}}(s) w_{i_2j_{i_2}}(s) \left(\frac{\alpha_{1i_1}\sigma_1}{\sigma_{i_1}} C_{1i_1}(s_{i_2j_{i_2}} - s_{i_1j_{i_1}}) - \frac{\alpha_{1i_1}\sigma_1}{\sigma_{i_1}} C_{1i_1}(s_{i_1j_{i_1}} - s) - \frac{\alpha_{1i_2}\sigma_1}{\sigma_{i_2}} C_{1i_2}(s - s_{i_2j_{i_2}}) + \sigma_1^2 \right). \quad (\text{A13})$$

References

- World Health Organization (WHO). Preventing disease through healthy environments. In *Exposure to Cadmium: A Major Public Concern*; World Health Organization: Geneva, Switzerland, 2019.
- European Food Safety Authority (EFSA). Cadmium in food - Scientific opinion of the Panel on Contaminants in the Food Chain. *EFSA J.* **2009**, *7*, 980.
- Rietra, R.P.J.; Mol, G.; Rietjens, I.M.C.M.; Römkens, P.F.A.M. *Cadmium in Soil, Crops and Resultant Dietary Exposure*; Wageningen Environmental Research: Wageningen, The Netherlands, 2017.
- Cressie, N. *Kriging for Spatial Data*; Wiley: New York, NY, USA, 1993.
- Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, USA, 1997.
- Wackernagel, H. *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed.; Springer: Berlin, Germany, 2003.
- Matheron, G. *Pour Une Analyse Krigeante des Données Regionalisées*; Centre de Géostatistique et de Morphologie Mathématique: Fontainebleau, France, 1982.
- Isaaks, E.; Srivastava, R.M. *Applied Geostatistics*; Oxford University Press: New York, NY, USA, 1989.
- Goulard, M.; Voltz, M. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Math. Geol.* **1992**, *24*, 269–282. [[CrossRef](#)]
- Yao, T. Nonparametric cross-covariance modeling as exemplified by soil heavy metal concentrations from the Swiss Jura. *Geoderma* **1999**, *88*, 13–38. [[CrossRef](#)]
- Ver Hoef, J.M.; Cressie, N.; Barry, R.P. Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier transform. *J. Comput. Graph. Stat.* **2004**, *13*, 265–282. [[CrossRef](#)]
- Gneiting, T.; Kleiber, W.; Schlather, M. Matérn cross-covariance functions for multivariate random fields. *J. Am. Stat. Assoc.* **2012**, *105*, 1167–1177. [[CrossRef](#)]
- Genton, M.G.; Kleiber, W. Cross-covariance functions for multivariate geostatistics. *Stat. Sci.* **2015**, *30*, 147–163. [[CrossRef](#)]
- Rojas-Avellaneda, D.; Martínez-Cervantes, J. Using the bivariate approach to spatial estimation of air pollution by ozone. *Procedia Environ. Sci.* **2011**, *3*, 20–25. [[CrossRef](#)]
- Halimi, M.; Farajzadeh, M.; Zarei, Z. Modeling spatial distribution of Tehran air pollutants using geostatistical methods incorporate uncertainty maps. *Pollution* **2011**, *2*, 375–386.
- Hooshmand, A.; Delghandi, M.; Izadi, A.; Aali, K.A. Application of kriging and cokriging in spatial estimation of groundwater quality parameters. *Afr. J. Agric. Res.* **2011**, *6*, 3402–3408.
- Gunduz, O.; Elci, A.; Simsek, C.; Baba, A. The use of cokriging algorithm for arsenic mapping in groundwater systems. In Proceedings of the Conference Paper, 5th International Perspective on Water Resources & the Environment, Marrakech, Morocco, 5–7 January 2012.
- Teixeira, D.B.; Bicalho, E.S.; Panosso, A.R.; Cerri, C.E.P.; Pereira, G.T.; La Scala, N. Spatial variability of soil CO₂ emission in a sugarcane area characterized by secondary information. *Sci. Agric.* **2013**, *70*, 195–203. [[CrossRef](#)]
- Fu, X.; Liu, X.; Li, Y.; Shen, J.; Wang, Y.; Zou, G.; Li, H.; Song, L.; Wu, J. Wet-season spatial variability in N₂O emissions from a tea field in subtropical central China. *Biogeosciences* **2015**, *12*, 3899–3911. [[CrossRef](#)]
- Bohorquez, M.; Giraldo, R.; Mateu, J. Multivariate functional random fields: Prediction and optimal sampling. *Stoch. Environ. Res. Risk Assess.* **2017**, *31*, 53–70. [[CrossRef](#)]

21. Giraldo, R.; Herrera, L.; Leiva, V. Cokriging prediction using as secondary variable a functional random field with application in environmental pollution. *Mathematics* **2020**, *8*, 1305. [[CrossRef](#)]
22. Solow, A.R. Geostatistical cross-validation: A cautionary note. *Math. Geol.* **1990**, *22*, 637–639. [[CrossRef](#)]
23. Yalcin, E. Cokriging and its effect on the estimation precision. *J. S. Afr. Ins. Min. Metall.* **2005**, *105*, 223–228.
24. Cellmer, R. The possibilities and limitations of geostatistical methods in real estate market analyses. *Real Estate Manag. Valuat.* **2014**, *22*, 54–62. [[CrossRef](#)]
25. Menezes, R.; García-Soidán, P.; Ferreira, C. Nonparametric spatial prediction under stochastic sampling design. *J. Nonparametr Stat.* **2010**, *22*, 363–377. [[CrossRef](#)]
26. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B* **1974**, *36*, 111–147. [[CrossRef](#)]
27. Terrell, G.; Scott, D.W. Variable kernel density estimation. *Ann. Stat.* **1992**, *20*, 1236–1265. [[CrossRef](#)]
28. Francisco-Fernández, M.; Quintela, A.; Fernández-Casal, R. Nonparametric methods for spatial regression. An application to seismic events. *Environmetrics* **2012**, *23*, 85–93. [[CrossRef](#)]
29. Kleiber, W.; Nychka, D. Nonstationary modeling for multivariate spatial processes. *J. Multivar. Anal.* **2012**, *112*, 76–91. [[CrossRef](#)]
30. Pebesma, E. gstat: Spatial and Spatio-Temporal Geostatistical Modeling, Prediction and Simulation. R Package Version 2.0-6. 2020. Available online: <https://CRAN.R-project.org/package=gstat> (accessed on 23 September 2020).
31. Schlather, M.; Malinowski, A.; Oesting, M.; Boecker, D.; Storkorb, K.; Engelke, S.; Martini, J.; Ballani, F.; Moreva, O. RandomFields: Simulation and Analysis of Random Fields. R Package Version 3.3.8. 2020. Available online: <https://CRAN.R-project.org/package=RandomFields> (accessed on 23 September 2020).
32. Fernández-Casal, R. npsp: Nonparametric Spatial Statistics. R Package Version 0.7-5. 2020. Available online: <https://CRAN.R-project.org/package=npsp> (accessed on 23 September 2020).
33. Ojeda-Cabrera, J.L. locpol: Kernel Local Polynomial Regression. R Package Version 0.7-0. 2018. Available online: <https://CRAN.R-project.org/package=locpol> (accessed on 23 September 2020).
34. Bivand, R.S.; Pebesma, E.; Gómez-Rubio, V. *Applied Spatial Data Analysis with R*; Springer: New York, NY, USA, 2008.
35. Loecher, M. RgoogleMaps: Overlays on Static Maps. R Package Version 1.4.5.3. 2020. Available online: <https://CRAN.R-project.org/package=RgoogleMaps> (accessed on 23 September 2020).
36. Rossiter, D.G. *Co-Kriging with the Gstat Package of the R Environment for Statistical Computing*; Technical Note; Cornell University, Section of Soil & Crop Sciences: Ithaca, NY, USA, 2018.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).