

Bootstrap approaches for spatial data

Pilar García-Soidán · Raquel Menezes ·
Óscar Rubiños

Published online: 11 October 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Generation of replicates of the available data enables the researchers to solve different statistical problems, such as the estimation of standard errors, the inference of parameters or even the approximation of distribution functions. With this aim, Bootstrap approaches are suggested in the current work, specifically designed for their application to spatial data, as they take into account the dependence structure of the underlying random process. The key idea is to construct nonparametric distribution estimators, adapted to the spatial setting, which are distribution functions themselves, associated to discrete or continuous random variables. Then, the Bootstrap samples are obtained by drawing at random from the estimated distribution. Consistency of the suggested approaches will be proved by assuming stationarity from the random process or by relaxing the latter hypothesis to admit a deterministic trend. Numerical studies for simulated data and a real data set, obtained from environmental monitoring, are included to illustrate the application of the proposed Bootstrap methods.

Keywords Distribution estimation · Resampling method · Spatial data · Stationarity · Trend

Mathematics Subject Classification 62G09 · 62M30

1 Introduction

Different statistical techniques can require extra work to approximate the unknown terms involved, to check the variance of the estimators employed or to derive their sampling distribution. Some of these problems have been solved in the curve estimation setting through the Bootstrap method (Efron 1979), whose key idea is that the relationship between the theoretical distribution and the observed sample is similar to the relationship between a distribution estimator and a secondary sample drawn from it. The original Bootstrap procedures were designed for independent data (De Angelis and Young 1992; Hall 1992) and an adaptation of these procedures to the spatial setting would have large applicability for addressing a variety of issues. A goodness-of-fit test for the variogram model could be implemented to extend the results in Maglione and Diblasi (2004) for gaussian data, and even inference on the underlying distribution of the random process could be developed. Bootstrap variants of the tests based on their asymptotic distribution, such as the one proposed in Li et al. (2007) for assessment of the properties of the covariance function, could be introduced to improve their speed of convergence. Also, the kernel approaches for characterization of the dependence structure (Hall and Patil 1994) or prediction (Menezes et al. 2010) could be accomplished by deriving the optimal bandwidths and additionally providing estimates of their accuracy.

P. García-Soidán (✉)
Department of Statistics and Operations Research, University of
Vigo, Vigo, Spain
e-mail: pgarcia@uvigo.es

R. Menezes
Department of Mathematics and Applications, University of
Minho, Braga, Portugal
e-mail: rmenezes@math.uminho.pt

Ó. Rubiños
Department of Signal Theory and Communications, University
of Vigo, Vigo, Spain
e-mail: oscar@com.uvigo.es

The parametric Bootstrap methodology can be easily extended to the spatial setting, thus enabling the researchers to develop techniques for variogram assessment (Olea and Pardo-Igúzquiza 2011), to construct confidence intervals for the parameters of a distribution estimator (Goovaerts et al. 2005) or to correct a test for the spectral density (Crujeiras et al. 2010). When independence can be assumed from the residuals, the traditional nonparametric resampling approaches can be applied for approximation of the variance of an estimator (Iranpanah et al. 2011) or probabilities derived from it (Hyun-Han and Young-Il 2006). However, for dependent data, it is necessary to design ad-hoc procedures in order to guarantee consistency of the results, such as the parametric Bootstrap method for small area estimation (Hall and Maiti 2006). Alternatives of more general use are those based on resampling blocks of data (Hall 1985), subsamples obtained by deleting portions of data (Politis et al. 1999) or marks assigned to the spatial points (Loh 2008).

The aim of this work is to introduce nonparametric Bootstrap approaches that allow us to generate replicates from the available data, at a set of locations selected, by first approximating the joint distribution in a nonparametric way and then randomly drawing samples from it. With this idea, different estimators of the multivariate distribution will be proposed, which are distribution functions themselves, associated to discrete or continuous random variables, so that they will be used as the basis for resampling. We will check that consistency follows for the suggested procedures, provided that the random process is strictly stationary or when this condition is relaxed by admitting a deterministic trend. In addition, numerical studies will be carried out to analyze the behavior of both Bootstrap methods for addressing different problems.

2 Multivariate distribution estimators

To derive the distribution estimators, we will assume that the random process $\{Z(s) \in \mathbb{R} : s \in D \subset \mathbb{R}^d\}$ can be modeled as:

$$Z(s) = \mu(s) + Y(s) \tag{1}$$

where $\{Y(s) \in \mathbb{R} : s \in D \subset \mathbb{R}^d\}$ is a zero-mean strictly stationary random process and $\mu(\cdot)$ represents the deterministic trend, namely, $E[Z(s)] = \mu(s)$, for all $s \in D$.

We will denote the multivariate distribution by:

$$F_{s_1, \dots, s_k}(x_1, \dots, x_k) = P(Z(s_1) \leq x_1, \dots, Z(s_k) \leq x_k)$$

for all sets of sites $s_1, \dots, s_k \in D$ and thresholds $x_1, \dots, x_k \in \mathbb{R}$, with $k \in \mathbb{N}$.

From the model established for the random process in (1), one has:

$$F_{s_1, \dots, s_k}(x_1, \dots, x_k) = F_{s_1+z, \dots, s_k+z}(x_1 + \mu(s_1 + z) - \mu(s_1), \dots, x_k + \mu(s_k + z) - \mu(s_k)) \tag{2}$$

for all $z \in \mathbb{R}^d$, on account of the stationarity condition of $Y(\cdot)$, since:

$$P(Y(s_1) \leq x_1, \dots, Y(s_k) \leq x_k) = P(Y(s_1 + z) \leq x_1, \dots, Y(s_k + z) \leq x_k) \tag{3}$$

Relation (2) yields that the multivariate distribution of $Z(\cdot)$ remains invariant when the spatial locations are subjected to the same translation, by vector z , and each threshold x_j is replaced by an appropriate correction, given by $x_j + \mu(s_j + z) - \mu(s_j)$.

Our aim is the estimation of $F_{s_1, \dots, s_k}(x_1, \dots, x_k)$, for a set of selected sites $s_1, \dots, s_k \in D$ and thresholds $x_1, \dots, x_k \in \mathbb{R}$, with $k \in \mathbb{N}$. This issue will be addressed by applying property (2), which yields:

$$F_{s_1, \dots, s_k}(x_1, \dots, x_k) = P(Z(s_1 + z) \leq x_1 + \mu(s_1 + z) - \mu(s_1), \dots, Z(s_k + z) \leq x_k + \mu(s_k + z) - \mu(s_k)) = P(Z(t_1) \leq x_1 + \mu(t_1) - \mu(s_1), \dots, Z(t_k) \leq x_k + \mu(t_k) - \mu(s_k)) \tag{2}$$

for $t_j = s_j + z$ and $z \in \mathbb{R}^d$.

The relations above allow us to conclude that:

$$F_{s_1, \dots, s_k}(x_1, \dots, x_k) = P(X(t_1) \leq x_1, \dots, X(t_k) \leq x_k) \tag{3}$$

for $t_j = s_j + z$ and $X(t_j) = Z(t_j) - \mu(t_j) + \mu(s_j)$. Consequently, if the set $\{t_1, \dots, t_k\}$ represents a translation of the selected locations $\{s_1, \dots, s_k\}$ by any vector $z \in \mathbb{R}^d$, the distribution of the random vector $(Z(s_1), \dots, Z(s_k))$ equals that of $(X(t_1), \dots, X(t_k))$.

Next, several approaches will be described for approximation of the distribution function, based on (3). With this aim, suppose that n data, $\{Z(t_1), \dots, Z(t_n)\}$, have been collected, at the respective locations $\{t_1, \dots, t_n\}$. Firstly, we propose constructing a weighted average of the indicator functions obtained for the possible k -combinations $\{t_{i_1}, \dots, t_{i_k}\}$ of the observed sites $\{t_1, \dots, t_n\}$, as follows:

$$\hat{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) = \sum_{i_1=1}^n \dots \sum_{i_k=1}^n p_{i_1, \dots, i_k} I_{\{X(t_{i_1}) \leq x_1\}} \dots I_{\{X(t_{i_k}) \leq x_k\}} \tag{4}$$

for some weights $p_{i_1, \dots, i_k} \geq 0$, with $\sum_{i_1=1}^n \dots \sum_{i_k=1}^n p_{i_1, \dots, i_k} = 1$, which will be established in (6) and (7), where I_A denotes the indicator function of the set A , for $t_{i_j} = s_j + z$ and $z \in \mathbb{R}^d$. In particular, since the latter equality holds for $j = 1$, we will take $z = t_{i_1} - s_1$, yielding:

$$X(t_{i_j}) = Z(t_{i_j}) - \mu(s_j + t_{i_1} - s_1) + \mu(s_j) \tag{5}$$

For selection of the values p_{i_1, \dots, i_k} , the key idea will be to assign more weight to the k -combination $\{t_{i_1}, \dots, t_{i_k}\}$ which

is closer to being a translation of $\{s_1, \dots, s_k\}$ by $z = t_{i_1} - s_1$. At first sight, this approach involves computing the lag between each pair of the selected locations, $s_j - s_{j'}$, and comparing it with that of the corresponding observed sites, $t_j - t_{j'}$, for j and j' varying from 1 to k , which amounts to $0.5k(k - 1)$ comparisons. For the sake of simplicity, we propose solely computing the indispensable lags needed to characterize the closeness of $\{t_{i_1}, \dots, t_{i_k}\}$ to being the aforementioned translation. This enables us to consider only the lags $s_j - s_{j+1}$ and $t_j - t_{j+1}$, for all $j \geq 1$, so the other ones would follow with simple sums or differences of vectors. Then, a nonparametric approximation of the multivariate distribution would be obtained by taking p_{i_1, \dots, i_k} in (4) as given below:

$$p_{i_1, \dots, i_k} = P_{i_1, \dots, i_k}^{(1)} = \frac{p_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1} \cdots p_{s_{k-1}, s_k}^{t_{i_{k-1}}, t_{i_k}, h_{k-1}}}{\sum_{i_1=1}^n \cdots \sum_{i_k=1}^n p_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1} \cdots p_{s_{k-1}, s_k}^{t_{i_{k-1}}, t_{i_k}, h_{k-1}}} \quad (6)$$

with $p_{s_j, s_{j+1}}^{t_j, t_{j+1}, h_j} = K\left(\frac{s_j - s_{j+1} - (t_j - t_{j+1})}{h_j}\right)$, where K represents a d -variate kernel function and h_j is a bandwidth parameter, for $j = 1, \dots, k - 1$. The resulting estimator will be referred to as $\hat{F}_{s_1, \dots, s_k}^{(1)}$.

One drawback related to estimator $\hat{F}_{s_1, \dots, s_k}^{(1)}$ is the combinatorial explosion that may occur for large k , when used for construction of Bootstrap replicates. This is mainly due to the fact that $\hat{F}_{s_1, \dots, s_k}^{(1)}$ is itself a discrete distribution function, conditional on the sample $\{Z(t_1), \dots, Z(t_n)\}$, which takes as many values $(X(t_{i_1}), \dots, X(t_{i_k}))$ (or, more precisely, vectors of size k) as combinations $\{t_{i_1}, \dots, t_{i_k}\}$ can be obtained from the set of the observed locations. Indeed, the number of combinations of this kind amount to n^k , so the resampling approach derived from this distribution estimator would entail drawing a vector of size k from a set of n^k vectors, whose probabilities $p_{i_1, \dots, i_k}^{(1)}$ require computing the terms $p_{s_j, s_{j+1}}^{t_j, t_{j+1}, h_j}$, for all j varying from 1 to $k - 1$ and all the combinations of size k from $\{t_1, \dots, t_n\}$.

In view of the latter, our suggestion will be to construct a new distribution estimator as given in (4), with weights:

$$p_{i_1, \dots, i_k} = P_{i_1, \dots, i_k}^{(2)} = \frac{p_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}}{\sum_{i_1=1}^n \sum_{i_2=1}^n p_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}} \frac{p_{s_2, s_3}^{t_{i_2}, t_{i_3}, h_2}}{\sum_{i_2=1}^n \sum_{i_3=1}^n p_{s_2, s_3}^{t_{i_2}, t_{i_3}, h_2}} \cdots \frac{p_{s_{k-1}, s_k}^{t_{i_{k-1}}, t_{i_k}, h_{k-1}}}{\sum_{i_{k-1}=1}^n \sum_{i_k=1}^n p_{s_{k-1}, s_k}^{t_{i_{k-1}}, t_{i_k}, h_{k-1}}} \quad (7)$$

The resulting estimator, denoted by $\hat{F}_{s_1, \dots, s_k}^{(2)}$, will allow us to derive a simpler Bootstrap approach, as we will describe in Sect. 3, for which valid probability statements will be made.

The two aforementioned approaches provide identical estimators of the bivariate distribution $F_{s, s'}$ and the univariate distribution F_s , as follows:

$$\hat{F}_{s, s'}(x_1, x_2) = \frac{\sum_{i=1}^n \sum_{j=1}^n \frac{K\left(\frac{s-s'-(t_i-t_j)}{h_1}\right) I_{\{X(t_i) \leq x_1\}} I_{\{X(t_j) \leq x_2\}}}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{s-s'-(t_i-t_j)}{h_1}\right)}}{\sum_{i=1}^n \sum_{j=1}^n \frac{K\left(\frac{t_j-t_i}{h_1}\right) I_{\{X(t_i) \leq x\}} I_{\{X(t_j) \leq x\}}}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{t_j-t_i}{h_1}\right)}} \quad (8)$$

The consistency of $\hat{F}_{s_1, \dots, s_k}^{(2)}$ can be established by considering a random design for the spatial locations and a mixed increasing-domain asymptotic structure for the random process, together with the appropriate convergence rates for the bandwidth parameters and the increasing scale. A sketch of the proof of this property is outlined in Appendix 1, which gives account of the dependence of the optimal bandwidths h_j on unknown terms. Then, we propose an alternative mechanism for their selection, based on computing h_j as the Euclidean distance from $s_j - s_{j+1}$ to the m -nearest difference $t_j - t_{j+1}$, for some m .

It is noteworthy that estimator $\hat{F}_{s_1, \dots, s_k}^{(2)}$ is itself a distribution function, conditional on the sample $\{Z(t_1), \dots, Z(t_n)\}$, which takes values $(X(t_{i_1}), \dots, X(t_{i_k}))$ with respective probabilities $p_{i_1, \dots, i_k}^{(2)}$, for $X(t_{i_j})$ defined in (5). However, $\hat{F}_{s_1, \dots, s_k}^{(2)}$ is a discrete distribution function. Hence, for a continuous random process, the use of a smoother version of the distribution estimator seems to be more appropriate, which can be derived by applying in (4) an integrand of a density, instead of an indicator function, as follows:

$$\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) = \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n p_{i_1, \dots, i_k}^{(2)} \mathcal{L}\left(\frac{x_1 - X(t_{i_1})}{h}\right) \cdots \mathcal{L}\left(\frac{x_k - X(t_{i_k})}{h}\right) \quad (9)$$

where $\mathcal{L}(x) = \int_{-\infty}^x L(u) du$, L is a univariate kernel function, h is a bandwidth parameter and $p_{i_1, \dots, i_k}^{(2)}$ is defined in (7). Consistency can also be derived for $\tilde{F}_{s_1, \dots, s_k}^{h_1, \dots, h_{k-1}}$, as outlined in Appendix 2.

Unlike $\hat{F}_{s_1, \dots, s_k}^{(i)}$, for $i = 1, 2$, estimator $\tilde{F}_{s_1, \dots, s_k}$ is a continuous distribution function, conditional on the sample, with density:

$$\tilde{f}_{s_1, \dots, s_k}(x_1, \dots, x_k) = \frac{1}{h^k} \sum_{i_1=1}^n \cdots \sum_{i_k=1}^n p_{i_1, \dots, i_k}^{(2)} L\left(\frac{x_1 - X(t_{i_1})}{h}\right) \cdots L\left(\frac{x_k - X(t_{i_k})}{h}\right)$$

The optimal h is dependent on the bandwidths h_j as well as on unknown moments from the random process. Hence, we suggest using a cross-validation procedure for its selection in practice, based on the results given in Bowman

et al. (1998) and adapted to the spatial data setting, which aims at providing an appropriate characterization of the performance of the multivariate distribution in the manner described below:

$$h_{CV,k} = \operatorname{argmin}_{h \in H} \sum_{i_1, \dots, i_k=1}^n \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \left(\tilde{F}_{-(i_1, \dots, i_k)}(x_1, \dots, x_k) - I_{\{Z(t_{i_1}) \leq x_1\}} \dots I_{\{Z(t_{i_k}) \leq x_k\}} \right)^2 \times dx_1 \dots dx_k$$

where H is an adequate set of positive numbers and $\tilde{F}_{-(i_1, \dots, i_k)}(x_1, \dots, x_k)$ is the result of implementing estimator $\tilde{F}_{t_{i_1}, \dots, t_{i_k}}$ at (x_1, \dots, x_k) when ignoring $\{Z(t_{i_1}), \dots, Z(t_{i_k})\}$. Proceeding in this way, $h_{CV, k}$ would provide us with a global bandwidth selector that could be applied at any (x_1, \dots, x_k) . For simplification of the approach to derive the cross-validation bandwidth, our alternative proposal will be based on the use of the univariate continuous estimator, by considering:

$$\tilde{F}_s(x) = \sum_{i=1}^n \sum_{j=1}^n \frac{K\left(\frac{y_j - t_i}{h_1}\right) \mathcal{L}\left(\frac{x - X(t_i)}{h}\right) \mathcal{L}\left(\frac{x - X(y_j)}{h}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{y_j - t_i}{h_1}\right)} \quad (10)$$

since the objective function should involve just one threshold and two bandwidths, to obtain:

$$h_{CV,1} = \operatorname{argmin}_{h \in H} \sum_{i=1}^n \int_{-\infty}^{+\infty} \left(\tilde{F}_{-(i)}(x) - I_{\{Z(t_i) \leq x\}} \right)^2 dx$$

where $Z(t_i)$ is left out in the implementation of $\tilde{F}_{t_i}(x)$ to produce $\tilde{F}_{-(i)}(x)$. Furthermore, the integral can be numerically approximated over a bounded subset instead of \mathbb{R} .

Remark 1 Application of any of the previous distribution approaches requires assuming that the trend function can be estimated. In this respect, different procedures have been proposed for approximation of $\mu(\cdot)$ (Goovaerts 1997), so that a parametric approach can be adopted for its estimation or the spatial interpolation techniques can be used to compute the trend.

Remark 2 When the trend function is supposed to be constant, relation (2) can be simplified to yield:

$$F_{s_1, \dots, s_k}(x_1, \dots, x_k) = F_{s_1+z, \dots, s_k+z}(x_1, \dots, x_k)$$

for all $z \in \mathbb{R}^d$. Then, no characterization of the constant trend is necessary for implementation of the distribution estimators, as we could take $X(t_{i_j}) = Z(t_{i_j})$ in $\tilde{F}_{s_1, \dots, s_k}$ and $\hat{F}_{s_1, \dots, s_k}^{(i)}$, for $i = 1, 2$.

Remark 3 For the specific case where the difference of trends depends on the lag between the locations involved, namely:

$$\mu(s) - \mu(s') = M(s - s') \quad (11)$$

for all $s, s' \in D$ and some function M , combination of (2) and (11) leads to:

$$F_{s_1, \dots, s_k}(x_1, \dots, x_k) = F_{s_1+z, \dots, s_k+z}(x_1 + M(z), \dots, x_k + M(z))$$

The equality above means that the translation of the spatial locations by vector z , together with the correction of the thresholds by $M(z)$, preserves the value of the distribution function. Then, the distribution approaches would hold for $X(t_{i_j}) = Z(t_{i_j}) - M(t_{i_j}) - s_1$, which demands approximation of function M . This issue can be addressed by adapting the different procedures that can be used for μ or even a nonparametric estimator can be derived as follows:

$$\hat{M}(z) = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j}(z) (Z(t_i) - Z(t_j))}{\sum_{i=1}^n \sum_{j=1}^n w_{i,j}(z)}$$

for some nonnegative values $w_{i,j}(z)$ satisfying that $\sum_{i=1}^n \sum_{j=1}^n w_{i,j}(z) > 0$. For instance, we can take $w_{i,j}(z) =$

$$I_{\{t_i - t_j \approx z\}} \text{ or } w_{i,j}(z) = G\left(\frac{z - (t_i - t_j)}{g}\right),$$

to yield an empirical or a kernel estimator, respectively, where G is a d -variate kernel function and g is a bandwidth parameter. By assuming appropriate hypotheses, consistency of \hat{M} could be proved by using similar arguments as those applied in the kernel variogram estimation (García-Soidán 2007).

Remark 4 None of the proposed distribution estimators fulfills Kolmogorov’s condition of symmetry, namely, that they remain invariant when the locations s_j and the corresponding thresholds x_j are subjected to the same permutation, for $j = 1, \dots, k$. Hence, a criterion must be established to assign an order to the spatial locations and, therefore, to the thresholds, previously to the implementation of the joint distribution estimator, so as to guarantee unicity of the result under permutation.

Among the different options, we propose proceeding in such a way that the sites will be organized in a decreasing order of their influence on the remainder, measured in terms of proximity, because of the underlying stationarity condition. With this idea, departing from the set $\{s_1, \dots, s_k\}$, we will take s_j , for j varying from 1 to k , as the closest location to the center of mass (or the d -dimensional mean of the coordinates) of the sites $\{s_j, s_{j+1}, \dots, s_k\}$, for $j = 1, \dots, k - 1$. To solve the problem of tied distances, preference can be given to the location with the smallest first coordinates. The thresholds x_j would also be reordered accordingly.

3 Bootstrap approaches

The distribution estimation approaches, introduced in Sect. 2, can be used to propose Bootstrap methods for spatial data, so that for a given set of selected sites, $\{s_1, \dots, s_k\}$, a Bootstrap sample $\{Z^*(s_1), \dots, Z^*(s_k)\}$ can be obtained.

The direct mechanism derived from the discrete estimator $\hat{F}_{s_1, \dots, s_k}^{(1)}$ entails producing each replicate by drawing from a random variable which takes values $(X(t_{i_1}), \dots, X(t_{i_k}))$ with probabilities $p_{i_1, \dots, i_k}^{(1)}$, for each $i_j = 1, \dots, n$ and $j = 1, \dots, k$, with $X(t_{i_j})$ as given in (5). Nevertheless, implementation of this approach, as mentioned in Sect. 2, can have a strong computational cost for large k .

Then, for construction of the replicates, we can take instead estimator $\hat{F}_{s_1, \dots, s_k}^{(2)}$, although there are again n^k probabilities $p_{i_1, \dots, i_k}^{(2)}$ to be considered in order to obtain a Bootstrap sample. However, we suggest an alternative option based on proceeding in a sequential way, which would be less computationally demanding. The resampling scheme would be performed in the manner described below:

- (1) Reorder the locations to take s_j as the closest location to the center of mass of $\{s_j, s_{j+1}, \dots, s_k\}$, for $j = 1, \dots, k - 1$.
- (2) Select the bandwidth h_1 as the Euclidean distance from $s_1 - s_2$ to the m -nearest difference $t_{i_1} - t_{i_2}$, for some m .
- (3) Obtain $(Z^*(s_1), Z^*(s_2))$ by drawing from a random variable that associates the probabilities:

$$\frac{P_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}}{\sum_{i_1=1}^n \sum_{i_2=1}^n P_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}}$$

to the respective pairs $(X(t_{i_1}), X(t_{i_2}))$.

Proceeding in this way, a couple of values $(X(t_{i_1}), X(t_{i_2}))$. (and, therefore, two indices i_1 and i_2) are selected in this step.

- (4) For $j = 3$, consider the index i_{j-1} derived previously and take h_{j-1} to be the Euclidean distance from $s_{j-1} - s_j$ to the m -nearest difference $t_{i_{j-1}} - t_{i_j}$.
- (5) For $j = 3$, consider indexes i_1 and i_{j-1} to obtain $Z^*(s_j)$ by resampling from the random variable which takes values $(X(t_{i_j}))$, with respective probabilities:

$$\frac{P_{s_{j-1}, s_j}^{t_{i_{j-1}}, t_{i_j}, h_{j-1}}}{\sum_{i_j=1}^n P_{s_{j-1}, s_j}^{t_{i_{j-1}}, t_{i_j}, h_{j-1}}}$$

An index i_j is chosen in this step.

- (6) Repeat steps 4 and 5 for all $j > 3$.

Validity of the preceding procedure follows straightforwardly from the fact that the resulting sample satisfies:

$$\begin{aligned} & \mathcal{P}^*(Z^*(s_1) = X(t_{i_1}), \dots, Z^*(s_k) = X(t_{i_k})) \\ &= \mathcal{P}^*(Z^*(s_1) = X(t_{i_1}), Z^*(s_2) = X(t_{i_2})) \\ & \cdot \prod_{j=3}^k \mathcal{P}^*(Z^*(s_j) = X(t_{i_j}) / Z^*(s_j) = X(t_{i_j}), j' < j) \\ &= \frac{P_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}}{\sum_{i_1=1}^n \sum_{i_2=1}^n P_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}} \prod_{j=3}^k \frac{P_{s_{j-1}, s_j}^{t_{i_{j-1}}, t_{i_j}, h_{j-1}}}{\sum_{i_{j-1}=1}^n \sum_{i_j=1}^n P_{s_{j-1}, s_j}^{t_{i_{j-1}}, t_{i_j}, h_{j-1}}} = p_{i_1, \dots, i_k}^{(2)} \end{aligned}$$

on account of the multiplication rule of probability, where \mathcal{P}^* denotes the probability, conditional on the sample $\{Z(t_1), \dots, Z(t_n)\}$. Then, proceeding as just indicated, a Bootstrap sample from $\hat{F}_{s_1, \dots, s_k}^{(2)}$ can be generated.

On the other hand, if the aim is that of drawing replicates from the continuous distribution estimator $\tilde{F}_{s_1, \dots, s_k}$, we should additionally extract a random sample of size k from the density L , denoted by $\{V_1, \dots, V_k\}$ and independent of $\{Z(t_1), \dots, Z(t_n)\}$. Then, the continuous version of the replicates for the random process Z , at locations $\{s_1, \dots, s_k\}$, would be constructed as:

$$\{X(t_{i_1}) + hV_1, \dots, X(t_{i_k}) + hV_k\}$$

To justify the generation of Bootstrap samples as described, for the continuous estimator, bear in mind that:

$$\begin{aligned} & \mathcal{P}^*(X(t_{i_1}) + hV_1 \leq x_1, \dots, X(t_{i_k}) + hV_k \leq x_k) \\ &= \mathcal{P}^*\left(V_1 \leq \frac{x_1 - X(t_{i_1})}{h}, \dots, V_k \leq \frac{x_k - X(t_{i_k})}{h}\right) \\ &= E^*\left[\mathcal{L}\left(\frac{x_1 - X(t_{i_1})}{h}\right) \dots \mathcal{L}\left(\frac{x_k - X(t_{i_k})}{h}\right)\right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathcal{P}^*(Z^*(s_1) = X(t_{i_1}), \dots, Z^*(s_k) = X(t_{i_k})) \\ & \cdot \mathcal{L}\left(\frac{x_1 - X(t_{i_1})}{h}\right) \dots \mathcal{L}\left(\frac{x_k - X(t_{i_k})}{h}\right) \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n p_{i_1, \dots, i_k}^{(2)} \mathcal{L}\left(\frac{x_1 - X(t_{i_1})}{h}\right) \dots \mathcal{L}\left(\frac{x_k - X(t_{i_k})}{h}\right) \\ &= \tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) \end{aligned}$$

by the conditions required from the variables V_j , where E^* denotes the expectation, conditional on the sample $\{Z(t_1), \dots, Z(t_n)\}$.

The Bootstrap approaches can be used to approximate unknown parameters, estimate standard errors, make inference on the correlation structure or on the distribution function of the random process. Suppose, for instance, that $T = T(Z(s_1), \dots, Z(s_k))$ is an estimator of interest, dependent on the data and on the underlying distribution F_{s_1, \dots, s_k} . Denote by T^* its Bootstrap counterpart, namely, $T^* = T(Z^*(s_1), \dots, Z^*(s_k))$, for a Bootstrap sample $\{Z^*(s_1), \dots, Z^*(s_k)\}$ obtained by either of the resampling methods proposed. Then, the unknown characteristic of T , depending on F_{s_1, \dots, s_k} , can be approximated by that of T^* , under the distribution estimator selected. For the latter aim in practice, we

can compute the corresponding sample characteristic of B values $T^{*(b)} = T(Z^{*(b)}(s_1), \dots, Z^{*(b)}(s_k))$, derived for B replicates $\{Z^{*(b)}(s_1), \dots, Z^{*(b)}(s_k)\}$, for $b = 1, \dots, B$ and large B .

4 Application examples

We now describe some examples of the practical usefulness of the methodology proposed in this manuscript to generate Bootstrap replicates from $\hat{F}_{s_1, \dots, s_k}^{(2)}$ (or its continuous counterpart $\tilde{F}_{s_1, \dots, s_k}$), which does not require the estimation of the corresponding multivariate distribution. Firstly, these methods are applied to simulated data and, then, an example with a real data set of air quality indicators is presented.

4.1 Numerical studies with simulated data

In order to analyze the performance of the Bootstrap approaches suggested in Sect. 3, we carried out several numerical studies with simulated data on the unit square $D = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. A complete spatial randomness design was assumed, so the sample locations were uniformly distributed on D . With the spatial locations t_i , obtained for $i = 1, \dots, n$ and $n = 50$, stationary gaussian data $Z(t_i)$ were generated, by assuming zero mean and by selecting a valid model for the variogram to specify the spatial dependency, as follows:

$$Z(s) = \mu(s) + Y(s), \quad \text{with } \mu(s) = 0 \\ \text{and } Y(s) \sim \text{SGP}(0, \sigma^2, \rho(\cdot; 0.2))$$

In particular, we considered the isotropic exponential and spherical variograms, with a partial sill σ^2 equaling 1 or 2.25 (or asymptotic partial sill, for the exponential model), a range $\phi = 0.2$ (or practical range, for the exponential model) and a null nugget effect or a nugget effect $\tau^2 = 0.09$.

For the implementation of the resampling algorithm given in the preceding section, we selected $k = 15$ sites, $\{s_1, \dots, s_k\}$, among the set of the sample locations $\{t_1, \dots, t_n\}$, when avoiding those points too close to the boundaries of the observation region. To generate a Bootstrap sample on these k locations, one needs to derive weights $p_{i_1, \dots, i_k}^{(2)}$, as explained in steps 3 and 5 of the algorithm described in Sect. 3. With this purpose, we took K as the Epanechnikov kernel and the bandwidths h_1, \dots, h_{k-1} , based on a balloon estimator, were computed by considering the m -nearest differences in the kernel function and by guaranteeing that 15 %, for h_1 , and 30 %, otherwise, of all distances were used. Given the probabilities $p_{i_1, \dots, i_k}^{(2)}$, the

indices i_j were then chosen by a classic accept–reject method. In this respect, note that the stochastic processes $X(\cdot)$ and $Z(\cdot)$ are the same when the trend function is constant, as pointed out in Remark 2.

The smoother Bootstrap version was acquired by applying the continuous distribution estimator $\tilde{F}_{s_1, \dots, s_k}$ in (9), where function \mathcal{L} was chosen as the standard normal distribution. The corresponding optimal bandwidth was elected by cross-validation, as explained in Sect. 2, among a reasonable set of bandwidth candidates. Then, Bootstrap replicates were generated for the simulated data, under the aforementioned conditions, aiming to analyze the performance of the proposed approaches for the following issues:

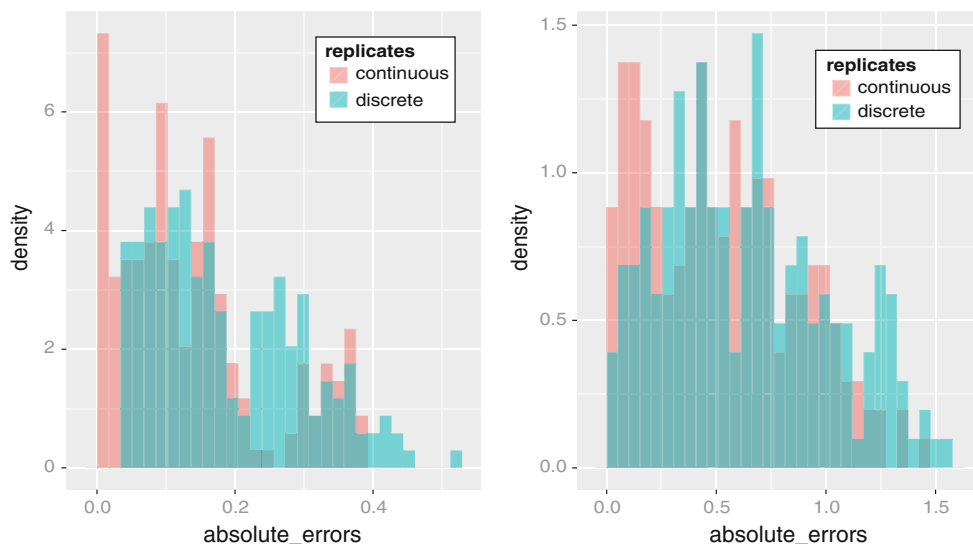
- (1) The estimation of the variance of the spatial process, as a common parameter for the overall process, $\text{Var}[Z(s)] = \text{E}[Z(s)^2] - \text{E}[Z(s)]^2$.
- (2) The comparison of the discrete and continuous estimators of the univariate distribution, denoted by \hat{F}_s and \tilde{F}_s , as given in (8) and (10), respectively.
- (3) The approximation of the variogram, as a function modeling the spatial dependence, $\gamma(t) = 0.5\text{Var}[Z(s) - Z(s+t)] = 0.5 \cdot \text{E}[(Z(s) - Z(s+t))^2]$.

The first numerical study was designed to compare the discrete and continuous resampling methods for approximation of the variance of the spatial process $Z(s)$. Proceeding as described above, the theoretical expectations involved were approximated through a sample average obtained from $B = 500$ Bootstrap replicates and 200 samples. The data were simulated under the exponential model, with σ^2 equal to 1 or 2.25, $\phi = 0.2$ and $\tau^2 = 0.09$, so the theoretical variance, whose true value is $\text{Var}[Z(s)] = \sigma^2 + \tau^2$, amounts to 1.09 or 2.34. The resulting absolute errors associated to the estimation of the variance of the spatial process are represented in Fig. 1.

The reduced values displayed in Fig. 1 show a good behavior of both Bootstrap approaches, although giving some advantage to the smooth version, despite the fact that an additional bandwidth must be estimated. Furthermore, one should bear in mind that the underlying continuous distribution asks for a continuous tool to make inference and, particularly, it avoids obtaining repeated values.

Aiming to proceed with a deeper analysis to compare the proposed Bootstrap approaches, a new simulation study was carried out, focused on the estimation of the unidimensional distribution F_s . Observe that $F_s(x) = F(x)$, for all s , since the trend function has been taken to equal 0 at all locations. Five thresholds x were selected, identifying the quantiles 5, 25, 50, 75 and 95 % as being representative of the distribution domain, respectively denoted by P_i , with $i = 5, 25, 50, 75, 95$. For each x , we derived the discrete

Fig. 1 Estimated absolute errors of $\text{Var}[Z(s)]$ for 500 discrete and 500 continuous Bootstrap replicates of 200 samples. The data were simulated with the exponential model and each sample size is 50. The true values of $\text{Var}[Z(s)]$ are 1.09 and 2.34 in the left and right panels, respectively



and continuous estimators from the sample locations $\{t_1, \dots, t_n\}$, given by:

$$\hat{F}(x) = \frac{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{t_j - t_i}{h_1}\right) I_{\{Z(t_i) \leq x\}} I_{\{Z(t_j) \leq x\}}}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{t_j - t_i}{h_1}\right)}$$

$$\tilde{F}(x) = \frac{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{t_j - t_i}{h_1}\right) \mathcal{L}\left(\frac{x - Z(t_i)}{h}\right) \mathcal{L}\left(\frac{x - Z(t_j)}{h}\right)}{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{t_j - t_i}{h_1}\right)}$$

The discrete and the continuous Bootstrap methods were used to approximate the errors of both estimators. With this aim, k locations s_i were chosen, where B discrete and continuous replicates were generated to derive the Bootstrap analogues of $\hat{F}(x)$ and $\tilde{F}(x)$, denoted by $\hat{F}^{*b}(x)$ and $\tilde{F}^{*b}(x)$ for the b -th replicate and $b = 1, \dots, B$. Then, we computed $(B^{-1} \sum_{b=1}^B (\hat{F}^{*b}(x) - \hat{F}(x))^2)^{1/2}$ and $(B^{-1} \sum_{b=1}^B (\tilde{F}^{*b}(x) - \tilde{F}(x))^2)^{1/2}$, which provide us with approximations of the Bootstrap standard errors of the discrete and continuous distribution estimators, respectively. Table 1 summarizes the results obtained from 500 Bootstrap replicates and 200 data sets simulated under the exponential model, with $\sigma^2 = 1$, $\phi = 0.2$ and $\tau^2 = 0.09$.

The small values presented in Table 1 make clear the good performance of the two Bootstrap approaches. In generic terms, it seems advantageous to adopt the continuous estimator \tilde{F} over the discrete estimator \hat{F} , regardless of the resampling method considered. On the other hand, both Bootstrap procedures provide similar estimates of the accuracy of \tilde{F} , while the difference is more evident for approximation of the distribution through \hat{F} .

Taking into account the foregoing results, the smoother Bootstrap version will be applied in the following studies. We will now compare the estimation of the total variance of

Table 1 Mean and standard deviation of the standard errors (SE) obtained for the discrete and continuous distribution estimators, for 500 discrete and 500 continuous Bootstrap replicates of 200 samples

Threshold x	Distribution estimator	Disc Boot SE Mean (SD)	Cont Boot SE Mean (SD)
P_5	Discrete	0.0203 (0.0046)	0.0051 (0.0026)
	Continuous	0.0025 (0.0017)	0.0024 (0.0014)
P_{25}	Discrete	0.0512 (0.0111)	0.0330 (0.0081)
	Continuous	0.0079 (0.0049)	0.0081 (0.0044)
P_{50}	Discrete	0.0971 (0.0238)	0.0708 (0.0211)
	Continuous	0.0151 (0.0082)	0.0158 (0.0078)
P_{75}	Discrete	0.0843 (0.0183)	0.0627 (0.0182)
	Continuous	0.0230 (0.0119)	0.0240 (0.0117)
P_{95}	Discrete	0.0598 (0.0090)	0.0385 (0.0124)
	Continuous	0.0297 (0.0155)	0.0310 (0.0158)

The data were simulated with the exponential model and each sample size is 50

$Z(s)$, with another spatial approach, such as the parametric estimation. With this idea, parametric variograms were obtained by selecting valid models and deriving maximum likelihood (ML) estimates to approximate the unknown parameters. The exponential and the spherical models were used for the latter purpose, thus providing two different settings, depending on whether the parametric candidate coincides with the theoretical model or it is affected by misspecification. The resulting values are shown in Table 2, where data were simulated under the exponential and the spherical models, with $\sigma^2 = 1$, $\phi = 0.2$ and $\tau^2 = 0$.

According to the values presented in Table 2, the Bootstrap replicates offer more accurate estimates of $\text{Var}[Z(s)]$ than the ML approaches, even when assuming knowledge of the true variogram model. Surprisingly, the prior knowledge of the parametric family is not always

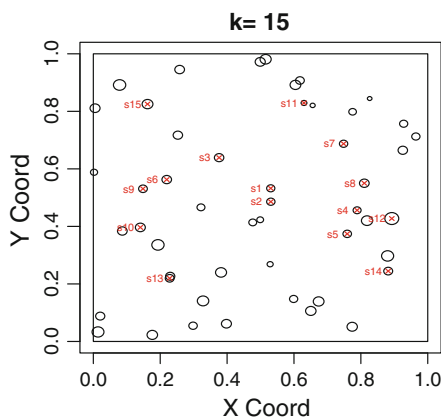
advantageous, as illustrated in the case presented in Fig. 2. The latter can be explained by the fact that, under the parametric approach, the approximation derived for the total variance is strongly dependent on the appropriate characterization of the model parameters rather than on the overall variogram function.

In view of the latter, a further step in this research was to analyze the behavior of the resampling methods when dealing with the estimation of the variogram function γ . To do the latter, the integrated quadratic error, $ISE = \int (\gamma(t) - \hat{\gamma}(t))^2 dt$, was approximated numerically, for each data set and for each of the estimators $\hat{\gamma}$ implemented, including a valid version obtained through the Bootstrap approach. We started by deriving the empirical Bootstrap estimates of the variogram and then fitting them, through an iterated weighted least squares criterion, to a class of permissible variograms, following the procedure developed in Shapiro and Botha (1991). Proceeding in this way, the validity of the resulting estimator was guaranteed with no prior specification of a parametric model. Figure 2 represents the example

Table 2 Mean and standard deviation of the absolute errors (AE) associated to the estimation of $\text{Var}[Z(s)]$, whose true value is 1, for 500 Bootstrap replicates and the ML estimates

Theoretical model	Type of estimator	AE Mean (SD)
Exponential	Bootstrap	0.149 (0.107)
	ML exponential	0.313 (0.160)
	ML spherical	0.372 (0.236)
Spherical	Bootstrap	0.148 (0.084)
	ML exponential	0.272 (0.277)
	ML spherical	0.233 (0.164)

The data were simulated with the exponential and spherical models. Total number of data sets is 200 and each sample size is 50



of a data set simulated with an exponential model together with the resulting variogram estimators, one given by the Bootstrap approach and the other two acquired by ML.

The procedure described above was repeated for 200 samples of size $n = 50$, for data simulated under the exponential and spherical models, with $\sigma^2 = 1$, $\phi = 0.2$ and $\tau^2 = 0$. The ISE values were computed and Table 3 summarizes the results obtained.

The results displayed in Table 3 give account of the good performance of the Bootstrap approach when addressed to estimate the spatial dependence of the random process in terms of the variogram. The ML estimates improve the behavior of the resampling procedure for small distances, although the Bootstrap method competes with the parametric approach for lags larger than 0.05.

4.2 Application to environmental monitoring data

In this section we derive an application of the Bootstrap methodology to a real data set concerning biomonitoring of arsenic pollution in the Central Region of Portugal, classified as NUTS II (NUTS stands for "Nomenclature of Units for Territorial Statistics"). The measured variable represents the concentrations in moss samples, in micrograms per gram dry weight. The typical procedure, alternative to the more expensive solution of determining the amount of pollutant directly, is to plant the moss and some time later to collect it, which allows the concentration of arsenic (and other heavy metals) to be measured. More details on this Portuguese project of air pollution analysis can be found in Martins et al. (2012).

The data set was collected in 2006 and it can be represented by $\{(t_i, Z(t_i)), i = 1, \dots, n\}$, with $n = 98$ and $Z(t_i)$ identifying the log-transformed concentration of arsenic (As) at location t_i . We adopted the log-transformation to

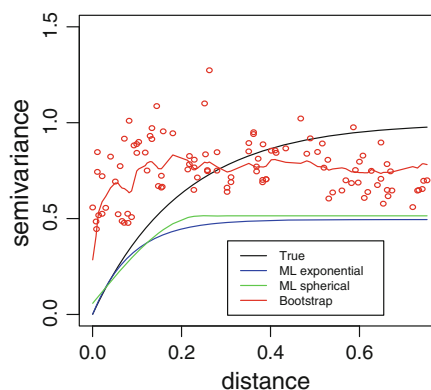


Fig. 2 The left panel represents one simulated data set with $n = 50$ sample locations, where each bullet size is proportional to the corresponding measured value. The selected $k = 15$ locations in the left panel are displayed after being reordered. The right panel

represents the variogram estimators, where the theoretical model chosen to generate the data is the exponential one. The bullets in the right panel represent the Bootstrap variogram estimates obtained for 500 replicates

Table 3 Mean and standard deviation of the ISE values obtained through the Bootstrap estimator combined with Shapiro and Botha’s method and the ML estimators

Theoretical model	Type of estimator	ISE-lags < 0.05	ISE-lags ≥ 0.05
		Mean (SD)	Mean (SD)
Exponential	Bootstrap	0.0286 (0.0144)	0.0497 (0.0331)
	ML exponential	0.0001 (0.0004)	0.0395 (0.0359)
	ML spherical	0.0001 (0.0004)	0.0450 (0.0445)
Spherical	Bootstrap	0.0355 (0.0154)	0.0337 (0.0256)
	ML exponential	0.0011 (0.0027)	0.0934 (0.1704)
	ML spherical	0.0004 (0.0009)	0.0506 (0.0729)

The data were simulated with the exponential and spherical models. Total number of data sets is 200 and each sample size is 50

reduce the impact of the presence of outliers. Afterwards, there were still three gross outliers, which were replaced by the average of the remaining values from that year’s survey. Table 4 gives the summary statistics for the resulting data, showing that the log-transformation leads to a more symmetric distribution. Furthermore, Fig. 3 presents the spatial representation of log-transformed data, where each bullet size is proportional to the corresponding measured value.

Aiming to exemplify the usefulness of the proposed Bootstrap approaches in this application, we first estimate a deterministic model for $E[Z(s)] = \mu(s)$ with $s = (UtmX, UtmY) \in D \subset \mathbb{R}^2$ and D identifying the region of NUTS II in Portugal. We then assume that the random process $Z(s)$ can be modeled as:

$$Z(s) = \mu(s) + Y(s)$$

where $\{Y(s) : s \in D\}$ is a zero-mean strictly stationary random process and $\mu(s) = \alpha_0 + \alpha_1 \times UtmY$. These regression coefficients were estimated, presenting statistically significant values $\alpha_0 = -37.8$ and $\alpha_1 = 0.0083$.

To derive weights $p_{i_1, \dots, i_k}^{(2)}$, associated to the multivariate distribution $\hat{F}_{s_1, \dots, s_k}^{(2)}$, one needs to select locations $s_1, \dots, s_k \in D$. So, we proceeded by choosing s_i , for $i = 1, \dots, k$ and $k = 40$, among the 98 sample locations, as represented in Fig. 3. The probabilities $p_{i_1, \dots, i_k}^{(2)}$ were acquired as explained in steps 3 and 5 of the resampling algorithm given in Sect. 3, which allowed us to generate Bootstrap samples by taking into account the dependence structure of the underlying random process.

Table 4 Summary statistics for arsenic pollution levels measured in the Central Region of Portugal (NUTS II)

	Mean	Median	SD	Minimum	Maximum
Untransformed	1.24	0.57	2.45	0.03	19.32
Log-transformed	-0.49	-0.55	0.98	-2.30	2.53

With 500 replicates, we estimated the total variance of the process $Y(\cdot)$, whose results are given in Table 5, where three different approaches were considered. According to the numerical studies presented in Sect. 4.1, the value 0.886 seems an accurate approximation of $\text{Var}[Y(s)]$.

To highlight the potentiality of the Bootstrap techniques within the scope of dependent data, one of their applications, pointed out in Sect. 1, will be addressed. In particular, we will focus on the estimation of the accuracy of a spatial approach, such as the nonparametric spatial predictor proposed in Menezes et al. (2010), under stochastic sampling design. The use of the latter kernel-based predictor demands an optimal bandwidth, which can be defined as dependent on the target location, offering better results than when a global optimal bandwidth is adopted. A drawback of the aforementioned predictor is that no estimation of the prediction error is available.

Here, we suggest to estimate standard errors through the Bootstrap approaches specified along this manuscript. In fact, if a prediction value is obtained for each Bootstrap sample (for a total of B replicates), then a simple method to approximate the unknown standard error is the standard deviation of those prediction values. Results are summarized in Table 6, for three points randomly chosen in the NUTS II region, represented as s_A, s_B and s_C in Fig. 3. The nonparametric predictions were computed as established in

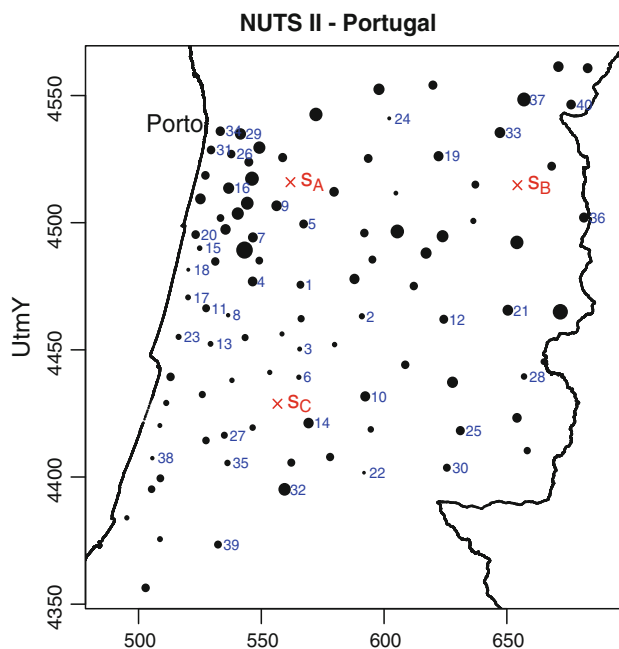


Fig. 3 Spatial representation of moss data in the Central Region of Portugal (NUTS II). The size of the bullets, representing the sampled locations, is proportional to the measured value. The points marked with numbers are used to generate the Bootstrap replicates. The points identifying the three locations $\{s_A, s_B, s_C\}$ are the goal of prediction for $Z(\cdot)$

Table 5 Estimates of the variance of the log-transformed arsenic pollution level, by considering the discrete and continuous versions of the Bootstrap replicates and assuming an exponential covariance model fitted to the observed data through ML

	Disc Bootstrap	Cont Bootstrap	ML exponential
Variance estimate	0.798	0.886	0.789

Table 6 Estimates of the prediction error of the log-transformed arsenic pollution level at locations s_A , s_B and s_C . For each nonparametric prediction (NP Pred), two different estimates of the standard errors (SE) are presented, by considering the discrete and continuous versions of the Bootstrap replicates. Results from OK are given in the two right columns

	NP Pred	Disc Boot SE	Cont Boot SE	OK Pred	OK SE
$\hat{Y}(s_A)$	0.270	0.229	0.234	0.299	0.820
$\hat{Y}(s_B)$	0.007	0.518	0.553	-0.064	0.854
$\hat{Y}(s_C)$	-0.183	0.191	0.199	-0.209	0.827

Menezes et al. (2010), by using optimal local bandwidths equal to 75.5, 36.9 and 90, for s_A , s_B and s_C , respectively.

For each nonparametric prediction, two different estimates of the standard errors are presented, obtained by considering either the discrete or the continuous versions of the Bootstrap replicates. As complementary information, we also present the prediction results derived by application of the ordinary kriging (OK) in Table 6. The latter can allow us to conclude that the Bootstrap estimators point out smaller values for the standard errors than those obtained through the OK. The largest standard errors are associated to the location s_B , where less information is available, since that area includes less sampled data.

As a last note, we have back-transformed the predicted values and added the trend information, to obtain the estimates of the process $Z(s)$ at those three target locations, leading to $\hat{Z}(s_A) = 1.153$, $\hat{Z}(s_B) = 0.878$ and $\hat{Z}(s_C) = 0.354$ (under OK, the corresponding values were 1.187, 0.818 and 0.345). Knowing that 50 % of the locations have a concentration of arsenic smaller than 0.57, it is possible to conclude that s_C is one of the locations with lower intensity of air pollution, as opposed to s_A and s_B .

5 Conclusions

In this paper consistent estimators of the multivariate distribution function have been proposed, which can be used as the basis for implementation of Bootstrap approaches in the spatial setting. The resampling method derived from the discrete estimator is an adaptation to this setting of the

naive Bootstrap described in Efron (1979) and has similar properties, such as consistency or the fact that nearly every sample derived from it contains repeated values. The alternative version, obtained in the current work by applying a continuous distribution estimator, is the analogue of the smoothed Bootstrap approach for independent data (Lejeune and Sarda 1992). An advantage of the second approach is that it entails resampling from a continuous distribution and, therefore, it avoids the aforementioned problem of providing repeated data in the replicates. However, an additional uncertainty is introduced, in terms of the bandwidth parameter that must be estimated. For independent data, the question of whether the smoothed Bootstrap is superior to the naive alternative, and for which smoothing parameter, has been analyzed by a few authors (Silverman and Young 1987; Hall 1992), but no definitive answers exist. Therefore, the difficulties to check the behavior of the resampling approaches in the spatial setting increase, because of the underlying dependence structure. In this respect, although further research should be developed, the numerical studies conducted in the current work give account of a good performance of both procedures, with a little superiority of the continuous spatial Bootstrap, when the bandwidth parameter is appropriately selected.

On the other hand, as pointed out in the introduction, the Bootstrap methodology allows the researchers to solve different statistical problems inherent to the estimation process. It must not be intended as a substitute of other techniques designed for addressing specific issues, but for complementing them and adding extra information. From this perspective, the Bootstrap proposals offer an attractive alternative for resampling in the spatial setting. Such approaches aim at reproducing the data dependence structure, before deriving subsamples. In this respect, the numerical studies developed with simulated data give account of the good behavior of the resampling methods, which can be even advantageous over other procedures for estimation of the variance or the variogram function, although the really important thing is that they help capture the main features of the underlying spatial process. Consequently, the generation of replicates offer an accurate alternative to derive estimates of the standard error or any other unknown characteristic of the statistical approach that can be considered.

Acknowledgments The authors would like to thank the helpful suggestions and comments from the Reviewers. The authors are also grateful to Dr. K. J. Duncan-Barlow (University of Vigo) for her contribution in the language revision. The first and third authors acknowledge financial support from the Project TEC2011-28683-C02-02 of the Spanish Ministry of Science and Innovation and the Project CN2012/279 from the European Regional Development Fund and the Galician Regional Government (Xunta de Galicia). The second author's work has been supported by the Project PTDC/MAT/112338/2009 (FEDER support included) of the Portuguese Ministry of Science, Technology and Higher Education.

Appendix 1: Consistency of $\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k)$

To check that consistency follows for $\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k)$, the hypotheses described below will be assumed:

- (i) $\{Z(s) \in \mathbb{R} : s \in D \subset \mathbb{R}^d\}$ can be modeled as given in (1).
- (ii) $D = \lambda D_0$, for some $\lambda = \lambda(n) \xrightarrow{n \rightarrow +\infty} +\infty$ and bounded $D_0 \subset \mathbb{R}^d$.
- (iii) $t_i = \lambda u_i$, for $1 \leq i \leq n$, where u_1, \dots, u_n denotes a realization of a random sample of size n drawn from a density function g_0 considered on D_0 .
- (iv) $Z(\cdot)$ is α -mixing, with $\alpha(r) = O(r^{-a})$, for $r > 0$ and some constant $a > 0$.
- (v) K is d -variate and symmetric density function with compact support.
- (vi) $\{h_1^2 + \dots + h_{k-1}^2 + \lambda^{-1} + n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d}\} \xrightarrow{n \rightarrow +\infty} 0$.
- (vii) $F_{s_1, \dots, s_k}(x_1, \dots, x_k)$ is three-times continuously differentiable as a function of (s_1, \dots, s_k) .

We will prove that the bias and the variance of $\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k)$ are of the respective orders $(h_1^2 + \dots + h_{k-1}^2)$ and $(n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d} + \lambda^{-d})$ and, therefore, tend to zero as the sample size n increases, which would yield the consistency of the distribution estimator. To do the latter, conditions (i)–(vii) will be applied and a similar procedure as in the proof of Theorem 3.1 in Hall and Patil (1994).

Write $A_{i_1, i_2} = \frac{P_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}}{\sum_{i_1=1}^n \sum_{i_2=1}^n P_{s_1, s_2}^{t_{i_1}, t_{i_2}, h_1}}$ and $A_{i_{j-1}, i_j} = \frac{P_{s_{j-1}, s_j}^{t_{i_{j-1}}, t_{i_j}, h_{j-1}}}{\sum_{i_{j-1}=1}^n \sum_{i_j=1}^n P_{s_{j-1}, s_j}^{t_{i_{j-1}}, t_{i_j}, h_{j-1}}}$ for $j = 3, \dots, k$. Firstly, we can take into account that, for large n :

$$\begin{aligned} \mathbb{E} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] &= \mathbb{E} \left[\mathbb{E} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) t_j, \forall j \right] \right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} \mathbb{E} \left[I_{\{X(t_{i_1}) \leq x_1\}} \dots I_{\{X(t_{i_k}) \leq x_k\}} t_j, \forall j \right] \right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} F_{t_{i_1}, \dots, t_{i_k}}(x_1 + \mu(s_1 + t_{i_1} - s_1) \right. \\ &\quad \left. - \mu(s_1), \dots, x_k + \mu(s_k + t_{i_1} - s_1) - \mu(s_k)) \right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-2}, i_{k-1}} \right. \\ &\quad \left. \cdot \mathbb{E} \left[A_{i_{k-1}, i_k} F_{t_{i_1} - t_{i_1} + s_1, \dots, t_{i_k} - t_{i_1} + s_1}(x_1, \dots, x_k) t_j, j \leq k - 1 \right] \right] \end{aligned}$$

on account of (3).

Now, the last conditional expectation will be approximated. With this aim, bear in mind that:

$$\begin{aligned} &\mathbb{E} \left[K \left(\frac{s_{k-1} - s_k - (t_{i_{k-1}} - t_{i_k})}{h_{k-1}} \right) F_{t_{i_1} - t_{i_1} + s_1, \dots, t_{i_k} - t_{i_1} + s_1}(x_1, \dots, x_k) \right] \\ &= \int K \left(\frac{s_{k-1} - s_k - (t_{i_{k-1}} - \lambda u)}{h_{k-1}} \right) \\ &\quad F_{t_{i_1} - t_{i_1} + s_1, \dots, \lambda u - t_{i_1} + s_1}(x_1, \dots, x_k) g_0(u) du \approx \lambda^d h_{k-1}^d g_0(0) \\ &\quad \int K(z_1) F_{t_{i_1} - t_{i_1} + s_1, \dots, s_k - s_{k-1} + t_{i_{k-1}} - t_{i_1} + s_1 + h_{k-1} z_1}(x_1, \dots, x_k) \\ &\quad dz_1 \mathbb{E} \left[K \left(\frac{s_{k-1} - s_k - (t_{i_{k-1}} - t_{i_k})}{h_{k-1}} \right) \right] \\ &= \int K \left(\frac{s_{k-1} - s_k - (t_{i_{k-1}} - \lambda u)}{h_{k-1}} \right) g_0(u) du \approx \lambda^d h_{k-1}^d g_0(0) \end{aligned}$$

From the previous relations, it follows that:

$$\begin{aligned} &\mathbb{E} \left[A_{i_{k-1}, i_k} F_{t_{i_1} - t_{i_1} + s_1, \dots, t_{i_k} - t_{i_1} + s_1}(x_1, \dots, x_k) / t_j, j \leq k - 1 \right] \\ &\quad \approx n^{-1} \int K(z_1) F_{t_{i_1} - t_{i_1} + s_1, \dots, s_k - s_{k-1} + t_{i_{k-1}} - t_{i_1} + s_1 + h_{k-1} z_1} \\ &\quad (x_1, \dots, x_k) dz_1 \end{aligned}$$

and, therefore:

$$\begin{aligned} &\mathbb{E} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] \\ &\quad \approx \sum_{i_1=1}^n \dots \sum_{i_{k-1}=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-2}, i_{k-1}} \int K(z_1) \right. \\ &\quad \left. \cdot F_{t_{i_1} - t_{i_1} + s_1, \dots, s_k - s_{k-1} - t_{i_{k-1}} - t_{i_1} + s_1 + h_{k-1} z_1}(x_1, \dots, x_k) dz_1 \right]. \end{aligned}$$

We can iterate the strategy above, based on applying an appropriate conditional expectation and developing the resulting term, to achieve that:

$$\begin{aligned} &\mathbb{E} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] \\ &\quad \approx \int \dots \int F_{s_1, s_2 + h_1 z_{k-1}, \dots, s_k + h_1 z_{k-1} + \dots + h_{k-1} z_1}(x_1, x_2, \dots, x_k) \\ &\quad \cdot K(z_1) \dots K(z_{k-1}) dz_1 \dots dz_{k-1} \\ &= F_{s_1, s_2, \dots, s_k}(x_1, x_2, \dots, x_k) + O(h_1^2 + \dots + h_{k-1}^2) \end{aligned}$$

With regard to the variance, one has for large n that:

$$\begin{aligned} &\text{Var} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] \\ &= \mathbb{E} \left[\left(\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) - \mathbb{E} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] \right)^2 \right] \\ &\quad \approx V_1 + V_2 \end{aligned}$$

where:

$$V_1 = \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2}^2 \dots A_{i_{k-1}, i_k}^2 \cdot \left(I_{\{X(t_{i_1}) \leq x_1\}} \dots I_{\{X(t_{i_k}) \leq x_k\}} - F_{s_1, \dots, s_k}(x_1, \dots, x_k)^2 \right) \right]$$

$$V_2 = \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \sum_{j_1=1}^n \dots \sum_{j_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} A_{j_1, j_2} \dots A_{j_{k-1}, j_k} \left(I_{\{X(t_{i_1}) \leq x_1\}} \dots I_{\{X(t_{i_k}) \leq x_k\}} \cdot I_{\{X(t_{j_1}) \leq x_1\}} \dots I_{\{X(t_{j_k}) \leq x_k\}} - F_{s_1, \dots, s_k}(x_1, \dots, x_k)^2 \right) \right]$$

By using similar arguments as above, we could check that:

$$V_1 \approx \frac{n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d} \left(F_{s_1, \dots, s_k}(x_1, \dots, x_k) - F_{s_1, \dots, s_k}(x_1, \dots, x_k)^2 \right)}{g_0(0)^k \left(\int K(z)^2 dz \right)^{k-1}}$$

$$V_2 \approx \lambda^{-d} \int \left(F_{s_1, \dots, s_k, s_1+t, \dots, s_k+t}(x_1, \dots, x_k, x_1, \dots, x_k) - F_{s_1, \dots, s_k}(x_1, \dots, x_k)^2 \right) dt$$

Consequently:

$$\text{Var} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] = O \left(n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d} + \lambda^{-d} \right)$$

We could derive the dominant terms of the bias and the variance of the distribution estimator as well as asymptotically minimize the mean squared error (MSE) of the distribution estimator, namely:

$$\text{MSE} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right] = \text{Bias} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right]^2 + \text{Var} \left[\hat{F}_{s_1, \dots, s_k}^{(2)}(x_1, \dots, x_k) \right]$$

to obtain the optimal bandwidths h_j , for $j = 1, \dots, k - 1$, which would be dependent on unknown terms, such as the multivariate distribution function itself and its second-order derivatives.

Appendix 2: Consistency of $\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k)$

To derive this proof, we will assume conditions (i)–(v), together with:

- (vi') $\{h_1^2 + \dots + h_{k-1}^2 + h^2 + \lambda^{-1} + n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d}\} \xrightarrow{n \rightarrow +\infty} 0$.
- (vii') $F_{s_1, \dots, s_k}(x_1, \dots, x_k)$ is three-times continuously differentiable as a function of (s_1, \dots, s_k) and as a function of (x_1, \dots, x_k) .
- (viii) L is a univariate and symmetric density function with compact support.

For large n , the aforementioned hypotheses yield that:

$$\begin{aligned} \mathbb{E} \left[\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) \right] &= \mathbb{E} \left[\mathbb{E} \left[\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) / t_{ij}, \forall j \right] \right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} \mathbb{E} \left[\mathcal{L} \left(\frac{x_1 - X(t_{i_1})}{h} \right) \dots \mathcal{L} \left(\frac{x_k - X(t_{i_k})}{h} \right) / t_{ij}, \forall j \right] \right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} \cdot \int \mathcal{L} \left(\frac{x_1 - u_1 + \mu(s_1 - t_{i_1} + s_1) - \mu(s_1)}{h} \right) \dots \mathcal{L} \left(\frac{x_k - u_k + \mu(s_k - t_{i_1} + s_1) - \mu(s_k)}{h} \right) \cdot f_{t_{i_1}, \dots, t_{i_k}}(u_1, \dots, u_k) du_1 \dots du_k \right] \end{aligned}$$

where f_{t_1, \dots, t_k} denotes the joint density function of $(Z(t_1), \dots, Z(t_k))$.

We can integrate by parts and apply relation (3) to obtain that:

$$\begin{aligned} \mathbb{E} \left[\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) \right] &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} \int L(y_1) \dots L(y_k) \cdot F_{t_{i_1}, \dots, t_{i_k}}(x_1 - hy_1 + \mu(s_1 - t_{i_1} + s_1) - \mu(s_1), \dots, x_k - hy_k + \mu(s_k - t_{i_1} + s_1) - \mu(s_k)) dy_1 \dots dy_k \right] \\ &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} \int L(y_1) \dots L(y_k) \cdot F_{t_{i_1} - t_{i_1} + s_1, \dots, t_{i_k} - t_{i_1} + s_1}(x_1 - hy_1, \dots, x_k - hy_k) dy_1 \dots dy_k \right] \end{aligned}$$

By proceeding with analogue arguments as those used for the bias of $\hat{F}_{s_1, \dots, s_k}^{(2)}$, it follows that:

$$\begin{aligned} \mathbb{E} \left[\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) \right] &\approx \int \dots \int \int \dots \int L(y_1) \dots L(y_k) \cdot K(z_1) \dots K(z_{k-1}) F_{s_1, s_2+h_1 z_{k-1}, \dots, s_k+h_1 z_{k-1}+\dots+h_{k-1} z_1}(x_1 - hy_1, \dots, x_k - hy_k) dz_1 \dots dz_{k-1} dy_1 \dots dy_k \\ &= F_{s_1, s_2, \dots, s_k}(x_1, \dots, x_k) + O(h_1^2 + \dots + h_{k-1}^2 + h^2) \end{aligned}$$

Finally, the approximation of the variance of the continuous estimator will be addressed as given below:

$$\text{Var} \left[\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k) \right] \approx W_1 + W_2$$

with:

$$\begin{aligned} W_1 &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2}^2 \dots A_{i_{k-1}, i_k}^2 \left(\mathcal{L} \left(\frac{x_1 - X(t_{i_1})}{h} \right) \dots \mathcal{L} \left(\frac{x_k - X(t_{i_k})}{h} \right) - F_{s_1, \dots, s_k}(x_1, \dots, x_k)^2 \right) \right] \\ W_2 &= \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \sum_{j_1=1}^n \dots \sum_{j_k=1}^n \mathbb{E} \left[A_{i_1, i_2} \dots A_{i_{k-1}, i_k} A_{j_1, j_2} \dots A_{j_{k-1}, j_k} \cdot \left(\mathcal{L} \left(\frac{x_1 - X(t_{i_1})}{h} \right) \dots \mathcal{L} \left(\frac{x_k - X(t_{i_k})}{h} \right) \mathcal{L} \left(\frac{x_1 - X(t_{j_1})}{h} \right) \dots \mathcal{L} \left(\frac{x_k - X(t_{j_k})}{h} \right) - F_{s_1, \dots, s_k}(x_1, \dots, x_k)^2 \right) \right] \end{aligned}$$

Now, we can combine the arguments used for the bias of the continuous estimator with those applied for the variance of the discrete estimator to check that both terms, W_1 and W_2 , are asymptotically negligible, as established next:

$$\begin{aligned}
 W_1 &\approx \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2}^2 \dots A_{i_{k-1}, i_k}^2 \int L(y_1) \dots L(y_k) \cdot (F_{s_1, \dots, s_k}(x_1 - hy_1, \dots, x_k - hy_k) - F_{s_1, \dots, s_k}(x_1, \dots, x_k))^2 dy_1 \dots dy_k \right] \\
 &\approx \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \mathbb{E} \left[A_{i_1, i_2}^2 \dots A_{i_{k-1}, i_k}^2 (F_{s_1, \dots, s_k}(x_1, \dots, x_k) - F_{s_1, \dots, s_k}(x_1, \dots, x_k))^2 \right] \\
 &\approx \frac{n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d} (x_1, \dots, x_k) (F_{s_1, \dots, s_k}(x_1, \dots, x_k) - F_{s_1, \dots, s_k}(x_1, \dots, x_k))^2}{g_0(0)^k \left(\int K(z)^2 dz \right)^{k-1}} = O\left(n^{-k} \lambda^{d(k-1)} h_1^{-d} \dots h_{k-1}^{-d}\right) W_2 \\
 &\approx \sum_{i_1=1}^n \dots \sum_{i_k=1}^n \sum_{j_1=1}^n \dots \sum_{j_k=1}^n \mathbb{E} [A_{i_1, i_2} \dots A_{i_{k-1}, i_k} A_{j_1, j_2} \dots A_{j_{k-1}, j_k}] \\
 &\quad \cdot \int \int \dots \int \int \dots \int L(y_1) \dots L(y_k) L(w_1) \dots L(w_k) (F_{s_1, \dots, s_k, s_1+t, \dots, s_k+t}(x_1 - hy_1, \dots, x_k - hy_k, x_1 - hw_1, \dots, x_k - hw_k) \\
 &\quad - F_{s_1, \dots, s_k}(x_1, \dots, x_k))^2 dt dy_1 \dots dy_k \cdot dw_1 \dots dw_k \\
 &\approx \lambda^{-d} \int (F_{s_1, \dots, s_k, s_1+t, \dots, s_k+t}(x_1, \dots, x_k, x_1, \dots, x_k) - F_{s_1, \dots, s_k}(x_1, \dots, x_k))^2 dt = O(\lambda^{-d})
 \end{aligned}$$

Then, consistency yields for $\tilde{F}_{s_1, \dots, s_k}(x_1, \dots, x_k)$, since its bias and variance tend to zero, as the sample size increases.

References

Bowman A, Hall P, Prvan T (1998) Bandwidth selection for the smoothing of distribution functions. *Biometrika* 85(4):799–808. doi:10.1093/biomet/85.4.799

Crujeiras RM, Fernández-Casal R, González-Manteiga W (2010) Goodness-of-fit tests for the spatial spectral density. *Stoch Environ Res Risk Assess* 24(1):67–79. doi:10.1007/s00477-008-0300-0

De Angelis D, Young GA (1992) Smoothing the bootstrap. *Int Stat Rev* 60(1):45–56

Efron B (1979) Bootstrap methods: another look at the Jackknife. *Ann Stat* 7(1):1–26. doi:10.1214/aos/1176344552

García-Soidán P (2007) Asymptotic normality of the Nadaraya–Watson semivariogram estimators. *TEST* 16(3):479–503. doi:10.1007/s11749-006-0016-8

Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York

Govaerts B, Beck B, Lecoutre E, Le Bailly C, VandenEeckaut P (2005) From monitoring data to regional distributions: a practical methodology applied to water risk assessment. *Environmetrics* 16(2):109–127. doi:10.1002/env.665

Hall P (1985) Resampling a coverage pattern. *Stoch Process Appl* 20(2):231–246. doi:10.1016/0304-4149(85)90212-1

Hall P (1992) *The bootstrap and edgeworth expansion*. Springer, New York

Hall P, Maiti T (2006) On parametric bootstrap methods for small area prediction. *J R Stat Soc B* 68(2):221–238. doi:10.1111/j.1467-9868.2006.00541.x

Hall P, Patil P (1994) Properties of nonparametric estimators of autocovariance for stationary random fields. *Probab Theory Relat Fields* 99(3):399–424. doi:10.1007/BF01199899

Hyun-Han K, Young-II M (2006) Improvement of overtopping risk evaluations using probabilistic concepts for existing dams. *Stoch Environ Res Risk Assess* 20(4):223–237. doi:10.1007/s00477-005-0017-2

Iranpanah N, Mansourianb A, Tashayob B, Haghghi F (2011) Spatial semi-parametric bootstrap method for analysis of kriging predictor of random field. *Procedia Environ Sci* 3:81–86. doi:10.1016/j.proenv.2011.02.015

Lejeune M, Sarda P (1992) Smooth estimators of distribution and density functions. *Comput Stat Data Anal* 14:457–471. doi:10.1016/0167-9473(92)90061-J

Li B, Genton M, Sherman M (2007) A nonparametric assessment of properties of space–time covariance functions. *JASA* 102(478):736–744. doi:10.1198/016214507000000202

Loh JM (2008) A valid and fast spatial Bootstrap for correlation functions. *Astrophys J* 681(1):726–734. doi:10.1086/588631

Maglione DS, Diblasi AM (2004) Exploring a valid model for the variogram of an isotropic spatial process. *Stoch Environ Res Risk Assess* 18(6):366–376. doi:10.1007/s00477-003-0143-7

Martins A, Figueira R, Sousa A, Sérgio C (2012) Spatio-temporal patterns of Cu contamination in mosses using geostatistical estimation. *Environ Pollut* 170:276–284. doi:10.1016/j.envpol.2012.07.004

Menezes R, García-Soidán P, Ferreira C (2010) Nonparametric spatial prediction under stochastic sampling design. *J Nonparametr Stat* 22(3):363–377. doi:10.1080/10485250903094294

Olea RA, Pardo-Igúzquiza E (2011) Generalized Bootstrap method for assessment of uncertainty in semivariogram inference. *Math Geosci* 43(2):203–228. doi:10.1007/s11004-010-9269-6

Politis DN, Romano JP, Wolf M (1999) *Subsampling*. Springer, Berlin

Shapiro A, Botha JD (1991) Variogram fitting with a general class of conditionally nonnegative definite functions. *Comput Stat Data Anal* 11(1):87–96. doi:10.1016/0167-9473(91)90055-7

Silverman BW, Young GA (1987) The bootstrap: to smooth or not to smooth? *Biometrika* 74(3):469–479. doi:10.1093/biomet/74.3.469