Supplementary materials for this article are available at https://doi.org/10.1007/s13253-023-00555-0.

Nonparametric Conditional Risk Mapping Under Heteroscedasticity

Rubén FERNÁNDEZ-CASAL, Sergio CASTILLO-PÁEZ, and
Mario FRANCISCO-FERNÁNDEZ

A nonparametric procedure to estimate the conditional probability that a nonstationary geostatistical process exceeds a certain threshold value is proposed. The method consists of a bootstrap algorithm that combines conditional simulation techniques with nonparametric estimations of the trend and the variability. The nonparametric local linear estimator, considering a bandwidth matrix selected by a method that takes the spatial dependence into account, is used to estimate the trend. The variability is modeled estimating the conditional variance and the variogram from corrected residuals to avoid the biasses. The proposed method allows to obtain estimates of the conditional exceedance risk in non-observed spatial locations. The performance of the approach is analyzed by simulation and illustrated with the application to a real data set of precipitations in the USA.

Supplementary materials accompanying this paper appear on-line.

Key Words: Bootstrap; Conditional simulation; Local linear estimation; Bias correction.

1. INTRODUCTION

Risk maps containing the probabilities that a certain variable of interest exceeds a given threshold or permissible value in an area of study are usually employed by environmental agencies to control different pollution levels (in soil, air or water) or to alert population of possible natural disasters (earthquakes, floods, etc.). The estimation of these exceeding probabilities using simple and reliable statistical methods is, therefore, an important practical

Rubén Fernández-Casal, Sergio Castillo-Páez, and Mario Francisco-Fernández have contributed equally to this work.

R. Fernández-Casal (✉) Departamento de Matemáticas, CITIC, Facultad de Informática, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain
(E-mail: *ruben.fcasal@udc.es*).
S. Castillo-Páez · M. Francisco-Fernández, Departamento de Ciencias Exactas, Universidad de las Fuerzas Armadas ESPE, Av. General Rumiñahui s/n, 171103 Sangolquí, Ecuador
S. Castillo-Páez (E-mail: *sacastillo@espe.edu.ec*)
M. Francisco-Fernández (E-mail: *mariofr@udc.es*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 29, Number 1, Pages 56–72
https://doi.org/10.1007/s13253-023-00555-0

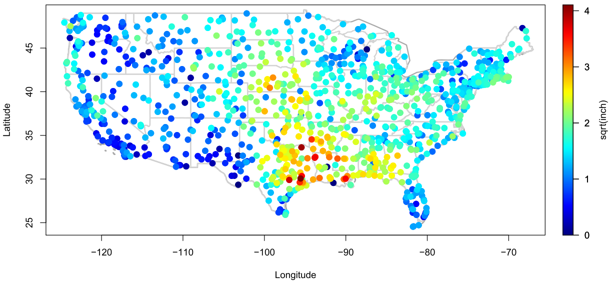

Figure 1. Spatial locations and observed values of the total precipitations (square-root of rainfall inches) during March 2016 recorded over 1053 locations on the continental part of USA .

issue. The resulting estimated maps can help governments to make decisions and to organize prevention policies in the near future.

In this paper, we develop a nonparametric methodology to produce risk maps and apply it to a data set containing the total precipitations (square-root of rainfall inches) during March 2016 recorded over 1053 locations on the continental part of the USA. (Fig. 1 contains the observed values). The goal is to estimate the conditional probability of occurring a total precipitation larger than or equal to a threshold value, which could have a direct application in agriculture or in flood prevention, for example.

Different geostatistical techniques have usually been employed to approximate exceeding probabilities. These methods include traditional approaches, such as the indicator kriging (IK) (e.g., Goovaerts et al. 1997) or the disjunctive kriging (DK) (e.g., Webster and Oliver 1989), or more recent procedures, such us those based on analysis of compositional data (e.g., Tolosana-Delgado et al. 2008). The IK consists in the application of the ordinary kriging linear predictor to indicator functions of the data. Although it is perhaps the most popular approach in this context, it has some drawbacks. First, the discretization of the data can lead to a loss of information. On the other hand, the estimated probabilities could be greater than one or negative. Moreover, it could present order-relation problems (see, e.g., Chilès and Delfiner 2012, Sect. 6.3.3). Some of these issues can be avoided with the use of the so-called simplicial indicator kriging (Tolosana-Delgado et al. 2008). This method employs a simplex approach for compositional data to estimate the conditional cumulative distribution function. Another alternative to the IK is the DK, a nonlinear estimation technique which usually assumes a Gaussian isofactorial model for the geostatistical process. However, there is no empirical evidence to recommend the DK in preference to the IK, or the opposite (Lark and Ferguson 2004).

The approaches previously described usually suppose stationarity and a parametric model. Therefore, if the assumed model is not appropriate, the conclusions drawn may be unreliable or even wrong. To avoid these problems, alternatively, nonparametric techniques could be used. For instance, García-Soidán and Menezes (2017) proposed two kernel-based estimators to approximate the local distribution under homoscedasticity. In this line, Fernández-Casal et al. (2018) proposed an unconditional bootstrap method to estimate the spatial risk, without assuming any parametric form for the trend function nor for the dependence structure of the process. They consider a homoscedastic model and used local linear estimates of the trend and the variogram, jointly with a procedure to correct the bias introduced by the direct use of the residuals in the variogram estimation. On the other hand, this nonparametric procedure was extended to a heteroscedastic context in Castillo-Páez et al. (2020). In both cases, although the use of nonparametric methods avoids misspecification problems, they focus on the estimation of the unconditional probability that the variable under study exceeds a threshold. Note that, as the replicas obtained by unconditional simulation will not necessarily match the observed sample values (see, e.g., Chilès and Delfiner 2012, Chapter 7), a direct comparison of these procedures with the ones described in the previous paragraph is not entirely appropriate, since they aim at the estimation of conditional exceeding probabilities.

In the present work, we propose a bootstrap method to estimate threshold exceeding conditional probabilities under heteroscedasticity of the spatial process. This approach uses the unconditional bootstrap method introduced in Castillo-Páez et al. (2020) as part of its implementation. The new procedure generates conditional replicates matching up the observed values at the sampled locations. The conditionalization of simulations is equivalent to choose among all possible unconditional simulations of the spatial process, those that coincide with the values obtained at the observation locations (Journel 1974).

The remainder of the paper is organized as follows. In Sect. 2, the spatial model considered in this research is presented. Additionally, the nonparametric estimators for the mean or trend function, the variance and the variogram employed in the conditional bootstrap method are introduced. In Sect. 3, the bootstrap algorithm to estimate the conditional risk is described (specifically, in Sect. 3.2). In this procedure, a slight modification of the bootstrap method proposed in Castillo-Páez et al. (2020) to approximate the unconditional risk (also discussed in Sect. 3.1) is used. A simulation study for assessing the performance of the new approach, considering stationary and nonstationary processes, under regular and non-regular sampling designs, is provided in Sect. 4. Section 5 discusses the application of the methods to the precipitation data introduced above. Finally, Sect. 6 contains some conclusions and finals remarks.

2. NONPARAMETRIC MODELING

Suppose that $\left\{Y(\mathbf{x}), \mathbf{x} \in D \subset \mathbb{R}^d\right\}$ is a spatial heteroscedastic process which can be modeled as follows:

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + \sigma(\mathbf{x})\varepsilon(\mathbf{x}), \tag{1}$$

where $\mu(\cdot)$ and $\sigma^2(\cdot)$ are the trend and variance functions, and $\varepsilon(\cdot)$ is a second-order stationary process with zero mean, unit variance and correlogram $\rho(\mathbf{u}) = \text{Cov}\left[\varepsilon(\mathbf{x}), \varepsilon(\mathbf{x}+\mathbf{u})\right]$, for $\mathbf{x}$ and $\mathbf{x}+\mathbf{u} \in D$. No specific forms will be assumed for $\mu(\cdot)$, $\sigma^2(\cdot)$ and $\rho(\cdot)$, although they should be smooth functions to be consistently estimated.

The goal is to estimate nonparametrically the conditional probability

$$r_c(\mathbf{x}_\alpha^e, \mathbf{Y}) = P\left[Y(\mathbf{x}_\alpha^e) \geq c \mid \mathbf{Y}\right], \tag{2}$$

where $\mathbf{Y} = [Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n)]^t$ are observed values of the process at certain sample locations, c is a threshold (critical) value, and $\left\{\mathbf{x}_\alpha^e\right\}_{\alpha=1}^{n_0}$ is a set of unobserved estimation locations.

It must be taken into account that the spatial dependence of the process Y depends on the variance and the correlogram of ε. For instance, the covariance matrix of the observations $\mathbf{Y}$ can be expressed as:

$$\boldsymbol{\Sigma} = \mathbf{DRD},$$

where $\mathbf{D} = \text{diag}\left[\sigma(\mathbf{x}_1), \ldots, \sigma(\mathbf{x}_n)\right]$ and $\mathbf{R}$ is the correlation matrix of the (unknown) errors $\boldsymbol{\varepsilon} = [\varepsilon(\mathbf{x}_1), \ldots, \varepsilon(\mathbf{x}_n)]^t$. The latter matrix is usually estimated from the semivariogram $\gamma(\mathbf{u}) = \frac{1}{2}Var\left[\varepsilon(\mathbf{x}) - \varepsilon(\mathbf{x}+\mathbf{u})\right] = 1 - \rho(\mathbf{u})$

The first step in the proposed approach consists in the nonparametric estimation of the trend, the variance and the dependence of the spatial process. Different nonparametric approaches have been used for the estimation of the model components, including kernel methods, splines or wavelets techniques. For instance, a comprehensive revision of trend estimation approaches can be found in Opsomer et al. (2001). In that paper, the authors focus on the framework of regression models with correlated homoscedastic errors. Nonparametric methods for the estimation of the functional variance have mainly based on the approximation of the mean of the squared residuals (see e.g., Fan and Yao 1998, for the independent case). In this line and among the available literature, we may highlight the works of Ruppert et al. (1997), who proposed a degrees-of-freedom correction of the bias due to the preliminary estimation of the trend with uncorrelated errors, and Vilar-Fernández and Francisco-Fernández (2006), who studied the properties of the squared residual estimator for one-dimensional correlated data. In the present paper, assuming model (1), a similar procedure to that proposed in Fernández-Casal et al. (2017) is used to nonparametrically estimate the model components.

To estimate the trend $\mu(\cdot)$, we consider a kernel-based method called the local linear estimator. This approach has shown a very good performance from a theoretical and a practical point of view (e.g., Fan and Gijbels 1996). In the spatial framework, given a sample $\{[\mathbf{x}_i, Y(\mathbf{x}_i)]\}_{i=1}^n$, the local linear estimator of $\mu(\mathbf{x})$ is given by:

$$\hat{\mu}_{\mathbf{H}}(\mathbf{x}) = \mathbf{e}_1^t \left(\mathbf{X}_{\mathbf{x}}^t \mathbf{W}_{\mathbf{x}} \mathbf{X}_{\mathbf{x}}\right)^{-1} \mathbf{X}_{\mathbf{x}}^t \mathbf{W}_{\mathbf{x}} \mathbf{Y} \equiv s_{\mathbf{x},\mathbf{H}}^t \mathbf{Y}, \tag{3}$$

where $\mathbf{e}_1$ is a vector with 1 in the first entry and all other entries 0, $\mathbf{X}_{\mathbf{x}}$ is a matrix whose i-th row is $[1, (\mathbf{x}_i - \mathbf{x})^t]$, $\mathbf{W}_{\mathbf{x}} = \mathtt{diag}\left[K_{\mathbf{H}}(\mathbf{x}_1 - \mathbf{x}), \ldots, K_{\mathbf{H}}(\mathbf{x}_n - \mathbf{x})\right]$, with $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}\mathbf{u})$, being K a multivariate kernel function, and $\mathbf{H}$ is a $d \times d$ nonsingular symmetric matrix called the bandwidth matrix. Note that $\hat{\mu}_{\mathbf{H}}(\mathbf{x})$ is a linear smoother, since the estimated values at the sample locations can be expressed as $\hat{\boldsymbol{\mu}} = \left[\hat{\mu}_{\mathbf{H}}(\mathbf{x}_1), \ldots, \hat{\mu}_{\mathbf{H}}(\mathbf{x}_n)\right]^t = \mathbf{S}_{\mathbf{H}}\mathbf{Y}$, being $\mathbf{S}_{\mathbf{H}}$ the smoothing matrix whose i-th row is equal to $s_{\mathbf{x}_i,\mathbf{H}}^t$ (the smoother vector for $\mathbf{x} = \mathbf{x}_i$).

On the other hand, the usual nonparametric procedure for estimation of the small-scale structure of the process is carried out from the residuals $\mathbf{r} = (r_1, \ldots, r_n)^t = \mathbf{Y} - \mathbf{S}_{\mathbf{H}}\mathbf{Y}$. A variance estimate is obtained by linear smoothing the squared residuals $\{[\mathbf{x}_i, r_i^2]\}_{i=1}^n$,

$$\hat{\sigma}^2_{\mathbf{r},\mathbf{H}_2}(\mathbf{x}) = s^t_{\mathbf{x},\mathbf{H}_2}\mathbf{r}^2, \tag{4}$$

where $\mathbf{r}^2 = \left(r_1^2, \ldots, r_n^2\right)^t$ and $\mathbf{H}_2$ is the corresponding bandwidth matrix. Likewise, a (pilot) residual semivariogram estimate $\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ is obtained by linear smoothing the sample semivariances

$$\left\{\left(||\mathbf{x}_i - \mathbf{x}_j||, \left[\hat{\varepsilon}(\mathbf{x}_i) - \hat{\varepsilon}(\mathbf{x}_j)\right]^2\right) : 1 \le i < j \le n\right\} \tag{5}$$

of the (estimated) standardized residuals $\hat{\boldsymbol{\varepsilon}} = [\hat{\varepsilon}(\mathbf{x}_1), \ldots, \hat{\varepsilon}(\mathbf{x}_n)]^t = \hat{\mathbf{D}}_0^{-1}\mathbf{r}$, being $\hat{\mathbf{D}}_0 = \operatorname{diag}\left[\hat{\sigma}_{\mathbf{r},\mathbf{H}_2}(\mathbf{x}_1), \ldots, \hat{\sigma}_{\mathbf{r},\mathbf{H}_2}(\mathbf{x}_n)\right]$. In this case, assuming isotropy for simplicity, it would only be necessary to consider a scalar bandwidth parameter h_3.

Bandwidth parameters play an important role in the performance of the previous local linear estimators, since they control the shape and the size of the local neighborhoods used to obtain the corresponding estimates, determining their smoothness. When the data are spatially correlated, as it is assumed in the present paper, we recommend the use of the "bias corrected and estimated" generalized cross-validation (CGCV) criterion, proposed in Francisco-Fernández and Opsomer (2005), to select the matrices $\mathbf{H}$ and $\mathbf{H}_2$, by minimizing

$$\mathrm{CGCV}(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^n \left[\frac{Y(\mathbf{x}_i) - \hat{\mu}_{\mathbf{H}}(\mathbf{x}_i)}{1 - \frac{1}{n}\operatorname{tr}\left(\mathbf{S}_{\mathbf{H}}\hat{\mathbf{R}}\right)}\right]^2$$

and

$$\mathrm{CGCV}(\mathbf{H}_2) = \frac{1}{n}\sum_{i=1}^n \left[\frac{r_i^2 - \hat{\sigma}^2_{\mathbf{r},\mathbf{H}_2}(\mathbf{x}_i)}{1 - \frac{1}{n}\operatorname{tr}\left(\mathbf{S}_{\mathbf{H}_2}\hat{\mathbf{R}}_{\mathbf{r}^2}\right)}\right]^2,$$

respectively, where $\operatorname{tr}(\mathbf{A})$ stands for the trace of a square matrix $\mathbf{A}$, and $\hat{\mathbf{R}}$ and $\hat{\mathbf{R}}_{\mathbf{r}^2}$ are estimates of the correlation matrices of the observations and of the squared residuals, respectively. A simpler approximation of the covariance matrix of the squared residuals $\boldsymbol{\Sigma}_{\mathbf{r}^2}$ can be obtained under the assumptions of normality and zero mean for the residuals. In that case,

$$\boldsymbol{\Sigma}_{\mathbf{r}^2} = 2\boldsymbol{\Sigma}_{\mathbf{r}} \odot \boldsymbol{\Sigma}_{\mathbf{r}},$$

where $\odot$ represents the Hadamard product and $\boldsymbol{\Sigma}_{\mathbf{r}} = Var(\mathbf{r})$ (Ruppert et al. 1997). Finally, the bandwidth parameter h_3 for the computation of the residual semivariogram estimate $\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ can be selected as the minimizer of the cross-validation relative squared error

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[\frac{\left(\hat{\varepsilon}(\mathbf{x}_i) - \hat{\varepsilon}(\mathbf{x}_j)\right)^2}{2\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}^{-(i,j)}\left(||\mathbf{x}_i - \mathbf{x}_j||\right)} - 1\right]^2,$$

where $\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}^{-(i,j)}$ is the estimate obtained when excluding the pair (i, j) in (5).

Nevertheless, it is well known that estimates based on direct use of residuals underestimate the variability of the spatial process:

$$\boldsymbol{\Sigma}_{\mathbf{r}} = \boldsymbol{\Sigma} + \mathbf{S}_{\mathbf{H}}\boldsymbol{\Sigma}\mathbf{S}_{\mathbf{H}}^t - \boldsymbol{\Sigma}\mathbf{S}_{\mathbf{H}}^t - \mathbf{S}_{\mathbf{H}}\boldsymbol{\Sigma}$$

(see, e.g., Cressie 1993, Sect. 3.4.3, for the linear case under homoscedasticity). Equivalently,

$$\begin{aligned}&\mathrm{Var}\,(r_i) = \sigma^2(\mathbf{x}_i)\,(1 + b_{ii})\,,\\ &\mathrm{Var}\left[r_i/\sigma(\mathbf{x}_i) - r_j/\sigma(\mathbf{x}_j)\right] = \mathrm{Var}\left[\varepsilon(\mathbf{x}_i) - \varepsilon(\mathbf{x}_j)\right] + b_{ii} + b_{jj} - 2b_{ij},\end{aligned}$$

where b_{ij} is the (i, j)-th element of

$$\mathbf{B} = \mathbf{D}^{-1}\left(\mathbf{S}_{\mathbf{H}}\boldsymbol{\Sigma}\mathbf{S}_{\mathbf{H}}^t - \boldsymbol{\Sigma}\mathbf{S}_{\mathbf{H}}^t - \mathbf{S}_{\mathbf{H}}\boldsymbol{\Sigma}\right)\mathbf{D}^{-1},$$

a square matrix representing the bias due to the direct use of the residuals. As it may have a significant impact on risk assessment, a slight modification of the iterative procedure proposed in Castillo-Páez et al. (2020) is used to obtain approximately unbiased estimates of the variance $\sigma^2(\cdot)$ and the error variogram $\gamma(\cdot)$. Starting with the residual estimates, $\hat{\sigma}^2_{\mathbf{r},\mathbf{H}_2}(\mathbf{x})$ and $\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$. At each iteration, the bias matrix $\mathbf{B}$ is approximated by $\hat{\mathbf{B}} = \hat{\mathbf{D}}^{-1}(\mathbf{S}_{\mathbf{H}}\hat{\boldsymbol{\Sigma}}\mathbf{S}_{\mathbf{H}}^t - \hat{\boldsymbol{\Sigma}}\mathbf{S}_{\mathbf{H}}^t - \mathbf{S}_{\mathbf{H}}\hat{\boldsymbol{\Sigma}})\hat{\mathbf{D}}^{-1}$. Then, an updated estimate $\hat{\sigma}^2(\cdot)$ is computed by replacing r_i^2 by $r_i^2/(1 + \hat{b}_{ii})$ in (4), and a "corrected" $\hat{\gamma}(\cdot)$ is derived by substituting $\left[\hat{\varepsilon}(\mathbf{x}_i) - \hat{\varepsilon}(\mathbf{x}_j)\right]^2$ for $\left[\hat{\varepsilon}(\mathbf{x}_i) - \hat{\varepsilon}(\mathbf{x}_j)\right]^2 - \hat{b}_{ii} - \hat{b}_{jj} + 2\hat{b}_{ij}$ in (5).

Note that the pilot local linear variogram estimates, $\hat{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ and $\hat{\gamma}(\cdot)$, obtained with the above procedure are not necessarily conditionally negative definite functions and cannot be directly used for prediction or simulation. Valid variogram estimates are obtained by fitting "nonparametric" isotropic Shapiro–Botha models (Shapiro and Botha 1991) to the pilot estimates (see e.g., Fernández-Casal et al. 2017, Sect. 4, for a description of this algorithm), which will be denoted by $\bar{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ and $\bar{\gamma}(\cdot)$, respectively.

3. UNCONDITIONAL AND CONDITIONAL BOOTSTRAP ALGORITHMS

In this section, the bootstrap algorithm to estimate the conditional risk under heteroscedasticity is presented. This method is based on a general conditional simulation method combining unconditional simulations with kriging predictions (see, e.g., Chilès and Delfiner 2012, Sect. 7.3.1). In a first step, the bootstrap algorithm studied in Castillo-Páez et al. (2020), and described below, is used to generate the unconditional replicas.

3.1. Unconditional Bootstrap Algorithm

The present bootstrap algorithm is used to generate unconditional replicates $Y^*_{NS}(\mathbf{x}^e_\alpha)$ at the different estimation locations $\left\{\mathbf{x}^e_\alpha : \alpha = 1, \ldots, n_0\right\}$, following these steps:

1. Using the procedure described in the previous section:
   (a) Obtain $\hat{\mu}_{\mathbf{H}}(\cdot)$, the corresponding residuals $\mathbf{r}$, the initial $\hat{\sigma}^2_{\mathbf{r},\mathbf{H}_2}(\mathbf{x})$ and final $\hat{\sigma}^2(\cdot)$ variance estimates, as well as the initial $\bar{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ and final $\bar{\gamma}(\cdot)$ semivariogram estimates.

   (b) Construct the matrix $\hat{\mathbf{R}}_0$ from the residual variogram $\bar{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ and obtain the Cholesky decomposition $\hat{\mathbf{R}}_0 = \mathbf{L}_0\mathbf{L}_0^t$.

   (c) Compute $\hat{\mathbf{R}}_\alpha$ corresponding to $\mathbf{x}_\alpha^e$ using $\bar{\gamma}(\cdot)$, and $\mathbf{L}_\alpha$ such that $\hat{\mathbf{R}}_\alpha = \mathbf{L}_\alpha\mathbf{L}_\alpha^t$.

   (d) Construct the "uncorrelated" errors $\mathbf{e} = \mathbf{L}_0^{-1}\hat{\mathbf{D}}_0^{-1}\mathbf{r}$ and standardize them.

2. Generate the unconditional bootstrap replicas as follows:
   (a) Obtain independent bootstrap residuals of size n_0 from $\mathbf{e}$, denoted by $\mathbf{e}^*$.

   (b) Compute the unconditional bootstrap residuals $\boldsymbol{\varepsilon}^*_{NC} = \mathbf{L}_\alpha\mathbf{e}^*$.

   (c)) Construct the unconditional bootstrap replicas

$$Y^*_{NC}(\mathbf{x}_\alpha^e) = \hat{\mu}_{\mathbf{H}}(\mathbf{x}_\alpha^e) + \hat{\sigma}(\mathbf{x}_\alpha^e)\boldsymbol{\varepsilon}^*_{NC}(\mathbf{x}_\alpha^e),\ \alpha = 1, \ldots, n_0,$$

   being $\boldsymbol{\varepsilon}^*_{NC}(\mathbf{x}_\alpha^e)$ the α-th component of the vector $\boldsymbol{\varepsilon}^*_{NC}$.

The previous algorithm produces bootstrap replicas that have their mean and variance–covariance matrix equal to the corresponding estimates of the spatial process $Y(\cdot)$ (see, e.g., Cressie 1993, Sect. 3.6.1). However, as these replicas mimic an unconditional realization of the process and their values at the observation positions are random, they will not necessarily match the observed values $\mathbf{Y}$ at the sample positions (for more details, see e.g., Chilès and Delfiner 2012, Chapter 7). Therefore, this algorithm is appropriate for estimating the unconditional risk, $P\left[Y(\mathbf{x}_\alpha^e) \geq c\right]$, but should not be used for the estimation of the conditional risk (2), unless it is modified properly, for example, as shown below.

3.2. Conditional Bootstrap Algorithm

Next, the proposed bootstrap algorithm to estimate the conditional risk (2) is described. The procedure uses unconditional replicas generated with the previous algorithm, although it would not be necessary to obtain replicas of the whole process (Step 2-c above), only of the heteroscedastic errors

$$\delta^*_{NC}(\mathbf{x}_\alpha^e) = \hat{\sigma}(\mathbf{x}_\alpha^e)\varepsilon^*_{NC}(\mathbf{x}_\alpha^e).$$

First of all, we will describe the principle of conditional simulation from a theoretical point of view. For this, we would have to assume that the components of model (1) (the trend, the variance and the variogram) are known, and the true errors $\boldsymbol{\varepsilon} = [\varepsilon(\mathbf{x}_1), \ldots, \varepsilon(\mathbf{x}_n)]^t$ are observed. Obviously, these assumptions are unrealistic in a practical situation. In fact, as described below, the theoretical components will be replaced by their estimates when using these ideas in the bootstrap method. The conditional simulation of the error at a location $\mathbf{x}_\alpha^e$

(see, e.g., Journel 1974) is based on the trivial decomposition

$$\delta(\mathbf{x}^e_\alpha) = \hat{\delta}(\mathbf{x}^e_\alpha) + \left[\delta(\mathbf{x}^e_\alpha) - \hat{\delta}(\mathbf{x}^e_\alpha)\right], \tag{6}$$

where $\hat{\delta}(\mathbf{x}^e_\alpha)$ is the simple kriging prediction at $\mathbf{x}^e_\alpha$ computed from $\boldsymbol{\delta} = [\sigma(\mathbf{x}_1)\varepsilon(\mathbf{x}_1), \ldots, \sigma(\mathbf{x}_n)\varepsilon(\mathbf{x}_n)]^t$. The idea is to substitute the unknown kriging error (the second term on the right-hand side of (6)) by a simulation of this error, obtained from an unconditional simulation $\delta_{NC}(\mathbf{x})$ of the error process. Then, a conditional simulation of this error is:

$$\delta_{CS}(\mathbf{x}^e_\alpha) = \hat{\delta}(\mathbf{x}^e_\alpha) + \left[\delta_{NC}(\mathbf{x}^e_\alpha) - \hat{\delta}_{NC}(\mathbf{x}^e_\alpha)\right], \tag{7}$$

where $\hat{\delta}_{NC}(\mathbf{x}^e_\alpha)$ is the kriging prediction at $\mathbf{x}^e_\alpha$ obtained from the unconditional simulations $\delta_{NC}(\mathbf{x}_i)$, $i = 1, \ldots, n$, at the sample locations. Proceeding in this way, it is easy to verify that $\delta_{CS}(\mathbf{x}_i) = \delta(\mathbf{x}_i)$ and, in the case of simple kriging, $Var\left[\delta_{CS}(\mathbf{x})\right] = \sigma^2(\mathbf{x})$ and $\mathrm{Corr}[\delta_{CS}(\mathbf{x}), \delta_{CS}(\mathbf{x}+\mathbf{u})] = \rho(\mathbf{u})$ (see, e.g., Chilès and Delfiner 2012, Sect. 7.3.1). These properties guarantee that the simulations reproduce the second-order structure of the spatial process (and the complete distribution if, for instance, Gaussian errors are assumed). Note also that the simple kriging predictor is not being used because of its properties as an optimal linear predictor. It is simply a tool to incorporate the conditional covariances, assuming that they are known, in the matrix computations.

In the bootstrap world, the estimates of the model components play the role of the theoretical ones (the trend, the variance, the variogram and the true errors are known) and the previous results can be applied. Taking this into account, the proposed bootstrap algorithm to estimate the conditional risk is as follows:

1. Generate the unconditional bootstrap replicates at the estimation locations $\delta^*_{NC}(\mathbf{x}^e_\alpha)$, $\alpha = 1, \ldots, n_0$, as well as in the sample locations $\delta^*_{NC}(\mathbf{x}_i)$, $i = 1, \ldots, n$.
2. Using simple kriging, obtain the predictions $\hat{\delta}(\mathbf{x}^e_\alpha)$ and $\hat{\delta}^*_{NC}(\mathbf{x}^e_\alpha)$ from the observed residuals $r(\mathbf{x}_i)$ and the unconditional heteroscedastic errors $\delta^*_{NC}(\mathbf{x}_i)$, respectively.
3. Calculate the conditional bootstrap heteroscedastic errors

$$\delta^*_{CS}(\mathbf{x}^e_\alpha) = \hat{\delta}(\mathbf{x}^e_\alpha) + \left[\delta^*_{NC}(\mathbf{x}^e_\alpha) - \hat{\delta}^*_{NC}(\mathbf{x}^e_\alpha)\right].$$

4. Construct the conditional bootstrap replicates $Y^*_{CS}(\mathbf{x}^e_\alpha) = \hat{\mu}_{\mathbf{H}}(\mathbf{x}^e_\alpha) + \delta^*_{CS}(\mathbf{x}^e_\alpha)$.
5. Repeat steps 1 to 4 a large number B of times, to get $Y^{*(1)}_{CS}(\mathbf{x}^e_\alpha), \ldots, Y^{*(B)}_{CS}(\mathbf{x}^e_\alpha)$.
6. Finally, estimate the conditional probability (2) by:

$$\hat{r}_c(\mathbf{x}^e_\alpha, \mathbf{Y}) = \frac{1}{B}\sum_{j=1}^{B} \mathbb{I}\left[Y^{*(j)}_{CS}(\mathbf{x}^e_\alpha) \geq c\right], \tag{8}$$

where $\mathbb{I}(\cdot)$ represents the indicator function.

4. SIMULATION STUDIES

In this section, the heteroscedastic conditional bootstrap procedure described in the previous section is numerically analyzed considering different scenarios. The R (R Development Core Team 2023) package `npsp` (Fernandez-Casal R 2023) was employed to carry out the simulation experiments. In each case, $N = 1000$ samples following the model (1) were generated on regular grids in the unit square of sizes $n_1 = 15 \times 15$, 20×20 and 30×30. The top right diagonal sites were set as the estimation locations $\mathbf{x}^e_\alpha$, $\alpha = 1, \ldots, n_0$, and the remaining ones as the observation sample $\mathbf{x}_i$, $i = 1, \ldots, n$ (note that $n = n_1 - n_0$). For example, Fig. 2a shows the estimation (triangles) and observation (circles) locations for the case of $n_1 = 20 \times 20$.

In order to take into account the effect of the functional form of the components of the model (1), the following theoretical trend and variance functions were considered: $\mu_1(x_1, x_2) = 2.5 + \sin(2\pi x_1) + 4(x_2 - 0.5)^2$ (nonlinear trend; see Fig. 2b), $\mu_2(x_1, x_2) = 5.8(x_1 - x_2 + x_2^2)$ (polynomial trend), $\mu_3(x_1, x_2) = 2$ (constant trend), $\sigma_1^2(x_1, x_2) = (\frac{15}{16})^2[1 - (2x_1 - 1)^2]^2[1 - (2x_1 - 1)^2]^2 + 0.1$ (nonlinear variance; see Fig. 2c), $\sigma_2^2(x_1, x_2) = 0.5(1 + x_1 + x_2)$ (linear variance) and $\sigma_3^2(x_1, x_2) = 1$ (constant variance, i.e., homocedastic case). The random errors $\varepsilon(\mathbf{x})$ were generated through a multivariate normal distribution with zero mean, unit variance and an isotropic Matérn variogram model given by:

$$\gamma(\mathbf{u}) = c_0 + (1 - c_0)\left[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(3\frac{||u||}{a}\right)^\nu \mathcal{K}_\nu\left(3\frac{||u||}{a}\right)\right], \tag{9}$$

where c_0 denotes the nugget ($1 - c_0$ is the partial sill), a is a scale parameter (proportional to the practical range), and $\mathcal{K}_\nu$ is the second kind modified Bessel function of order ν, being ν a smoothness parameter. In order to analyze the effect of the spatial dependence, the following parameters have been considered: $c_0 = 0, 0.2, 0.4, 0.8$, $a = 0.3, 0.6, 0.9$ and $\nu = 0.25, 0.5, 1$. Parameter ν determines the shape of the semivariogram at small lags. For instance, $\gamma(\cdot)$ corresponds to an exponential model when $\nu = 0.5$ (being a its practical range). Figure 2d shows the theoretical semivariograms corresponding to $c_0 = 0.2$, $a = 0.6$ and the different values ν considered.

To apply the conditional bootstrap algorithm described in Sect. 3.2, firstly, it is necessary to estimate the trend $\mu(\cdot)$. For this, we employed the local linear estimator (3), with a multiplicative triweight kernel. To avoid the bandwidth selection effect in the results, the bandwidth $\mathbf{H}_{\text{MASE}}$ minimizing the mean average squared error,

$$\text{MASE}(\mathbf{H}) = \frac{1}{n}(\mathbf{S_H}\boldsymbol{\mu} - \boldsymbol{\mu})^t(\mathbf{S_H}\boldsymbol{\mu} - \boldsymbol{\mu}) + \frac{1}{n}\text{tr}(\mathbf{S_H}\boldsymbol{\Sigma}\mathbf{S}^t_\mathbf{H}),$$

where $\boldsymbol{\mu} = [\mu(\mathbf{x_1}), \ldots, \mu(\mathbf{x_n})]^t$, was employed for trend estimation. An analogous approach was applied to select $\mathbf{H}_2$ for the variance estimation. This also considerably reduced the computing time, as the smoothing matrices $\mathbf{S_H}$ and $\mathbf{S}_{\mathbf{H}_2}$ only needed to be computed once. Note that these optimal bandwidths cannot be used in practice, since their calculations depend on the unknown trend $\mu(\cdot)$ and covariance matrix $\boldsymbol{\Sigma}$. In that case, we recommend

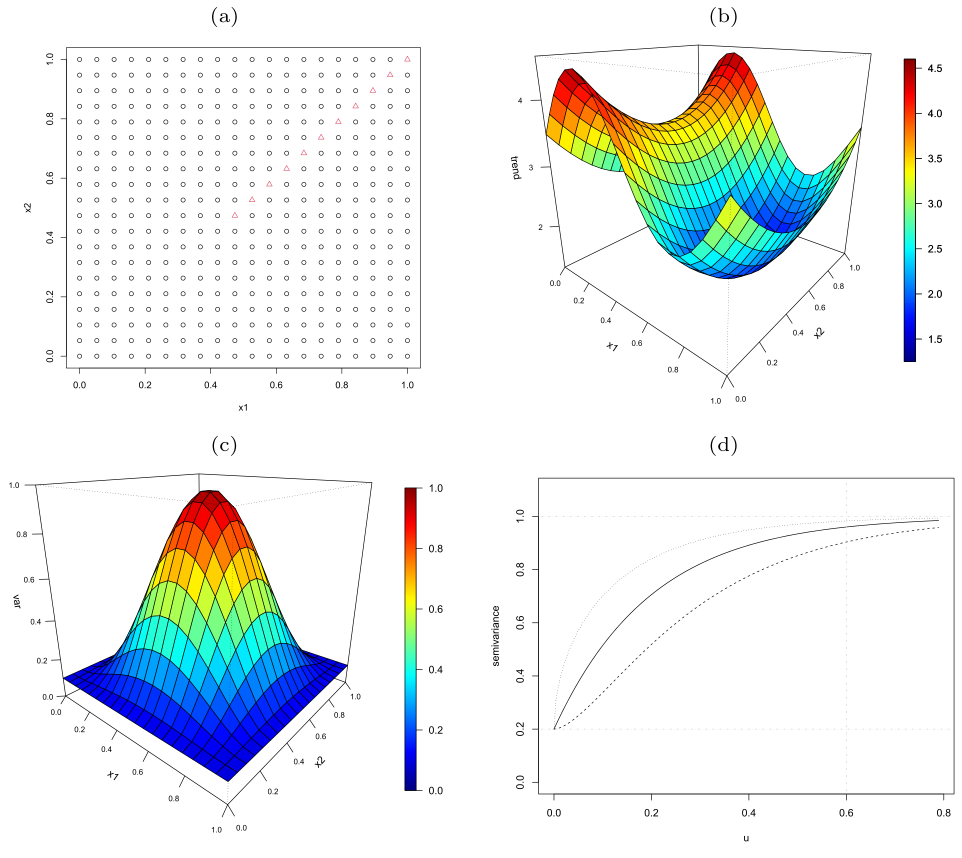

Figure 2. **a** Sample and estimation locations (circles and triangles, respectively) for $n_1 = 20 \times 20$, **b** theoretical nonlinear trend $\mu_1(x_1, x_2)$, **c** theoretical nonlinear variance $\sigma_1^2(x_1, x_2)$, and **d** semivariogram models for $c_0 = 0.2$, $a = 0.6$ and $\nu = 1$ (dashed), $\nu = 0.5$ (solid), $\nu = 0.25$ (dotted line) .

the use of the CGCV criterion (described in Sect. 2) that provided very similar results to those obtained with $\mathbf{H}_{\text{MASE}}$ in some simulation experiments (but it required a much longer computation time because, in addition to the selection of the bandwidths, the corresponding smoothing matrices must be computed in each iteration).

Once that the trend estimate $\hat{\mu}_{\mathbf{H}_{\text{MASE}}}(\mathbf{x})$ is obtained, the iterative algorithm described in Sect. 2 is employed to obtain the final variance estimates $\hat{\sigma}^2(\cdot)$, the residual semivariogram $\bar{\gamma}_{\hat{\boldsymbol{\varepsilon}}}(\cdot)$ and its bias-corrected version $\bar{\gamma}(\cdot)$.

Next, at each simulation the algorithm proposed in Sect. 3 was applied with $B = 1,000$ bootstrap replicas. For each $\alpha = 1, \ldots, n_0$, the conditional probabilities $r_c(\mathbf{x}_\alpha^e, \mathbf{Y})$ were estimated by $\hat{r}_c(\mathbf{x}_\alpha^e, \mathbf{Y})$, given in (8), considering threshold values $c = 2$, 3 and 4. At each estimation location $\mathbf{x}_\alpha^e$, $\alpha = 1, \ldots, n_0$, the squared errors $\left[\hat{r}_c(\mathbf{x}_\alpha^e, \mathbf{Y}) - r_c(\mathbf{x}_\alpha^e, \mathbf{Y})\right]^2$ were computed to evaluate the performance of the proposed procedure.

Note that, taking into account that the responses are normally distributed, the theoretical probabilities $r_c(\mathbf{x}_\alpha^e, \mathbf{Y})$ can be obtained as:

$$1 - \Phi\left[\frac{c - \hat{Y}_{SK}(\mathbf{x}_\alpha^e)}{\hat{\sigma}_{SK}(\mathbf{x}_\alpha^e)}\right],$$

Table 1. Mean, median and standard deviations of the squared errors ($\times 10^{-2}$) of the conditional probability estimates, for μ_1 (nonlinear), σ_1^2 (nonlinear), $c_0 = 0.2$, $a = 0.6$, and $\nu = 0.5$

	$n_1 = 15 \times 15$			$n_1 = 20 \times 20$			$n_1 = 30 \times 30$		
c	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
2	0.35	0.06	0.74	0.29	0.05	0.62	0.21	0.04	0.45
3	0.66	0.03	3.58	0.46	0.02	2.44	0.28	0.01	1.36
4	0.11	0.00	0.73	0.08	0.00	0.73	0.05	0.00	0.39

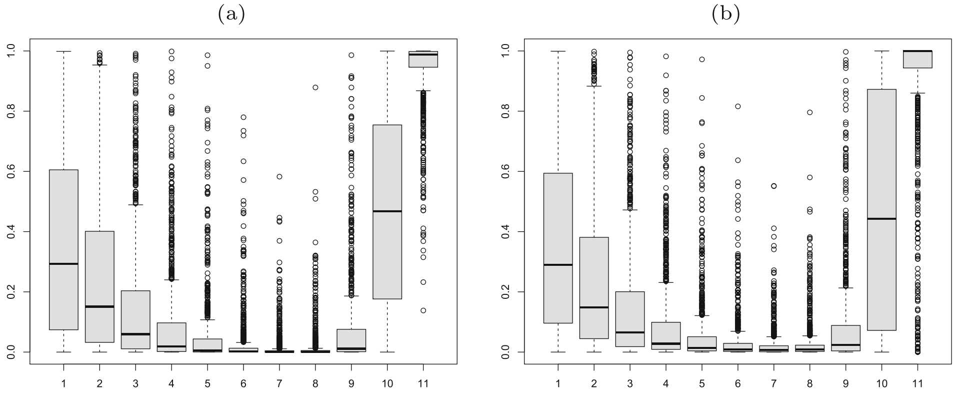

Figure 3. Boxplots of the theoretical (**a**) and estimated (**b**) conditional probabilities of exceeding a threshold of $c = 3$ using the bootstrap method, for μ_1 (nonlinear), σ_1^2 (nonlinear), $n_1 = 20 \times 20$, $c_0 = 0.2$, $a = 0.6$ and $\nu = 0.50$, at the different estimation locations $\mathbf{x}_\alpha^e$, $\alpha = 1, \ldots, 11$.

being Φ the standard normal cumulative distribution function, $\hat{Y}_{SK}(\mathbf{x}_\alpha^e)$ the simple kriging prediction of $Y(\mathbf{x}_\alpha^e)$, obtained using the theoretical trend and covariance matrix, and $\hat{\sigma}_{SK}^2(\mathbf{x}_\alpha^e)$ the corresponding simple kriging variance.

For the sake of brevity, only some representative results are shown here. For example, Table 1 shows the mean, median and standard deviations of the squared errors ($\times 10^{-2}$) of the estimates obtained with the proposed bootstrap approach, for μ_1 (nonlinear), σ_1^2 (nonlinear), $c_0 = 0.2$, $a = 0.6$, $\nu = 0.5$, and the different threshold values and sample sizes considered. It can be observed that, for the different values of c, the mean squared error (MSE) decreases as the sample size n increases, suggesting the consistency of the conditional probability estimator.

The good performance of the proposed approach can also be observed in Fig. 3. It contains boxplots of the theoretical (left panel) and estimated (right panel) conditional probabilities of exceeding a threshold of $c = 3$ at the estimation locations $\mathbf{x}_\alpha^e$, using the proposed method and considering μ_1 (nonlinear), σ_1^2 (nonlinear), $n_1 = 20 \times 20$, $c_0 = 0.2$, $a = 0.6$ and $\nu = 0.5$. A very similar pattern of the corresponding boxplots for the theoretical and the estimated conditional risks at all estimation locations is observed in this figure.

The effect of the spatial dependence was also studied by comparing the results obtained with the different values for a and c_0. In general, an interaction in the effect of these parameters was observed. For example, Table 2 shows that, for a given level of nugget

Table 2. Mean, median and standard deviations of the squared errors ($\times 10^{-2}$) of the conditional probability estimates, for μ_1 (nonlinear), σ_1^2 (nonlinear), $n_1 = 20 \times 20$, $c = 3$ and $\nu = 0.5$

	$a = 0.3$			$a = 0.6$			$a = 0.9$		
c_0	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
0	0.57	0.03	2.39	0.51	0.01	2.60	0.40	0.00	2.34
0.2	0.48	0.03	2.29	0.46	0.02	2.44	0.41	0.01	2.30
0.4	0.43	0.03	2.40	0.43	0.02	2.40	0.42	0.02	2.36
0.8	0.44	0.04	2.44	0.43	0.04	2.47	0.44	0.04	2.48

effect, the error means decrease as the practical range increases. As expected, this effect decreases when the nugget is larger (lower spatial dependence), resulting in similar errors with the different range values.

Additionally, in the case of a stationary process considering $\mu_3(\cdot)$ and $\sigma_3^2(\cdot)$, the IK method, briefly mentioned in the Introduction, was also used to estimate the conditional exceeding probabilities and compared with the method proposed in this paper. In the IK method, only the observed values of the indicator variable $I_{\{Y(\mathbf{x}_0)\geq c\}}$ are considered and, assuming that they are stationary, ordinary kriging is performed to compute predictions at the estimation locations, since

$$P\left[Y(\mathbf{x}_0) \geq c \mid \mathbf{Y}\right] = \mathbb{E}\left[I_{\{Y(\mathbf{x}_0)\geq c\}} \mid \mathbf{Y}\right],$$

(see, e.g., Chilès and Delfiner 2012, Sect. 6.3.3, for further details). In practice, this parametric approach was implemented using the `geoR` package (Ribeiro et al. 2020), fitting an exponential variogram model to the indicator variable (assuming a constant trend). The performance of this method was compared with that obtained by the proposed approach in this scenario. Note that the nonparametric procedure was applied assuming (wrongly) the presence of non-constant trend and variance functions.

As an illustrative example of the comparison results, Table 3 shows a summary of the squared errors ($\times 10^{-2}$) of the estimates obtained with the IK and the proposed approach (denoted by NP in this table), for $n_1 = 20 \times 20$, $a = 0.6$, $c_0 = 0.2$ and different values of ν. In general, the errors obtained by the proposed method are lower than those provided by the IK approach. This could be because of another limitation of the IK method, due to the loss of information by using only the discretized response values (see, e.g., Tolosana-Delgado et al. 2008). Therefore, even when a non-constant trend and variance are incorrectly assumed, the nonparametric proposed method seems to be a better alternative for estimating the conditional probability.

As a final study, the case of irregular sampling was analyzed. The previous scenarios were considered, but now n_1 spatial locations were randomly generated following a bidimensional uniform distribution over the unit square. The same n_0 estimation locations as in the regular sampling design were chosen. In general, very similar results to those achieved under regular design were obtained, although the errors were slightly smaller with the irregular design. For instance, Table 4 shows the MSE for $n_1 = 20 \times 20$, $c_0 = 0.2$, $a = 0.6$, $c = 3$ and

Table 3. Mean, median and standard deviations of the squared errors ($\times 10^{-2}$) obtained with the IK and the proposed (denoted by NP) methods, for $\mu_3(\cdot) = 2$ (constant), $\sigma_3^2(\cdot) = 1$ (homoscedastic), $n_1 = 20 \times 20$, $c_0 = 0.2$, $a = 0.6$ and the different ν values

Method		IK			NP		
ν	c	Mean	Median	SD	Mean	Median	SD
	2	0.88	0.39	1.29	0.22	0.07	0.42
0.25	3	0.60	0.20	1.13	0.12	0.03	0.34
	4	0.13	0.01	0.58	0.02	0.00	0.15
	2	1.07	0.42	1.64	0.23	0.06	0.49
0.5	3	0.73	0.14	1.55	0.13	0.01	0.40
	4	0.16	0.00	0.81	0.03	0.00	0.23
	2	0.85	0.25	1.52	0.17	0.03	0.49
1	3	0.56	0.05	1.44	0.10	0.01	0.36
	4	0.12	0.00	0.70	0.03	0.00	0.29

Table 4. Averaged squared errors ($\times 10^{-2}$) of the conditional probability estimates under irregular sampling, for $n_1 = 20 \times 20$, $c_0 = 0.2$, $a = 0.6$ and $c = 3$

	σ_1^2 (nonlinear)			σ_2^2 (linear)			σ_3^2 (constant)		
ν (Matérn model)	0.25	0.50	1.00	0.25	0.50	1.00	0.25	0.50	1.00
μ_1 (nonlinear)	0.50	0.46	0.39	0.21	0.18	0.14	0.23	0.20	0.16
μ_2 (polynomial)	0.16	0.14	0.10	0.09	0.09	0.09	0.07	0.07	0.08
μ_3 (constant)	0.12	0.13	0.11	0.13	0.15	0.16	0.12	0.13	0.15

the different values of the smoothness parameter ν. As it can be seen, by increasing the smoothness of the process ν the errors tend to be smaller (similar results are observed in Table 3). Additionally, errors tend to be larger when more complex models are considered.

5. APPLICATION TO PRECIPITATION DATA

In order to illustrate the performance in practice of the proposed methodology, the precipitation data set briefly mentioned in the Introduction is considered. This data set is supplied with the R package npsp. The trend and variogram estimates were obtained using the iterative algorithm described in Sect. 2. The final trend and variance function estimates are shown in Fig. 4a, b, respectively. Figure 4c shows the pilot residual variogram $\hat{\gamma}_{\hat{\varepsilon}}(\cdot)$ (circles) and the bias-corrected estimate $\bar{\gamma}(\cdot)$ (solid line). Using these estimates, the kriging predictions were computed (Fig. 4d).

Estimated probability maps were computed for different threshold values (square-root of rainfall inches), $c = \{1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0\}$, by applying the two bootstrap algorithms (unconditional and conditional) described in Sects. 3.1 and 3.2, respectively, with $B = 1000$ bootstrap replicas in each case. For reason of space, only the case of threshold $c = 2.0$ is included here (Fig. 5).

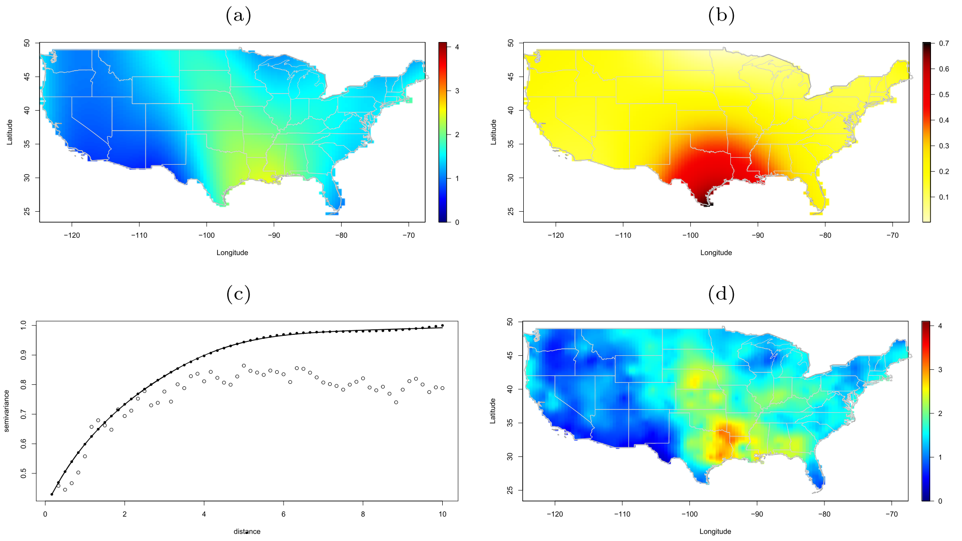

Figure 4. Nonparametric trend estimates (**a**), conditional variance estimates (**b**), semivariogram estimates (**c**), and kriging predictions (**d**) of the precipitation data (in root-squared rainfall inches).

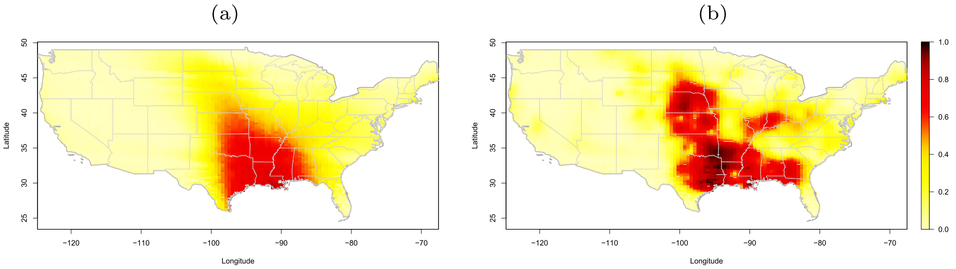

Figure 5. Estimated unconditional (**a**) and conditional (**b**) risk maps for $c = 2.0$.

In Fig. 5, it can be observed that the unconditional bootstrap provides smoother estimates than those obtained with the conditional approach, emphasizing the dominant effect of the trend estimate (large-scale variability). On the other hand, the conditional map shows higher variability, reaching the extreme values 0 and 1 at sample locations. These differences are more evident in the northern regions of the central area of the map, where the proposed method produces estimates in line with the observed values (shown in Fig. 1), due to the stronger effect of the spatial dependence (small-scale variability) on the conditional estimates.

6. DISCUSSION AND FURTHER REMARKS

A bootstrap algorithm to estimate threshold exceeding conditional probabilities under heteroscedasticity of the spatial process is proposed and numerically analyzed in this paper. The probabilities are approximated from bootstrap conditional replicates obtained in a two-stage procedure. In the first step, the unconditional bootstrap method proposed in Castillo-

Páez et al. (2020) is used. In the second step, the unconditional replicates are combined with kriging predictions so that the resulting values coincide with the observed values at the sample locations.

Unlike traditional methods, such as IK or DK, the new approach is designed to be applied for processes that are not stationary (in the mean or in the variance). Moreover, the new approach is fully nonparametric and, therefore, problems due to model misspecification are at least partially avoided. However, to ensure the consistency of the local linear estimator, smooth functions for the trend, variance and variogram are being implicitly assumed. Additionally, a "bias-corrected" method to jointly estimate the variance function and the spatial dependence is employed. In this way, the proposed bootstrap algorithm takes into account that the variability of the residuals is not equal to that of the true errors. Note that although the local linear estimator has been considered in this research due to its good properties, other linear smoothers could also be used.

The complete simulation study shows a good behavior of the new method and its appropriate performance in different scenarios, considering several degrees of spatial dependence and functional forms for the spatial trend and variance. Simulations considering regular and non-regular designs were performed. The results obtained in both frameworks were very similar and analogous conclusions could be deduced from them. For this reason, for the sake of brevity, only some representative scenarios in the case of irregular designs are included in the paper. Note that the case of non-regular design is also illustrated in the real data application in Sect. 5. It is important to remark that in the case of non-regular designs the computational cost of the simulations is much larger, as the optimal bandwidths and the corresponding smoothing matrices, for trend and variance estimation, have to be computed in each iteration (see the comments about bandwidth selection in Sect. 4).

The numerical analysis carried out in this research was performed with the statistical environment R (R Development Core Team 2023), using the functions for nonparametric regression and variogram estimation supplied with the npsp package (Fernandez-Casal R 2023).

ACKNOWLEDGEMENTS

The authors thank the Editor, an Associate Editor and two anonymous referees for numerous useful comments that significantly improved this article.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work has been supported by Grant PID2020-113578RB-I00, funded by MCIN/AEI/10.13039/501100011033. The research of Rubén Fernández-Casal and Mario Francisco-Fernández has also been supported by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020/14) and by CITIC that is supported by Xunta de Galicia, convenio de colaboración entre la Consellería de Cultura, Educación, Formación Profesional e Universidades y las universidades gallegas para el refuerzo de los centros de investigación del Sistema Universitario de Galicia (CIGUS). The research of Sergio Castillo Páez has also been supported by the Universidad de las Fuerzas Armadas ESPE, from Ecuador.

Data Availability The original data are supplied with the R package npsp, as the precipitation data set, and can also be downloaded from http://www.ncdc.noaa.gov/pclcd. The code used to apply the proposed methodology to the precipitation data and the results generated are included in the supplementary files.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Author Contributions All authors contributed equally to this work.

[Received October 2022. Revised May 2023. Accepted May 2023. Published Online July 2023.]

REFERENCES

Castillo-Páez S, Fernández-Casal R, García-Soidán P (2020) Nonparametric bootstrap approach for unconditional risk mapping under heteroscedasticity. Spat Stat 40(100):389

Chilès J, Delfiner P (2012) Geostatistics: modeling spatial uncertainty, 2nd edn. Wiley, New York

Cressie N (1993) Statistics for spatial data. Wiley, New York

Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Chapman & Hall, London

Fan J, Yao Q (1998) Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85(3):645–660

Fernandez-Casal R (2023) npsp: Nonparametric spatial (geo)statistics. R package version 0.7-11, https://rubenfcasal.github.io/npsp

Fernández-Casal R, Castillo-Páez S, García-Soidán P (2017) Nonparametric estimation of the small-scale variability of heteroscedastic spatial processes. Spat Stat 22:358–370

Fernández-Casal R, Castillo-Páez S, Francisco-Fernández M (2018) Nonparametric geostatistical risk mapping. Stoch Environ Res Risk Assess 32(3):675–684

Francisco-Fernández M, Opsomer JD (2005) Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Can J Stat 33:539–558

García-Soidán P, Menezes R (2017) Nonparametric construction of probability maps under local stationarity. Environmetrics 28(3):e2438

Goovaerts P, Webster R, Dubois JP (1997) Assessing the risk of soil contamination in the swiss jura using indicator geostatistics. Environ Ecol Stat 4(1):49–64

Journel AG (1974) Geostatistics for conditional simulation of ore bodies. Econ Geol 69(5):673–687

Lark R, Ferguson R (2004) Mapping risk of soil nutrient deficiency or excess by disjunctive and indicator kriging. Geoderma 118:39–53

Opsomer JD, Wang Y, Yang Y (2001) Nonparametric regression with correlated errors. Stat Sci 16:134–153

R Development Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org

Ribeiro Jr PJ, Diggle PJ, Schlather M, et al (2020) geoR: Analysis of geostatistical data. R package version 1.8-1, https://CRAN.R-project.org/package=geoR

Ruppert D, Wand MP, Holst U et al (1997) Local polynomial variance-function estimation. Technometrics 39(3):262–273

Shapiro A, Botha JD (1991) Variogram fitting with a general class of conditionally nonnegative definite functions. Comput Stat Data Anal 11(1):87–96

Tolosana-Delgado R, Pawlowsky-Glahn V, Egozcue JJ (2008) Indicator kriging without order relation violations. Math Geosci 40(3):327–347

Vilar-Fernández JM, Francisco-Fernández M (2006) Nonparametric estimation of the conditional variance function with correlated errors. J Nonparametr Stat 18(4–6):375–391

Webster R, Oliver M (1989) Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability. J Soil Sci 40(3):497–512

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.