Environmetrics 00, 1–31

DOI: 10.1002/env.XXXX

Nonparametric construction of probability maps under local stationarity

P. García-Soidán[a]* and R. Menezes[b]

Summary: The environmental contamination risk can be evaluated in a specific area by approximating the probability that the pollutant under study exceeds a critical value. This issue requires the estimation of the distribution function involved, which can be addressed by applying the indicator kriging methodology or by approximating the sill of the variogram of the underlying indicator process. These approaches demand an appropriate characterization of the indicator variogram, which in turn requires a previous specification of the trend function, if the latter is suspected to be non-constant. Since accuracy of the results will be strongly dependent on the adequate approximation of both functions, we suggest proceeding in a different way to avoid these requirements. Thus, in the current paper, two kernel-type estimators are proposed, based on first approximating the distribution at the sampled sites and then obtaining a weighted average of the resulting values, to derive a valid estimator at each (sampled or unsampled) location. Consistency of the kernel approaches is proved under rather general conditions, such as local stationarity and the existence of derivatives up to the second order of the distribution function. Numerical studies have been carried out to illustrate the performance of our proposals when compared to those procedures requiring the approximation of the indicator variogram. In a final step, the kernel-type estimation of the distribution function has been applied to map the risk of contamination by arsenic in the Central Region of Portugal. With this aim, biomonitoring data of arsenic concentrations were used to detect those zones with higher risk of arsenic accumulation, which is mainly located on the northern part of the region.

Keywords: Distribution estimation; Kernel method; Stationarity; Trend.

[a] *Faculty of Social Sciences and Communication, University of Vigo, Campus A Xunqueira, Pontevedra, 36005, Spain*
[b] *Center of Mathematics (CMAT), University of Minho, Campus de Azurém, Guimarães, 4800-058, Portugal*
[*] *Correspondence to: P. García-Soidán, Department of Statistics and Operations Research, University of Vigo, Spain. E-mail: pgarcia@uvigo.es*

This paper has been submitted for consideration for publication in *Environmetrics*

1. INTRODUCTION

Assessment of human exposure to toxic elements is increasingly concerning the authorities and the agents responsible for it, since contamination poses a threat to the population health and may lead to the payment of significative fines when surpassing the regulatory thresholds. Arsenic (As) is one of those pollutants under control, as it can cause adverse health effects and has even been linked to cancer (IARC, 2004). The maximum admissible concentrations of As have been established in different regulations, such as the European Directives for drinking water (EC, 1998) or food (EC, 2006; amended in EC, 2015), among others. These are some of the reasons why in practice monitoring data are regularly collected, at a number of spatial sites. The information obtained can be used to estimate the level of As at an unsampled location or to approximate the probability that it exceeds (or does not exceed) a given threshold. In the current study, we will deal with the ultimate goal, which will allow us to construct a probability map of the observation region, showing the distribution function of the pollutant at a fixed maximum or its complementary value, depending on the issue of interest. The probability map in the second case is usually called a risk map in the environmental setting, as it displays the contamination risk, namely, the probability of surpassing a fixed maximum value.

To construct a probability map, the distribution function of a spatial random process $\{Z(\mathrm{s}) : \mathrm{s} \in D \subset \mathbb{R}^d\}$ must be approximated, where $Z(\mathrm{s})$ represents the variable of interest (in this case, As) and D is the observation region. We will write $F_{\mathrm{s}}(x)$ for the distribution function of $Z(\mathrm{s})$ at x, given by:

$$F_{\mathrm{s}}(x) = \mathcal{P}\left(Z(\mathrm{s}) \leq x\right) \tag{1}$$

Then, $1 - F_{\mathrm{s}}(x)$ will denote the probability that the random process exceeds threshold x at location s.

The classical methods applied to estimate the distribution function have originally been designed

for independent data. These procedures are referred to as parametric or nonparametric methods, depending on whether they assume or avoid knowledge of the distribution model, respectively. The main drawback of the former ones is that they require selection of an existing model, which cannot always be supported by the data provided. Under independence, nonparametric methods, such as the empirical distribution or the kernel estimator, enjoy good properties, as proved in Sanov (1961) and Nadaraya (1964), although neither method incorporates the spatial correlation of data.

Other approaches for approximating the distribution function have been specifically designed for spatial data and, therefore, they take into account the underlying dependence structure. Some of these methods are based on estimating the indicator variogram, through the sample variogram, and then deriving the required value by computing the sill (Journel, 1983) or by applying the indicator kriging techniques (Goovaerts, 1997). An alternative is introduced in García-Soidán and Menezes (2012), which suggests using a kernel-type estimator in the first step, as it provides a smoother approximation of the indicator variogram than the sample estimator. In environmental sciences, the indicator kriging is the typical tool employed to establish the probability of exceeding critical and/or regulatory thresholds, using the sample variogram. Examples of some studies derived through the latter approach have been adopted to estimate the risk contamination by As (Hassan and Atkins, 2011; Antunes and Albuquerque, 2013). The application of this methodology can be extended to the assessment of other elements in a variety of settings, such as nitrates (Pardo-Igúzquiza et al., 2015) or other toxic elements (Cinti et al., 2015) in groundwater, the human exposure to dioxins (Augusto et al., 2007) or different airborne pollutants (Finazzi et al., 2013), as well as for the analysis of heavy metal concentrations in soil (Ihl et al., 2015; Reza et al., 2015), among other examples.

A drawback of the aforementioned methods for construction of probability maps is that their accuracy strongly relies on the appropriate specification of the indicator variogram, which characterizes the dependence structure of the indicator process involved. Furthermore, the referred techniques have been designed for stationary processes, although they can be adapted to more

general settings. For instance, if a deterministic trend $\mu(\mathrm{s}) = \mathrm{E}[Z(\mathrm{s})]$ can be assumed from the underlying process, this trend function must be approximated and removed from the data, prior to deriving the distribution estimates. However, care must be taken when proceeding in this way, since these attempts can lead to biased results (Papritz, 2009). On the other hand, when approximation of the entire distribution is needed, the resulting function can be affected by the order relation problem, namely, it may not satisfy the monotonic property of the theoretical distribution, so that $F_{\mathrm{s}}(x) \leq F_{\mathrm{s}}(x')$, whenever $x < x'$. To solve this issue, additional tools (Sullivan, 1984) must be applied to correct the achieved values in order to yield to a non-decreasing function.

In view of the above-mentioned problems, a different strategy to approximate the distribution function will be suggested in the current paper. This task will be accomplished through a nonparametric method and, more specifically, by using the kernel methodology, which has been extensively applied to tackle a variety of problems on random fields. For instance, the approximation of the density or the regression functions has been dealt with the classical kernel approaches in Tran (1990), Hallin et al. (2004) or Carbon et al. (2007), to investigate the properties of these estimators in the spatial setting. Other studies incorporate the spatial dependency to derive new kernel-based proposals, through distinct procedures. If the issue of interest is not referred to any particular site of the observation region, as for characterizing the dependence structure of a stationary random process (Hall et al., 1994), the kernel-type estimator can be obtained in a simple way, by assigning appropriate weights to the information provided by the sampled locations. However, sometimes the problem under study requires deriving an estimator at a specific site. For such a situation, two kernel-based procedures have been applied, which are mainly dependent on whether the value of the random process at the target location is needed. When this observation is not required, as in the derivation of a kernel predictor at a specific site (Menezes et al., 2010), the kernel approach can be designed so as to account for the lags between the target site and the sampled locations. This way of proceeding offers the advantage that it can be applied to random processes departing from the stationarity condition. Other approaches demand using the value of

the process at the specified location and they are solely applicable to the sampled sites, as suggested for density estimation in Dabo-Niang et al. (2014) and even extended to functional data in Ternynck (2014).

Our proposals are focused on approximating the distribution function through kernel-type estimators that take into account the spatial dependence. With this idea, some of the aforementioned strategies have been combined in a two-step procedure, where we first derive kernel-type approaches at the sampled sites and then use them to obtain the distribution estimator at each generic location. Consistency of the resulting kernel estimators will be proved under rather general conditions. Thus, instead of a restrictive (global) stationarity condition, we will simply require local stationarity from the random process, so that close locations follow similar distributions and distant sites tend to present uncorrelated patterns. The existence of second-order derivatives of the distribution function will be also assumed, which is not so demanding. Among the main advantages of the kernel proposals herein presented, we can highlight that neither knowledge of the trend is necessary for their implementation, nor an approximation of the indicator variogram is needed. Furthermore, our approaches overcome the order relation problem, as they yield non-decreasing functions.

A further step in our research is the application of the kernel distribution estimation to determine those zones with higher risk of As accumulation in the Central Region of Portugal, as detection of these hot spots seems crucial, in terms of health prevention. The data set considered for this study was not collected by conventional sampling of the pollutant, using monitoring stations spread over the ecosystem of interest (water, air, sediments, etc.), but from organisms used as biomonitors. In particular, the assessment was made through the moss technique, developed in Sweden in the late 1960s, which is considered a valuable means of identifying sources of airborne pollution (Figueira et al., 2007). It provides biomonitoring data obtained by measuring the concentrations of the element under study (As) that mosses absorb (Ruhling and Steinnes, 1998). An advantage of this sampling method, over the traditional ones, is the larger number of sites that can be included

in the same survey with a smaller cost (Szczepaniak and Biziuk, 2003).

This paper is organized as follows. Section 2 presents the methods proposed for approximation of the distribution function. In Section 2.1, we introduce the new discrete and continuous distribution estimators, whose consistency will be checked in the Appendices. The specification of the bandwidth parameters involved is addressed in Section 2.2. The numerical studies carried out to analyze the behavior of our proposals are described in Section 3. With the new distribution estimators, the risk contamination by As is evaluated in Section 4, from biomonitoring data collected in the Central Region of Portugal. Finally, the main conclusions are summarized in Section 5.

2. METHODS

Let us assume that $\{Z(\mathrm{s}) : \mathrm{s} \in D \subset \mathbb{R}^d\}$ is a spatial random process, which can be modeled as:

$$Z(\mathrm{s}) = \mu(\mathrm{s}) + Y(\mathrm{s}) \tag{2}$$

where $\{Y(\mathrm{s}) \in \mathbb{R} : \mathrm{s} \in D \subset \mathbb{R}^d\}$ is a zero-mean strictly stationary random process and $\mu(\cdot)$ represents the deterministic trend, namely, $\mathrm{E}[Z(\mathrm{s})] = \mu(\mathrm{s})$, for all $\mathrm{s} \in D$. Function $\mu(\cdot)$ is usually unknown and the estimators provided in the current work will not demand a specific characterization of the trend component, unlike other proposals.

Our aim is to estimate the univariate distribution function of $Z(\mathrm{s})$ at threshold x, defined in (1) and denoted by $F_{\mathrm{s}}(x)$, for all $\mathrm{s} \in D$ and $x \in \mathbb{R}$. This problem will be addressed in Section 2.1, where discrete and continuous kernel-type distribution estimators are introduced. Their convergence in probability will be established under similar hypotheses as those required in Hall et al. (1994), related to the following issues:

- A mixed increasing domain asymptotic structure for the sampling design, where the observation region grows to infinity and the distance between neighboring sampling sites

tends to zero.

- An α-mixing condition for the spatial process, to guarantee an adequate decreasing rate of the data correlation, as the distance between locations grows.
- Some assumptions on the kernel functions, as well as on the convergence rates of the bandwidth parameters involved.
- The existence and continuity of second-order derivatives of the m-variate distribution function, for different values of m.

An appropriate choice of the bandwidths is required for implementation of our kernel proposals. In Section 2.2, some guidelines are given for selection of these smoothing parameters.

2.1. Approximating the distribution function of $Z(\mathrm{s})$

Suppose that n observations, $Z(\mathrm{s}_1)$, $Z(\mathrm{s}_2)$, ..., $Z(\mathrm{s}_n)$, have been collected, at spatial locations s_1, s_2, ..., s_n. A first attempt to derive a kernel-type estimator of $F_{\mathrm{s}}(x)$ can lead us to the following weighted average of the indicator functions:

$$\hat{F}_{\mathrm{s},h}(x) = \frac{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) I_{\{Z(\mathrm{s}_i)\leq x\}}}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)} \tag{3}$$

so that the closer s_i is to s, the more weight is assigned to $I_{\{Z(\mathrm{s}_i)\leq x\}}$. Function K represents a d-variate kernel function, h is the bandwidth parameter and I_A denotes the indicator function of the set A. Estimator (3) incorporates the data correlation, by taking into account the lags between sites. However, it fails to produce a consistent approach, since $\hat{F}_{\mathrm{s},h}(x)$ converges in probability to the random variable $I_{\{Z(\mathrm{s})\leq x\}}$, rather than to the theoretical distribution $F_{\mathrm{s}}(x)$. In fact, we can check that the means of $\hat{F}_{\mathrm{s},h}(x) - F_{\mathrm{s}}(x)$ and $\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\leq x\}}$ are asymptotically negligible, while their respective variances tend to $F_{\mathrm{s}}(x) - F_{\mathrm{s}}(x)^2$ and 0, as the sample size increases. A proof of these results is outlined in Appendix A.

In view of the above, we suggest proceeding in an alternative way, by taking advantage of the

available data at the observed sites in the initial step, so as to implement the univariate distribution at them, and then using this information as the basis for the overall estimation. Hence, our proposal will start from approximating the distribution at each sampled location s_i by:

$$\tilde{F}_{1,\mathrm{s}_i,h_1}(x) = \frac{\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) I_{\{Z(\mathrm{s}_j)\leq x\}}}{\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)} \tag{4}$$

Secondly, the resulting terms will be combined in a weighted average, with weights incorporating the spatial dependency, to obtain:

$$\hat{F}_{1,\mathrm{s},h,h_1}(x) = \frac{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\tilde{F}_{1,\mathrm{s}_i,h_1}(x)}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)} = \sum_i \sum_j \frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) I_{\{Z(\mathrm{s}_j)\leq x\}}}{\sum_{i'} K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right)\sum_{j'} K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_{j'})}{h_1}\right)} \tag{5}$$

where K and K_1 denote a d-variate kernel and a univariate kernel functions, respectively, and the bandwidth parameters are represented by h and h_1.

The idea behind the implementation of $\hat{F}_{1,\mathrm{s},h,h_1}(x)$ is to replace each term $I_{\{Z(\mathrm{s}_i)\leq x\}}$ in (3) by a consistent estimator of $F_{\mathrm{s}_i}(x)$. To derive such an estimator at s_i, as defined in (4), we use a weighted average of the indicator functions at the sampled sites, whose weights take into account the differences between each of the observed values and the one at s_i, rather than the distances between the respective sites, unlike estimator (3). Then, $\hat{F}_{1,\mathrm{s},h,h_1}(x)$ in (5) provides an approximation of:

$$\frac{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) F_{\mathrm{s}_i}(x)}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)}$$

which converges in probability to the target value $F_{\mathrm{s}}(x)$, thus solving the original inconsistency that affects $\hat{F}_{\mathrm{s},h}(x)$ in (3). A sketch of this proof is outlined in Appendix B, where we check that the bias and variance of $\hat{F}_{1,\mathrm{s},h,h_1}(x)$ in (5) tend to 0, for large n.

Appropriate bandwidths h and h_1 are required for implementation of estimator (5). Some ideas for selection of both parameters are given in Section 2.2.

Consequently, for construction of the probability map at a threshold x, we suggest proceeding as follows:

- For each sampled site s_i, select the bandwidth h_1 and compute the distribution of $Z(\mathrm{s}_i)$ at x through $\tilde{F}_{1,\mathrm{s}_i,h_1}(x)$.
- Select the target locations $\mathrm{s} \in D$ for approximating the distribution of $Z(\mathrm{s})$ at x.
- For each s, obtain the bandwidth h and estimate $F_{\mathrm{s}}(x)$ by using $\hat{F}_{1,\mathrm{s},h,h_1}(x)$.

We should remark that $\hat{F}_{1,\mathrm{s},h,h_1}$ is a non-decreasing function, as it involves indicator functions satisfying this property, and consequently this proposal avoids the order relation problem. Furthermore, the referred distribution estimator is itself a distribution function, conditional on the sample $\{Z(\mathrm{s}_1),...,Z(\mathrm{s}_n)\}$, which takes values $Z(\mathrm{s}_j)$ with probabilities p_j defined as:

$$p_j = \sum_i \frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)}{\sum_{i'} K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right) \sum_{j'} K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_{j'})}{h_1}\right)}$$

However, since $\hat{F}_{1,\mathrm{s},h,h_1}$ is a discrete distribution function, the use of a smoother version seems to be more appropriate for estimation of a continuous distribution. With this aim, an alternative approach can be derived by applying the integral of a density in (4), rather than an indicator function, and by replacing the resulting estimator for $\tilde{F}_{1,\mathrm{s}_i,h_1}$ in (5). In other words, we could construct a distribution estimator at each sampled site as:

$$\tilde{F}_{2,\mathrm{s}_i,h_1,h_2}(x) = \frac{\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) \mathcal{K}_2\left(\frac{x-Z(\mathrm{s}_j)}{h_2}\right)}{\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)} \tag{6}$$

and then obtain a weighted average of the values achieved:

$$\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x) = \frac{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) \tilde{F}_{2,\mathrm{s}_i,h_1,h_2}(x)}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)} = \sum_i \sum_j \frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) \mathcal{K}_2\left(\frac{x-Z(\mathrm{s}_j)}{h_2}\right)}{\sum_{i'} K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right) \sum_{j'} K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_{j'})}{h_1}\right)} \tag{7}$$

where $\mathcal{K}_2(x) = \int_{-\infty}^{x} K_2(y)dy$, K_2 is a univariate kernel function and h_2 is a new bandwidth parameter.

Consistency holds again for estimator (7), since its bias and variance are asymptotically null. The proof of the latter results, outlined in Appendix C, follows similar arguments as those used with estimator $\hat{F}_{1,\mathrm{s},h,h_1}$. The main advantage of $\hat{F}_{2,\mathrm{s},h,h_1,h_2}$ over $\hat{F}_{1,\mathrm{s},h,h_1}$ is that, by way of construction, the former function is itself a continuous distribution, conditional on the sample, whose associated density is given by:

$$\hat{f}_{2,\mathrm{s},h,h_1,h_2}(x) = \sum_i \sum_j \frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) K_2\left(\frac{x-Z(\mathrm{s}_j)}{h_2}\right)}{\sum_{i'} K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right) \sum_{j'} K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_{j'})}{h_1}\right)}$$

We deal with the selection of the smoothing parameters involved in Section 2.2. Then, the procedure for constructing the probability map at a threshold x, when the continuous distribution estimator is considered, would consist of the following steps:

- For each observed location s_i, select the bandwidth h_1 and use it to obtain h_2, which will allow us to approximate the distribution of $Z(\mathrm{s}_i)$ at x by $\tilde{F}_{2,\mathrm{s}_i,h_1,h_2}(x)$.
- Select the set of sites $\mathrm{s} \in D$, where the distribution of $Z(\mathrm{s})$ at x will be computed.
- For each s, compute the bandwidth h and estimate $F_{\mathrm{s}}(x)$ through $\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)$.

2.2. Guidelines for selection of the bandwidths

Firstly, we address the selection of the smoothing parameters involved in the estimation of the distribution function through (5). Thus, for the selection of h and h_1, we could start by asymptotically minimizing the mean squared error (MSE) or the mean integrated squared error (MISE) of (5). These procedures would yield optimal bandwidths (Liu, 2001), although they would be unknown in practice due to their dependence on the underlying distribution. Hence, we explore other alternatives for the selection of the bandwidth parameters, more easily attainable for a given

data set, such as those based on the cross-validation methods (Hall et al., 1992; Menezes et al., 2010) or on the balloon estimation (Terrell and Scott, 1992; García-Soidán and Menezes, 2012).

The balloon approach consists of taking the bandwidth as the distance from the target value to the k-nearest of the remainder values, for some $k \in \mathbb{N}$ or, equivalently, as the m-th percentile of the distances between the target value and each of the other observations, for some $m \in (0,1)$. This mechanism provides local bandwidths. For implementation of (4) at the sampled site s_i, a bandwidth $h_1 = h_1(\mathrm{s}_i)$ is required, which could be obtained through the balloon approach as the percentile of order $m_1 = m_1(\mathrm{s}_i)$ of the positive values $|Z(\mathrm{s}_i) - Z(\mathrm{s}_j)|$, for all $j \neq i$ and some $m_1 \in (0,1)$. Regarding h, knowledge of which is necessary to compute (5) at location s, the balloon estimator $h = h(\mathrm{s})$ would be given by the percentile of the order $m = m(\mathrm{s})$ of the distances $\|\mathrm{s} - \mathrm{s}_i\|$, for all i and some $m \in (0,1)$.

The cross-validation methodology can give rise to global or local bandwidths, although the latter ones typically demand the implementation of accurate replicates and, therefore, the use of consistent resampling methods. Taking this into account, we focus on the global bandwidths obtained through the classic cross-validation approach, based on the idea of omitting the information at one sampled site and then trying to derive the estimation at that location with the remaining data. This procedure is not applicable to the selection of bandwidth h_1, needed to derive (4), as this estimator cannot be computed at site s_i when $Z(\mathrm{s}_i)$ is left out.

The aforementioned problem, in turn, also affects the selection of the smoothing parameter h. Consequently, we propose to select h_1 through the balloon approach that can then be used to compute $\tilde{F}_{1,\mathrm{s}_i,h_1}(x)$ at each site s_i. At a last stage, the bandwidth h can be taken as the value that minimizes the following expression, so that:

$$h = \mathrm{argmin}_{h \in \mathcal{H}} \left\{ \sum_{i=1}^{n} \left| \hat{F}^{(-i)}_{1,\mathrm{s}_i,h,h_1}(x) - I_{\{Z(\mathrm{s}_i) \leq x\}} \right| \right\}$$

where $\hat{F}^{(-i)}_{1,s_i,h,h_1}(x)$ stands for estimator (5) at s_i, when the i-th location is omitted, namely:

$$\hat{F}^{(-i)}_{1,s_i,h,h_1}(x) = \frac{\sum_{j\neq i} K\left(\frac{s_i - s_j}{h}\right) \tilde{F}_{1,s_j,h_1}(x)}{\sum_{j\neq i} K\left(\frac{s_i - s_j}{h}\right)}$$

and $\mathcal{H}$ is an adequate set of positive numbers, when taking into account the spatial distribution of the sample locations.

When using estimator (7), selection of h and h_1 is also required, as well as the specification of a new smoothing parameter h_2. We first focus on the last bandwidth h_2, for which the optimal choices would be again dependent on unknown terms, so different strategies are proposed. A first attempt to obtain a balloon bandwidth h_2, for a threshold x, would lead to take h_2 as the m_2-percentile of the positive values $|x - Z(s_j)|$, for all j and some $m_2 \in (0,1)$. Nevertheless, this procedure could give rise to an inappropriate null estimate of the distribution at some or all sampled sites s_i through $\tilde{F}_{2,s_i,h_1,h_2}(x)$, since the terms $Z(s_j)$ involved in the choice of h_2 could correspond to null values of $K_1\left(\frac{Z(s_i)-Z(s_j)}{h_1}\right)$. Hence, to guarantee a reliable approximation of the distribution at each observed location s_i, an alternative bandwidth should be considered instead, dependent on s_i and h_1, as well as restricted to the data $Z(s_j)$ for which $K_1\left(\frac{Z(s_i)-Z(s_j)}{h_1}\right) \neq 0$. Thus, we propose using the bandwidth $h_2 = h_2(s_i, h_1)$ obtained as the percentile of the order $m_2 = m_2(s_i, h_1)$ of the resulting positive values $|x - Z(s_j)|$, for some $m_2 \in (0,1)$.

Selection of h_2 by the cross-validation method presents the same problem as that indicated for h_1, since implementation of (6) at site s_i requires the observation $Z(s_i)$. Thus, again we restrict the application of the cross-validation method to the bandwidth h and proceed by a similar approach as that suggested above for deriving the discrete distribution (5). This idea entails choosing the balloon selectors for h_1 and h_2 and using them to obtain $\tilde{F}_{2,s_i,h_1,h_2}(x)$ at each s_i. Thus, the bandwidth h can be taken as the minimizer of:

$$h = \text{argmin}_{h\in\mathcal{H}} \sum_{i=1}^{n} \left| \hat{F}^{(-i)}_{2,s_i,h,h_1,h_2}(x) - I_{\{Z(s_i)\leq x\}} \right| \tag{8}$$

where $\hat{F}^{(-i)}_{2,\mathrm{s}_i,h,h_1,h_2}(x)$ represents the result of (7) at s_i when this site is omitted, given by:

$$\hat{F}^{(-i)}_{2,\mathrm{s}_i,h,h_1,h_2}(x) = \frac{\sum_{j\neq i} K\left(\frac{\mathrm{s}_i-\mathrm{s}_j}{h}\right)\tilde{F}_{2,\mathrm{s}_j,h_1,h_2}(x)}{\sum_{j\neq i} K\left(\frac{\mathrm{s}_i-\mathrm{s}_j}{h}\right)}$$

Remark 2.1 *It is important to notice that $d \times d$ bandwidth matrices could have been considered instead of the single smoothing parameters that we propose for the estimators $\hat{F}_{1,\mathrm{s},h,h_1}$ and $\hat{F}_{2,\mathrm{s},h,h_1,h_2}$. The range of each optimal bandwidth matrix should equal the d-dimension of the observation region and its efficient implementation would demand approximating the corresponding covariance (Liu, 2001), dependent on the distribution function. A simple alternative is given by a d-diagonal bandwidth matrix, where each term in the diagonal would control the amount of smoothing in each direction. Under isotropy, a similar performance of the underlying random process is assumed in all directions, thus making it reasonable to take a diagonal bandwidth matrix with equal terms in the diagonal or, equivalently, to reduce the bandwidth to a single parameter. When isotropy fails, the selection of one bandwidth parameter can be considered as a compromise between efficiency of the estimation and simplicity of the method considered for selection of the bandwidth. Hence, our choice of a single smoothing parameter, which is expected to provide a bandwidth in the range of the selectors that would be derived in the different directions, although this issue requires further research.*

3. NUMERICAL STUDIES

In this section we describe the numerical studies carried out to analyze the performance of the kernel-type estimators of the spatial distribution. Simulated data were used, with the aim of comparing the continuous estimator (7) with the two approximations of the distribution function given in García-Soidán and Menezes (2012). The latter ones are based on the indicator kriging and the sill approaches when using kernel estimators, hereafter referred to as *IK* and *Sill* approaches.

For the sake of simplicity, throughout this section, we will write $\hat{F}_{2,\mathrm{s},h,h_1,h_2} = \hat{F}_2$.

Data were simulated on the unit square $D = [0,1] \times [0,1] \subset \mathbb{R}^2$, assuming a complete spatial randomness design, so that sample locations were uniformly distributed on D. With the spatial locations s_i obtained, for $i = 1, \ldots, n$ and $n = 60$, stationary Gaussian data $Z(\mathrm{s}_i)$ following model (2) were generated, with a linear trend and variance 2.25. To specify the spatial dependency, we initially considered an isotropic exponential variogram, with an asymptotic sill of 2.25, a practical range of 0.9 and a nugget effect equal to 0.36.

3.1. Study 1

Firstly, isotropy was assumed, and balloon estimators were adopted to obtain the different bandwidth estimates, as described in Section 2.2. In particular, h_1 in (6) was approximated as a local bandwidth for each s_i, being constructed as the 20% percentile of the positive values $|Z(\mathrm{s}_i) - Z(\mathrm{s}_j)|$. The bandwidth h_2, also needed in (6), was locally chosen for each x, s_i and h_1, by following the proposed criterion. Thus, we restricted to the observed data $Z(\mathrm{s}_j)$ for which $K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) \neq 0$ and then derived the 10% percentile of the positive values $|x - Z(\mathrm{s}_j)|$. Finally, the bandwidth h was taken to equal the 20% percentile of the total sampled distances.

In the preliminary numerical studies, the central point $s = (0.5, 0.5)$ was taken as the target location to approximate $F_{\mathrm{s}}(x)$. Five thresholds x were selected, given by the quantiles 5%, 25%, 50%, 75% and 95%, as being representative of the distribution domain, which will be respectively denoted by P_k, with $k = 5, 25, 50, 75, 95$ (%). A total of 150 independent data sets were generated and the corresponding mean square error (MSE) was approximated, for each threshold and each procedure. Results are summarized in Table 1, through the values of the mean and the standard deviation derived for the MSE.

[Table 1 about here.]

Table 1 shows that the smallest estimates of the MSE, for both the mean and standard deviation, are achieved by the continuous distribution estimator (7), thus confirming its better performance to

approximate the univariate distribution function of $Z(s)$ than the other methods.

3.2. Study 2

Further goals of our numerical studies include assessing the performance of estimator (7) for distinct variogram models and parameters, evaluating the sensitivity of bandwidth parameters and analyzing the effect of anisotropy. Taking into account the high computational cost of the two estimators given in García-Soidán and Menezes (2012), we decided to restrict the current comparison studies to the *Sill* approach. This way of proceeding allowed us to run a larger number of independent simulations. In particular, the results presented in Table 2 were obtained by generating 500 samples.

To evaluate the sensitivity of the bandwidth parameter h, we first considered the cross-validation method described in section 2.2 and minimized the expression given in (8). As expected, more accurate estimates for (7) were obtained when compared to the foregoing approach of taking a specific percentile of the total sampled distances. So, when the computational cost is acceptable, we would advise adopting the cross-validation method to select h. In the numerical study whose results are presented in Table 2, since many replicates are involved, we followed the pragmatic option of using a percentile of the total sampled distances, which allows for a lower computational cost. The possibility of adopting percentiles 10%, 15%, 20% or 30% was tested and we ended up with percentile 15%, as it originates more accurate estimates.

Model (2) was assumed, where $\mu(\cdot)$ is a linear trend and $Y(\cdot)$ is a zero-mean Gaussian process. Under isotropy, data were simulated from the spherical variogram model, given by $\gamma(\mathrm{t}) = \tau^2 + \sigma^2 - C(\mathrm{t})$, where $C(\mathrm{t} = C_{\sigma^2,\phi}(\|\mathrm{t}\|)$, for all $\mathrm{t} \in I\!R^2$, with:

$$C_{\sigma^2,\phi}(z) = \begin{cases} \sigma^2 \left(1 - \dfrac{3z}{2\phi} + \dfrac{z^3}{2\phi^3}\right), & \text{if } z \leq \phi \\ 0, & \text{if } z > \phi \end{cases}$$

Under anisotropy, instead of the previous covariance function, we considered $C = C_{\sigma^2,\phi,r}$, with:

$$C_{\sigma^2,\phi,r}(\mathrm{t}) = C_{\sigma^2,\phi}\left(\sqrt{t_1^2 + r t_2^2}\right)$$

where $\mathrm{t} = (t_1, t_2) \in \mathbb{R}^2$ and r identifies an anisotropy ratio. In particular, the numerical study was developed by taking r as 0.2, the partial sill σ^2 as 2.25, and the range ϕ as 0.3.

The performance of estimator (7) was assessed at the central point, for different values of the relative nugget effect, by considering a measure of the percentage of the total variability not spatially structured and two distinct values of τ^2, namely, 0.36 and 0.64. Consequently, the relative nugget effects, given by τ^2/σ^2, equaled 0.16 and 0.28, respectively.

[Table 2 about here.]

Table 2 summarizes the results derived for the second numerical study, regarding the mean and the standard deviation estimates of the MSE, based on 500 samples and obtained for two distribution estimators, the *Sill* approach and $\hat{F}_2$ in (7). As in the first study, the selected thresholds were quantiles 5%, 25%, 50%, 75% and 95%. In general, estimator (7) offers more accurate estimates, with smaller means and standard deviations, under both isotropic and anisotropic cases. The exception is observed at threshold P_{95}, although not much difference is exhibited between the two distribution estimators in the distinct scenarios. It is worth noticing the effect of the spatial dependence degree on the results; an increment of it (smaller nugget) yields larger values of the MSE estimates, regardless of the setting considered. Furthermore, the MSE means and variances achieved for $\hat{F}_2$, with the stronger dependent data (nugget equal to 0.36), suffer an increment from the isotropic case to the anisotropic one.

Similar numerical studies were conducted to approximate the distribution function over different points of the observation region $D = [0,1] \times [0,1] \subset \mathbb{R}^2$, including points close to the region borders. Alternative trend models, such as a quadratic trend and a model with covariates, were also tested. Nonetheless, all of them provided similar results to those presented and, therefore, we

decided not to include them.

From the numerical studies derived in Section 3, we can conclude that they support the benefits of the new distribution approach, over the estimators presented in García-Soidán and Menezes (2012). In fact, estimator $\hat{F}_2$ avoids the needs of detrending data and characterizing the dependence structure of the underlying indicator process. As a major advantage of our current proposal, we should emphasize its lower computational effort compared to the other two approaches considered in the numerical studies developed previously.

4. ASSESSMENT OF RISK CONTAMINATION BY AS

In this section, we describe the results obtained when applying the continuous approach (7) to a biomonitoring data set, regarding As levels that were taken in the Central Region of Portugal. The aim is to derive the risk map of the zone as well as estimates of the underlying standard deviations. The sample was collected in the Central Region of Portugal (left panel of Figure 1), classified as NUTS II (NUTS stands for "Nomenclature of Units for Territorial Statistics"). The measured variable represents the concentrations of As in moss samples, in micrograms per gram dry weight.

The use of plants as biomonitors is frequent for ecosystem quality assessment, due to their sensitivity to chemical changes in environmental composition. Other benefits of this use include, among others, low costs, the possibility of long-term sampling, and high availability. Lower plant organisms, like mosses, are often used in the analysis of atmospheric depositions, soil quality and water purity. This measurement system has an additional advantage, as these plants have the capacity to accumulate and store heavy metals and other toxins (Gadzala-Kopciuch et al., 2004).

This particular data set was collected in 2006 and it can be represented by $\{(\mathrm{s}_i, Z(\mathrm{s}_i)), i = 1, ..., n\}$, with $n = 98$ and $Z(\mathrm{s}_i)$ identifying the log-transformed concentration of As at location s_i. We adopted the log-transformation to reduce the impact of outliers. Following the transformation, three evidently gross outliers were assumed as incorrect measurements, so

they were replaced by the average of the remaining values from that year's survey, as suggested in Diggle et al. (2010). Table 3 gives the summary statistics for the resulting data, showing that the log-transformation leads to a more symmetric distribution. Furthermore, Figure 1 illustrates the spatial representation of the log-transformed data, where each bullet size is proportional to the corresponding measured value.

[Table 3 about here.]

To highlight the usefulness of the new kernel estimator within the scope of environmental sciences, or whenever one intends to quantify the risk of some variable indexed in a continuous space exceeding a given threshold, we now proceed with the construction of a probability map, also referred to as a risk map. The new proposal is then applied to the log-transformed As data from Portugal, so that estimates of $\mathcal{P}[Z(\mathrm{s}) > x]$ are calculated on a regular grid of locations s, with 10-km spacing, over the target region (right panel of Figure 1). According to García-Soidán et al. (2014), it is possible to identify an increasing linear trend for these data, when one moves from south to north in NUTS II region. As explained before, the spatial distribution function will be approximated, without requiring the estimation of the trend $\mu(s)$. Our threshold was defined similarly as in Figueira et al. (2007), since neither regulatory critical values for As biomonitoring data have been established, nor correspondence between As concentrations in moss and in other ecosystems was found. In addition, we aimed to determine those areas with higher risk of As accumulation, as being crucial in terms of health prevention, leading us to take the third quartile as the cutoff in the current study, corresponding to $Q_3 = 0.0928$. Hence, we approximated $F_{\mathrm{s}}(Q_3) = \mathcal{P}[Z(\mathrm{s}) \leq Q_3]$ and then plotted the estimate $1 - \hat{F}_{\mathrm{s}}(Q_3)$.

[Figure 1 about here.]

[Figure 2 about here.]

The left panel of Figure 2 displays the pollution risk map of NUTS II region, where the probabilities that As values exceed the quartile Q_3 are represented. The darker colors identify

the high risk areas, mainly located on the northern part of the region and, particularly, close to the western and eastern borders. In addition, accuracy maps of the probability estimates were constructed by applying the bootstrap approach given in García-Soidán et al. (2014). With this idea, we generated 100 replicates of the available data, taking into account the dependence structure of the underlying random process. Then, the probability $\mathcal{P}[Z(\mathrm{s}) > Q_3]$ was approximated for each bootstrap sample and for each location in the regular grid considered. From the total replicates, the standard deviations of the probability estimates were derived, whose values are represented in the right panel of Figure 2. As expected, the lowest standard deviations are associated to the northwest zone, where a dense data set was collected.

5. CONCLUSIONS

Different alternatives have been proposed in the statistics literature to approximate the spatial distribution $F_{\mathrm{s}}(x)$ (or its complementary), mainly based on first characterizing the indicator variogram and then deriving the estimates from the resulting sill or the indicator kriging methodology. These procedures can be directly applied either to stationary data or in the presence of a deterministic trend, although the latter case demands a previous specification of the trend itself. Our kernel-type approaches have been provided to deal with the estimation of the spatial distribution, which require the appropriate selection of distinct bandwidth parameters that is also addressed in the current work. The results obtained enable us to understand the benefits of the new distribution estimators, which do not require an analysis of the dependence structure of the indicator process. When the underlying random process departs from the stationary condition and presents some trend, our studies highlight the advantage of working with the original data instead of the detrended data, in order to achieve more accurate estimates. The new approaches are quite competitive in terms of computational effort and have valuable applications in environmental sciences. Indeed, they allow for the construction of risk maps, a visual tool for assessing

compliance with the environmental quality indicators regulated by the governments, as well as for detecting hot spots.

ACKNOWLEDGMENTS

The authors would like to thank the helpful suggestions and comments from the Editor, the Associate Editor and the Reviewers. The authors are also grateful to Karen J. Duncan for her contribution in the language revision. The first author's work has been partially supported by the Spanish National Research and Development Program project [TEC2015-65353-R], by the European Regional Development Fund (ERDF), and by the Galician Regional Government under project GRC 2015/018 and under agreement for funding AtlantTIC (Atlantic Research Center for Information and Communication Technologies). The second author acknowledges financial support from the Portuguese Funds through FCT-"Fundação para a Ciência e a Tecnologia", within the Project UID/MAT/00013/2013.

REFERENCES

Antunes IMHR, Albuquerque MTD. 2011. Using indicator kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal). *Science of the Total Environment* **442**: 545–552. DOI: 10.1016/j.scitotenv.2012.10.010

Augusto S, Pereira MJ, Soares A, Branquinho C. 2007. The contribution of environmental biomonitoring with lichens to assess human exposure to dioxins. *International Journal of Hygiene and Environmental Health* **210**: 433–438. DOI: 10.1016/j.ijheh.2007.01.017

Carbon M, Francq C, Tran LT. 2007. Kernel regression estimation for random fields. *Journal of Statistical Planning and Inference* **137**: 778–798. DOI: 10.1016/j.jspi.2006.06.008

Cinti D, Poncia PP, Brusca L, Tassi F, Quattrocchia F, Vaselli O. 2015. Spatial distribution of arsenic, uranium and vanadium in the volcanic-sedimentary aquifers of the Vicano-Cimino Volcanic District (Central Italy). *Journal of Geochemical Exploration* **152**: 123–133. DOI: 10.1016/j.gexplo.2015.02.008

Dabo-Niang S, Hamdad L, Ternynck C, Yao AF. 2014. A kernel spatial density estimation allowing for the analysis of spatial clustering. Application to Monsoon Asia Drought Atlas data. *Stochastic Environmental Research and Risk Assessment* **28**: 2075–2099. DOI: 10.1007/s00477-014-0903-6

Diggle P, Menezes R, Ting-Li S. 2010. Geostatistical Inference under Preferential Sampling (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**: 191–232. DOI: 10.1111/j.1467-9876.2009.00701.x

European Commission 1998. Directive 98/83/EC on the quality of water intended for human consumption. Official Journal of the European Communities. L 330: 32–53

European Commission 2006. Directive (EC) 1881/2006 setting maximum levels for certain contaminants in foodstuffs. Official Journal of the European Communities. L 364: 5–24

European Commission 2015. Directive (EU) 2015/1006 amending Regulation (EC) 1881/2006 as regards maximum levels of inorganic arsenic in foodstuffs. Official Journal of the European Communities. L 161: 14–16

Figueira R,Sérgio C, Lopes JL, Sousa AJ. 2007. Detection of exposition risk to arsenic in Portugal assessed by air deposition in biomonitors and water contamination. *International Journal of Hygiene and Environmental Health* **210**: 393–397. DOI: 10.1016/j.ijheh.2007.01.003

Finazzi F, Scott EM, Fasso A. 2013. A model-based framework for air quality indices and population risk evaluation, with an application to the analysis of Scottish air quality data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **62**: 287–308. DOI: 10.1111/rssc.12001

García-Soidán P, Menezes R. 2012. Estimation of the spatial distribution through the kernel indicator variogram. *Environmetrics* **23**: 535–548. DOI: 10.1002/env.2151

García-Soidán P, Menezes R, Rubiños O. 2014. Bootstrap approaches for spatial data. *Stochastic Environmental Research and Risk Assessment* **28**: 1207–1219. DOI: 10.1007/s00477-013-0808-9

Gadzala-Kopciuch R, Berecka B, Bartoszewicz J, Buszewski B. 2004. Some considerations about bioindicators in environmental monitoring. *Polish Journal of Environmental Studies* **13**: 453–462.

Goovaerts P. 1997. *Geostatistics for natural resources evaluation* 1st edition. New York: Oxford University Press.

Hall P, Fisher I, Hoffmann B. 1994. On the nonparametric estimation of covariance functions. *Annals of Statistics* **22**: 2115–2134. DOI: 10.1214/aos/1176325774

Hall P, Marron JS, Park BU. 1992. Smoothed Cross-validation. *Probability Theory and Related Fields* **92**: 1–20. DOI: 10.1007/BF01205233

Hallin M, Lu Z, Tran LT. 2004. Kernel density estimation for spatial processes: The L_1 theory. *Journal of Multivariate Analysis* **88**: 61–75. DOI: 10.1016/S0047-259X(03)00060-5

Hassan MM, Atkins PJ. 2011. Application of geostatistics with Indicator Kriging for analyzing spatial variability of groundwater arsenic concentrations in Southwest Bangladesh. *Journal of environmental science and health. Part A, Toxic/hazardous substances & environmental engineering* **46**: 1185–1196. DOI: 10.1080/10934529.2011.598771

International Agency for Research on Cancer. 2004. *IARC monographs on the evaluation of carcinogenic risks to humans. Some drinking-water disinfectants and contaminants, including arsenic*. Volume 84. Lyon: IARC Press.

Ihl T, Bautista F, Cejudo-Ruíz FR, Delgado MC, Quintana-Owen P, Aguilar D, Goguitchaichvili A. 2015. Concentration of toxic elements in topsoils of the metropolitan area of Mexico City: a spatial analysis using ordinary kriging and indicator kriging. *Revista Internacional de Contaminación Ambiental* **31**: 47–62.

Journel AG. 1983. Nonparametric estimation of spatial distribution. *Mathematical Geology* **15**: 445–468. DOI: 10.1007/BF01031292

Liu XH. 2001. *Kernel smoothing for spatially correlated data*. PhD thesis. Ames, Iowa: Department of Statistics, Iowa State University.

Menezes R, García-Soidán, Ferreira C. 2010. Nonparametric Spatial Prediction under Stochastic Sampling Design. *Journal of Nonparametric Statistics* **22**: 363–377. DOI: 10.1080/10485250903094294

Nadaraya EA. 1964. Some new estimates for distribution functions. *Theory of Probability and its Application* **9**: 497–500. DOI: 10.1137/1109069

Papritz A. 2009. Why indicator kriging should be abandoned, *Pedometron* **26**: 4–7.

Pardo-Igúzquiza E, Chica-Olmo M, Luque-Espinar JA, Rodríguez-Galiano V. 2015. Compositional cokriging formapping the probability risk of groundwater contamination by nitrates. *Science of the Total Environment* **532**: 162–175. DOI: 10.1016/j.scitotenv.2015.06.004

Reza SK, Baruah U, Singh SK, Das TH. 2015. Geostatistical and multivariate analysis of soil heavy metal contamination near coal mining area, Northeastern India. *Environmental Earth Sciences* **73**: 5425–5433. DOI: 10.1007/s12665-014-3797-1

Rühling Å, Steinnes E. 1998. *Atmospheric heavy metal deposition in Europe 1995-1996*. Copenhagen: Nordic Council of Ministers.

Sanov IN. 1961. On the probability of large deviations of random variables. *IMS and AMS Translations of Probability and Statistics* from *Matematicheskii Sbornik* **42**: 11–44.

Sullivan J. 1984. *Conditional recovery estimation through probability kriging: Theory and practice* In Verly G, David M, Journel AG, Marechal A (eds) *Geostatistics for Natural Resources Characterization*, Part I, 365–384. Amsterdam: D. Reidel Publishing Company. DOI: 10.1007/978-94-009-3699-7_22

Szczepaniak K, Biziuk M. 2003. Aspects of the biomonitoring studies using mosses and lichens as indicators of metal pollution. *Environmental Research* **93**: 221–230. DOI: 10.1016/S0013-9351(03)00141-5

Ternynck C. 2014. Spatial regression estimation for functional data with spatial dependency. *Journal de la Société Française de Statistique* **155**: 138–160.

Terrell G, Scott DW. 1992. Variable kernel density estimation. *Annals of Statistics* **20**: 1236–1265. DOI: 10.1214/aos/1176348768

Tran LT. 1990. Kernel density estimation on random fields. *Journal of Multivariate Statistics* **34**: 37–53. DOI: 10.1016/0047-259X(90)90059-Q

APPENDIX

Next, we check the properties of the proposed distribution estimators, where similar arguments as those used in the proof of Theorem 3.1 of Figueira et al. (2007) are followed.

Hereafter, the use of "$\approx$", instead of "=", means that only the dominant part of the second term is specified. In addition, the distribution function and the density function of $(Z(\mathrm{t}_1),...,Z(\mathrm{t}_i))$ at $(x_1,...,x_i)$ are respectively denoted by:

$$F_{\mathrm{t}_1,...,\mathrm{t}_i}(x_1,...,x_i) = \mathcal{P}(Z(\mathrm{t}_1) \leq x_1,...,Z(\mathrm{t}_i) \leq x_i)$$
$$f_{\mathrm{t}_1,...,\mathrm{t}_i}(x_1,...,x_i) = \frac{\partial^i F_{\mathrm{t}_1,...,\mathrm{t}_i}(x_1,...,x_i)}{\partial x_1 \cdot ... \cdot \partial x_i}$$

for $\mathrm{t}_i \in \mathbb{R}^d$, $x_i \in \mathbb{R}$ and $i \in \mathbb{N}$.

A. PROPERTIES OF $\hat{F}_{\mathrm{s},h}(x)$

To derive the dominant terms of the bias and the variance of $\hat{F}_{\mathrm{s},h}(x)$, we assume the following conditions:

(i) $F_{\mathrm{t}_1,\mathrm{t}_2}$ is absolutely continuous for all $\mathrm{t}_i \in \mathbb{R}^d$ and $i = 1,2$.

(ii) $f_{\mathrm{t}_1,\mathrm{t}_2}(x_1,x_2)$ is continuously differentiable as a function of $\mathrm{t}_i \in \mathbb{R}^d$ and $x_i \in \mathbb{R}$, for $i = 1,2$.

(iii) $Z(\cdot)$ is α-mixing, with $\alpha(r) = O(r^{-a})$, for $r > 0$ and some constant $a > 0$.

(iv) $D = \beta D_0$, for some $\beta = \beta_n \overset{n\to+\infty}{\longrightarrow} +\infty$ and bounded $D_0 \subset \mathbb{R}^d$.

(v) $\mathrm{s}_i = \beta \mathrm{u}_i$, for $1 \leq i \leq n$, where u_1, ..., u_n denotes a realization of a random sample of size n drawn from a density function g_0 considered on D_0.

(vi) K is a d-variate density function, which is symmetric and compactly supported.

(vii) $\{h + \beta^{-1} + n^{-2}h^{-d}\beta^d\} \overset{n\to+\infty}{\longrightarrow} 0$.

Firstly, take into account that:

$$\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)\right] = \mathrm{E}\left[\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)\big/ \mathrm{s}_k, \forall k\right]\right] = \sum_i \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) F_{\mathrm{s}_i}(x)}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)}\right] \approx \frac{A_1(\mathrm{s},x)}{A_2(\mathrm{s})} \approx F_{\mathrm{s}}(x)$$

with:

$$A_1(\mathrm{s},x) = \int K\left(\frac{\mathrm{s}-\beta\mathrm{u}}{h}\right) F_{\beta\mathrm{u}}(x) g_0(\mathrm{u})\, d\mathrm{u} \approx h^d \beta^{-d} g_0(0) F_{\mathrm{s}}(x) \int K(\mathrm{v}) d\mathrm{v} = h^d \beta^{-d} g_0(0) F_{\mathrm{s}}(x)$$
$$A_2(\mathrm{s}) = \int K\left(\frac{\mathrm{s}-\beta\mathrm{u}}{h}\right) g_0(\mathrm{u})\, d\mathrm{u} \approx h^d \beta^{-d} g_0(0) \int K(\mathrm{v}) d\mathrm{v} = h^d \beta^{-d} g_0(0)$$

Then, the bias of $\hat{F}_{\mathrm{s},h}(x)$ is asymptotically negligible, namely, $\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)\right] - F_{\mathrm{s}}(x) \approx 0$.

Now, bear in mind that:

$$\begin{aligned}\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)^2\right] &= \mathrm{E}\left[\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)^2\big/ \mathrm{s}_k, \forall k\right]\right] \approx \\ &\approx \sum_i \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)^2 F_{\mathrm{s}_i}(x)}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\right] + \sum_i \sum_{i'} \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right) F_{\mathrm{s}_i,\mathrm{s}_{i'}}(x,x)}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\right] \approx \\ &\approx \frac{B_1(\mathrm{s},x)}{nA_2(\mathrm{s})^2} + \frac{B_2(\mathrm{s},x)}{A_2(\mathrm{s})^2}\end{aligned}$$

where:

$$B_1(\mathrm{s},x) = \int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}}{h}\right)^2 F_{\beta\mathrm{u}}(x) g_0(\mathrm{u})\, d\mathrm{u} \approx g_0(0)\, h^d \beta^{-d} F_{\mathrm{s}}(x) \int K(\mathrm{v})^2 d\mathrm{v}$$

$$B_2(\mathrm{s},x) = \int\int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_1}{h}\right) K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_2}{h}\right) F_{\beta\mathrm{u}_1,\beta\mathrm{u}_2}(x,x) g_0(\mathrm{u}_1)\, g_0(\mathrm{u}_2)\, d\mathrm{u}_1 d\mathrm{u}_2 \approx$$
$$\approx h^{2d} \beta^{-2d} g_0(0)^2 F_{\mathrm{s}}(x)$$

Hence, $\hat{F}_{\mathrm{s},h}(x)$ does not lead to a consistent estimator, because its variance does not necessarily tend to zero as the sample size increases, since:

$$\mathrm{Var}\left[\hat{F}_{\mathrm{s},h}(x)\right] = \mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)^2\right] - \left(\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x)\right]\right)^2 \approx F_{\mathrm{s}}(x) - F_{\mathrm{s}}(x)^2$$

Next, we establish the convergence in probability of $\hat{F}_{\mathrm{s},h}(x)$ to $I_{\{Z(\mathrm{s})\le x\}}$. For this purpose, observe that $\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\le x\}}\right] = \mathrm{Bias}\left[\hat{F}_{\mathrm{s},h}(x)\right] \approx 0$, together with:

$$\mathrm{Var}\left[\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\le x\}}\right] = \mathrm{E}\left[\left(\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\le x\}}\right)^2\right] - \left(\mathrm{E}\left[\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\le x\}}\right]\right)^2 \approx$$
$$\approx \mathrm{E}\left[\mathrm{E}\left[\left(\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\le x\}}\right)^2 \Big/ \mathrm{s}_k, \forall k\right]\right] - 0 \approx$$
$$\approx \textstyle\sum_i \mathrm{E}\left[\dfrac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)^2 \left(F_{\mathrm{s}_i}(x) - 2F_{\mathrm{s}_i,\mathrm{s}}(x,x) + F_{\mathrm{s}}(x)\right)}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\right] +$$
$$+ \textstyle\sum_i \sum_{i'} \mathrm{E}\left[\dfrac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right) \left(F_{\mathrm{s}_i,\mathrm{s}_{i'}}(x,x) - F_{\mathrm{s}_i,\mathrm{s}}(x,x) + F_{\mathrm{s},\mathrm{s}_{i'}}(x,x) - F_{\mathrm{s}}(x)\right)}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\right] \approx$$
$$\approx \frac{B_3(\mathrm{s},x)}{nA_2(\mathrm{s})^2} + \frac{B_4(\mathrm{s},x)}{A_2(\mathrm{s})^2}$$

with:

$$B_3(\mathrm{s},x) = \int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}}{h}\right)^2 \left(F_{\beta\mathrm{u}}(x) - 2F_{\beta\mathrm{u},\mathrm{s}}(x,x) + F_{\mathrm{s}}(x)\right) g_0(\mathrm{u})\, d\mathrm{u} \approx$$
$$\approx g_0(0)\, h^d \beta^{-d} \left(F_{\mathrm{s}}(x) - 2F_{\mathrm{s}}(x) + F_{\mathrm{s}}(x)\right) \int K(\mathrm{v})^2 d\mathrm{v} = o\left(h^d \beta^{-d}\right)$$

$$B_4(\mathrm{s},x) = \int\int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_1}{h}\right) K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_2}{h}\right) \left(F_{\beta\mathrm{u}_1,\beta\mathrm{u}_2}(x,x) - F_{\beta\mathrm{u}_1,\mathrm{s}}(x,x) + F_{\mathrm{s},\beta\mathrm{u}_2}(x,x) - F_{\mathrm{s}}(x)\right) \cdot$$
$$\cdot g_0(\mathrm{u}_1)\, g_0(\mathrm{u}_2)\, d\mathrm{u}_1 d\mathrm{u}_2 \approx g_0(0)^2 h^{2d} \beta^{-2d} \left(F_{\mathrm{s}}(x) - F_{\mathrm{s}}(x) + F_{\mathrm{s}}(x) - F_{\mathrm{s}}(x)\right) =$$
$$= o\left(h^{2d} \beta^{-2d}\right)$$

Hence, $\mathrm{Var}\left[\hat{F}_{\mathrm{s},h}(x) - I_{\{Z(\mathrm{s})\leq x\}}\right] \approx 0$, since $A_2(\mathrm{s})$ is of the exact order $h^d\beta^{-d}$, so we may conclude that $\hat{F}_{\mathrm{s},h}(x) \xrightarrow{P} I_{\{Z(\mathrm{s})\leq x\}}$.

B. PROPERTIES OF $\hat{F}_{1,\mathrm{s},h,h_1}(x)$

We will check that the bias and the variance of $\hat{F}_{\mathrm{s},h,h_1}(x)$ tend to zero as the sample size n increases, which would state the consistency of the distribution estimator, under conditions (i'), (ii'), (iii)-(vi), (vii') and (viii), where:

(i') $F_{\mathrm{t}_1,\mathrm{t}_2,\mathrm{t}_3,\mathrm{t}_4}$ is absolutely continuous for all $\mathrm{t}_i \in \mathbb{R}^d$ and all $i \leq 4$.

(ii') $f_{\mathrm{t}_1,\mathrm{t}_2,\mathrm{t}_3,\mathrm{t}_4}(x_1,x_2,x_3,x_4)$ is continuously differentiable as a function of $\mathrm{t}_i \in \mathbb{R}^d$ and $x_i \in \mathbb{R}$, for all $i \leq 4$.

(vii') K_1 is a univariate density function, which is symmetric and compactly supported.

(viii) $\{h + h_1 + \beta^{-1} + n^{-2}h_1^{-1}h^{-d}\beta^d\} \overset{n\to+\infty}{\longrightarrow} 0$.

Starting with the bias, $\mathrm{Bias}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right] = \mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right] - F_{\mathrm{s}}(x)$. In addition:

$$\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right] = \mathrm{E}\left[\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x) \middle/ \mathrm{s}_k, \forall k\right]\right] =$$
$$= \sum_i \sum_j \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)} \mathrm{E}\left[\frac{K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right) I_{\{Z(\mathrm{s}_j)\leq x\}}}{\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)} \middle/ \mathrm{s}_k, \forall k\right]\right] \approx$$
$$\approx \sum_i \sum_j \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)}{\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)} \frac{C_1(\mathrm{s}_i,\mathrm{s}_j,x)}{nC_2(\mathrm{s}_i,\mathrm{s}_j)}\right] \approx \frac{C_3(\mathrm{s},x)}{C_4(\mathrm{s})} \approx \int I_{\{z\leq x\}} \frac{C_5(\mathrm{s},z)}{C_6(\mathrm{s})} dz$$

with:

$$
\begin{aligned}
C_1\left(\mathrm{s}_i,\mathrm{s}_j,x\right) &= \int\int K_1\left(\tfrac{z_1-z_2}{h_1}\right) I_{\{z_2\le x\}} f_{\mathrm{s}_i,\mathrm{s}_j}(z_1,z_2)dz_1dz_2 \approx h_1\int I_{\{z\le x\}} f_{\mathrm{s}_i,\mathrm{s}_j}(z,z)dz \\
C_2\left(\mathrm{s}_i,\mathrm{s}_j\right) &= \int\int K_1\left(\tfrac{z_1-z_2}{h_1}\right) f_{\mathrm{s}_i,\mathrm{s}_j}(z_1,z_2)z_1dz_2 \approx h_1\int f_{\mathrm{s}_i,\mathrm{s}_j}(z,z)dz \\
C_3(\mathrm{s},x) &= \int\int\int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_1}{h}\right) I_{\{z\le x\}} f_{\beta\mathrm{u}_1,\beta\mathrm{u}_2}(z,z)g_0\left(\mathrm{u}_1\right)g_0\left(\mathrm{u}_2\right)dzd\mathrm{u}_1d\mathrm{u}_2 \approx \\
&\approx h^d\beta^{-d}g_0\left(0\right)\int\int I_{\{z\le x\}} f_{\mathrm{s},\beta\mathrm{u}}(z,z)g_0\left(\mathrm{u}\right)dzd\mathrm{u} = h^d\beta^{-d}g_0\left(0\right)\int I_{\{z\le x\}} C_5(\mathrm{s},z)dz \\
C_4(\mathrm{s}) &= \int\int\int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_1}{h}\right) f_{\mathrm{s},\beta\mathrm{u}_2}(z,z)g_0\left(\mathrm{u}_1\right)g_0\left(\mathrm{u}_2\right)dzd\mathrm{u}_1d\mathrm{u}_2 \approx \\
&\approx h^d\beta^{-d}g_0\left(0\right)\int\int f_{\mathrm{s},\beta\mathrm{u}}(z,z)g_0\left(\mathrm{u}\right)dzd\mathrm{u} = h^d\beta^{-d}g_0\left(0\right)C_6(s) \\
C_5(\mathrm{s},z) &= \int f_{\mathrm{s},\beta\mathrm{u}}(z,z)g_0\left(\mathrm{u}\right)d\mathrm{u} \\
C_6(\mathrm{s}) &= \int\int f_{\mathrm{s},\beta\mathrm{u}}(z,z)g_0\left(\mathrm{u}\right)dzd\mathrm{u}
\end{aligned}
$$

Observe that $C_5(\mathrm{s},z)$ is the bivariate density of $\left(Z\left(\mathrm{s}\right)-Z\left(\beta\mathrm{U}\right),Z\left(\beta\mathrm{U}\right)\right)$ at $(0,z)$ and that $C_6(\mathrm{s})$ is the univariate density of $Z\left(\mathrm{s}\right)-Z\left(\beta\mathrm{U}\right)$ at 0, where U denotes a random variable with density g_0. Then, $\frac{C_5(\mathrm{s},z)}{C_6(\mathrm{s})}$ equals the density of $Z\left(\beta\mathrm{U}\right)$ at z, conditional on $Z\left(\mathrm{s}\right)-Z\left(\beta\mathrm{U}\right)=0$. In other words, $\frac{C_5(\mathrm{s},z)}{C_6(\mathrm{s})}$ is the density of $Z\left(\mathrm{s}\right)$ at z, which leads to:

$$
\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right] \approx \int I_{\{z\le x\}}\frac{C_5(\mathrm{s},z)}{C_6(\mathrm{s})}dz = \int I_{\{z\le x\}} f_{\mathrm{s}}(z)dz = F_{\mathrm{s}}(x) \tag{A.1}
$$

The latter yields the convergence of $\mathrm{Bias}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right]$ to zero.

We now deal with the variance of $\hat{F}_{1,\mathrm{s},h,h_1}(x)$, by considering that:

$$
\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)^2\right] = \mathrm{E}\left[\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)^2\big/\,\mathrm{s}_k,\forall k\right]\right] \approx D_1(\mathrm{s},x)+D_2(\mathrm{s},x)
$$

where:

$$D_1(\mathrm{s},x)=\sum_i\sum_j \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)^2}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\mathrm{E}\left[\frac{K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)^2 I_{\{Z(\mathrm{s}_j)\leq x\}}}{\left(\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)\right)^2}\Big/ \mathrm{s}_k,\forall k\right]\right]$$

$$D_2(\mathrm{s},x)=\sum_i\sum_j\sum_{i'}\sum_{j'}\mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right)}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\cdot\right.$$
$$\left.\cdot\mathrm{E}\left[\frac{K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)K_1\left(\frac{Z(\mathrm{s}_{i'})-Z(\mathrm{s}_{j'})}{h_1}\right)I_{\{Z(\mathrm{s}_j)\leq x\}}I_{\{Z(\mathrm{s}_{j'})\leq x\}}}{\sum_j K_1\left(\frac{Z(\mathrm{s}_i)-Z(\mathrm{s}_j)}{h_1}\right)\sum_{j'}K_1\left(\frac{Z(\mathrm{s}_{i'})-Z(\mathrm{s}_{j'})}{h_1}\right)}\Big/ \mathrm{s}_k,\forall k\right]\right]$$

Similar arguments would allow us to check that:

$$D_1(\mathrm{s},x)\approx\sum_i\sum_j\mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)^2}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2}\frac{D_3\left(\mathrm{s}_i,\mathrm{s}_j,x\right)}{n^2C_2\left(\mathrm{s}_i,\mathrm{s}_j\right)^2}\right]\approx\frac{D_4(\mathrm{s},x)}{n^2h_1^2C_4(\mathrm{s})^2}$$

with:

$$D_3\left(\mathrm{s}_i,\mathrm{s}_j,x\right)=\int\int K_1\left(\tfrac{z_1-z_2}{h_1}\right)^2 I_{\{z_2\leq x\}}f_{\mathrm{s}_i,\mathrm{s}_j}(z_1,z_2)dz_1dz_2\approx$$
$$\approx h_1\int I_{\{z\leq x\}}f_{\mathrm{s}_i,\mathrm{s}_j}(z,z)dz$$
$$D_4(\mathrm{s},x)=h_1\int\int\int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_1}{h}\right)^2 I_{\{z\leq x\}}f_{\beta\mathrm{u}_1,\beta\mathrm{u}_2}(z,z)g_0\left(\mathrm{u}_1\right)g_0\left(\mathrm{u}_2\right)dzd\mathrm{u}_1d\mathrm{u}_2\approx$$
$$\approx n^2h_1h^d\beta^{-d}g_0\left(0\right)\int K(\mathrm{v})^2d\mathrm{v}\int\int I_{\{z\leq x\}}f_{\mathrm{s},\beta\mathrm{u}}(z,z)g_0\left(\mathrm{u}\right)dzd\mathrm{u}$$

In consequence:

$$D_1(\mathrm{s},x)\approx\frac{D_4(\mathrm{s},x)}{n^2h_1^2C_4(\mathrm{s})^2}=\frac{h_1h^d\beta^{-d}g_0\left(0\right)\int K(\mathrm{v})^2d\mathrm{v}\int I_{\{z\leq x\}}C_5(\mathrm{s},z)dz}{n^2h_1^2h^{2d}g_0\left(0\right)^2\beta^{-2d}C_6(\mathrm{s})}=$$
$$=\frac{\int K(\mathrm{v})^2d\mathrm{v}\int I_{\{z\leq x\}}f_{\mathrm{s}}(z)dz}{n^2h_1h^d\beta^{-d}g_0\left(0\right)}=\frac{F_{\mathrm{s}}(x)\int K(\mathrm{v})^2d\mathrm{v}}{n^2h_1h^d\beta^{-d}g_0\left(0\right)C_6(\mathrm{s})}=O\left(n^{-2}h_1^{-1}h^{-d}\beta^d\right)$$

Finally, we focus on the approximation of $D_2(\mathrm{s},x)$, by proceeding as above to obtain:

$$D_2(\mathrm{s},x) \approx \sum_i \sum_j \sum_{i'} \sum_{j'} \mathrm{E}\left[\frac{K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right) K\left(\frac{\mathrm{s}-\mathrm{s}_{i'}}{h}\right)}{\left(\sum_i K\left(\frac{\mathrm{s}-\mathrm{s}_i}{h}\right)\right)^2} \frac{D_5\left(\mathrm{s}_i,\mathrm{s}_j,\mathrm{s}_{i'},\mathrm{s}_{j'},x\right)}{n^2 C_2\left(\mathrm{s}_i,\mathrm{s}_j\right) C_2\left(\mathrm{s}_{i'},\mathrm{s}_{j'}\right)}\right] \approx$$
$$\approx \frac{D_6(\mathrm{s},x)}{C_4(\mathrm{s})^2} \approx \frac{h^{2d}\beta^{-2d} g_0(0)^2 \left(D_7(\mathrm{s},x)+D_8(\mathrm{s},x)\right)}{C_4(\mathrm{s})^2}$$

with:

$$D_5\left(\mathrm{s}_i,\mathrm{s}_j,\mathrm{s}_{i'},\mathrm{s}_{j'},x\right) = \int\int\int\int K_1\left(\tfrac{z_1-z_2}{h_1}\right) K_1\left(\tfrac{z_3-z_4}{h_1}\right) I_{\{z_2\le x\}} I_{\{z_4\le x\}} \cdot$$
$$\cdot f_{\mathrm{s}_i,\mathrm{s}_j,\mathrm{s}_{i'},\mathrm{s}_{j'}}\left(z_1,z_2,z_3,z_4\right) dz_1 dz_2 dz_3 dz_4 \approx h_1^2 \int\int I_{\{z\le x\}} I_{\{z'\le x\}} f_{\mathrm{s}_i,\mathrm{s}_j,\mathrm{s}_{i'},\mathrm{s}_{j'}}\left(z,z,z',z'\right) dz dz'$$
$$D_6(\mathrm{s},x) = \int\int\int\int\int K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_1}{h}\right) K\left(\tfrac{\mathrm{s}-\beta\mathrm{u}_2}{h}\right) I_{\{z\le x\}} I_{\{z'\le x\}} \cdot$$
$$\cdot f_{\beta\mathrm{u}_1,\beta\mathrm{u}_3,\beta\mathrm{u}_2,\beta\mathrm{u}_4}(z,z,z',z') g_0\left(\mathrm{u}_1\right) g_0\left(\mathrm{u}_2\right) dz dz' d\mathrm{u}_1 d\mathrm{u}_2 g_0\left(\mathrm{u}_3\right) g_0\left(\mathrm{u}_4\right) d\mathrm{u}_3 d\mathrm{u}_4$$
$$D_7(\mathrm{s},x) = \int_S \left(\int\int I_{\{z\le x\}} I_{\{z'\le x\}} f_{\mathrm{s},\beta\mathrm{u},\mathrm{s},\beta\mathrm{u}'}(z,z,z',z') dz dz'\right) g_0\left(\mathrm{u}\right) g_0\left(\mathrm{u}'\right) d\mathrm{u} d\mathrm{u}'$$
$$D_8(\mathrm{s},x) = \int_{S^C} \left(\int\int I_{\{z\le x\}} I_{\{z'\le x\}} f_{\mathrm{s},\beta\mathrm{u},\mathrm{s},\beta\mathrm{u}'}(z,z,z',z') dz dz'\right) g_0\left(\mathrm{u}\right) g_0\left(\mathrm{u}'\right) d\mathrm{u} d\mathrm{u}'$$

where $S = \{(\mathrm{u},\mathrm{u}') \in I\!R^{2d} : \|\mathrm{u}-\mathrm{u}'\| \le \beta^{-1/2}\}$.

From the definition of S, it is easy to see that $D_7(\mathrm{s},x) = O\left(\beta^{-d/2}\right)$. On the other hand:

$$f_{\mathrm{s},\beta\mathrm{u},\mathrm{s},\beta\mathrm{u}'}(z,z,z',z') = D_9\left(\mathrm{s},\mathrm{u},\mathrm{u}',z,z'\right) D_{10}\left(\mathrm{s},\mathrm{u}\right) D_{10}\left(\mathrm{s},\mathrm{u}'\right) \tag{A.2}$$

where $D_9\left(\mathrm{s},\mathrm{u},\mathrm{u}',z,z'\right)$ denotes the density of $\left(Z\left(\beta\mathrm{u}\right), Z\left(\beta\mathrm{u}'\right)\right)$, conditional on $Z\left(\mathrm{s}\right) - Z\left(\beta\mathrm{u}\right) = 0$ and $Z\left(\mathrm{s}\right) - Z\left(\beta\mathrm{u}'\right) = 0$, at (z,z') and $D_{10}\left(\mathrm{s},\mathrm{u}\right)$ represents the density of $Z\left(\mathrm{s}\right) - Z\left(\beta\mathrm{u}\right)$ at 0.

Now, for $(\mathrm{u},\mathrm{u}') \in S^C$, one has that $\beta\|\mathrm{u}-\mathrm{u}'\| > \beta^{1/2}$. Hence, from hypothesis (iii), the random variables $Z\left(\beta\mathrm{u}\right)$ and $Z\left(\beta\mathrm{u}'\right)$ are asymptotically uncorrelated to yield that:

$$D_9\left(\mathrm{s},\mathrm{u},\mathrm{u}',z,z'\right) \approx D_{11}\left(\mathrm{s},\mathrm{u},z\right) D_{11}\left(\mathrm{s},\mathrm{u}',z'\right) \tag{A.3}$$

where $D_{11}(\mathrm{s},\mathrm{u},z)$ equals the density of $Z(\beta\mathrm{u})$, conditional on $Z(\mathrm{s})-Z(\beta\mathrm{u})=0$, at z. Then, from relations (A.2) and (A.3), $f_{\mathrm{s},\beta\mathrm{u},\mathrm{s},\beta\mathrm{u}'}(z,z,z',z') \approx f_{\mathrm{s},\beta\mathrm{u}}(z,z) f_{\mathrm{s},\beta\mathrm{u}'}(z',z')$, for $(\mathrm{u},\mathrm{u}') \in S^C$.

In view of the latter, it follows that:

$$D_2(\mathrm{s},x) \approx \frac{h^{2d}\beta^{-2d} g_0(0)^2 \left(D_7(\mathrm{s},x)+D_8(\mathrm{s},x)\right)}{C_4(\mathrm{s})^2} = \frac{h^{2d}\beta^{-2d} g_0(0)^2 D_8(\mathrm{s},x)}{C_4(\mathrm{s})^2} + O\left(\beta^{-d/2}\right) \approx$$
$$\approx \frac{h^{2d}\beta^{-2d} g_0(0)^2 \int_{S^C} \left(\int\int I_{\{z\le x\}} I_{\{z'\le x\}} f_{\mathrm{s},\beta\mathrm{u}}(z,z) f_{\mathrm{s},\beta\mathrm{u}'}(z',z') dz dz'\right) g_0(\mathrm{u})\, g_0(\mathrm{u}')\, d\mathrm{u} d\mathrm{u}'}{h^{2d}\beta^{-2d} g_0(0)^2 C_6(s)^2} \approx$$
$$\approx \int\int I_{\{z\le x\}} I_{\{z'\le x\}} \frac{C_5(\mathrm{s},z)}{C_6(\mathrm{s})} \frac{C_5(\mathrm{s},z')}{C_6(\mathrm{s})} dz dz' = \left(\int I_{\{z\le x\}} f_{\mathrm{s}}(z) dz\right)^2 = F_{\mathrm{s}}(x)^2$$

Therefore, $\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)^2\right] \approx D_1(\mathrm{s},x)+D_2(\mathrm{s},x) \approx F_{\mathrm{s}}(x)^2$, which leads us to conclude that:

$$\mathrm{Var}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right] = \mathrm{E}\left[\left(\hat{F}_{1,\mathrm{s},h,h_1}(x)\right)^2\right] - \left(\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)\right]\right)^2 \approx 0$$

on account of (A.1).

C. PROPERTIES OF $\hat{F}_{2,\mathrm{s},h,h_2}(x)$

We will give just a sketch of the procedure to derive the dominant terms of the bias and the variance of $\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)$, which requires assuming hypotheses (i'), (ii'), (iii)-(vi), (vi'), (vii'), (viii') and (ix), where:

(viii') K_2 is a univariate density, which is symmetric and compactly supported.

(ix) $\{h+h_1+h_2+\beta^{-1}+n^{-2}h_1^{-1}h^{-d}\beta^d\} \overset{n\to+\infty}{\longrightarrow} 0$.

Take into account that:

$$\int \mathcal{K}_2\left(\tfrac{x-z}{h_2}\right) f_{\mathrm{s}}(z) dz = \tfrac{1}{h_2}\int K_2\left(\tfrac{x-z}{h_2}\right) F_{\mathrm{s}}(z) dz = \int K_2(y) F_{\mathrm{s}}(x-h_2 y)\, dy \approx F_{\mathrm{s}}(x) \qquad \text{(A.4)}$$

From (A.4), we could proceed as in section B to obtain the analogue of (A.1), adapted to this setting, to yield:

$$\mathrm{E}\left[\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)\right] \approx \int \mathcal{K}_2\left(\tfrac{x-z}{h_2}\right) f_{\mathrm{s}}(z)dz \approx F_{\mathrm{s}}(x)$$

so that the bias of $\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)$ is asymptotically negligible.

On the other hand, relation (A.4) together with the application of similar arguments as those used to derive $\mathrm{E}\left[\hat{F}_{1,\mathrm{s},h,h_1}(x)^2\right]$, in terms of $D_1(\mathrm{s},x)$ and $D_2(\mathrm{s},x)$, lead to:

$$\mathrm{E}\left[\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)^2\right] \approx E_1(\mathrm{s},x) + E_2(\mathrm{s},x)$$

with:

$$E_1(\mathrm{s},x) = \frac{\int K(\mathrm{v})^2 d\mathrm{v} \int \mathcal{K}_2\left(\tfrac{x-z}{h_2}\right) f_{\mathrm{s}}(z)dz}{n^2 h_1 h^d \beta^{-d} g_0(0)} \approx \frac{F_{\mathrm{s}}(x)\int K(\mathrm{v})^2 d\mathrm{v}}{n^2 h_1 h^d \beta^{-d} g_0(0)} = O\left(n^{-2}h_1^{-1}h^{-d}\beta^d\right)$$
$$E_2(\mathrm{s},x) = \left(\int \mathcal{K}_2\left(\tfrac{x-z}{h_2}\right) f_{\mathrm{s}}(z)dz\right)^2 \approx F_{\mathrm{s}}(x)^2$$

In consequence, $\mathrm{Var}\left[\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)\right] = \mathrm{E}\left[\left(\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)\right)^2\right] - \left(\mathrm{E}\left[\hat{F}_{2,\mathrm{s},h,h_1,h_2}(x)\right]\right)^2 \approx 0.$

FIGURES

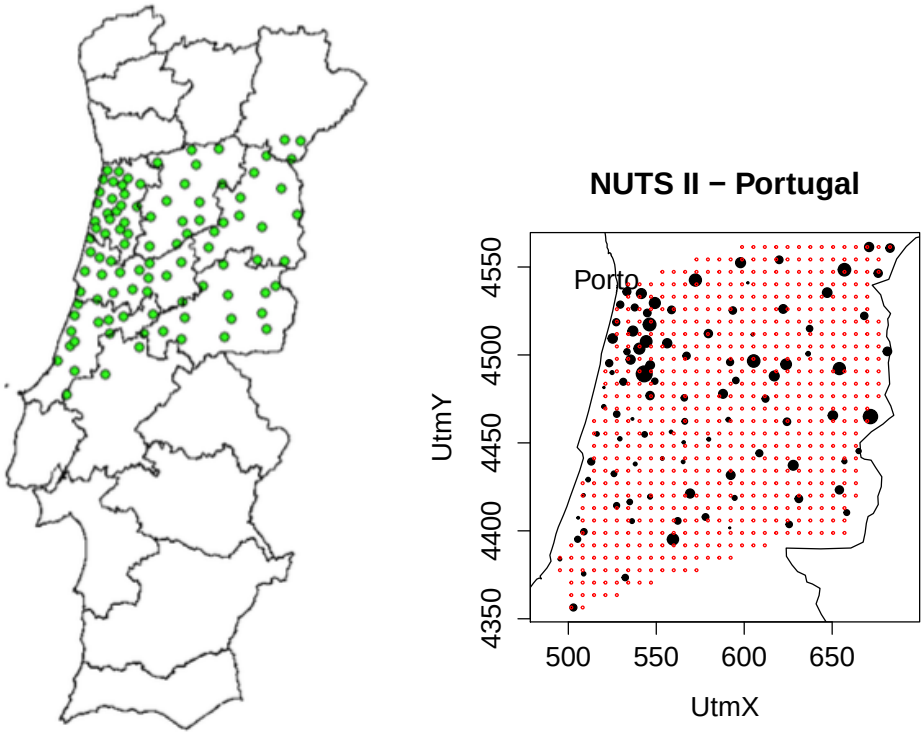

Figure 1. The left panel displays the spatial representation of moss locations in the Central Region of Portugal (NUTS II). In the right panel, the black bullets also represent the sampled locations, whose size is proportional to the measured values, while the red bullets identify the regular grid considered for construction of the maps presented in Figure 2.

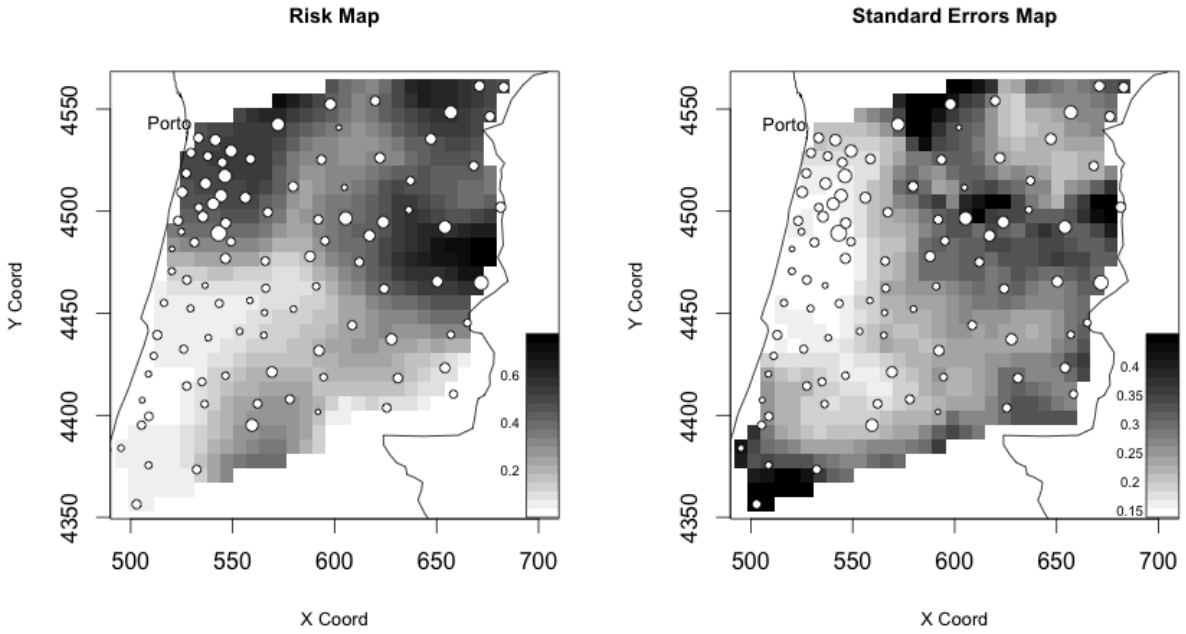

Figure 2. The left panel presents the risk map derived for As pollution data in the NUTS II region, providing the probabilities $\mathcal{P}[Z(s) > Q_3]$. The right panel shows the standard deviations of the estimated probabilities that were acquired from 100 bootstrap replicates.

TABLES

Table 1. Mean and standard deviation (in brackets) estimates of the MSE, based on 150 samples and obtained for three distribution estimators, *IK* approach, *Sill* approach and $\hat{F}_2$ in (7). The selected thresholds are quantiles 5%, 25%, 50%, 75% and 95%. All values were multiplied by 10^2. Data were generated considering an isotropic exponential model for the spatial dependency. The approximation of the distribution function was done over $\mathrm{s} = (0.5, 0.5)$.

Distrib. estim.	P_5	P_{25}	P_{50}	P_{75}	P_{95}
IK	1.52 (2.81)	3.70 (4.39)	3.07 (4.19)	3.87 (4.37)	1.42 (2.53)
Sill	0.24 (0.43)	2.53 (2.59)	2.54 (2.71)	2.83 (3.02)	0.31 (0.33)
$\hat{F}_2$ in (7)	0.20 (0.37)	0.54 (0.89)	1.06 (1.30)	0.59 (0.89)	0.12 (0.27)

Table 2. Mean and standard deviation (in brackets) estimates of the MSE, based on 500 samples and obtained for two distribution estimators, *Sill* approach and $\hat{F}_2$ in (7). The selected thresholds are quantiles 5%, 25%, 50%, 75% and 95%. All values were multiplied by 10^2. Data were generated considering an isotropic or an anisotropic spherical model for the spatial dependency. Two different values for the nugget effect (τ^2) were considered, namely 0.36 and 0.64, fixing the partial sill (σ^2) as 2.25.

Isotropy						
Distr.estim.	τ^2/σ^2	P_5	P_{25}	P_{50}	P_{75}	P_{95}
Sill	0.16	0.07 (0.14)	4.46 (7.43)	9.17 (9.98)	3.49 (2.34)	0.21 (0.47)
$\hat{F}_2$ in (7)	0.16	0.06 (0.09)	2.66 (4.30)	4.36 (5.27)	1.85 (2.47)	0.43 (0.29)
Sill	0.28	0.07 (0.14)	4.07 (6.66)	8.23 (9.52)	3.11 (2.08)	0.15 (0.29)
$\hat{F}_2$ in (7)	0.28	0.07 (0.11)	3.17 (4.42)	5.19 (6.58)	1.71 (2.1)	0.47 (0.32)
Anisotropy						
Distr.estim.	τ^2/σ^2	P_5	P_{25}	P_{50}	P_{75}	P_{95}
Sill	0.16	0.12 (0.28)	4.14 (5.89)	9.17 (9.97)	3.45 (2.34)	0.18 (0.35)
$\hat{F}_2$ in (7)	0.16	0.12 (0.28)	3.25 (4.87)	5.07 (5.87)	2.15 (2.79)	0.51 (0.38)
Sill	0.28	0.06 (0.12)	3.43 (4.34)	8.07 (9.04)	2.97 (1.98)	0.15 (0.33)
$\hat{F}_2$ in (7)	0.28	0.06 (0.11)	2.22 (3.60)	4.69 (6.21)	2.07 (2.81)	0.56 (0.42)

Table 3. Summary statistics for As pollution levels measured in the Central Region of Portugal (NUTS II).

Type of data	Mean	Median	St. dev.	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