Convergence in distribution of the L_2-deviations of the kernel-type variogram estimators with applications

Pilar García-Soidán[a,*], Tomás R. Cotos-Yáñez[a]

[a] *Dept. of Statistics and Operations Research, University of Vigo (Spain)*

Abstract

In this work, some properties of the L_2-deviations of the Nadaraya-Watson variogram estimators are analyzed, for both the anisotropic and the isotropic settings. Their convergence in distribution is established, which provides the basis for addressing practical problems, such as the construction of goodness of fit tests for the variogram and, therefore, for modeling the spatial dependence. However, the development of the latter application requires solving different issues, such as the approximation of the model parameters and the critical points. For estimation of the former ones, we propose proceeding through the least squares criteria, whose consistency will be proved, together with a reformulation of the global measures for the kernel-type estimators. Then, the resulting critical points can be approximated by appealing to the bootstrap approaches. Numerical studies with simulated and real data have been developed to illustrate the potentiality of our results, in order to check the appropriateness of a variogram model selected for the variogram.

Keywords: Global measure, goodness of fit test, intrinsic stationarity, variogram, kernel estimation

2010 MSC: 62G05, 62G10

*Corresponding author at Faculty of Social Sciences and Communication, CP 36005 Pontevedra, Spain

Email addresses: pgarcia@uvigo.es (Pilar García-Soidán), cotos@euvigo.es (Tomás R. Cotos-Yáñez)

1. Introduction

The kriging techniques allow the researchers to reconstruct a phenomenon over the whole observation region, from a finite set of data, with applications in a large spectrum of areas, such as hydrology, atmospheric science, geology, etc. However, the aforementioned procedures demand an appropriate characterization of the second-order structure of the underlying random process, which can be addressed through the variogram or the covariance functions.

Let $\{Z(\ s) \in \mathbb{R} : \mathrm{s} \in D \subset \mathbb{R}^d\}$ be a spatial random process, where D denotes the observation region. We will focus our attention on the variogram of an intrinsically stationary random process, so the following conditions are assumed:

(I1) $\mathrm{E}[Z(\mathrm{s})] = \mu$, for all $\mathrm{s} \in D$ and some constant μ.

(I2) $\mathrm{Var}[Z(\mathrm{s}) - Z(\mathrm{s}')] = 2\gamma(\mathrm{s} - \mathrm{s}')$, for all $\mathrm{s}, \mathrm{s}' \in D$ and some function γ, which is referred to as the semivariogram (2γ is the variogram).

Furthermore, an intrinsic random process is known as isotropic, when condition (I2) is replaced by:

(I2') $\mathrm{Var}[Z(\mathrm{s}) - Z(\mathrm{s}')] = 2\gamma(\|\mathrm{s} - \mathrm{s}'\|)$, for all $\mathrm{s}, \mathrm{s}' \in D$ and some function γ, where $\|\cdot\|$ denotes the Euclidean norm.

Hypotheses I1 and I2 impose that the first two moments of $Z(\mathrm{s}) - Z(\mathrm{s}')$ depend only on $\mathrm{s} - \mathrm{s}'$ and, therefore, on the distance and the relative orientation of the lag between the locations involved. The more restrictive condition I2' avoids the influence of the lag-orientation on both moments.

For approximation of the variogram, nonparametric procedures may be used in a first step, providing us with the empirical estimator and more robust alternatives, studied in Matheron [1], Cressie [2] or Genton [3]. On the other hand, kernel-type approaches can be derived by adapting the Nadaraya-Watson estimation or the local linear method to the spatial setting, as suggested in Hall and Patil [4] or in García-Soidán et al. [5], respectively. Alternative mechanisms for

approximating the semivariogram include the constant and the variable nearest-neighbor estimators, given in Yu and Mateu [6], which yield generalizations of the Matheron and the Nadaraya-Watson semivariograms, respectively.

The performance of different nonparametric semivariogram estimators is analyzed in Menezes et al. [7], which have been put into comparison in a numerical study covering a range of dependence situations. Nevertheless, the above-mentioned approaches are not necessarily valid for their direct application to spatial prediction. In fact, they typically fail to fulfill the conditionally negative definiteness property, which is satisfied by the theoretical semivariogram and must be required from their estimators, in order to guarantee a solution for the kriging equations. We can cope with this problem by first choosing a valid parametric family and then selecting within it the variogram which best fits the data, as described in Cressie [2] or extended in Shapiro and Botha [8] to a broad class of valid variograms, not depending on a small number of parameters.

The usage of a parametric estimator may be attractive at first because of its simplicity and validity, although one of its main drawbacks is the procedure followed for the choice of a parametric model, typically addressed through graphical diagnostics, since the shape of some variogram models is similar. An alternative is developed in Gorsich and Genton [9], based on the fact that, unlike the variogram models, their derivatives are often quite different; hence, an estimation of the first variogram derivative may help to select from different models.

The current study is aimed to obtain global measures for the Nadaraya-Watson variogram deviations, involving the L_2-norm. These results will provide the basis for developing applications, such as that of testing the goodness of fit of a variogram model, which requires solving the general contrast:

$$\begin{aligned} &H_0 : \gamma \in \Gamma_\theta = \{\gamma_\theta(\cdot) : \theta \in \Theta \subset \mathbb{R}^p\} \text{ versus} \\ &H_1 : \ \gamma \notin \Gamma_\theta \end{aligned} \tag{1}$$

In the statistics literature, an approach for the aforementioned goal was proposed in Maglione and Diblasi [10], specifically designed for gaussian random

processes. An extensive option is analyzed in Crujeiras et al. [11], based on constructing tests for the spectral density of spatial processes observed on a regular grid. However, a different via could be explored, which tried to mimic the goodness of fit tests suggested for curve estimation with independent data, such as those given in Fan [12] or Härdle and Mammen [13], for the density or the regression settings. With this aim, the convergence in distribution of the L_2-deviations of the Nadaraya-Watson estimators of the variogram must be established. In addition, the approximation of the model parameters and the critical points should be addressed, together with the selection of the bandwidth in the variogram estimators or the threshold involved in each of the statistics considered. Then, the adequateness of a parametric variogram model could be tested through the kernel-type estimation and this procedure would be valid for their application to intrinsically stationary random processes, under stochastic sampling design. In the current work, we will focus on the estimation of the model parameters and the resulting critical points, as well as provide some ideas for choice of the other elements, the thresholds and the bandwidth parameters.

For specifying the parameters of the variogram model, different alternatives have been proposed, based on any of the following criteria: maximum likelihood, minimum norm quadratic, minimum variance or least squares; see Cressie [2], Müller [14], Stein [15] and references therein. We suggest proceeding through the least squares approaches, since they require the fewest distributional assumptions about the random process and their consistency follows from the results in Lahiri et al. [16]. To derive our results, we will deal with the general variogram, corresponding to the anisotropic setting, as well as with the specific case of isotropy. Additional tests provided in the literature would allow us to check whether or not the isotropic condition is satisfied, as the one described in Maity and Sherman [17], aiming us to take a decision about this issue, previously to select the appropriate goodness of fit test to be applied in each case.

Regarding the critical points, they can be approximated from the normal limit distribution that could be established for the functionals considered, although this procedure is not recommended in general due to its slow speed

of convergence, particularly when the sample size is small. A second strategy might consist of deriving the aforementioned values from the resulting asymptotic functionals; however, their dependence on unknown terms would involve the approximation of additional parameters and, consequently, could lead to an increment in the errors of the final estimates derived. Therefore, we propose solving this problem by appealing to the resampling techniques. A discussion about the accuracy of the bootstrap approximations for independent random variables can be found in Bose and Babu [18]. With spatial data, the dependence structure of the stochastic process asks for the selection of a bootstrap approach specifically designed for this setting. Under the assumption of knowledge of the underlying distribution of the random process, the parametric bootstrap methodology can be easily extended for its application to the spatial setting, by estimating the unknown parameters of the distribution model from the observed data. In addition, nonparametric resampling approaches could be used, as that introduced in García-Soidán et al. [19].

This paper is organized as follows. Section 2 introduces the main hypotheses that will be imposed along this work. The properties of the L_2-deviations of the kernel-type variograms are studied in Section 3, where we also address the estimation of the model parameters. Section 4 presents some applications of these properties, focused on checking the validity of a parametric variogram. To illustrate their performance in practice, numerical studies with simulated and real data have been derived, whose results are described in Sections 5 and 6, respectively. The main conclusions are summarized in Section 7.

2. Main hypothesis

Suppose that n data, $\mathbf{Z} = (Z(\mathrm{s}_1), ..., Z(\mathrm{s}_n))$, have been collected, at the respective spatial locations s_1, ..., s_n. The Nadaraya-Watson semivariogram is a kernel-type estimator obtained as a weighted average of the square differences

$(Z(\mathrm{s}_i) - Z(\mathrm{s}_j))^2$, given by:

$$\hat{\gamma}_1(\mathrm{t}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right)\left(Z\left(\mathrm{s}_i\right) - Z\left(\mathrm{s}_j\right)\right)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right)},\ \mathrm{t} \in \mathbb{R}^d \qquad (2)$$

where K_d represents a d-variate symmetric density and h_d is the bandwidth parameter.

Estimator (2) can be adapted for its specific use in the isotropic setting, leading us to:

$$\hat{\gamma}_2(x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right)\left(Z\left(\mathrm{s}_i\right) - Z\left(\mathrm{s}_j\right)\right)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right)},\ x \geq 0 \qquad (3)$$

In the current work, we analyze the convergence in distribution of the L_2-deviations of the Nadaraya-Watson semivariograms (2) and (3), as well as the specification of the model parameters involved. To simplify the conditions that will be imposed on the two different settings, an index δ_k will be introduced, which takes values d or 1, for $k = 1$ (anisotropy) or $k = 2$ (isotropy), respectively.

In addition, an α-mixing condition will be established for the random process, similar to that imposed in Zhu and Lahiri [20]. For this purpose, given $\mathrm{S}, \mathrm{S}' \subset \mathbb{R}^d$, take $Z[\mathrm{S}]$ as the σ-field generated by $\{Z(\mathrm{s})/\mathrm{s} \in \mathrm{S}\}$ and $d\left(\mathrm{S}, \mathrm{S}'\right) = \inf\{\|\mathrm{s} - \mathrm{s}'\| : \mathrm{s} \in \mathrm{S}, \mathrm{s}' \in \mathrm{S}'\}$, where $\|\cdot\|$ denotes the L_1-norm on $\mathbb{R}^d$. Write $\alpha_1\left(\mathrm{S}, \mathrm{S}'\right) = \sup\{|\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)| : A \in Z[\mathrm{S}], B \in Z[\mathrm{S}']\}$. The α-mixing coefficient is defined as:

$$\alpha\left(m, b\right) = \sup\left\{\alpha_1\left(\mathrm{S}, \mathrm{S}'\right) : \mathrm{S}, \mathrm{S}' \in \mathcal{R}_p(b), d\left(\mathrm{S}, \mathrm{S}'\right) \geq m\right\}$$

with $\mathcal{R}_p(b) = \{\cup_{i=1}^{p} D_i : D_i's \text{ are disjoint and } \sum_{i=1}^{p} \|D_i\| \leq b\}$.

Next we enumerate the different hypotheses that will be required:

(H1) Let $\{Z(\mathrm{s}) \in \mathbb{R} : \mathrm{s} \in D \subset \mathbb{R}^d\}$ be a random process satisfying condition I1, together with I2 or I2', for k equal to 1 or 2, respectively.

(H2) $\alpha(m, b) \leq a_1 m^{-a_2} b^{a_3}$, for some positive constants a_1, a_2, a_3, satisfying that $a_2 > 2d$ and $a_3 < a_2 d^{-1}$.

(H3) $\mathrm{E}[Z(\mathrm{s})^8] < \infty$, for all $\mathrm{s} \in D$.

(H4) $D = D_n = \lambda D_0$, for some $\lambda = \lambda_n$ diverging to $+\infty$ and some bounded region $D_0 \subset \mathbb{R}^d$ containing a sphere with positive d-dimensional volume.

(H5) Let f_0 be a density function considered on D_0. The spatial locations will be taken as $\mathrm{s}_i = \lambda\,\mathrm{r}_i$, for $1 \leq i \leq n$, where $(\mathrm{r}_1, ..., \mathrm{r}_n)$ represents a realization of a random sample random sample $(\mathrm{R}_1, ..., \mathrm{R}_n)$ drawn from f_0.

(H6) K_{δ_k} is a δ_k-variate, compactly supported, symmetric and bounded density function, satisfying that $K_{\delta_k}(0) > 0$ and its covariance matrix is $c_{K_{\delta_k}} 1_{\delta_k}$, for some $c_{K_{\delta_k}} > 0$, where 1_{δ_k} denotes the $\delta_k \times \delta_k$ diagonal matrix of 1, for k equal to 1 or 2.

(H7) $\{h_{\delta_k} + \lambda^{-1} + \lambda^d n^{-1} + \lambda^d h_{\delta_k}^4 + n^{-2}\lambda^{2d} h_{\delta_k}^{-\delta_k}\} \overset{n\to\infty}{\longrightarrow} 0$, for k equal to 1 or 2.

(H8) γ_θ is twice continuously differentiable with respect to θ, for all $\theta \in \Theta \subset \mathbb{R}^p$, and satisfies that $\sup\{\gamma_\theta(\mathrm{t}) : \mathrm{t} \in \mathbb{R}^d, \theta \in \Theta\} < \infty$ or $\sup\{\gamma_\theta(x) : x \geq 0, \theta \in \Theta\} < \infty$, for k equal to 1 or 2, respectively.

(H9) For each threshold $y > 0$ and for any $\varepsilon_1 > 0$, there exists $\varepsilon_2 > 0$ such that $\inf\{\int_{\|\mathrm{t}\| \leq y} (\gamma_{\theta_1}(\mathrm{t}) - \gamma_{\theta_2}(\mathrm{t}))^2\, d\mathrm{t} \;:\; \|\theta_1 - \theta_2\| > \varepsilon_1\} > \varepsilon_2$ or $\inf\{\int_0^y (\gamma_{\theta_1}(x) - \gamma_{\theta_2}(x))^2\, dx \;:\; \|\theta_1 - \theta_2\| > \varepsilon_1\} > \varepsilon_2$, for k equaling 1 or 2, respectively.

(H10) Given $w_{k,\theta} : \mathbb{R}^{\delta_k} \to \mathbb{R}$, $w_{k,\theta}$ is continuously differentiable with respect to θ, for all $\theta \in \Theta$ and k equal to 1 or 2. In addition, $\sup\{\int_{\|\mathrm{t}\| \leq y} \left(|w_{k,\theta}(\mathrm{t})| + |w_{k,\theta}(\mathrm{t})^{-1}|\right) d\mathrm{t} : \theta \in \Theta\} < \infty$ or $\sup\{\int_0^y \left(|w_{k,\theta}(x)| + |w_{k,\theta}(x)^{-1}|\right) dx : \theta \in \Theta\} < \infty$, for k equal to 1 or 2, respectively.

Given an intrinsic random process, as established in H1, index δ_k determines whether we deal with the general case (k=1) or with the isotropic one (k=2). Hypotheses H2 and S3 are imposed to guarantee finite variance for the functionals considered, although the resulting asymptotic terms will only involve

the moments of the random process up to the fourth-order. Both conditions are satisfied for a stationary gaussian process with a bounded variogram that has either finite-range or an asymptotic range with an exponentially decreasing rate of convergence. They can also be fulfilled by smooth functions of stationary gaussian processes with the latter requirements, together with additional arguments leading to H3.

Our proposals will be obtained under the assumptions of a mixed increasing domain asymptotic structure for the random process and a stochastic sampling design for the spatial locations, as derived from H4 and H5. Indeed, the α-mixing condition adopted in H4 yields the properties required from the dependence structure in Hall and Patil [4] and in Lahiri et al. [16], whose results will be applied in the current work. An extension to the fixed design case asks for extra hypotheses regarding the sample size and the dependence structure, similar to those considered in Bandyopadhyay and Lahiri [21] to develop properties of the Fourier transform.

The usual requirements for a δ_k-variate kernel density are imposed in H6 and consistency of the results is guaranteed under H7, aiming to specify the order of the dominant terms involved.

Hypotheses H8-H10 are included to ensure the regularity conditions established in Lahiri et al. [16], adapted to this setting. In this respect, H8 and H9 enable us to apply the ordinary least squares approach for approximation of the variogram parameters in an accurate way. The selection of a weight function, satisfying the additional condition H10, extends the consistency of the results to the weighted least squares estimator. Hypothesis H8 is satisfied by sufficiently smooth variogram models, under the assumption of a compact space Θ. The latter requirement together with the continuity condition of γ_θ in H8 would be enough to check H9.

3. Properties of the L_2-deviations of the kernel variograms

The kernel-type estimation provides consistent estimators for the dependence structure of a stationary random process, as proved in Hall and Patil [4], by requiring appropriate conditions. With similar arguments, the convergence rates of (2) and (3) have been derived in García-Soidán [22], under hypotheses H1-H7 for k equal to 1 and 2, respectively, where even the asymptotic normality of the Nadaraya-Watson semivariograms has been established. The resulting variance of the kernel estimators is of the exact order λ^{-d}, for both the anisotropic and the isotropic settings. In view of the latter, we deal with the rescaled L_2-deviations of the kernel semivariograms, given by:

$$\begin{aligned} T_1 &= \lambda^d \int_{\|\mathrm{t}\|\leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma(\mathrm{t})\right)^2 d\mathrm{t} \\ T_2 &= \lambda^d \int_0^y \left(\hat{\gamma}_2(x) - \gamma(x)\right)^2 dx \end{aligned} \tag{4}$$

under anisotropy or isotropy, respectively, for some threshold $y > 0$.

The convergence in distribution of the functionals defined in (4) will be established in the following theorems. We will start by considering T_1, in order to derive results for the general setting.

Theorem 3.1. *Assume that hypotheses H1-H7 are satisfied for k=1 and that γ is three-times continuously differentiable in a neighborhood of* t*, for all* t *satisfying that $0 < \|\mathrm{t}\| < y$ and some $y > 0$. Then, $T_1 \xrightarrow{D} \int_{\|\mathrm{t}\|\leq y} X_1(\mathrm{t})^2 d\mathrm{t}$, for a zero mean gaussian process $\{X_1(\mathrm{t}) : \mathrm{t} \in \mathbb{R}^d,\ \|\mathrm{t}\| < y\}$, with covariance function C_1 given in (A.2).*

Now we will address the convergence of functional T_2, specifically proposed in (4) for the isotropic setting.

Theorem 3.2. *Suppose that conditions H1-H7 are satisfied for k=2 and that γ admits three continuous derivatives in a neighborhood of x, for all $x \in (0, y)$ and some $y > 0$. Then, $T_2 \xrightarrow{D} \int_0^y X_2(x)^2 dx$, for a zero mean gaussian process $\{X_2(x) : x \in \mathbb{R},\ x \in (0, y)\}$, with covariance function C_2, defined in (B.3).*

From the previous properties, tests will be derived to check whether or not the theoretical variogram can be taken to equal γ_{θ_0}, for a fixed parameter $\theta_0 \in \Theta$. We will refer to this issue in the following section. However, an additional step would be required to address the more general contrast (1). In fact, we need an estimator $\hat{\theta}$ of the true parameter, under the null hypothesis H_0 in (1), together with a reformulation of the L_2-deviations of the kernel variograms, whose consistency must be guaranteed.

We first adopt an ordinary least squares approach for approximation of the true parameter, so the estimation of θ_0 will be addressed by using:

$$\hat{\theta}_{OLS,k} = \begin{cases} \operatorname{argmin} \left\{ \displaystyle\int_{\|\mathrm{t}\| \le y} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_\theta(\mathrm{t})\right)^2 d\mathrm{t} : \theta \in \Theta \right\}, & \text{if } k = 1 \\ \operatorname{argmin} \left\{ \displaystyle\int_0^y \left(\hat{\gamma}_2(x) - \gamma_\theta(x)\right)^2 dx : \theta \in \Theta \right\}, & \text{if } k = 2 \end{cases} \tag{5}$$

for each threshold $y > 0$.

Next, we will establish the consistency of $\hat{\theta}_{OLS,k}$ and also prove that it keeps the same rate of convergence as $\hat{\gamma}_k$, for $k = 1, 2$. With this aim, denote by $A(\theta_0)$ the $p \times p$ matrix $A(\theta_0) = (a^{(j_1,j_2)}(\theta_0))$, with:

$$a^{(j_1,j_2)}(\theta_0) = \begin{cases} \displaystyle\int_{\|\mathrm{t}\| \le y} \left.\frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j_1)}}\right|_{\theta_0} \left.\frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j_2)}}\right|_{\theta_0} d\mathrm{t} & \text{if } k = 1, \\ \displaystyle\int_0^y \left.\frac{\partial \gamma_\theta(x)}{\partial \theta^{(j_1)}}\right|_{\theta_0} \left.\frac{\partial \gamma_\theta(x)}{\partial \theta^{(j_2)}}\right|_{\theta_0} dx & \text{if } k = 2 \end{cases}$$

Write $B(\theta_0)$, $C(\theta_0)$, $D(\theta_0)$ and $E(\theta_0)$ for the p-dimensional vectors $B(\theta_0) = (b^{(j)}(\theta_0))$ and $D(\theta_0) = (d^{(j)}(\theta_0))$ and the $p \times p$ matrices $C(\theta_0) = (c^{(j_1,j_2)}(\theta_0))$ and $E(\theta_0) = (e^{(j_1,j_2)}(\theta_0))$, with $B(\theta_0) = A(\theta_0)^{-1} D(\theta_0)$ and $C(\theta_0) = A(\theta_0)^{-1} E(\theta_0) A(\theta_0)^{-1}$, as well as:

$$d^{(j)}(\theta_0) = \int_{\|\mathrm{t}\| \le y} B_1(\mathrm{t}) \left.\frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j)}}\right|_{\theta_0} d\mathrm{t}$$

$$e^{(j_1,j_2)}(\theta_0) = \int_{\|\mathrm{t}\| \le y} \int_{\|\mathrm{t}'\| \le y} C_1(\mathrm{t}, \mathrm{t}') \left.\frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j_1)}}\right|_{\theta_0} \left.\frac{\partial \gamma_\theta(\mathrm{t}')}{\partial \theta^{(j_2)}}\right|_{\theta_0} d\mathrm{t} d\mathrm{t}'$$

for k=1 or, alternatively, for k=2:

$$d^{(j)}(\theta_0) = \int_0^y B_2(\mathrm{t}) \left.\frac{\partial \gamma_\theta(x)}{\partial \theta^{(j)}}\right|_{\theta_0} dx$$

$$e^{(j_1,j_2)}(\theta_0) = \int_0^y \int_0^y C_2(x, x') \left.\frac{\partial \gamma_\theta(x)}{\partial \theta^{(j_1)}}\right|_{\theta_0} \left.\frac{\partial \gamma_\theta(x')}{\partial \theta^{(j_2)}}\right|_{\theta_0} dx dx'$$

where functions B_1 and C_1 are defined in (A.2) and functions B_2 and C_2 are respectively given in (B.1) and (B.3).

Theorem 3.3. *Assume the hypotheses required in Theorem 3.1 or in Theorem 3.2, depending on whether anisotropy (k=1) or isotropy (k=2) is assumed, together with condition H8. If* $\det A(\theta_0) \neq 0$ *and hypothesis* H_0 *in (1) holds, it follows that:*

$$\mathrm{E}\left[\hat{\theta}^{(j)}_{OLS,k}\right] = \theta_0^{(j)} + h^2_{\delta_k} b^{(j)}(\theta_0) + o\left(h^2_{\delta_k}\right)$$
$$\mathrm{Cov}\left[\hat{\theta}^{(j_1)}_{OLS,k}, \hat{\theta}^{(j_2)}_{OLS,k}\right] = \lambda^{-d} c^{(j_1,j_2)}(\theta_0) + o\left(h^4_{\delta_k} + \lambda^{-d}\right)$$

for $k = 1, 2$ *and* $j, j_1, j_2 \in \mathbb{N}$*, with* $1 \leq j, j_1, j_2 \leq p$*.*

Now, we will reformulate the L_2-deviations of the kernel-type variogram estimators, previously analyzed, by substituting $\gamma_{\hat{\theta}_{OLS,k}}$ for γ, as given below:

$$U_1 = \lambda^d \int_{\|\mathrm{t}\| \leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t})\right)^2 d\mathrm{t}$$
$$U_2 = \lambda^d \int_0^y \left(\hat{\gamma}_2(x) - \gamma_{\hat{\theta}_{OLS,2}}(x)\right)^2 dx$$

Under the null hypothesis H_0 in (1), the convergence in distribution of U_k and T_k follow similar patterns, as stated below.

Theorem 3.4. *Assume the conditions required in Theorem 3.3. Then,* $U_1 \xrightarrow{D} \int_{\|\mathrm{t}\| \leq y} Y_1(\mathrm{t})^2 d\mathrm{t}$ *or* $U_2 \xrightarrow{D} \int_0^y Y_2(x)^2 dx$*, for some zero mean gaussian processes* $\{Y_1(\mathrm{t}) : \mathrm{t} \in \mathbb{R},\ \|\mathrm{t}\| < y\}$ *or* $\{Y_2(x) : x \in \mathbb{R},\ x \in (0, y)\}$*, respectively.*

Theorem 3.4 enables us to check whether or not the theoretical variogram belongs to a parametric model, as we will describe in the following section. It is important to notice that the covariance functions of the gaussian processes $Y_k(\cdot)$ have not been specified in Theorem 3.4. Thus, we will propose to proceed through a bootstrap approach for approximation of the critical points involved.

An alternative estimator of the true parameter can be derived from the weighted least squares criterion. It will be denoted by $\hat{\theta}_{WLS,k}$ and obtained in

the manner described below:

$$\hat{\theta}_{WLS,k} = \begin{cases} \operatorname{argmin}\left\{ \int_{\|\mathrm{t}\|\leq y} w_{1,\theta}(\mathrm{t}) \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_\theta(\mathrm{t})\right)^2 d\mathrm{t} : \theta \in \Theta \right\}, & \text{if } k = 1 \\ \operatorname{argmin}\left\{ \int_0^y w_{2,\theta}(x) \left(\hat{\gamma}_2(x) - \gamma_\theta(x)\right)^2 dx : \theta \in \Theta \right\}, & \text{if } k = 2 \end{cases} \tag{6}$$

for each threshold $y > 0$ and some weight functions $w_{k,\theta}$, for $k = 1, 2$.

A similar result to Theorem 3.2 in Lahiri et al. [16] holds in the current setting, under hypotheses H8-H10, which guarantee the regularity conditions required there. The latter ensures consistency of the weighted least squares estimator of the true parameter, with the same rate of convergence, aiming to derive a limit distribution analogue to that established in Theorem 3.4 by replacing $\hat{\theta}_{OLS,k}$ for $\hat{\theta}_{WLS,k}$, defined in (6).

Theorem 3.5. *Assume the conditions required in Theorem 3.4 and that hypothesis H10 holds. Then, one has that:*

$$V_1 = \lambda^d \int_{\|\mathrm{t}\|\leq y} w_{1,\theta}(\mathrm{t}) \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\hat{\theta}_{WLS,1}}(\mathrm{t})\right)^2 d\mathrm{t} \xrightarrow{D} \int_{\|\mathrm{t}\|\leq y} Y_1(\mathrm{t})^2 d\mathrm{t}$$
$$V_2 = \lambda^d \int_0^y w_{2,\theta}(x) \left(\hat{\gamma}_2(x) - \gamma_{\hat{\theta}_{WLS,2}}(x)\right)^2 dx \xrightarrow{D} \int_0^y Y_2(x)^2 dx$$

for some zero mean gaussian processes $\{Y_1(\mathrm{t}) : \mathrm{t} \in \mathbb{R},\ \|\mathrm{t}\| < y\}$ *or* $\{Y_2(x) : x \in \mathbb{R},\ x \in (0, y)\}$.

This proof would be developed similarly to that of Theorem 3.4 and, therefore, it will be omitted.

Usually, $w_{1,\theta}(\mathrm{t})$ and $w_{2,\theta}(x)$ are taken to be proportional to the inverse of $\mathrm{Var}[\hat{\gamma}_1(\mathrm{t})]$ and $\mathrm{Var}[\hat{\gamma}_2(x)]$, respectively. From this perspective, the weighted least squares estimators for the general or the isotropic cases are respectively constructed so as to assign more weight to the lag vectors or distances with smaller estimated errors. For selection of the latter functions, bear in mind the variances of the kernel variograms, derived in the proofs of Theorems 3.1 and 3.2, through relations (A.3) and (B.2), respectively.

So, our first choice for the weight functions could be given by sample versions of the functions involved, leading to:

$$w_{1,\theta}(\mathrm{t}) = \frac{w_{1,1}(\mathrm{t})^2}{w_{1,2,\theta}(\mathrm{t})^2}, \quad w_{2,\theta}(x) = \frac{w_{2,1}(x)^2}{w_{2,2,\theta}(x)^2} \tag{7}$$

for the general (k=1) or the isotropic (k=2) settings, with:

$$w_{1,1}(\mathrm{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right)$$
$$w_{1,2,\theta}(\mathrm{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right)\left(\left(Z\left(\mathrm{s}_i\right) - Z\left(\mathrm{s}_j\right)\right)^2 - 2\gamma_\theta(\mathrm{t})\right)$$
$$w_{2,1}(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right)$$
$$w_{2,2,\theta}(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right)\left(\left(Z\left(\mathrm{s}_i\right) - Z\left(\mathrm{s}_j\right)\right)^2 - 2\gamma_\theta(x)\right)$$

To avoid dependence on the unknown parameter θ in (7), the variogram function could be replaced by its kernel counterpart as follows:

$$w_{1,\theta}(\mathrm{t}) = w_1(\mathrm{t}) = \frac{w_{1,1}(\mathrm{t})^2}{w_{1,3}(\mathrm{t})^2}, \;\; w_{2,\theta}(x) = w_2(x) = \frac{w_{2,1}(x)^2}{w_{2,3}(x)^2} \tag{8}$$

where functions $w_{k,3}$ are obtained by substituting $\hat{\gamma}_k$ for γ_θ in $w_{k,2,\theta}$, for $k = 1, 2$.

For large sample sizes, implementation of the latter proposals entails a high computational cost and, therefore, other approaches for specification of the weight functions could be advisable. For instance, we can adapt the alternative described in Cressie [2], based on the assumption of gaussianity from the random process, yielding that $\mathrm{Var}\left[\left(Z\left(0\right) - Z\left(\mathrm{t}\right)\right)^2\right] = 8\gamma(\mathrm{t})^2$, as well as on the fact that:

$$\mathrm{Cov}\left[\left(Z\left(0\right) - Z\left(\mathrm{t}\right)\right)^2, \left(Z\left(\mathrm{r}\right) - Z\left(\mathrm{t}' + \mathrm{r}\right)\right)^2\right] \approx 0 \tag{9}$$

for large values $\|\mathrm{r}\|$ and small lags $\|\mathrm{t}\|$ and $\|\mathrm{t}'\|$.

The previous statements applied on (7) lead to the following weight functions:

$$\begin{aligned} w_{1,\theta}(\mathrm{t}) &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right)}{\gamma_\theta(\mathrm{t})^2} \\ w_{2,\theta}(x) &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right)}{\gamma_\theta(x)^2} \end{aligned} \tag{10}$$

From (8), we would obtain instead:

$$\begin{aligned} w_{1,\theta}(\mathrm{t}) = w_1(\mathrm{t}) &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right)}{\hat{\gamma}_1(\mathrm{t})^2} \\ w_{2,\theta}(x) = w_2(x) &= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right)}{\hat{\gamma}_2(x)^2} \end{aligned} \tag{11}$$

An additional strategy is introduced in Das et al. [23], which can be supported by the hypothesis of variance stabilization of the random process along the observation region. This proposal consists of reducing the weight functions given in (11) to their numerators:

$$\begin{aligned} w_{1,\theta}(\mathrm{t}) = w_1(\mathrm{t}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{(\mathrm{s}_i - \mathrm{s}_j) - \mathrm{t}}{h_d}\right) \\ w_{2,\theta}(x) = w_2(x) &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\|\mathrm{s}_i - \mathrm{s}_j\| - x}{h_1}\right) \end{aligned} \tag{12}$$

As remarked in Das et al. [23], the accuracy of the estimators $\hat{\theta}_{WLS,k}$, with weights provided by (10), (11) or (12), relies on the validity of the assumption of gaussianity from the random process and on condition (9), which is quite restrictive in practice.

Remark 3.6. *Boundary kernel functions could be used instead of symmetric kernel densities for implementation of the variogram estimators (2) and (3). This alternative would avoid the bias increment close to the endpoints, which might have a significative effect on the tests derived, particularly for small sample sizes.*

Remark 3.7. *In practice, we should proceed through numerical approximations to compute the integrals involved in the L_2-deviations of the Nadaraya-Watson estimators or in the estimation of the true parameter of the variogram model.*

Remark 3.8. *Taking into account the previous comments, other alternatives that could be explored for approximation of the true parameter include the maximum likelihood approaches. For their application, the form of the distribution of the underlying random process must be known and a finite number of locations must be selected for specification of the likelihood function. The asymptotic normality of the resulting estimators has been derived for gaussian processes with linear trend, whose accuracy will be dependent on the fulfillment of the required hypotheses. In particular, the normal limit distribution has been obtained for the maximum likelihood estimator in Mardia and Marshall [24],*

under the spatial regression model. A similar convergence can be extended to the variance-component model and the restricted maximum likelihood estimator, as established in Cressie and Lahiri [25].

Remark 3.9. *We have not tackled the curse of dimensionality problem in the current research, due to the convergence rate of the proposed tests and to the fact that the spatial data sets in practice include at most three coordinates for the locations in the observation region ($d \leq 3$). However, suppose that we apply alternative mechanisms to the Nadaraya-Watson estimation of the semivariogram and/or to the least squares criteria for approximating the true parameter. The resulting tests might be affected by the curse of dimensionality, thus making it advisable to introduce some strategy to overcome this problem. The previous issue could be addressed by adapting approaches introduced in other settings, such as the one suggested in Verleysen and Francois [26] to reduce the dimensionality in time series prediction or the alternative provided in Lavergne and Patilea [27] to calibrate the level of a test to check the validity of a parametric regression model.*

4. Applications

Next, some potential applications of the results provided in Section 3 are outlined, focused on checking the adequateness of a selected variogram. With this aim, the critical points of the resulting tests must be approximated, as we will describe next.

Firstly, as an immediate consequence of Theorems 3.1 and 3.2, the following contrast can be performed:

$$H_0 : \ \gamma = \gamma_{\theta_0} \quad \text{versus}$$
$$H_1 : \ \gamma \neq \gamma_{\theta_0}$$

for a fixed parameter $\theta_0 \in \mathbb{R}^p$, at an approximate level α. This issue demands substituting γ_{θ_0} for γ in T_k, defined in (4), for $k = 1, 2$, as well as approximating the quantiles of the resulting functional. Then, the test would be rejected when $T_k \geq t_{k,1-\alpha}$, with $t_{k,\beta}$ denoting the β-quantile of T_k.

Now, we consider the more general problem of deriving a goodness of fit test for the variogram, through the contrast (1), which requires an appropriate characterization of the unknown true parameter θ_0, under the null hypothesis H_0 in (1). For such a purpose, we introduced in Section 3 the least squares estimators, $\hat{\theta}_{OLS,k}$ and $\hat{\theta}_{WLS,k}$, defined in (5) and (6), respectively. Using these approaches, from Theorems 3.4 or 3.5, the rejection region for the corresponding test at a level α would be given by $U_k \geq u_{k,1-\alpha}$ or $V_k \geq v_{k,1-\alpha}$, respectively, where $u_{k,\beta}$ and $v_{k,\beta}$ stand for the β-quantile of U_k and V_k.

To complete the specification of the critical points, $t_{k,\beta}$, $u_{k,\beta}$ and $v_{k,\beta}$, an appealing option is provided by the Bootstrap techniques. The key idea would be the selection of n locations s_i^* of the observation region, for $i = 1, ..., n$, as well as the generation of B replicates of the original data at the selected locations, $\mathbf{Z}^{*(b)} = (Z^*(\mathrm{s}_1^*), ..., Z^*(\mathrm{s}_n^*))$, for $b = 1, ..., B$. Each replicate would be used to compute the Bootstrap analogue of the functional involved, T_k, U_k or V_k, which will be respectively represented by $T_k^{*(b)}$, $U_k^{*(b)}$ or $V_k^{*(b)}$. By arranging the resulting B values in an increasing order, the one in the $[B\beta]$-th position would provide us with an approximation of its β-quantile, where $[a]$ denotes the nearest integer to a.

For this approximation of the critical points, we can make use of the parametric resampling approaches, whose accuracy will be strongly dependent on the characteristics of the specific setting to be applied. In fact, when the form of the distribution of the spatial process is assumed to be known, the model parameters can be estimated from the observed data, so the Bootstrap samples can be generated from the resulting distribution. Furthermore, this issue is reduced to applying the well-known methodology of the structural analysis (Cressie [2]), when only the dependence structure of the random process is unknown. A nonparametric alternative for construction of Bootstrap samples is introduced in García-Soidán et al. [19], which was designed for its application to the spatial data setting.

Remark 4.1. *The application of our proposals asks for the choice of the thresh-*

old $y > 0$, the scale parameter λ and the bandwidth parameter involved in the kernel estimation of the variogram. With regard to the threshold, we can start by considering an estimate of the variogram range (or the asymptotic range) or a percentage of it, although this issue needs further research to put into comparison different options. On the other hand, the specification of the scale parameter λ, included to guarantee consistency of the results, can be easily avoided in practice, by simply omitting this term in each of the functionals considered and by implementing the critical points accordingly. In what respects to the bandwidth estimation, we could proceed through the cross-validation approaches, as suggested in Yao and Tong [28] for the regression setting Other alternatives could be based on using plug-in selectors derived from the optimal bandwidths, which could be obtained by minimizing the corresponding mean squares errors, as those proposed in García-Soidán [22].

5. Numerical studies with simulated data

In this section, we describe the results of the simulation studies developed to illustrate the behavior of our proposals, when applied to check the adequateness of a parametric variogram. With this aim, the statistic given below will be considered:

$$S = \int_0^y w_\theta(x) \left(\hat{\gamma}(x) - \gamma_{\hat{\theta}}(x)\right)^2 dx \tag{13}$$

under different scenarios.

Bear in mind that S represents the analogue of V_2, obtained by omitting λ^d (see Remark 4.1), as well as by considering these terms: a generic weight function w_θ, some estimator $\hat{\gamma}$ of the semivariogram and the estimator $\hat{\theta}$ derived for θ, when using w_θ and $\hat{\gamma}$ in the weighted least squares procedure (6), for k=2. This general statistic S will allow us to check the performance of the tests introduced in the current work and to compare them with other alternatives.

The following notation will be used:

- S_1 and S_2 are the statistics derived from (13) by taking the Matheron empirical semivariogram and the weight functions suggested in Cressie

[29], for the OLS and WLS methods, respectively, given by $w_\theta(x) = 1$ and $w_\theta(x) = n_x \gamma_{\hat{\theta}}(x)^{-2}$, where n_x is the number of distinct pairs $(\mathrm{s}_i, \mathrm{s}_j)$, with $\|\mathrm{s}_i - \mathrm{s}_j\| \approx x$.

- S_3 and S_4 denote the analogues of S_1 and S_2, respectively, when using the kernel semivariogram (3) instead of the Matheron estimator.

- S_5, S_6 and S_7 represent the statistics obtained by employing again the kernel semivariogram in (13), as well as by selecting the weights (8), (11) and (12), respectively.

In the first numerical studies, we used the parametric resampling approaches. More specifically, a gaussian Bootstrap method was applied on the locations $\mathrm{s}_\mathrm{i}^* = \mathrm{s}_\mathrm{i}$, for $i = 1, ..., n$. Then, the replicates $\mathbf{Z}^{*(b)} = (Z^*(\mathrm{s}_1), ..., Z^*(\mathrm{s}_n))$, for $b = 1, ..., B$ and $B = 500$, have been generated from a n-dimensional normal distribution, whose mean vector and variance matrix were approximated from the observed data.

Two scenarios were designed for application of the parametric Bootstrap to approximate the quantiles, with data drawn from gaussian and non-gaussian random processes. This way of proceeding has enabled us to check the effect of misspecification of the underlying distribution of the random process on each test. For both settings, the simulated data were used to approximate the significance levels of the statistics S_j, for $j = 1, ..., 7$, by simulation and to obtain the resulting power functions.

We started by generating data from spatial gaussian processes with isotropic exponential and spherical variograms, over a grid of $n = 100$ locations on the unit square $D = [0, 1] \times [0, 1]$. Indeed, 500 samples of size n were simulated under the assumption of gaussianity. The critical points were approximated through the gaussian Bootstrap, by drawing 500 replicates from a gaussian process with a covariance matrix derived from the selected model. In both cases, the true variogram model was taken as the null hypothesis.

Table 1 summarizes the percentage of rejections achieved for the exponential model. In general, the worst performance is observed for the statistics

based on the OLS method, S_1 and S_3, at the significance levels considered, $\alpha = 1\%, 5\%, 10\%$. On the other hand, the best behavior is observed for the statistics S_5 and S_6, as they provide good approximations of the different significance levels. The smoothing parameter in the nonparametric semivariogram was selected by cross validation.

$\sigma^2 = 1$	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\alpha = 1\%$	1.2	1.6	1.4	1.0	0.4	1.0	1.2
$\alpha = 5\%$	6.0	5.0	6.4	6.6	4.4	4.4	5.6
$\alpha = 10\%$	12.4	10.8	13.0	13.0	9.4	10.6	11.4
$\sigma^2 = 2.25$	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\alpha = 1\%$	0.2	1.4	1.4	1.0	0.2	1.0	1.0
$\alpha = 5\%$	0.6	3.6	6.4	6.0	4.4	4.4	5.4
$\alpha = 10\%$	12.6	10.4	13.2	9.0	12.0	10.6	11.8

Table 1: Percentage of rejections obtained for 500 samples of size $n = 100$, $\tau^2 = 0$ and $\phi = 0.2$. Data were generated from a gaussian process with exponential variogram. The true model was taken as the null hypothesis. The critical points were derived from 500 gaussian Bootstrap replicates.

When using the spherical variogram in the previous studies instead of the exponential one, similar conclusions remain valid, in terms of approximation of the significance level, as shown in Table 2.

To check the effect of misspecification of the variogram model, the following studies were carried out by simulating gaussian data from the isotropic Matérn model, with parameter κ varying in the interval $[0.5, 2.5]$. The isotropic exponential variogram, corresponding to $\kappa = 0.5$, was considered again for the null hypothesis, with the same variance, nugget and range as those selected for the data generation. The right pannel of Figure 1 displays the Matérn semivariograms for different values of κ. The power function of each test is represented in the left pannel of Figure 1. An advantageous behavior is observed for the statistics S_2, S_4 and S_5, with an increment of the percentage of rejections, as κ departs from the assumed value 0.5.

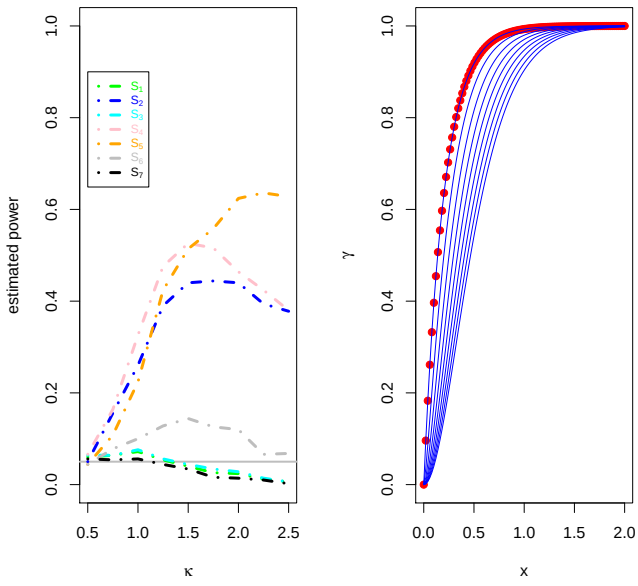

Figure 1: Left: Estimated power at the significance level $\alpha = 5\%$, for $\kappa \in [0.5, 2.5]$, 500 samples of size $n = 100$, $\sigma^2 = 1$, $\tau^2 = 0$ and $\phi = 0.2$. Data were generated from a gaussian process with Matérn variogram. The exponential model was taken as the null hypothesis. The critical points were derived from 500 gaussian Bootstrap replicates. Right: Isotropic Matérn semivariograms for several values of κ, with $\sigma^2 = 1$, $\tau^2 = 0$ and $\phi = 0.2$ (the red bullets identify the isotropic exponential semivariogram).

$\sigma^2 = 1$	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\alpha = 1\%$	0.2	1.2	0.2	1.2	1.0	0.6	0.8
$\alpha = 5\%$	3.0	6.4	3.6	6.8	4.0	3.6	4.4
$\alpha = 10\%$	8.8	9.6	8.2	11.4	8.8	9.4	9.2
$\sigma^2 = 2.25$	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\alpha = 1\%$	0.8	0.8	0.4	0.4	0.6	0.6	1.0
$\alpha = 5\%$	0.4	3.4	4.4	3.8	4.0	4.8	4.4
$\alpha = 10\%$	8.0	7.2	8.2	8.8	8.0	8.2	8.6

Table 2: Percentage of rejections obtained for 500 samples of size $n = 100$, $\tau^2 = 0$ and $\phi = 0.2$. Data were generated from a gaussian process with spherical variogram. The true model was taken as the null hypothesis. The critical points were derived from 500 gaussian Bootstrap replicates.

In a further step, non-gaussian stationary processes have been considered, through the model:

$$Z(\mathrm{s}) = Y(\mathrm{s})^2 \tag{14}$$

where $Y(\mathrm{s})$ is a zero mean stationary gaussian process. Denoting by C_Y and γ_Y the covariance and semivariogram function of Y, it follows that:

(J1) $\mathrm{E}[Z(\mathrm{s})] = \sigma^2$, for all $\mathrm{s} \in D$.

(J2) $\mathrm{Cov}(Z(\mathrm{s}), Z(\mathrm{s}')) = 2C_Y\left(\|\mathrm{s} - \mathrm{s}'\|\right)^2$, for all $\mathrm{s}, \mathrm{s}' \in D$.

(J3) $\mathrm{Var}(Z(\mathrm{s}) - Z(\mathrm{s}')) = 4\gamma_Y\left(\|s - s'\|\right)\left(2\sigma^2 - \gamma_Y\left(\|s - s'\|\right)\right)$, for all $\mathrm{s}, \mathrm{s}' \in D$.

More specifically, we assumed an isotropic exponential model for the spatial process Y, with parameters σ^2, ϕ and $\tau^2 = 0$. Then, properties J1-J3 yield that the resulting variogram of Z, given in (14), is also isotropic exponential, with variance $2\sigma^4$, range $\frac{\phi}{2}$ and $\tau^2 = 0$.

In this setting, 500 samples of size $n = 100$ were generated to check the validity of the exponential model as the null hypothesis. The critical values were approximated through the gaussian Bootstrap approach, thus enabling us to evaluate the effect of a bad specification of the underlying distribution. The

results achieved are summarized in Table 3. As expected, the misspecification of the distribution of the underlying process, which is non-gaussian, gives rise to inappropriate approximations of the significance levels in all cases. However, in this scenario, the test derived from S_5 has a clear superiority over the others, with a conservative behavior.

σ^2=1	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\alpha = 1\%$	17.4	18.8	20.0	18.8	0.6	14.8	16.2
$\alpha = 5\%$	22.8	32.8	29.6	33.4	1.8	28.2	25.2
$\alpha = 10\%$	35.0	41.4	35.8	40.8	5.6	33.0	32.8

Table 3: Percentage of rejections obtained for 500 samples of size $n = 100$, with $\sigma^2 = 1$, $\phi = 0.2$ and $\tau^2 = 0$. Data were generated from a non-gaussian process with exponential variogram. The true model was taken as the null hypothesis. The critical points were derived from 500 gaussian Bootstrap replicates.

The final simulation studies have been developed by deriving the bootstrap replicates, $\left\{\left(\mathrm{s}_1^*, ..., \mathrm{s}_n^*\right), \mathbf{Z}^{*(b)}\right\}$, through the nonparametric resampling approach introduced in García-Soidán et al. [19], for $b = 1, ..., B$ and $B = 500$. This mechanism asks for the selection of different bandwidth parameters, needed to estimate the conditional distributions required in the sequential procedure. These parameters must be appropriately taken, as they have an important effect in the spatial dependence structure generated in the Bootstrap process. We have chosen the bandwidths as balloon estimators, although further research is needed for their optimal selection.

To illustrate the results obtained when applying this Bootstrap method, Figure 2 displays the performance of the statistics S_j for testing the goodness of fit, at a significance level $\alpha = 0.05$ and a sequence of values for the bandwidth ranging from 0.05 to 0.20. Indeed, 500 samples of size 100 were generated on $D = [0, 1] \times [0, 1]$ from a gaussian process, with isotropic exponential variogram.

These numerical studies were carried out by using the same bandwidth h_i for generating all the bootstrap locations $s_i^{*(b)}$, $i = 1, ..., n$, which was taken

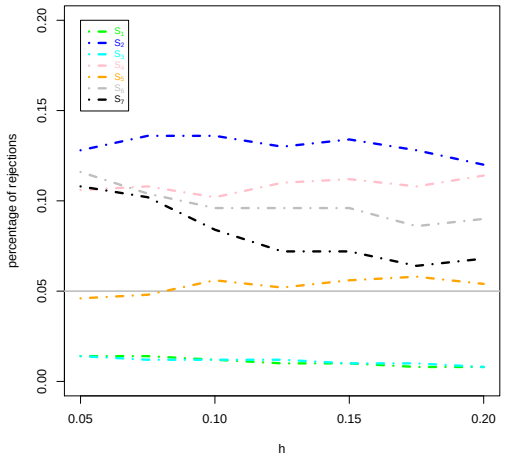

Figure 2: Percentage of rejections obtained for different values of $h \in [0.05, 0.20]$, with 500 samples of size $n = 100$, generated from the exponential semivariogram with $\tau^2 = 0$, $\sigma^2 = 1$ and$\phi = 0.2$). Data were generated from a gaussian process with exponential variogram. The true model was taken as the null hypothesis. The critical points were derived from 500 nonparametric Bootstrap replicates.

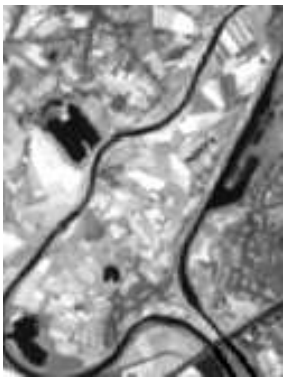

Figure 3: Geographical region associated to the Meuse data set.

as $h_i = h, i = 1, ..., n-1$, with h varying in a grid. The best performance is observed for estimator S_5, as it exhibits more robustness to small variations of the bandwidth. The statistics based on the OLS method, S_1 and S_3, tend to underestimate the power of the test, whereas the procedure based on S_7 seems to be more affected by changes in the bandwidth parameter.

6. Assessment of the variogram model in practice

To exemplify the application of our proposals in practice, we have considered the Meuse data set, provided with the statistical package *sp* of the R library, which is described in Bivand et al. [30]. It contains measures of heavy metal pollutants on the topsoil of a flood plain along the Meuse river, near the village of Stein (Netherlands). Figure 3 displays a map of the region, taken from Hengl [31], where 155 spatial locations were considered.

We focused our attention on the concentrations of zinc. This data set has been widely analyzed and its log-transformed version has been modeled through a stationary random process with isotropic exponential variogram; see, for instance, Bivand et al. [30, Section 8.4]. In view of the latter, our goal has been to apply the proposed tests to determine whether they lead to confirm

the assumed model. As well, a different parametric family was considered, the isotropic spherical variogram, with the aim of checking the performance of our approaches to reject this second model.

By proceeding as in the numerical studies with simulated data, we computed the statistics S_j, by first implementing the required nonparametric variograms and then using them for approximating the model parameters $\hat{\theta}_j$, through the least squares methods, for $j = 1, ..., 7$. Figure 4 plots the nonparametric and the parametric variograms derived for the different statistics S_j, under both null hypothesis (exponential and spherical models). The specific values derived for the model parameter are shown in Table (6).

The p-values associated to the different tests were estimated by applying the gaussian Bootstrap approach. The resulting values are summarized in Table 6. When the exponential variogram is taken as the null hypothesis, no test leads to reject the exponential model at the significance level $\alpha = 5\%$, as expected, since all the p-values are larger than α. Under the assumption of a spherical variogram, again the p-values surpass the target value, although the result derived for the statistic S_5 is considerably smaller and close to the significance level. Then, the test based on S_5 seems a good candidate to take a decision on the validity of a parametric model.

7. Conclusions

This research is focused on deriving the asymptotic distribution of the L_2 deviations of the Nadaraya-Watson variograms, for both the anisotropic and the isotropic settings. These results can be applied to construct goodness of fit tests for the dependence structure of an intrinsic random process, although they require the specification of the model parameters. The latter issue has been addressed through the ordinary and the weighted least squares, because we have proved that the statistics based on them keep the same rate of convergence as the original functionals. To complete the specification of the tests, the critical points must be approximated and the Bootstrap methods provide different alternatives

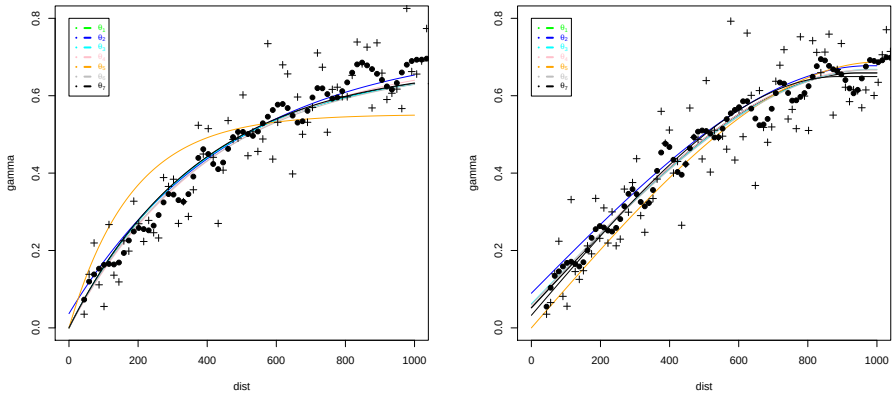

Figure 4: Estimates of the variogram obtained by the Matheron (plus signs) and the Nadaraya-Watson (bullets) approaches, as well as the parametric variograms, with the resulting model parameters $\hat{\theta}_j$, derived for the statistics S_j, for $j = 1, ..., 7$. Left: The exponential model was taken as the null hypothesis. Right: The spherical model was taken as the null hypothesis.

for this purpose.

The numerical studies, developed with simulated data, show the good performance of those procedures based on statistics with weight functions inversely proportional to the variance of the variogram estimators, as given in (8). Indeed, the tests implemented in this way help to decide about the appropriateness of a parametric model, regardless of the following issues: the theoretical variogram used for data generation, the distribution of the random process or the Bootstrap approach considered. The use of simpler weights, such as (11) and (12), provide tests with good behavior, when both the theoretical variogram model and the (gaussian) distribution of the spatial process have been adequately specified; indeed, those weights were derived from the assumptions of gaussianity and variance stabilization of the underlying process.

Anyway, further research is needed to improve our proposals and to extend their application in practice. The latter involves investigating different aspects, addressed to determine optimal selectors of the bandwidth of the variogram estimator or the thresholds involved in the tests, as well as to provide more

	Exponential			Spherical		
	τ	σ_p	ϕ	τ	σ_p	ϕ
$\hat{\theta}_1$	0.000	0.681	384.43	0.052	0.607	946.55
$\hat{\theta}_2$	0.037	0.702	473.66	0.090	0.588	977.36
$\hat{\theta}_3$	0.000	0.680	384.62	0.062	0.606	983.94
$\hat{\theta}_4$	0.000	0.708	426.86	0.060	0.607	966.91
$\hat{\theta}_5$	0.000	0.552	180.86	0.000	0.687	1007.17
$\hat{\theta}_6$	0.000	0.687	404.26	0.060	0.607	984.82
$\hat{\theta}_7$	0.000	0.683	382.90	0.032	0.617	888.15

Table 4: Model parameter estimates $\hat{\theta}_j$ derived for the statistics S_j, for $j = 1, ..., 7$. Left: The exponential model was taken as the null hypothesis. Right: The spherical model was taken as the null hypothesis.

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\theta}_7$
Exponential	0.562	0.636	0.546	0.694	0.334	0.752	0.562
Spherical	0.748	0.570	0.738	0.762	0.078	0.780	0.624

Table 5: Bootstrap p-values approximated for the tests based on S_j, with the derived model parameter estimates $\hat{\theta}_j$, for $j = 1, ..., 7$, under both null hypothesis (exponential and spherical models).

general mechanisms for approximation of the critical points. In fact, an ad hoc Bootstrap approach for estimation of the quantiles would be desirable.

Acknowledgements

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).

References

[1] G. Matheron, Principles of geostatistics, Economic Geology 58 (8) (1963) 1246–1266. doi:10.2113/gsecongeo.58.8.1246.

[2] N. Cressie, Statistics for spatial data, Wiley series in probability and mathematical statistics: Applied probability and statistics, J. Wiley, 1993.

[3] M. Genton, Highly robust variogram estimation, Mathematical Geology 30 (2) (1998) 213–221.

[4] P. Hall, P. Patil, Properties of nonparametric estimators of autocovariance for stationary random fields, Probability Theory and Related Fields 99 (3) (1994) 399–424. doi:10.1007/BF01199899.

[5] P. García-Soidán, W. González-Manteiga, M. Febrero-Bande, Local linear regression estimation of the variogram, Statistics & Probability Letters 64 (2) (2003) 169–179. doi:10.1016/S0167-7152(03)00149-4.

[6] K. Yu, J. Mateu, Nonparametric nearest-neighbour variogram estimation, in: J. Mateu, F. Montes (Eds.), Spatial Statistics through Applications, Witpress, 2002, Ch. 5, pp. 103–125.

[7] R. Menezes, P. García-Soidán, M. Febrero-Bande, A comparison of approaches for valid variogram achievement, Computational Statistics 20 (4) (2006) 623–642. doi:10.1007/BF02741319.

[8] A. Shapiro, J. Botha, Variogram fitting with a general class of conditionally nonnegative definite functions, Computational Statistics and Data Analysis 11 (1) (1991) 87–96. doi:10.1016/0167-9473(91)90055-7.

[9] D. Gorsich, M. Genton, Variogram model selection via nonparametric derivative estimation, Mathematical Geology 32 (3) (2000) 249–270. doi:10.1023/A:1007563809463.

[10] D. Maglione, A. Diblasi, Exploring a valid model for the variogram of an isotropic spatial process, Stochastic Environmental Research and Risk Assessment 18 (6) (2004) 366–376. doi:10.1007/s00477-003-0143-7.

[11] R. Crujeiras, R. Fernández-Casal, W. González-Manteiga, Goodness-of-fit tests for the spatial spectral density, Stochastic Environmental Research and Risk Assessment 24 (1) (2010) 67–79. doi:10.1007/s00477-008-0300-0.

[12] Y. Fan, Testing the goodness of fit of a parametric density function by kernel method, Econometric Theory 10 (2) (1994) 316–356. doi:10.1017/S0266466600008434.

[13] W. Härdle, E. Mammen, Comparing nonparametric versus parametric regression fits, Annals of Statistics 21 (1993) 1926–1947.

[14] W. Müller, Least-squares fitting from the variogram cloud, Statistics and Probability Letters 43 (1) (1999) 93–98.

[15] M. L. Stein, Statistical Interpolation of Spatial Data: Some Theory for Kriging, Springer, 1999.

[16] S. Lahiri, Y. Lee, N. Cressie, On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters, Journal of Statistical Planning and Inference 103 (1-2) (2002) 65–85. doi:10.1016/S0378-3758(01)00198-7.

[17] A. Maity, M. Sherman, Testing for spatial isotropy under general designs, Journal of Statistical Planning and Inference 142 (5) (2012) 1081–1091. doi:10.1016/j.jspi.2011.11.013.

[18] A. Bose, G. Babu, Accuracy of the bootstrap approximation, Probability Theory and Related Fields 90 (3) (1991) 301–316. doi:10.1007/BF01193748.

[19] P. García-Soidán, R. Menezes, O. Rubiños, Bootstrap approaches for spatial data, Stochastic Environmental Research and Risk Assessment 28 (5) (2014) 1207–1219. doi:10.1007/s00477-013-0808-9.

[20] J. Zhu, S. Lahiri, Bootstrapping the empirical distribution function of a spatial process, Statistical Inference for Stochastic Processes 10 (2) (2007) 107–145.

[21] S. Bandyopadhyay, S. Lahiri, Asymptotic properties of discrete fourier transforms for spatial data, Sankhyā: The Indian Journal of Statistics, Series A (2008-) (2009) 221–259.

[22] P. García-Soidán, Asymptotic normality of the nadaraya-watson semivariogram estimators, Test 16 (3) (2007) 479–503. doi:10.1007/s11749-006-0016-8.

[23] S. Das, T. Subba-Rao, G. N. Boshnakov, On the estimation of parameters of variograms of spatial stationary isotropic random processes, Research Report 2, Probability and Statistics Group, School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK (2012).

[24] K. Mardia, R. Marshall, Maximum likelihood estimation of models for residual covariance in spatial regression, Biometrika 71 (1) (1984) 135–146. doi:10.1093/biomet/71.1.135.

[25] N. Cressie, S. Lahiri, The asymptotic distribution of reml estimators, Journal of Multivariate Analysis 45 (2) (1993) 217–233. doi:10.1006/jmva.1993.1034.

[26] M. Verleysen, D. Franois, The curse of dimensionality in data mining and time series prediction, in: J. Cabestany, A. Prieto, F. S. Hernndez (Eds.), IWANN, Vol. 3512 of Lecture Notes in Computer Science, Springer, 2005, pp. 758–770.

[27] P. Lavergne, V. Patilea, Breaking the curse of dimensionality in nonparametric testing, Journal of Econometrics 143 (1) (2008) 103–122. doi:10.1016/j.jeconom.2007.08.014.

[28] Q. Yao, H. Tong, Cross-validatory bandwidth selections for regression estimation based on dependent data, Journal of Statistical Planning and Inference 68 (2) (1998) 387–415.

[29] N. Cressie, Fitting variogram models by weighted least squares, Journal of the International Association for Mathematical Geology 17 (5) (1985) 563–586. doi:10.1007/BF01032109.

[30] R. S. Bivand, E. J. Pebesma, V. Gmez-Rubio, Applied spatial data analysis with R, Springer, New York; London, 2008.

[31] T. Hengl, A Practical Guide to Geostatistical Mapping, 2nd Edition, University of Amsterdam, 2009.

Appendix A. Proof of Theorem 3.1

Firstly, we should check that:

$$\begin{aligned} &\text{Bias}\left[\hat{\gamma}_1(\mathrm{t})\right] = h_d^2 B_1\left(\mathrm{t}\right) + o\left(h_d^2\right) \\ &\text{Cov}\left[\hat{\gamma}_1\left(\mathrm{t}\right), \hat{\gamma}_1(\mathrm{t}')\right] = \lambda^{-d} C_1\left(\mathrm{t}, \mathrm{t}'\right) + o\left(h_d^4 + \lambda^{-d}\right) \end{aligned} \tag{A.1}$$

for $\mathrm{t}, \mathrm{t}' \neq 0$, with:

$$\begin{aligned} B_1\left(\mathrm{t}\right) &= \frac{1}{2} c_{K_d} \sum_{k=1}^{d} \left.\frac{\partial^2 \gamma(\mathrm{t})}{\partial t_k^2}\right|_{\mathrm{t}} \\ C_1\left(\mathrm{t}, \mathrm{t}'\right) &= \frac{\int f_0(\mathrm{r})^4 d\mathrm{r}}{\left(\int f_0(\mathrm{r})^2 d\mathrm{r}\right)^2} \cdot \\ &\quad \cdot \int \text{Cov}\left[\left(Z\left(0\right) - Z\left(\mathrm{t}\right)\right)^2, \left(Z\left(\mathrm{r}\right) - Z\left(\mathrm{t}' + \mathrm{r}\right)\right)^2\right] d\mathrm{r} \end{aligned} \tag{A.2}$$

For the latter purpose, we can bear in mind condition H5, so that the spatial locations have been taken as $\mathrm{s}_i = \lambda \mathrm{r}_i$, where $(\mathrm{r}_1, ..., \mathrm{r}_n)$ represents a realization

of a random sample $(\mathrm{R}_1, ..., \mathrm{R}_n)$ drawn from f_0. The following notation will be used:

$$a_1(\mathrm{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{\lambda\left(\mathrm{R}_i - \mathrm{R}_j\right) - \mathrm{t}}{h_d}\right)$$
$$a_2(\mathrm{t}) = \sum_{i=1}^{n}\sum_{j=1}^{n} K_d\left(\frac{\lambda\left(\mathrm{R}_i - \mathrm{R}_j\right) - \mathrm{t}}{h_d}\right)\left(\gamma\left(\lambda\left(\mathrm{R}_i - \mathrm{R}_j\right)\right) - \gamma(\mathrm{t})\right)$$

together with the fact that:

$$\mathrm{Bias}\left[\hat{\gamma}_1(\mathrm{t})\right] = \mathrm{E}\left[\mathrm{Bias}\left[\hat{\gamma}_1(\mathrm{t})\,/\mathrm{R}_1, ..., \mathrm{R}_n\right]\right] = \mathrm{E}\left[\frac{a_2(\mathrm{t})}{a_1(\mathrm{t})}\right]$$

As proved in García-Soidán [22], the orders given below hold for $\mathrm{t} \neq 0$:

$$a_1(\mathrm{t}) = n^2\lambda^{-d}h_d^d \int f_0(\mathrm{r})^2 d\mathrm{r} + o\left(n^2\lambda^{-d}h_d^d\right) \ \ a.s.$$
$$a_2(\mathrm{t}) = \frac{1}{2}n^2\lambda^{-d}h_d^{d+2}c_{K_d}\sum_{k=1}^{d}\left.\frac{\partial^2\gamma(\mathrm{t})}{\partial t_k^2}\right|_{\mathrm{t}} \int f_0(\mathrm{r})^2 d\mathrm{r} + o\left(n^2\lambda^{-d}h_d^{d+2}\right) \ \ a.s.$$

Then, the first relation in (A.1) follows for $\mathrm{t} \neq 0$.

Now, write:

$$a_3\left(\mathrm{t}, \mathrm{t}'\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} K_d\left(\frac{\lambda\left(\mathrm{R}_i - \mathrm{R}_j\right) - \mathrm{t}}{h_d}\right) K_d\left(\frac{\lambda\left(\mathrm{R}_k - \mathrm{R}_l\right) - \mathrm{t}'}{h_d}\right) \cdot$$
$$\cdot \mathrm{Cov}\left[\left(Z\left(\lambda\mathrm{R}_i\right) - Z\left(\lambda\mathrm{R}_j\right)\right)^2, \left(Z\left(\lambda\mathrm{R}_k\right) - Z\left(\lambda\mathrm{R}_l\right)\right)^2\right]$$

and consider that:

$$\begin{aligned}\mathrm{Cov}\left[\hat{\gamma}_1\left(\mathrm{t}\right), \hat{\gamma}_1\left(\mathrm{t}'\right)\right] &= \mathrm{E}\left[\mathrm{Cov}\left[\hat{\gamma}_1\left(\mathrm{t}\right), \hat{\gamma}_1\left(\mathrm{t}'\right)/\mathrm{R}_1, ..., \mathrm{R}_n\right]\right] + o\left(h_d^4\right) = \\ &= \mathrm{E}\left[\frac{a_3(\mathrm{t}, \mathrm{t}')}{a_1(\mathrm{t})a_1(\mathrm{t}')}\right] + o\left(h_d^4\right)\end{aligned} \tag{A.3}$$

We could check that:

$$a_3\left(\mathrm{t}, \mathrm{t}'\right) = n^4\lambda^{-3d}h_d^{2d}\int f_0(\mathrm{r})^4 d\mathrm{r} \cdot$$
$$\cdot \int \mathrm{Cov}\left[\left(Z\left(0\right) - Z\left(\mathrm{t}\right)\right)^2, \left(Z\left(\mathrm{r}\right) - Z\left(\mathrm{t}' + \mathrm{r}\right)\right)^2\right] d\mathrm{r} + o\left(n^4\lambda^{-3d}h_d^{2d}\right) \ \ a.s.$$

for $\mathrm{t}, \mathrm{t}' \neq 0$, which would yield the second relation in (A.1). Then, by proceeding as in the proof of Theorem 3.1 of Hall and Patil [4], the result derived for the covariance in (A.1) would lead to:

$$\lambda^d \int_{\|\mathrm{t}\| \leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \mathrm{E}\left[\hat{\gamma}_1(\mathrm{t})\right]\right)^2 d\mathrm{t} \xrightarrow{D} \int_{\|\mathrm{t}\| \leq y} X_1(\mathrm{t})^2 d\mathrm{t}$$

for a zero mean gaussian process $\{X_1(\mathrm{t}) : \|\mathrm{t}\| < y\}$, with covariance function C_1.

Finally, the previous relation and the bias achieved for $\hat{\gamma}_1$ would lead to:

$$\begin{aligned}\lambda^d \int_{\|\mathrm{t}\|\leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma(\mathrm{t})\right)^2 d\mathrm{t} = \\ = \lambda^d \int_{\|\mathrm{t}\|\leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \mathrm{E}\left[\hat{\gamma}_1(\mathrm{t})\right] + \mathrm{E}\left[\hat{\gamma}_1(\mathrm{t})\right] - \gamma(\mathrm{t})\right)^2 d\mathrm{t} = \\ = \lambda^d \int_{\|\mathrm{t}\|\leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \mathrm{E}\left[\hat{\gamma}_1(\mathrm{t})\right]\right)^2 d\mathrm{t} + O\left(\lambda^d h_d^4 + \lambda^{d/2} h_d^2\right)\end{aligned}$$

aiming to conclude this proof.

Appendix B. Proof of Theorem 3.2

This proof would be similar to that of Theorem 3.1, provided that the dominant terms of $\mathrm{E}\left[\hat{\gamma}_2(x)\right]$ and $\mathrm{Cov}\left[\hat{\gamma}_2\left(x\right), \hat{\gamma}_2\left(x'\right)\right]$ were derived, for $x, x' > 0$. These issues are addressed below, yielding relations (B.1) and (B.3). We first deal with the bias, by considering that:

$$\mathrm{Bias}\left[\hat{\gamma}_2(x)\right] = \mathrm{E}\left[\mathrm{Bias}\left[\hat{\gamma}_2(x)\,/\mathrm{R}_1, ..., \mathrm{R}_n\right]\right] = \mathrm{E}\left[\frac{a_2(x)}{a_1(x)}\right]$$

for $x > 0$, where:

$$\begin{aligned}a_1(x) &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\lambda\|\mathrm{R}_i - \mathrm{R}_j\| - x}{h_1}\right) \\ a_2(x) &= \sum_{i=1}^{n}\sum_{j=1}^{n} K_1\left(\frac{\lambda\|\mathrm{R}_i - \mathrm{R}_j\| - x}{h_1}\right)\left(\gamma\left(\lambda\|\mathrm{R}_i - \mathrm{R}_j\|\right) - \gamma(x)\right)\end{aligned}$$

From the results in García-Soidán [22], it follows that:

$$\begin{aligned}a_1(x) = n^2\lambda^{-d} h_1 x^{d-1} \int f_0(\mathrm{r})^2 d\mathrm{r} \cdot \\ \cdot \int_0^{\pi} ... \int_0^{\pi} \int_0^{2\pi} J_d\left(\beta_1, ..., \beta_{d-1}\right) d\beta_1 ... d\beta_{d-2} d\beta_{d-1} + o\left(n^2\lambda^{-d} h_1\right) \ \ a.s. \\ a_2(x) = \frac{1}{2} n^2\lambda^{-d} h_1^3 c_{K_1} x^{d-1} \gamma''(x) \int f_0(\mathrm{r})^2 d\mathrm{r} \cdot \\ \cdot \int_0^{\pi} ... \int_0^{\pi} \int_0^{2\pi} J_d\left(\beta_1, ..., \beta_{d-1}\right) d\beta_1 ... d\beta_{d-2} d\beta_{d-1} + o\left(n^2\lambda^{-d} h_1^3\right) \ \ a.s.\end{aligned}$$

with $J_d\left(\beta_1, ..., \beta_{d-1}\right) = \left(\sin\beta_1\right)^{d-2}\left(\sin\beta_2\right)^{d-3} ... \sin\beta_{d-2}$, $\vec{\beta}_i = \left(\vec{\beta}_i^{(1)}, ..., \vec{\beta}_i^{(d)}\right)$.

Consequently:

$$\mathrm{Bias}\left[\hat{\gamma}_2(x)\right] = h_1^2 B_2(x) + o\left(h_1^2\right), \text{ where } B_2(x) = \frac{1}{2} c_{K_1} \gamma''(x) \qquad \text{(B.1)}$$

With regard to the covariance, one has:

$$\begin{aligned}\mathrm{Cov}\left[\hat{\gamma}_2\left(x\right),\hat{\gamma}_2\left(x'\right)\right] &= \mathrm{E}\left[\mathrm{Cov}\left[\hat{\gamma}_2\left(x\right),\hat{\gamma}_2\left(x'\right)/\mathrm{R}_1,...,\mathrm{R}_n\right]\right]+o\left(h_1^4\right)= \\ &= \mathrm{E}\left[\frac{a_3(x,x')}{4a_1(x)a_1(x')}\right]+o\left(h_1^4\right)\end{aligned} \tag{B.2}$$

where:

$$\begin{aligned}a_3\left(x,x'\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}K_1\left(\frac{\lambda\|\mathrm{R}_i-\mathrm{R}_j\|-x}{h_1}\right)K_1\left(\frac{\lambda\|\mathrm{R}_k-\mathrm{R}_l\|-x'}{h_1}\right)\cdot \\ \cdot\mathrm{Cov}\left[\left(Z\left(\lambda\mathrm{R}_i\right)-Z\left(\lambda\mathrm{R}_j\right)\right)^2,\left(Z\left(\lambda\mathrm{R}_k\right)-Z\left(\lambda\mathrm{R}_l\right)\right)^2\right]\end{aligned}$$

We could check that:

$$a_3\left(x,x'\right)=n^4\lambda^{-3d}h_1^2x^{d-1}x'^{d-1}\left(\int f_0(\mathrm{r})^2d\mathrm{r}\right)^2C_2\left(x,x'\right)+o\left(n^4\lambda^{-3d}h_1^2\right)\ a.s.$$

with:

$$\begin{aligned}C_2\left(x,x'\right)=\frac{\int f_0(\mathrm{r})^4d\mathrm{r}}{\left(\int f_0(\mathrm{r})^2d\mathrm{r}\right)^2}\int\int_0^{\pi}...\int_0^{\pi}\int_0^{2\pi}\int_0^{\pi}...\int_0^{\pi}\int_0^{2\pi}\cdot \\ \cdot\mathrm{Cov}\left[\left(Z\left(0\right)-Z\left(x\vec{\beta}_1\right)\right)^2,\left(Z\left(\mathrm{r}\right)-Z\left(\mathrm{r}+x\vec{\beta}_2\right)\right)^2\right]\cdot \\ \cdot J_d\left(\beta_{1,1},...,\beta_{d-1,1}\right)J_d\left(\beta_{1,2},...,\beta_{d-1,2}\right)\cdot \\ \cdot d\mathrm{r}d\beta_{1,1}...d\beta_{d-2,1}d\beta_{d-1,1}d\beta_{1,2}...d\beta_{d-2,2}d\beta_{d-1,2}\end{aligned}$$

where $\vec{\beta}_i^{(j)}=\cos\left(\beta_{j,i}\right)\prod_{k=0}^{j-1}\sin\left(\beta_{k,i}\right)$, for $i=1,2$ and $j=1,...,d$, and $\sin\left(\beta_{0,i}\right)=\cos\left(\beta_{d,i}\right)=1$. Therefore:

$$\mathrm{Cov}\left[\hat{\gamma}_2\left(x\right),\hat{\gamma}_2\left(x'\right)\right]=\lambda^{-d}C_2\left(x,x'\right)+o\left(\lambda^{-d}+h_1^4\right) \tag{B.3}$$

Appendix C. Proof of Theorem 3.3

In the proofs of Theorems 3.1 and 3.2, we have derived the exact orders of the bias and the variance of the kernel variogram estimators, which amount to $h_{\delta_k}^2$ and λ^{-d}, respectively. Furthermore, the normal limit distribution of the kernel variogram estimators has been established in García-Soidán [22]. Then, we could apply analogue arguments as those used in Theorem 3.2 in Lahiri et al. [16], to check that $\lambda^{d/2}(\hat{\theta}_{OLS,k}-\theta_0)$ asymptotically converges to a zero mean normal distribution, where $\hat{\theta}_{OLS,k}$ is defined in (5), for $k=$

1, 2. Next we go into details in order to characterize the dominant terms of $\text{Bias}[\hat{\theta}^{(j)}_{OLS,k}]$ and $\text{Cov}[\hat{\theta}^{(j_1)}_{OLS,k}, \hat{\theta}^{(j_2)}_{OLS,k}]$ for $k = 1$. The results for isotropy ($k = 2$) would be derived in a similar way.

If anisotropy is assumed, the least squares criterion for deriving $\hat{\theta}_{OLS,1}$ leads to:

$$\int_{\|\mathrm{t}\| \leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t}) \right) \left. \frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j)}} \right|_{\hat{\theta}_{OLS,1}} d\mathrm{t} = 0$$

for $j = 1, ..., p$. Therefore, Taylor expanding about θ_0, it follows that:

$$\begin{aligned} \sum_{j_1=1}^{p} \left(\hat{\theta}^{(j_1)}_{OLS,1} - \theta_0^{(j_1)} \right) \int_{\|\mathrm{t}\| \leq y} \left. \frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j)}} \right|_{\theta_0} \left. \frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j_1)}} \right|_{\theta_0} d\mathrm{t} \approx \\ \approx \int_{\|\mathrm{t}\| \leq y} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\theta_0}(\mathrm{t}) \right) \left. \frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j)}} \right|_{\theta_0} d\mathrm{t} \end{aligned}$$

or, equivalently, in matrix notation:

$$\begin{aligned} \left(\hat{\theta}_{OLS,1} - \theta_0 \right) &= \begin{pmatrix} \hat{\theta}^{(1)}_{OLS,1} - \theta_0^{(1)} \\ ... \\ \hat{\theta}^{(p)}_{OLS,1} - \theta_0^{(p)} \end{pmatrix} \approx \\ &\approx A\left(\theta_0\right)^{-1} \begin{pmatrix} \int_{\|\mathrm{t}\| \leq z} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\theta_0}(\mathrm{t}) \right) \left. \frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(1)}} \right|_{\theta_0} d\mathrm{t} \\ ... \\ \int_{\|\mathrm{t}\| \leq z} \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\theta_0}(\mathrm{t}) \right) \left. \frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(p)}} \right|_{\theta_0} d\mathrm{t} \end{pmatrix} \end{aligned} \tag{C.1}$$

under the null hypothesis in (1). Then, the desired results for $\text{Bias}[\hat{\theta}^{(j)}_{OLS,1}]$ and $\text{Cov}[\hat{\theta}^{(j_1)}_{OLS,1}, \hat{\theta}^{(j_2)}_{OLS,1}]$ could be obtained from (A.1) and (C.1).

Appendix D. Proof of Theorem 3.4

This proof could be developed similarly as those of Theorems 1 and 2, by deriving the dominant terms of the bias and the covariance involved in each case. Indeed, for the anisotropic setting ($k = 1$), we could check that:

$$\begin{aligned} &\mathrm{E}\left[\hat{\gamma}_1(\mathrm{t}) - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t}) \right] = O\left(h_d^2\right) \\ &\text{Cov}\left[\hat{\gamma}_1(\mathrm{t}) - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t}), \hat{\gamma}_1(\mathrm{t}') - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t}') \right] = O\left(\lambda^{-d}\right) \end{aligned} \tag{D.1}$$

The first equality in (D.1) follows straightforwardly from the results in Theorems 3.1 and 3.3. On the other hand, Taylor expanding about θ_0 yields that:

$$\begin{aligned}
&\left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t})\right)\left(\hat{\gamma}_1(\mathrm{t}') - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t}')\right) = \\
&\quad = \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\theta_0}(\mathrm{t}) + \gamma_{\theta_0}(\mathrm{t}) - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t})\right) \cdot \\
&\qquad \cdot \left(\hat{\gamma}_1(\mathrm{t}') - \gamma_{\theta_0}(\mathrm{t}') + \gamma_{\theta_0}(\mathrm{t}') - \gamma_{\hat{\theta}_{OLS,1}}(\mathrm{t}')\right) \approx \\
&\quad \approx \left(\hat{\gamma}_1(\mathrm{t}) - \gamma_{\theta_0}(\mathrm{t}) - \sum_{j_1} \left.\frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j_1)}}\right|_{\theta_0} \left(\hat{\theta}^{(j_1)}_{OLS,1} - \theta_0^{(j_1)}\right)\right) \cdot \\
&\qquad \cdot \left(\hat{\gamma}_1(\mathrm{t}') - \gamma_{\theta_0}(\mathrm{t}') - \sum_{j_2} \left.\frac{\partial \gamma_\theta(\mathrm{t})}{\partial \theta^{(j_2)}}\right|_{\theta_0} \left(\hat{\theta}^{(j_2)}_{OLS,1} - \theta_0^{(j_2)}\right)\right)
\end{aligned}$$

Then, the order established in the second equality of (D.1) holds from the previous relation and (C.1), together with the covariances obtained in the proofs of Theorems 3.1 and 3.2.

Analogue arguments could be applied for the isotropic case ($k = 2$).