A computational validation for nonparametric assessment of spatial trends

Meilán-Vila, A. · **Fernández-Casal, R.** ·
Crujeiras, R.M. · **Francisco-Fernández, M.**

Abstract The analysis of continuously spatially varying processes usually considers two sources of variation, namely, the large-scale variation collected by the trend of the process, and the small-scale variation. Parametric trend models on latitude and longitude are easy to fit and to interpret. However, the use of parametric models for characterizing spatially varying processes may lead to misspecification problems if the model is not appropriate. Recently, Meilán-Vila et al. (2020) proposed a goodness-of-fit test based on an L_2-distance for assessing a parametric trend model with correlated errors, under random design, comparing parametric and nonparametric trend estimates. The present work aims to provide a detailed computational analysis of the behavior of this approach using different bootstrap algorithms for calibration, one of them including a procedure that corrects the bias introduced by the direct use of the residuals in the variogram estimation, under a fixed design geostatistical framework. Asymptotic results for the test are provided and an extensive simulation study, considering complexities that usually arise in geostatistics, is carried out to illustrate the performance of the proposal. Specifically, we analyze the impact of the sample size,

Andrea Meilán-Vila
Research group MODES, Department of Mathematics, CITIC
Universidade da Coruña
E-mail: andrea.meilan@udc.es

Rubén Fernández-Casal
Research group MODES, Department of Mathematics, CITIC
Universidade da Coruña
E-mail: rfcasal@udc.es

Rosa M. Crujeiras
Research group MODESTYA, Department of Statistics, Mathematical Analysis and Optimization
Universidade de Santiago de Compostela
E-mail: rosa.crujeiras@usc.es

Mario Francisco-Fernández
Research group MODES, Department of Mathematics, CITIC
Universidade da Coruña
E-mail: mariofr@udc.es

the spatial dependence range and the nugget effect on the empirical calibration and power of the test.

Keywords Parametric spatial trends · Bootstrap algorithm · Nonparametric fit · Goodness-of-fit test · Bias correction

1 Introduction

Continuously varying spatial processes are usually described through the analysis of their trend and their dependence structure. The trend component captures the large-scale variability of the process, usually modeled by parametric functions on latitude and longitude (the so-called trend-surface models), and possibly altitude (see Cressie, 1993). For a proper trend estimation, the dependence structure of the process (although not being of primary interest) must be accounted for, usually employing iterative least squares procedures or maximum likelihood approaches under Gaussian and stationary assumptions (see, for instance, Diggle et al., 2010; Cressie, 1993).

However, the consideration of an inadequate parametric trend model may lead to wrong conclusions on the process behavior. In a regression setting with independent errors, to prevent misspecification of the regression function, different testing procedures for assessing a certain parametric regression model have been introduced. For instance, the proposals by Härdle and Mammen (1993); Alcalá et al. (1999); Opsomer and Francisco-Fernández (2010) and Li (2005) are based on the comparison of a parametric and a nonparametric estimator using an L_2-distance. Following this idea, Meilán-Vila et al. (2020) introduced a testing procedure to check if the trend function of a spatial process belongs to a certain class of parametric models. The authors considered a multiple regression model with random design and spatially correlated errors. They used the local linear estimator as the nonparametric fit. In the present paper, focused on a geostatistical framework with fixed design, a thorough analysis of the behavior of a similar test considering different trend models, sample sizes and dependence patterns, is provided. In this case, for simplicity, the Nadaraya–Watson estimator is employed as the nonparametric fit.

The proposed test statistic shows a slow rate of convergence to its asymptotic distribution, motivating the use of resampling methods to approximate its distribution under the parametric null hypothesis. It should be noted that, in order to mimic the process behavior under the null hypothesis, not only the parametric form of the trend must be considered, but also the spatial dependence of the data. This has to be recovered from a single realization of the spatial process under certain stationarity conditions. In the presence of spatial correlation, resampling methods may not be accurate enough for mimicking the spatial dependence structure under the null hypothesis from a single realization of the process. This is the reason why a thorough analysis of the impact of the spatial dependence configuration in the distribution approximation of the test is required and provided in this work.

Traditional resampling procedures for test calibration designed for independent data should not be used for spatial processes, as they do not account for the correlation structure. One of the aims of this paper is to present and analyze three different proposals for test calibration which take the dependence of the data into account:

a parametric residual bootstrap (PB), a nonparametric residual bootstrap (NPB) and a bias-corrected nonparametric bootstrap (CNPB). Parametric bootstrap procedures, following the ideas introduced by Solow (1985), are a usual strategy in geostatistics, since they directly involve the dependence structure (see, for example, Olea and Pardo-Iguzquiza, 2011). In the PB approach, the residuals are obtained from a parametric trend fit and the spatial dependence structure is estimated parametrically. If the assumption that the trend function belongs to the parametric family considered in the null hypothesis holds, then the residuals obtained with this approach will be *similar* to the theoretical errors, and it is expected that the PB method presents a good performance. A possible drawback of this procedure is the misspecification of the parametric model selected for the dependence estimation, although this issue could be avoided by using a nonparametric dependence estimator. In the NPB method, to increase the power of the test, the residuals are obtained from the nonparametric fit (see González-Manteiga and Cao, 1993) and the dependence structure is estimated without considering parametric assumptions. It is clear that the NPB resampling method avoids the misspecification problem both for the trend and the dependence. However, no matter the method used to remove the trend, either parametric or nonparametric, the direct use of residuals gives rise to biased variogram estimates, especially at large lags (see Cressie, 1993, Section 3.4.3), an issue that is corrected with the CNPB approach. This procedure is a modification of the NPB method, including a bias-corrected algorithm for the dependence estimation (see Fernández-Casal and Francisco-Fernández, 2014; Castillo-Páez et al., 2019).

This paper is organized as follows. In Section 2, the parametric and nonparametric trend estimation methods employed in our testing procedure are briefly described. The L_2-test statistic measuring the discrepancy between both fits, as well as its asymptotic distribution (under the null hypothesis and under local alternatives) are also included in Section 2. A detailed description of the calibration algorithms considered is given in Section 3. An exhaustive simulation study to assess the performance of the test, when the PB, NPB and CNPB resampling approaches are used, is presented is Section 4. Finally, Section 5 includes some discussion and further considerations.

2 Inference for spatial trends

Important issues frequently addressed by spatial modeling are the estimation of a surface/map describing the trend of the process and the prediction of the variable of interest at certain unobserved locations. To deal with these problems, traditional approaches in geostatistics consist in assuming (generally simple) parametric models for the trend and, then, reconstructing the whole trend surface using parametric techniques and making predictions by spatial interpolation methods, such as kriging (see Cressie, 1993, Section 3). However, in some situations, the complex interactions between the possible factors affecting the variable of interest make it difficult to write down a simple parametric model for its trend over a large geographic region. In the absence of such a model, the trend can be represented as a smooth spatial function, establishing the relation between the process values and the location coordinates (latitude, longitude and maybe altitude in a three-dimensional setting). This

approach provides a useful first step to characterize important features of the variable of interest, or can be helpful in the development of more complete models including additional significant covariates.

Consider a real-valued spatial process $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathbb{R}^d\}$, observed at fixed locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$. From a model-based perspective (see Diggle and Ribeiro, 2007), the spatial process can be assumed to be decomposed as:

$$Z_i = m(\boldsymbol{s}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

being $Z_i = Z(\mathbf{s}_i)$, with $i = 1, \ldots, n$, a realization of the process at a collection of locations within the observation domain. The trend of the process is given by m, which is an unknown (but smooth) function modeling the expectation of the process, and ε_i denotes the error at location $\boldsymbol{s}_i$, for $i = 1, \ldots, n$, so these values can be viewed as a realization of a spatially varying error process. In order to estimate the trend in (1) from a single realization of the process, stationary conditions must be assumed. Usually, the error process is supposed to be zero-mean with covariance structure satisfying:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \rho_n(\boldsymbol{s}_i - \boldsymbol{s}_j), \quad i, j = 1, \ldots, n, \tag{2}$$

with σ^2 being the variance of the process and ρ_n a continuous stationary correlation function satisfying $\rho_n(\mathbf{0}) = 1$, $\rho_n(\boldsymbol{s}) = \rho_n(-\boldsymbol{s})$, and $|\rho_n(\boldsymbol{s})| \leq 1$, $\forall \boldsymbol{s}$. The subscript n in ρ_n allows the correlation function to shrink as $n \to \infty$. In a spatial context, the dependence structure is typically characterized through the variogram function, γ_n, which satisfies that $\gamma_n(\boldsymbol{s}) = \sigma^2[1 - \rho_n(\boldsymbol{s})]$, $\forall \boldsymbol{s} \in \mathbb{R}^d$. For simplicity, the subscript n will be sometimes omitted. In the previous expression for the covariance of the errors (2), it is supposed that there is no nugget effect. Otherwise, the variance of the errors is written as the sum of two terms, $\text{Var}(\varepsilon) = \sigma^2 = c_0 + c_e$, the nugget effect ($c_0$) and the partial sill ($c_e$), and $\text{Cov}(\varepsilon_i, \varepsilon_j) = c_e \rho_n(\boldsymbol{s}_i - \boldsymbol{s}_j)$, if $i \neq j$. In what follows, the covariance matrix of the errors is denoted by $\boldsymbol{\Sigma}$, being $\Sigma(i, j) = \text{Cov}(\varepsilon_i, \varepsilon_j)$ its (i, j)-entry. For the sake of simplicity, no nugget is considered in the theoretical result given in Section 2.2. However, its effect is analyzed in the simulation study presented in Section 4.

In model (1), the trend function m can be characterized using parametric or nonparametric models. Parametric models are easy to compute and allow for a direct interpretation of the parameter values (e.g. variation of the process along latitude and longitude). On the other hand, nonparametric models also provide a global view of the large-scale behavior of the process. Their flexibility allows to model complex relations beyond a parametric form. Therefore, a question of interest in spatial modeling is focused on characterizing the large-scale variability of the process Z, checking if the trend function belongs to a parametric family by solving the following testing problem:

$$H_0 : m \in \mathscr{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \mathscr{B}\} \quad \text{vs.} \quad H_a : m \notin \mathscr{M}_{\boldsymbol{\beta}}, \tag{3}$$

with $\mathscr{B} \subset \mathbb{R}^p$ a compact set, and p denoting the dimension of the parameter space $\mathscr{B}$.

A test statistic to address (3) is proposed and studied in this paper. Following similar ideas to those in Meilán-Vila et al. (2020), the proposed test consists in using a nonparametric fit as a pilot estimator to assess if a certain parametric family is suitable

for fitting the observed data, comparing with an L_2-distance the nonparametric fit with a parametric one.

In the following section, the parametric and the nonparametric estimators of the spatial trend m, in model (1), used in our L_2-test statistic are described. Subsequently, the asymptotic distribution of this test is derived and its empirical performance is analyzed in a comprehensive simulation study, under different spatial dependence scenarios.

2.1 Spatial trend estimation

Spatial trend estimation in (1) can be performed parametrically by different methods, being least squares and maximum likelihood approaches the most frequently used (Diggle and Ribeiro, 2007). Next, we briefly describe the parametric least squares trend estimator used in our test statistic. On the other hand, nonparametric methods can also be employed for this task. Among the different alternatives, the multivariate Nadaraya–Watson estimator is applied in the goodness-of-fit test proposed. This nonparametric trend estimator is also formulated and discussed below in the context of this paper.

Parametric estimation

As pointed out previously, the goodness-of-fit test proposed in this paper makes use of a parametric estimator of the trend function. As it will be remarked in the next section, the test statistic can be applied considering any parametric estimator of m satisfying a $\sqrt{n}$-consistency property. Specifically, if $m_{\boldsymbol{\beta}_0}$ denotes the "true" regression function under the null hypothesis, and $m_{\hat{\boldsymbol{\beta}}}$ the corresponding parametric estimator, it is required that the difference $m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}) - m_{\boldsymbol{\beta}_0}(\boldsymbol{s}) = O_p(n^{-1/2})$ uniformly in $\boldsymbol{s}$. A suitable parametric estimator satisfying this property is, for example, the one introduced by Crujeiras and Van Keilegon (2010) for nonlinear trends. The steps of the parametric estimation method employed for the practical application of the test are the following:

1. Obtain an initial estimator of $\boldsymbol{\beta}$ by least squares regression:

$$\tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(\boldsymbol{Z} - \boldsymbol{m}_{\boldsymbol{\beta}})^{\top}(\boldsymbol{Z} - \boldsymbol{m}_{\boldsymbol{\beta}}), \tag{4}$$

   where $\boldsymbol{Z} = (Z_1, \ldots, Z_n)^{\top}$ and $\boldsymbol{m}_{\boldsymbol{\beta}} = [m_{\boldsymbol{\beta}}(\boldsymbol{s}_1), \ldots, m_{\boldsymbol{\beta}}(\boldsymbol{s}_n)]^{\top}$.
2. Using the residuals obtained with the estimation in (4), $\tilde{\varepsilon}_i = Z_i - m_{\tilde{\boldsymbol{\beta}}}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$, estimate the covariance matrix of the errors, $\tilde{\boldsymbol{\Sigma}}$.
3. Update the regression parameter estimates, introducing the estimated covariance matrix $\tilde{\boldsymbol{\Sigma}}$ in the least squares minimization problem:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}(\boldsymbol{Z} - \boldsymbol{m}_{\boldsymbol{\beta}})^{\top}\tilde{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{Z} - \boldsymbol{m}_{\boldsymbol{\beta}}). \tag{5}$$

Finally, take $m_{\hat{\boldsymbol{\beta}}}$ as the parametric estimator for the regression function.

Covariance matrix estimation in Step 2 could be carried out using different methods. Firstly, using a parametric methodology and assuming that the variogram belongs to a valid parametric family $\{2\gamma_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^q\}$, a parameter estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ can be obtained. Following a classical approach, $\boldsymbol{\theta}$ could be estimated by fitting the parametric model considered for the variogram to a set of empirical semivariogram estimates, computed using the residuals $\tilde{\varepsilon}_i$, applying the weighted least squares method (Cressie, 1985). With this parametric approximation, the variance-covariance matrix of the errors can be denoted by $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$, and replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$, a parametric estimation of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ (denoted by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$) can be obtained.

On the other hand, instead of using a parametric approach, flexible nonparametric estimators can be employed to approximate the dependence structure, avoiding misspecification problems. For instance, an estimate of the variogram of the residuals could be obtained as follows. First, compute a nonparametric pilot variogram estimator (Hall and Patil, 1994). A first attempt could be to use the empirical semivariogram estimator. Nevertheless, in practice, empirical variogram estimates could be deficient. For instance, the assumption of isotropy (or geometric anisotropy) could be not appropriate (Fernández-Casal et al., 2003a). Therefore, it would be desirable to have models with enough flexibility. Nonparametric kernel semivariogram estimators could be used instead, producing significantly better results than those obtained with the empirical estimator (Fernández-Casal et al., 2003b). However, these estimators do not necessarily satisfy the conditionally negative definiteness property of a valid semivariogram. For that reason, a valid model should be fitted to the nonparametric pilot estimates. For example, a flexible Shapiro–Botha variogram approach (Shapiro and Botha, 1991) could be employed at this step. The combination of the Shapiro–Botha method with nonparametric kernel semivariogram pilot estimation provides an efficient variogram estimator which can be used to estimate the corresponding covariance matrix.

Nonparametric estimation

Kernel methods can also be used to estimate the trend function m in model (1), providing more flexible approaches than the parametric fits. In this work, a multivariate Nadaraya–Watson estimator (Härdle and Müller, 2012; Liu, 2001) is considered. For a certain location $\boldsymbol{s} \in D$, this estimator is given by:

$$\hat{m}_{\boldsymbol{H}}^{NW}(\boldsymbol{s}) = \frac{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) Z_i}{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})}, \tag{6}$$

where $K_{\boldsymbol{H}}(\boldsymbol{s}) = |\boldsymbol{H}|^{-1} K(\boldsymbol{H}^{-1}\boldsymbol{s})$ is the rescaled version of a multivariate kernel function K and $\boldsymbol{H}$ is a $d \times d$ symmetric positive definite matrix. The kernel function K can be obtained as the product of univariate kernels (see, for example, Wand and Jones, 1994; Fan and Gijbels, 1996). The bandwidth matrix $\boldsymbol{H}$ controls the shape and the size of the local neighborhood used to estimate m at a location $\boldsymbol{s}$, and its selection plays an important role in the estimation process. In the presence of spatially correlated errors, traditional data-driven bandwidth selection methods, such as cross-validation and generalized cross-validation, fail to provide good bandwidth values (Opsomer

et al., 2001). Asymptotic results for (6) as well as the proposal of different bandwidth selection methods under the assumption of spatially correlated errors can be found in Liu (2001), extending the results for independent data given in Härdle and Müller (2012).

The estimator given in (6) can be seen as a particular case of a wider class of nonparametric estimators, the so-called local polynomial estimators, assuming that the polynomial degree is equal to zero (local constant). Since this work aims to provide a deeper analysis of the practical performance of a version of the test studied in Meilán-Vila et al. (2020), the local constant fit was chosen given its reduced computational cost compared with other nonparametric approaches.

2.2 Trend model assessment

Following the ideas by Härdle and Mammen (1993) and Alcalá et al. (1999), in Meilán-Vila et al. (2020), the testing problem (3) is addressed by constructing a weighted L_2-test statistic, comparing a smoothed version of a parametric and a nonparametric regression estimates. These authors considered a model like (1), but assuming a random design. As the parametric fit, they used the least squares estimator described in Section 2.1, estimating the covariance matrix of the errors in Step 2 of that algorithm using a parametric approach. As the nonparametric fit, they employed the local linear estimator. In the present paper, we assume a geostatistical fixed design and consider the same type of testing approach. As the parametric fit, we also employ the method described in Section 2.1, but the Nadaraya–Watson estimator introduced in (6) is used as the nonparametric fit. Specifically, the test statistic considered is given by:

$$T_n = n|\boldsymbol{H}|^{1/2} \int_D \left[\hat{m}_{\boldsymbol{H}}^{NW}(\boldsymbol{s}) - \hat{m}_{\boldsymbol{H},\hat{\boldsymbol{\beta}}}^{NW}(\boldsymbol{s}) \right]^2 w(\boldsymbol{s}) d\boldsymbol{s}, \tag{7}$$

where w is a weight function that helps in mitigating possible boundary effects and $\hat{m}_{\boldsymbol{H},\hat{\boldsymbol{\beta}}}^{NW}$ is a smoothed version of $m_{\hat{\boldsymbol{\beta}}}$, which is defined by:

$$\hat{m}_{\boldsymbol{H},\hat{\boldsymbol{\beta}}}^{NW}(\boldsymbol{s}) = \frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i)}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})}. \tag{8}$$

If the null hypothesis holds, then the parametric and nonparametric estimators in (7) will tend to be *similar* and the value of T_n will be *small*. Conversely, if the null hypothesis does not hold, major differences between both fits are expected and, therefore, the value of T_n will be *large*. So, H_0 will be rejected if the distance between both fits exceeds a critical value. For example, as a visual illustration of the performance of the test, suppose that a sample of size $n = 400$ is generated on a regular grid in the unit square $D = [0,1] \times [0,1]$, following model (1), with trend function (9) and $c = 0$. The random errors ε_i are normally distributed with zero mean and covariance function (11), with values $\sigma^2 = 0.16$, $c_0 = 0.04$ and $a_e = 0.6$. If we want to test if $m(\boldsymbol{s}) \in \{\beta_0 + \beta_1(s_1 - 0.5)^3, \beta_0, \beta_1 \in \mathbb{R}\}$, being $\boldsymbol{s} = (s_1, s_2)$, using the test statistic given in (7), both $\hat{m}_{\boldsymbol{H}}^{NW}$ and $\hat{m}_{\boldsymbol{H},\hat{\boldsymbol{\beta}}}^{NW}$ fits must be computed. Figure 1 shows the

theoretical spatial trend (top left panel), the simulated observations of the spatial process (top right panel), the Nadaraya–Watson trend estimation (bottom left panel) and the smoothed version of the parametric fit (bottom right panel). A multiplicative triweight kernel and the optimal bandwidth obtained by minimizing the mean average squared error (MASE) of the Nadaraya–Watson estimator (see Francisco-Fernandez and Opsomer, 2005, p. 288) are considered for $\hat{m}_{\boldsymbol{H}}^{NW}$ and $\hat{m}_{\boldsymbol{H},\hat{\boldsymbol{\beta}}}^{NW}$. In this case, from a visual comparison, one may argue that given that both estimates at bottom left and right panels are very similar, the value of the test statistic T_n would presumably be *small*, and consequently, there may be no evidences against the assumption of parametric trend $m_{\boldsymbol{\beta}}(\boldsymbol{s}) = \beta_0 + \beta_1(s_1 - 0.5)^3$. However, apart from getting some insight to what might occur when using exploratory methods, in order to formally test the model using T_n, given in (7), it is essential to approximate the distribution of the test statistic under the null hypothesis.

A result providing the asymptotic distribution of the test statistic T_n under the null hypothesis and also under local alternatives is provided below. Specifically, we consider alternatives of the form $m(\boldsymbol{s}) = m_{\boldsymbol{\beta}_0}(\boldsymbol{s}) + c_n g(\boldsymbol{s})$, where c_n is a sequence, such that $c_n \to 0$ and g is a deterministic function collecting the deviation direction from the null model. Additionally, it is assumed that the function g is bounded (uniformly in $\boldsymbol{s}$ and n) and $c_n = n^{-1/2}|\boldsymbol{H}|^{-1/4}$. In particular, this contains the null hypothesis when $g(\boldsymbol{s}) = 0$. The following assumptions are required:

(A1) The regression function m is twice continuously differentiable.

(A2) The weight function w is continuously differentiable.

(A3) For the correlation function ρ_n, there exist two constants ρ_M and ρ_c, such that, $n \int |\rho_n(\boldsymbol{s})| d\boldsymbol{s} < \rho_M$ and $\lim_{n\to\infty} n \int \rho_n(\boldsymbol{s}) d\boldsymbol{s} = \rho_c$. Moreover, for any sequence $\varepsilon_n > 0$ satisfying $n^{1/d}\varepsilon_n \to \infty$,

$$n \int_{\|\boldsymbol{s}\| \geq \varepsilon_n} |\rho_n(\boldsymbol{s})| d\boldsymbol{s} \to 0 \quad \text{as} \quad n \to \infty.$$

(A4) For any i, j, k, l,

$$\text{Cov}(\varepsilon_i\varepsilon_j, \varepsilon_k\varepsilon_l) = \text{Cov}(\varepsilon_i, \varepsilon_k)\text{Cov}(\varepsilon_j, \varepsilon_l) + \text{Cov}(\varepsilon_i, \varepsilon_l)\text{Cov}(\varepsilon_j, \varepsilon_k).$$

(A5) The errors are a geometrically strong mixing sequence with mean zero and $\mathbb{E}|\varepsilon(\boldsymbol{s})|^r < \infty$ for all $r > 4$.

(A6) The kernel K is a spherically symmetric density function, twice continuously differentiable and with compact support (for simplicity with a nonzero value only if $\|\boldsymbol{u}\| \leq 1$). Moreover, $\int \boldsymbol{u}\boldsymbol{u}^\top K(\boldsymbol{u}) d\boldsymbol{u} = \mu_2(K)\boldsymbol{I}_d$, where $\mu_2(K)$ is a constant real value different from zero and $\boldsymbol{I}_d$ is the $d \times d$ identity matrix.

(A7) K is Lipschitz continuous. That is, there exists $\mathfrak{L} > 0$, such that

$$|K(\boldsymbol{s}_1) - K(\boldsymbol{s}_2)| \leq \mathfrak{L}\|\boldsymbol{s}_1 - \boldsymbol{s}_2\|, \quad \forall \boldsymbol{s}_1, \boldsymbol{s}_2 \in D.$$

(A8) The bandwidth matrix $\boldsymbol{H}$ is symmetric and positive definite, with $\boldsymbol{H} \to 0$ and $n|\boldsymbol{H}|\lambda_{\min}^2(\boldsymbol{H}) \to \infty$, when $n \to \infty$. The ratio $\lambda_{\max}(\boldsymbol{H})/\lambda_{\min}(\boldsymbol{H})$ is bounded above, where $\lambda_{\max}(\boldsymbol{H})$ and $\lambda_{\min}(\boldsymbol{H})$ are the maximum and minimum eigenvalues of $\boldsymbol{H}$, respectively.
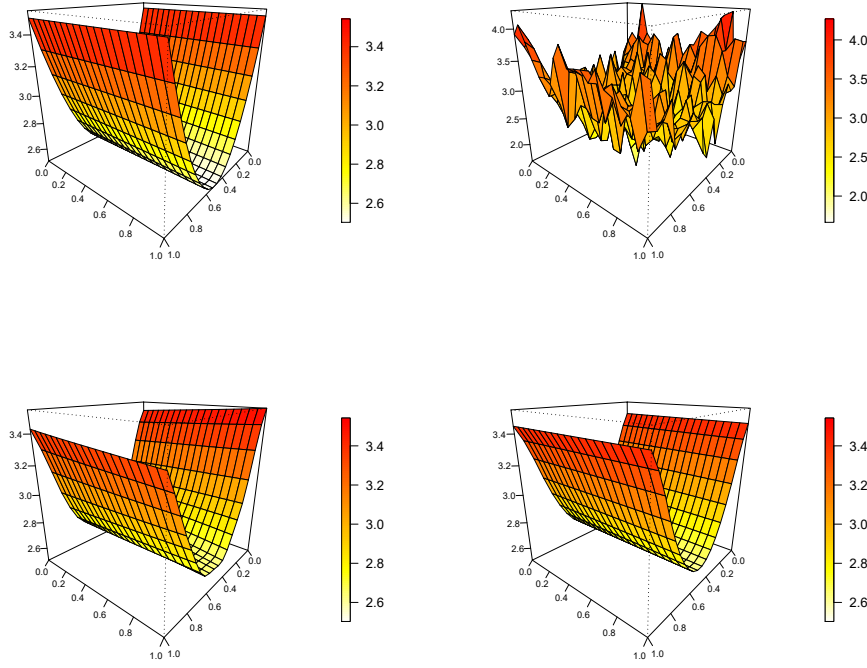
Fig. 1 Spatial trend (top left), spatial process (top right), Nadaraya–Watson trend estimator (bottom left) and smoothed version of the parametric fit (bottom right). Sample of size $n = 400$ generated on a regular grid in the unit square $D = [0,1] \times [0,1]$, following model (1), with trend function $m(\boldsymbol{s}) = 2.5 + 4(s_1 - 0.5)^3$, $\boldsymbol{s} = (s_1, s_2)$, and exponential covariance structure with $\sigma^2 = 0.16$, $c_0 = 0.04$ and $a_e = 0.6$

Assumption (A3) implies that the correlation function depends on n, and the integral $\int |\rho_n(\boldsymbol{s})| d\boldsymbol{s}$ should vanish as $n \to \infty$. The vanishing speed should not be slower than $O(n^{-1})$. This assumption also implies that the integral of $|\rho_n(\boldsymbol{s})|$ is dominated by the values of $\rho_n(\boldsymbol{s})$ near to the origin $\boldsymbol{0}$. Hence, the correlation is short-range and decreases as $n \to \infty$. This can be considered as a case of increasing-domain spatial asymptotics (see Cressie, 1993), since this setup can be transformed to one in which the correlation function ρ_n is fixed with respect to the sample size, but the support D for $\boldsymbol{s}$ expands. The current setup with fixed domain D and shrinking ρ_n is more natural to consider when the primary purpose of the estimation is a fixed regression function m defined over a spatial domain, not the correlation function itself. Two examples of commonly used correlation functions that satisfy the conditions of assumption (A3) are the exponential and rational quadratic models (see Cressie, 1993). Assumption (A4) is satisfied, for example, for Gaussian errors. Assumption (A5) is needed to apply the central limit theorem for reduced U-statistics under dependence given by Kim et al. (2013). In assumption (A8), $\boldsymbol{H} \to 0$ means that every entry of $\boldsymbol{H}$ goes to 0. Moreover, since $\boldsymbol{H}$ is symmetric and positive definite, $\boldsymbol{H} \to 0$ is equivalent to $\lambda_{\max}(\boldsymbol{H}) \to 0$. It can also be deduced that $|\boldsymbol{H}|$ is a quantity of order $O[\lambda_{\max}^d(\boldsymbol{H})]$ given that $|\boldsymbol{H}|$ is equal to the product of all eigenvalues of $\boldsymbol{H}$.

Regarding the parametric estimator, the assumption of being a $\sqrt{n}$-consistent estimator is required. This is guaranteed if the least squares estimator $m_{\hat{\beta}}$ described in Section 2.1 is used in the statistic (7). A different parametric estimator of the trend function could be used as long as the $\sqrt{n}$-consistency property is fulfilled.

The following theorem provides the asymptotic distribution of the proposed test statistic T_n given in (7). The proof of the result can be found in the final Appendix.

Theorem 1 *Under Assumptions* (A1)–(A8)*, and if* $0 < V < \infty$*, it can be proved that*

$$V^{-1/2}(T_n - b_{0\boldsymbol{H}} - b_{1\boldsymbol{H}}) \to_{\mathscr{L}} N(0,1) \text{ as } n \to \infty$$

where $\to_{\mathscr{L}}$ *denotes convergence in distribution, with*

$$b_{0\boldsymbol{H}} = |\boldsymbol{H}|^{-1/2}\sigma^2 K^{(2)}(\boldsymbol{0}) \left[\int w(\boldsymbol{s})d\boldsymbol{s} + \rho_c \int w(\boldsymbol{s})d\boldsymbol{s}\right],$$

$$b_{1\boldsymbol{H}} = \int \left[K_{\boldsymbol{H}} * g(\boldsymbol{s})\right]^2 w(\boldsymbol{s})d\boldsymbol{s},$$

and

$$V = \sigma^4 K^{(4)}(0) \int w^2(\boldsymbol{s})d\boldsymbol{s} \left(1 + \rho_c + 2\rho_c^2\right),$$

where $K^{(j)}$ *denotes the j-times convolution product of K with itself.*

3 Testing proposal in practice

Notice that the asymptotic distribution of the test obtained in Theorem 1, as in other nonparametric testing procedures (see, for example, Härdle and Mammen, 1993), may not be sufficiently precise to approximate the test statistic distribution under the null hypothesis in practice, for small or moderate sample sizes. A brief simulation experiment (not shown here for reasons of space) was conducted to analyze the performance of the asymptotic distribution of the test under the null hypothesis. In this example, we observed that the sample size n needed to get a reliable approximation of the sampling distribution by the Gaussian limit distribution has to be considerably large. In addition, this size depends on other process characteristics, such as the strength of the dependence structure. For this particular example, we noticed that values of n rather larger than 2500 are needed to obtain accurate approximations. Therefore, it could be deduced that to get sufficiently precise approximations in a general framework, it would be necessary to have a very large sample size, which is not always the case for geostatistical data. Moreover, the limit distribution of the test statistic depends on unknown quantities that must be estimated. This is a common problem in smoothing-based tests, as already noted by González-Manteiga and Crujeiras (2013). This issue is usually overcome using resampling methods for approximating the distribution of the test statistics, specifically, employing bootstrap algorithms that try to mimic the data structure under the null hypothesis.

In what follows, a detailed description of the different bootstrap proposals designed to perform the calibration of the test (namely PB, NPB and CNPB) is presented. The main difference between the proposals is how the resampling residuals (required for mimicking the dependence structure) are computed. In PB, the residuals are obtained from the parametric trend estimator. Alternately, in NPB, the residuals are drawn from the nonparametric trend estimator (see González-Manteiga and Cao, 1993). In this way, the error variability could be reproduced consistently both under the null and the alternative hypotheses, increasing the power of the test. Finally, the CNPB procedure is a modification of the NPB, where the residuals are also obtained from the nonparametric trend estimator, but, additionally, the variability is estimated with an iterative algorithm to correct the bias due to the use of the residuals (Fernández-Casal and Francisco-Fernández, 2014).

In order to describe the PB and NPB resampling approaches, a common bootstrap algorithm is firstly introduced. The specific algorithm for the CNPB method is subsequently described. In what follows, no matter the method used, either parametric or nonparametric, $\hat{m}$ and $\hat{\boldsymbol{\Sigma}}$ denote the trend and the covariance matrix estimates, respectively.

Algorithm 1

1. Compute a parametric or a nonparametric trend estimator (described in Section 2.1), namely $\hat{m}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$, depending if a parametric (PB) or a nonparametric (NPB) bootstrap procedure is employed.

2. Obtain an estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ of the residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)^\top$, where $\hat{\varepsilon}_i = Z_i - \hat{m}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$.

3. Find the matrix $\boldsymbol{L}$, such that $\hat{\boldsymbol{\Sigma}} = \boldsymbol{L}\boldsymbol{L}^\top$, using Cholesky decomposition.

4. Compute the *independent* variables, $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$, given by $\boldsymbol{e} = \boldsymbol{L}^{-1}\hat{\boldsymbol{\varepsilon}}$.

5. The previous *independent* variables are centered and an *independent* bootstrap sample of size n, denoted by $\boldsymbol{e}^* = (e_1^*, \ldots, e_n^*)^\top$, is obtained.

6. The bootstrap errors $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)^\top$ are computed as $\boldsymbol{\varepsilon}^* = \boldsymbol{L}\boldsymbol{e}^*$, and the bootstrap samples are $Z_i^* = m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i) + \varepsilon_i^*$, being $m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i)$ the parametric trend estimator.

7. Using the bootstrap sample $\{(\boldsymbol{s}_i, Z_i^*)\}_{i=1}^n$, the bootstrap test statistic T_n^* is computed as in (7).

8. Repeat Steps 5-7 a large number of times B.

The empirical distribution of the B bootstrap test statistics can be employed to approximate the finite sample distribution of the test statistic T_n under the null hypothesis. Thus, denoting by $\{T_{n,1}^*, \ldots, T_{n,B}^*\}$ the sample of the B bootstrap test statistics, and defining t_α^* as its $(1 - \alpha)$-quantile, the null hypothesis in (3) is rejected if $T_n > t_\alpha^*$. Additionally, the *p*-value of the test statistic can be approximated by:

$$p\text{-value} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}_{\{T_{n,b}^* > T_n\}}.$$

Some specific steps of the previous algorithm are discussed below for the PB and NPB methods. The main differences between the procedures are highlighted.

3.1 Parametric residual bootstrap (PB)

The PB method extends to the case of spatial trends the parametric residual bootstrap discussed in Vilar-Fernández and González-Manteiga (1996). In Step 1 of the previous algorithm, the trend is estimated parametrically, employing the iterative least squares estimator described in Section 2.1. In Step 2, the covariance matrix is computed from the parametric residuals also using a parametric approach (see Section 2.1 for further details). Notice that if the trend and the semivariogram belong to the assumed parametric families, then this procedure should provide good results. However, a drawback of this method is the possible misspecification of the trend function and/or the variogram. In addition, as it was pointed out in the Introduction, the direct use of the parametric residuals introduces a bias in the estimation of the variability of the process in Step 2.

3.2 Nonparametric residual bootstrap (NPB)

The NPB approach tries to avoid the misspecification problems by using more flexible trend and dependence estimation methods than those employed in PB. In Step 1 of Algorithm 1, to increase the power of the test, following González-Manteiga and Cao (1993), the Nadaraya–Watson estimator given in (6) is employed. In addition, in Step 2, a flexible procedure is considered to estimate the covariance matrix. The Shapiro–Botha variogram approach (Shapiro and Botha, 1991), combined with a nonparametric kernel semivariogram pilot estimation provides efficient variogram estimates, which are used to approximate the corresponding covariance matrix. For more details see Section 2.1.

Regardless the methodology used to remove the trend in Step 2 of Algorithm 1, either parametric or nonparametric, the direct use of residuals in variogram estimation introduces a bias in the approximation of the process variability (see Cressie, 1993, Section 3.4.3, for the case of parametric linear trends). The CNPB method is a modification of the NPB approach, including a bias-corrected algorithm for the dependence estimation. The specific bias correction will be briefly described in the next section.

3.3 Corrected nonparametric residual bootstrap (CNPB)

As pointed out before, the CNPB method considers a procedure to correct the resulting bias in the nonparametric estimator of the variogram. In the geostatistical framework, more accurate results have been obtained using this technique (Castillo-Páez et al., 2019). The following adjustments are performed in the bootstrap Algorithm 1 for the CNPB method. In Step 1, the trend is estimated using the Nadaraya–Watson estimator given in (6), whereas in Step 2, a (nonparametric) bias-corrected pilot variogram estimator is obtained using an iterative algorithm (Fernández-Casal and Francisco-Fernández, 2014). In this estimator, the squared differences of the residuals $(\hat{\varepsilon}_i - \hat{\varepsilon}_j)^2$, for $1 \leq i < j \leq n$, are replaced by $(\hat{\varepsilon}_i - \hat{\varepsilon}_j)^2 - \hat{b}_{ii} - \hat{b}_{jj} - 2\hat{b}_{ij}$, where

$\hat{b}_{ij}$ is the (i,j)-entry of $\hat{\boldsymbol{B}}$, being $\hat{\boldsymbol{B}}$ an approximation of $\boldsymbol{B} = \boldsymbol{S\Sigma S}^{\top} - \boldsymbol{\Sigma S}^{\top} - \boldsymbol{S\Sigma}$ (a square matrix representing the bias obtained due to the use of the residuals in the nonparametric variogram estimator), and $\boldsymbol{S}$ the *smoothing matrix* of the Nadaraya-Watson estimator (6), whose (i,j)-entry is $\boldsymbol{S}(i,j) = K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}_j)/[\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}_j)]$. Therefore, in Algorithm 1, additional computations are included in Step 2 and 3. In addition, Step 6 needs to be modified. The complete algorithm for the CNPB method is presented below.

Algorithm 2

---

1. Compute a nonparametric trend estimator (described in Section 2.1), denoted by $\hat{m}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$.

2a. Obtain an estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ of the residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)^{\top}$, where $\hat{\varepsilon}_i = Z_i - \hat{m}(\boldsymbol{s}_i)$, $i = 1, \ldots, n$.

2b. Obtain a bias-corrected estimate of the variogram, as explained before, using the residuals from the nonparametric fit, and calculate the corresponding (estimated) covariance matrix $\tilde{\boldsymbol{\Sigma}}$ of the errors.

3a. Find the matrix $\boldsymbol{L}$, such that $\hat{\boldsymbol{\Sigma}} = \boldsymbol{L}\boldsymbol{L}^{\top}$, using Cholesky decomposition.

3b. Find the matrix $\tilde{\boldsymbol{L}}$, such that $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}^{\top}$, using Cholesky decomposition.

4. Compute the *independent* variables, $\boldsymbol{e} = (e_1, \ldots, e_n)^{\top}$, given by $\boldsymbol{e} = \boldsymbol{L}^{-1}\hat{\boldsymbol{\varepsilon}}$.

5. The previous *independent* variables are centered and an *independent* bootstrap sample of size n, denoted by $\boldsymbol{e}^* = (e_1^*, \ldots, e_n^*)^{\top}$, is obtained.

6. The bootstrap errors $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)^{\top}$ are $\boldsymbol{\varepsilon}^* = \tilde{\boldsymbol{L}}\boldsymbol{e}^*$, and the bootstrap samples are $Z_i^* = m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i) + \varepsilon_i^*$, where $m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i)$ is computed using the procedure described in Section 2.1.

7. Using the bootstrap sample $\{(\boldsymbol{s}_i, Z_i^*)\}_{i=1}^{n}$, the bootstrap test statistic T_n^* is computed as in (7).

8. Repeat Steps 5-7 a large number of times B.

---

4 Simulation study

In this section, the practical performance of the proposed test statistic is analyzed through a simulation study comparing the different bootstrap procedures described in Section 3.

Although different regression models were considered in this study, for reasons of space, only the results obtained for two of them are presented here. First, the parametric trend family $\mathcal{M}_{1,\boldsymbol{\beta}} = \{\beta_0 + \beta_1(s_1 - 0.5)^3, \beta_0, \beta_1 \in \mathbb{R}\}$ is assumed for the null hypothesis. In this case, the theoretical trend functions are given by:

$$m_1(\boldsymbol{s}) = 2.5 + 4(s_1 - 0.5)^3 + c\sin(2\pi s_2), \quad \boldsymbol{s} = (s_1, s_2). \tag{9}$$

The second parametric trend family considered for the null hypothesis is $\mathcal{M}_{2,\boldsymbol{\beta}} = \{\beta_0 + \beta_1\cos(\pi s_1), \beta_0, \beta_1 \in \mathbb{R}\}$, and the trend functions are:

$$m_2(\boldsymbol{s}) = 1 + 2\cos(\pi s_1) + c\sin(2\pi s_2), \quad \boldsymbol{s} = (s_1, s_2). \tag{10}$$

In both cases, the parameter c controls whether the null ($c = 0$) or the alternative ($c \neq 0$) hypotheses hold. For different values of this parameter, 500 samples of sizes n ($n = 100, 225$ and 400) are generated on a regular grid in the unit square $D = [0,1] \times [0,1]$, following model (1), with regression functions (9) or (10). The random errors ε_i are normally distributed with zero mean and isotropic exponential covariogram:

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = c_e \left[\exp(-\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/a_e) \right], \quad \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \neq 0, \tag{11}$$

where c_e is the partial sill and a_e the practical range, whereas the variance of the errors (also called the sill) is $\sigma^2 = c_0 + c_e$, being c_0 the nugget effect. Different degrees of spatial dependence are studied, considering values of $a_e = 0.3, 0.6$ and 0.9, $\sigma^2 = 0.16, 0.32$ and 0.32, and nugget values of $0\%, 25\%$ and 50% of σ^2.

The behavior of the test statistic given in (7) is analyzed in the different scenarios. The parametric fit used to construct T_n is computed using the iterative least squares procedure described in Section 2.1. The nonparametric fit is obtained using the multivariate Nadaraya–Watson estimator, given in (6), with a multiplicative triweight kernel. The bandwidth selection problem is addressed by employing the same procedure as that used in Härdle and Mammen (1993), Alcalá et al. (1999), Opsomer and Francisco-Fernández (2010), or Meilán-Vila et al. (2020), among others, analyzing the performance of the test statistic T_n in (7) for a range of bandwidths. This allows to check how sensitive the results are to variations in $\boldsymbol{H}$. Initially, to simplify the calculations, the bandwidth matrix is restricted to a diagonal matrix with both equal elements (scalar matrix), $\boldsymbol{H} = \text{diag}(h, h)$, and different values of h in the interval $[0.25, 1.50]$ are chosen. The weight function employed in (7) to avoid the possible boundary effect (González-Manteiga and Cao, 1993) is $w(\mathbf{s}) = \mathbb{I}_{\{\mathbf{s} \in [1/\sqrt{n}, 1 - 1/\sqrt{n}] \times [1/\sqrt{n}, 1 - 1/\sqrt{n}]\}}$, where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function.

The bootstrap procedures described in Section 3 are applied using $B = 500$ replicates. In the nonparametric residual bootstrap procedures, NPB and CNPB, the multivariate Nadaraya–Watson estimator is computed in Step 1 using the optimal bandwidth that minimizes the MASE. Similar results were obtained when the corrected generalized cross-validation (CGCV) bandwidth (Francisco-Fernandez and Opsomer, 2005) is employed. However, the use of the MASE bandwidth matrix reduces the computing time and avoids the effect of the bandwidth selection for the trend estimation on the results. Regarding the variogram, the (uncorrected) variogram estimates and the bias-corrected version are computed on a regular grid up to the 55% of the largest sample distance between observations. In this case, the bandwidths are selected applying the cross-validation relative squared error criterion.

Proportions of rejections (under the null hypothesis, $c = 0$) using PB (solid black line), NPB (dashed red line) and CNPB (dotted blue line) bootstrap procedures, for several values of h are plotted in Fig. 2 (models (9) and 10 in left and right panels, respectively). In this study, a significance level $\alpha = 0.05$, and values of $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $n = 400$ are considered. Under the null hypothesis, the trend function belongs to the parametric family and, as expected, the resampling procedure with a better performance is the parametric one (PB). On the other hand, although resampling methods following the rationale of the NPB approach have provided good results in similar testing problems in other frameworks, for example, for independent and univariate data (González-Manteiga and Cao, 1993), this is not the case in our
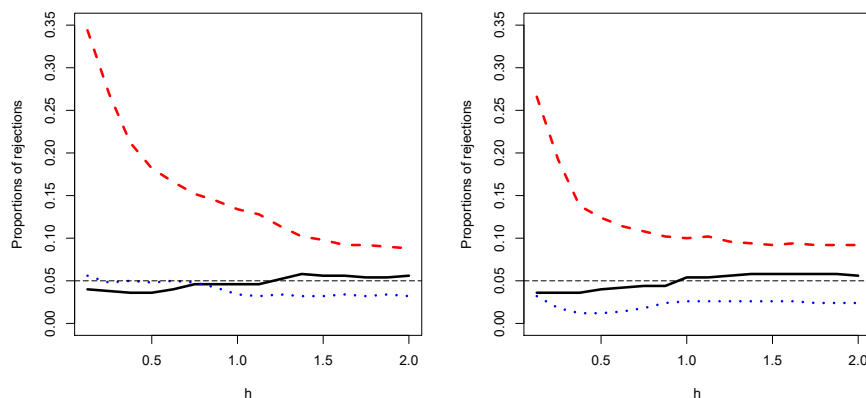
Fig. 2 Proportions of rejections under the null hypothesis ($c = 0$) using PB (solid black line), NPB (dashed red line) and CNPB (dotted blue line) bootstrap procedures, considering trend functions (9) (left panel) and (10) (right panel), for different values of h. Model parameters: $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $n = 400$. Horizontal dashed black line represents the significance level $\alpha = 0.05$

geostatistical context. Using the NPB method, the proportions of rejections of the null hypothesis are increased by the fact that the variability is underestimated (as noted by Fernández-Casal and Francisco-Fernández, 2014), advising against the use of this procedure. Finally, the benefits of correcting the bias can be observed, being the results for the CNPB procedure much better than those obtained with the NPB approach.

The effect of the sample size as well as the impact of the spatial dependence on the behavior of the test, under the null and under some alternative hypotheses, are analyzed below. In the different scenarios considered, a comparison of the proposed bootstrap procedures (PB, NPB and CNPB) is presented. For the sake of brevity, only some representative results employing the parametric family $\mathcal{M}_{1,\boldsymbol{\beta}}$ are shown here. Similar conclusions were obtained when the parametric family $\mathcal{M}_{2,\boldsymbol{\beta}}$ was considered.

4.1 Sample size effect

In this section, the performance of the bootstrap procedures is analyzed for different sample sizes ($n = 100, 225$ and 400). Proportions of rejections of the null hypothesis, for a significance level $\alpha = 0.05$, considering the parameters $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$ in model (11), and different sample sizes, are displayed in Table 1. Under the null hypothesis (i.e., when $c = 0$), it can be observed that the test has a reasonable behavior when using the PB and CNPB resampling methods. With both procedures, the test seems to preserve the nominal significance level of 5%, given that, under the null hypothesis, it lies within the corresponding 95% confidence interval for the proportion. It should be noted that the proportions of rejections are slightly affected by the value of h. In fact, for the CNPB method, the proportions of rejections are smaller

Table 1 Proportions of rejections of the null hypothesis for the parametric family $\mathscr{M}_{1,\boldsymbol{\beta}}$ with different sample sizes. Model parameters: $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$. Significance level: $\alpha = 0.05$

c	n	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
0	100	PB	0.056	0.042	0.054	0.066	0.070	0.068
		NPB	0.340	0.216	0.170	0.142	0.102	0.088
		CNPB	0.064	0.050	0.040	0.028	0.018	0.018
	225	PB	0.070	0.068	0.060	0.070	0.082	0.082
		NPB	0.268	0.192	0.176	0.144	0.116	0.108
		CNPB	0.078	0.058	0.046	0.042	0.030	0.028
	400	PB	0.038	0.036	0.046	0.046	0.052	0.056
		NPB	0.270	0.182	0.152	0.134	0.114	0.098
		CNPB	0.048	0.048	0.048	0.034	0.034	0.032
0.5	100	PB	0.000	0.002	0.006	0.022	0.026	0.036
		NPB	1.000	0.996	0.978	0.960	0.898	0.818
		CNPB	0.740	0.574	0.380	0.198	0.076	0.050
	225	PB	0.018	0.012	0.020	0.046	0.066	0.074
		NPB	1.000	0.994	0.976	0.938	0.846	0.734
		CNPB	0.692	0.550	0.398	0.218	0.102	0.056
	400	PB	0.004	0.004	0.016	0.036	0.058	0.070
		NPB	1.000	1.000	0.984	0.940	0.852	0.716
		CNPB	0.384	0.316	0.208	0.076	0.034	0.028
1	100	PB	0.056	0.010	0.018	0.038	0.074	0.110
		NPB	1.000	1.000	1.000	1.000	0.994	0.982
		CNPB	0.996	0.972	0.894	0.584	0.294	0.146
	225	PB	0.002	0.002	0.010	0.030	0.078	0.098
		NPB	1.000	1.000	1.000	1.000	0.994	0.962
		CNPB	0.990	0.938	0.848	0.564	0.296	0.126
	400	PB	0.000	0.000	0.000	0.024	0.064	0.098
		NPB	1.000	1.000	1.000	1.000	1.000	0.968
		CNPB	0.990	0.944	0.824	0.578	0.304	0.170

when the bandwidth value is larger. The opposite effect is observed when the PB approach is employed. For alternative assumptions ($c = 0.5$ and $c = 1$), the performance of the PB method is unsatisfactory. A much better behavior is observed for the CNPB approach. A decreasing power is obtained when the value of h increases. As expected, the power of the test becomes larger when the value of c gets bigger. Note that although it may seem that the NPB procedure presents a better behavior in terms of power, this is due to the underestimation of the variability, which induced really poor results under the null hypothesis.

4.2 Range of dependence effect

In this section, the performance of the bootstrap procedures is analyzed for different spatial dependence degrees ($a_e = 0.3$, $a_e = 0.6$ and $a_e = 0.9$). Values of $n = 400$, $\sigma^2 = 0.16$ and $c_0 = 0.04$ are considered. Fig. 3 shows exponential variogram models with $\sigma^2 = 0.16$ and $c_0 = 0.04$, for $a_e = 0.3$ (solid black line), $a_e = 0.6$ (dashed red line) and $a_e = 0.9$ (dotted blue line). Table 2 contains the proportions of rejections of the null hypothesis for $\alpha = 0.05$. Notice that results for $a_e = 0.6$ have already been shown in Table 1, but for the sake of comparison they are also included in Table 2. Again, it can be observed that the CNPB method provides good results for the null
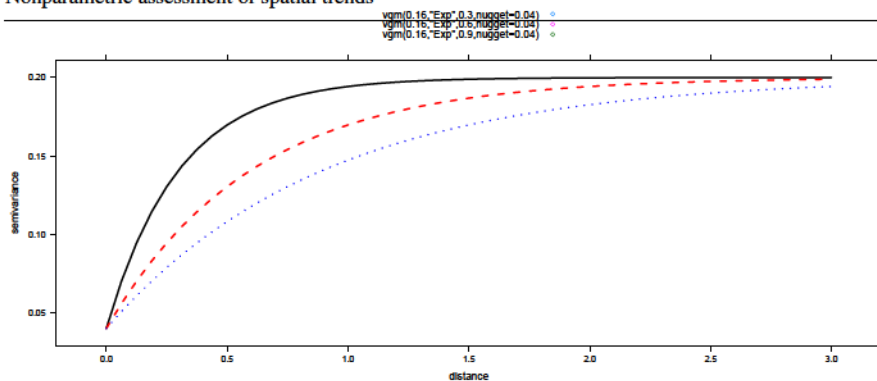
Fig. 3 Exponential variogram models (the corresponding exponential covariogram models are given in equation (11)) with $\sigma^2 = 0.16$ and $c_0 = 0.04$, for $a_e = 0.3$ (solid black line), $a_e = 0.6$ (dashed red line) and $a_e = 0.9$ (dotted blue line)

and the alternative hypotheses. As expected, for larger values of the practical range a_e, the bandwidth values providing an effective calibration of the test should also be larger. Regarding the PB approach, this resampling method works properly under the null hypothesis (for appropriate values of the bandwidth parameters h), but its performance under the alternatives is very poor. On the other hand, although the NPB method has a very high power, the proportions of rejections under the null hypothesis are very large.

4.3 Nugget effect

The performance of the proposed bootstrap procedures is now presented for different values of the nugget, $0\%, 25\%$ and 50% of σ^2. Proportions of rejections of the null hypothesis are shown in Table 3, for $\alpha = 0.05$, considering $n = 400$, $\sigma^2 = 0.16$ and $a_e = 0.6$. The best behavior is observed when the CNPB approach is employed, showing a good performance for the null and the different alternative hypotheses. For larger values of the variogram at zero distance, smaller bandwidths should be taken to calibrate the test properly. On the other hand, no reliable results are obtained for the NPB method under the null and the alternative hypotheses. Finally, regarding the PB procedure, the power of the test is very small in all the scenarios considered.

4.4 More general bandwidth matrices

This section contains additional simulations considering diagonal bandwidths with different elements, $\boldsymbol{H} = \text{diag}(h_1, h_2)$. Notice that for the parametric trends chosen in this work, there is no interaction effect between the spatial coordinates and, therefore, the selection of a diagonal bandwidth matrix seems reasonable. In this simulation scenario, values of $n = 400$, $\sigma^2 = 0.16$, $c_0 = 0.04$ and $a_e = 0.6$ are set. Firstly, in order to reduce computing times, the test is applied when considering the diagonal optimal bandwidth matrix that minimizes the MASE to compute $\hat{m}_{\boldsymbol{H}}^{NW}$ and $\hat{m}_{\boldsymbol{H},\boldsymbol{\beta}}^{NW}$. Note that

Table 2 Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{1,\boldsymbol{\beta}}$, with $\alpha = 0.05$, considering $c_0 = 0.04$, $\sigma^2 = 0.16$, $n = 400$ and different range values

c	a_e	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
0	0.3	PB	0.056	0.050	0.040	0.044	0.044	0.050
		NPB	0.158	0.090	0.076	0.062	0.050	0.052
		CNPB	0.030	0.024	0.018	0.012	0.010	0.010
	0.6	PB	0.038	0.036	0.046	0.046	0.052	0.056
		NPB	0.270	0.182	0.152	0.134	0.114	0.098
		CNPB	0.048	0.048	0.048	0.034	0.034	0.032
	0.9	PB	0.002	0.004	0.018	0.030	0.040	0.042
		NPB	0.328	0.266	0.232	0.186	0.162	0.144
		CNPB	0.070	0.070	0.068	0.060	0.048	0.048
0.5	0.3	PB	0.000	0.002	0.006	0.022	0.026	0.036
		NPB	1.000	1.000	0.998	0.976	0.878	0.716
		CNPB	0.346	0.270	0.170	0.042	0.018	0.016
	0.6	PB	0.004	0.004	0.016	0.038	0.058	0.066
		NPB	1.000	1.000	0.984	0.940	0.852	0.716
		CNPB	0.384	0.316	0.208	0.076	0.034	0.028
	0.9	PB	0.008	0.010	0.020	0.048	0.062	0.072
		NPB	1.000	0.996	0.982	0.954	0.880	0.750
		CNPB	0.526	0.432	0.284	0.146	0.074	0.052
1	0.3	PB	0.000	0.000	0.000	0.010	0.050	0.062
		NPB	1.000	1.000	1.000	1.000	1.000	0.986
		CNPB	1.000	0.990	0.906	0.648	0.380	0.194
	0.6	PB	0.000	0.000	0.000	0.024	0.064	0.098
		NPB	1.000	1.000	1.000	1.000	1.000	0.968
		CNPB	0.990	0.944	0.824	0.578	0.304	0.170
	0.9	PB	0.000	0.000	0.008	0.030	0.080	0.110
		NPB	1.000	1.000	1.000	1.000	1.000	0.968
		CNPB	1.000	0.962	0.858	0.662	0.360	0.206

this bandwidth matrix can not be calculated in a real practical situation. However, as pointed out before, we have checked that very similar results are obtained when using the data-driven CGCV matrix bandwidth (Francisco-Fernández and Opsomer, 2005), but with a considerable increase in computing time. Proportions of rejections of the null hypothesis when $c = 0, 0.5$ and 1, for $\alpha = 0.05$, jointly with the diagonal elements of the corresponding MASE bandwidth matrices, are shown in Table 4. Results for NPB are omitted due to its deficient calibration. For this scenario, it can be observed that the test preserves the nominal significance level when using both methods (PB and CNPB). For alternative hypotheses, the CNPB approach, unlike the PB one, provides reasonable results. It seems that similar conclusions to those obtained in the previous sections can be deduced when the diagonal MASE bandwidth matrix is used. However, it should be noted that an optimal bandwidth for estimation may not be an optimal one for testing (being not even clear what optimal means). A more reliable comparison between scalar and diagonal bandwidth matrices can be performed by considering different combinations of h_1 and h_2. Proportions of rejections (under the null hypothesis, $c = 0$) for $\alpha = 0.05$, are plotted in Fig. 4. Left panel of Fig. 4 shows the results for the PB approach and right panel for the CNPB one. In this scenario, it can be observed that for the PB method there are not relevant differences in terms of proportions of rejections if $\boldsymbol{H} = \mathrm{diag}(h, h)$ or $\boldsymbol{H} = \mathrm{diag}(h_1, h_2)$ (with

Table 3 Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{1,\boldsymbol{\beta}}$, with $\alpha = 0.05$, considering $n = 400$, $\sigma^2 = 0.16$, $a_e = 0.6$ and different nugget effect values

| c	c_0	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
0	0%	PB	0.014	0.008	0.020	0.034	0.046	0.046
		NPB	0.338	0.230	0.182	0.154	0.132	0.110
		CNPB	0.060	0.058	0.052	0.044	0.040	0.038
	25%	PB	0.038	0.036	0.046	0.046	0.052	0.056
		NPB	0.270	0.182	0.152	0.134	0.114	0.098
		CNPB	0.048	0.048	0.048	0.034	0.034	0.032
	50%	PB	0.050	0.048	0.042	0.048	0.052	0.056
		NPB	0.254	0.172	0.144	0.116	0.092	0.082
		CNPB	0.050	0.050	0.046	0.036	0.030	0.028
0.5	0%	PB	0.006	0.014	0.034	0.052	0.062	0.068
		NPB	1.000	0.990	0.972	0.912	0.834	0.686
		CNPB	0.418	0.336	0.240	0.098	0.050	0.036
	25%	PB	0.004	0.004	0.016	0.038	0.058	0.066
		NPB	1.000	1.000	0.984	0.940	0.852	0.716
		CNPB	0.640	0.534	0.384	0.206	0.096	0.058
	50%	PB	0.000	0.002	0.010	0.026	0.048	0.056
		NPB	1.000	1.000	0.998	0.976	0.890	0.776
		CNPB	0.880	0.780	0.652	0.468	0.264	0.164
1	0%	PB	0.104	0.074	0.092	0.132	0.182	0.204
		NPB	1.000	1.000	1.000	1.000	0.992	0.922
		CNPB	0.926	0.830	0.676	0.378	0.194	0.080
	25%	PB	0.000	0.000	0.000	0.024	0.064	0.098
		NPB	1.000	1.000	1.000	1.000	1.000	0.968
		CNPB	0.990	0.944	0.824	0.578	0.304	0.170
	50%	PB	0.000	0.000	0.004	0.014	0.052	0.076
		NPB	1.000	1.000	1.000	1.000	1.000	0.996
		CNPB	1.000	1.000	0.970	0.840	0.576	0.378
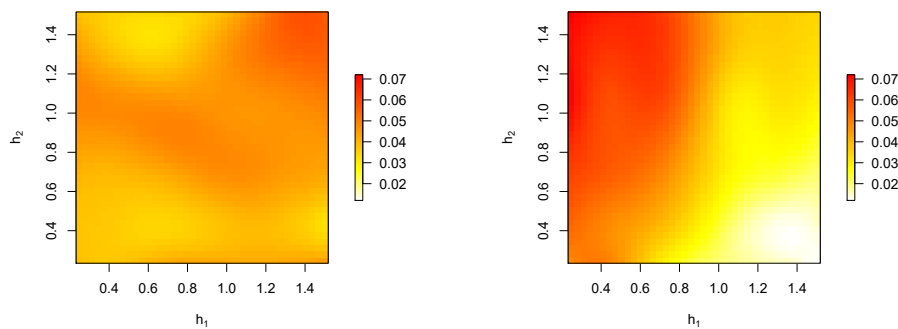
Fig. 4 Proportions of rejections of the null hypothesis ($c = 0$), for $\alpha = 0.05$, considering $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $n = 400$, using PB (left) and CNPB (right), for several values of h_1 and h_2

$h_1 \neq h_2$) are considered. Regarding the CNPB approach, the use of a more general bandwidth matrix does not provide better results with respect to using scalar bandwidth matrices. On the other hand, although it is omitted here, similar conclusions can be obtained for alternative hypotheses ($c \neq 0$).

Table 4 Proportions of rejections of the null hypothesis for the parametric family $\mathscr{M}_{1,\boldsymbol{\beta}}$, with $\alpha = 0.05$, considering $n = 400$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $c_0 = 0.04$, obtained with the optimal MASE bandwidth matrix $\boldsymbol{H} = \mathrm{diag}(h_1, h_2)$.

		$c = 0$	$c = 0.5$	$c = 1$
Proportion of rejections	PB	0.044	0.004	0.000
	CNPB	0.070	0.620	1.000
MASE bandwidth matrix	h_1	0.190	0.282	0.316
	h_2	1.500	0.291	0.151

5 Discussion

A goodness-of-fit test to assess a parametric trend surface for a geostatistical process is studied in this work. An exhaustive analysis of the behavior of the test considering different trends and dependence configurations is provided. The proposed test statistic measures the difference between parametric and nonparametric fits using an L_2-distance. An iterative least squares procedure has been used to obtain the parametric trend estimates, but other approaches such as maximum likelihood methods, could be also used, as long as a $\sqrt{n}$-consistency property is satisfied. Regarding the nonparametric fit, the multivariate Nadaraya–Watson estimator was employed. Other kernel-type estimator specifically defined for fixed design, such as the Gasser–Müller (Gasser and Müller, 1979) or the Priestley–Chao (Priestley and Chao, 1972) estimators could be used instead. However, the computation of the Gasser–Müller estimator involves sorting and taking middle points in the design space, which is not computationally trivial in a multidimensional setting. That makes it unsuitable to use in the setting of the present paper. The Priestley–Chao estimator has a simple expression and can be directly defined in a multidimensional framework. Moreover, although this estimator was initially introduced for equally spaced fixed and uniformly random designs, it could be adapted to the context of this paper (i.e., for non-equally spaced fixed designs). Nevertheless, its use would not allow us to extend the asymptotic theory to the non-uniform random design case, since this estimator is asymptotically biased in that setting. With these elections of parametric and nonparametric fits, the asymptotic distribution of the test statistic, under the null and under local alternatives, was derived considering the assumption of increasing-domain spatial asymptotics.

For a practical implementation, due to the slow convergence to the limit distribution, resampling methods were used to calibrate the test. Specifically, three bootstrap procedures were designed and applied in practice: PB, NPB and CNPB. The CNPB resampling method avoids model selection and, therefore, prevents against misspecification problems in the estimation of the trend and/or the dependence structure, unlike the PB approach. The CNPB method also corrects the bias induced by the use of the residuals in the approximation of the dependence, using an iterative method, providing good results of the test under the null and alternative hypotheses. As it was pointed out in Fernández-Casal and Francisco-Fernández (2014), a similar tool for bias adjustment could be included in the parametric semivariogram estimation in the PB approach (see Davison and Hinkley, 1997). However, this way of proceeding would not avoid misspecification problems in the parametric semivariogram estimation.

The performance of the goodness-of-fit test, in terms of calibration and power, has been explored in a grid of different bandwidths to check how it is affected by the bandwidth choice. In the vast majority of scenarios considered in the simulation study, the best performance is observed when the CNPB approach is employed, showing a reasonable behavior for both null and alternative hypotheses. It seems that the use of non-scalar bandwidths has not provided better results for the CNPB procedure, at least for the scenarios considered in this work. The PB proposal works properly for calibration, but it shows a limited capacity to detect alternatives. On the other hand, although similar resampling methods to the NPB proposal have given good results when used in goodness-of-fit tests considering regression models with independent and univariate data, this is not the case in the geostatistical context. In this case, the proportions of rejections under the null hypothesis are very large compared with the significance level considered, due to the underestimation of the variability of the process. For the sake of brevity, the performance of the test was only shown for non-linear trends, although a linear model has been also considered as a null hypothesis. In general, no matter the parametric family considered, results obtained by the CNPB approach improve those achieved by the PB and NPB methods, for the null and the different alternative hypotheses.

The three resampling approaches compared in this paper are based on computing the residuals from a pilot fit, estimating the corresponding covariance matrix of the errors and, finally, using a Cholesky decomposition, obtaining a vector of *independent* errors to generate bootstrap resamples. Other resampling procedures, such as the block bootstrap (see Lahiri, 2013), could be used to calibrate the test. This method, unlike parametric and nonparametric bootstrap based methods, requires an appropriate partition of the observation region. These procedures fail to reproduce the variability of the process, thus leading to an underestimation of the semivariogram, possibly caused by the selection of the blocks. In Castillo-Páez et al. (2019), parametric, corrected nonparametric and block bootstrap mechanisms were compared by checking their performance in the approximation of the bias and the variance of two variogram estimators. For inference on geostatistical processes and, particularly, on dependence structure estimation, those authors recommend the use of corrected nonparametric bootstrap methods. For these reasons, block bootstrap approaches were not employed in the present research.

The procedures used in the simulation study were implemented in the statistical environment R (R Development Core Team, 2020), using functions included in the `npsp` and `geoR` packages (Fernández-Casal, 2019; Ribeiro and Diggle, 2020) to estimate the variogram and the spatial regression functions. In particular, the bias correction in CNPB bootstrap algorithm is implemented in the function `np.svariso.corr` of the `npsp` R package.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alcalá J, Cristóbal J, González-Manteiga W (1999) Goodness-of-fit test for linear models based on local polynomials. Stat Probabil Lett 42:39–46

Castillo-Páez S, Fernández-Casal R, García-Soidán P (2019) A nonparametric bootstrap method for spatial data. Comput Stat Data Anal 137:1–15

Cressie N (1985) Fitting variogram models by weighted least squares. J Int Ass Math Geol 17(5):563–586

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

Crujeiras RM, Van Keilegon I (2010) Least squares estimation of nonlinear spatial trends. Comput Stat Data Anal 54(2):452–465

Davison AC, Hinkley DV (1997) Bootstrap methods and their application, vol 1. Cambridge University Press

Diggle P, Ribeiro PJ (2007) Model-based geostatistics. Springer, New York

Diggle P, Menezes R, Su Tl (2010) Geostatistical inference under preferential sampling. J R Stat Soc Ser C-Appl Stat 59(2):191–232

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

Fernández-Casal R (2019) npsp: Nonparametric spatial (geo)statistics. URL http://cran.r-project.org/package=npsp, R package version 0.7-5

Fernández-Casal R, Francisco-Fernández M (2014) Nonparametric bias-corrected variogram estimation under non-constant trend. Stoch Env Res Risk A 28(5):1247–1259

Fernández-Casal R, González-Manteiga W, Febrero-Bande M (2003a) Flexible spatio-temporal stationary variogram models. Stat Comput 13(2):127–136

Fernández-Casal R, González-Manteiga W, Febrero-Bande M (2003b) Space-time dependency modeling using general classes of flexible stationary variogram models. J Geophys Res-Atmos 108(D24)

Francisco-Fernandez M, Opsomer JD (2005) Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. Can J Stat-Rev Can Stat 33(2):279–295

Gasser T, Müller HG (1979) Kernel estimation of regression functions. In: Smoothing techniques for curve estimation, Springer, pp 23–68

González-Manteiga W, Cao R (1993) Testing the hypothesis of a general linear model using nonparametric regression estimation. Test 2(1-2):161–188

González-Manteiga W, Crujeiras RM (2013) An updated review of Goodness-of-Fit tests for regression models. Test 22(3):361–411

Hall P, Patil P (1994) Properties of nonparametric estimators of autocovariance for stationary random fields. Probab Theory Rel 99(3):399–424

Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. Ann Stat 21(4):1926–1947

Härdle W, Müller M (2012) Multivariate and semiparametric kernel regression. In: Smoothing and regression: approaches, computation, and application, John Wiley & Sons, New Jersey

Kim TY, Ha J, Hwang SY, Park C, Luo ZM (2013) Central limit theorems for reduced U-statistics under dependence and their usefulness. Aust N Z J Stat 55(4):387–399

Lahiri SN (2013) Resampling methods for dependent data. Springer Science & Business Media

Li CS (2005) Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. Stat Methodol 2(4):267–284

Liu XH (2001) Kernel smoothing for spatially correlated data. PhD thesis, Department of Statistics, Iowa State University

Meilán-Vila A, Opsomer JD, Francisco-Fernández M, Crujeiras RM (2020) A goodness-of-fit test for regression models with spatially correlated errors. TEST 29:728–749

Olea RA, Pardo-Iguzquiza E (2011) Generalized bootstrap method for assessment of uncertainty in semivariogram inference. Math Geosci 43(2):203–228

Opsomer J, Francisco-Fernández M (2010) Finding local departures from a parametric model using nonparametric regression. Stat Pap 51:69–84

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

Priestley MB, Chao M (1972) Non-parametric function fitting. J R Stat Soc Series B Stat Methodol 34(3):385–392

R Development Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org

Ribeiro PJ, Diggle PJ (2020) geoR: Analysis of geostatistical data. URL https://cran.r-project.org/package=geoR, R package version 1.7-5.2.2

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

Solow AR (1985) Bootstrapping correlated data. Math Geol 17(7):769–775

Vilar-Fernández J, González-Manteiga W (1996) Bootstrap test of goodness of fit to a linear model when errors are correlated. Commun Stat-Theory Methods 25(12):2925–2953

Wand MP, Jones MC (1994) Kernel smoothing. Chapman and Hall/CRC

Appendix. Proof of the main theorem

In this appendix, under assumptions (A1)–(A8), Theorem 1 is proved. The asymptotic distribution of the test statistic, given in (7), is derived. This test compares the nonparametric and the smooth parametric estimators, given in (6) and (8), respectively, using an L_2-distance.

Proof The test statistic (7) can be decomposed as

$$T_n = n|\boldsymbol{H}|^{1/2} \int \left[\hat{m}_{\boldsymbol{H}}^{NW}(\boldsymbol{s}) - \hat{m}_{\boldsymbol{H},\hat{\boldsymbol{\beta}}}^{NW}(\boldsymbol{s}) \right]^2 w(\boldsymbol{s}) d\boldsymbol{s}$$

$$= n|\boldsymbol{H}|^{1/2} \int \left[\frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})Z_i}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})} - \frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i)}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})} \right]^2 w(\boldsymbol{s}) d\boldsymbol{s}$$

$$= n|\boldsymbol{H}|^{1/2} \int \frac{\left\{ \sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \left[m(\boldsymbol{s}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i) \right] \right\}^2}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \right]^2} w(\boldsymbol{s}) d\boldsymbol{s}.$$

Now, taking into account that the trends considered are of the form $m = m_{\boldsymbol{\beta}_0} + n^{-1/2}|\boldsymbol{H}|^{-1/4}g$, one gets:

$$T_n = n|\boldsymbol{H}|^{1/2} \int \frac{\left\{ \sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \left[m_{\boldsymbol{\beta}_0}(\boldsymbol{s}_i) + n^{-1/2}|\boldsymbol{H}|^{-1/4}g(\boldsymbol{s}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i) \right] \right\}^2}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \right]^2} w(\boldsymbol{s}) d\boldsymbol{s}$$

$$= n|\boldsymbol{H}|^{1/2} \int \left[I_1(\boldsymbol{s}) + I_2(\boldsymbol{s}) + I_3(\boldsymbol{s}) \right]^2 w(\boldsymbol{s}) d\boldsymbol{s},$$

where

$$I_1(\boldsymbol{s}) = \frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \left[m_{\boldsymbol{\beta}_0}(\boldsymbol{s}_i) - m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i) \right]}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})},$$

$$I_2(\boldsymbol{s}) = \frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) n^{-1/2}|\boldsymbol{H}|^{-1/4}g(\boldsymbol{s}_i)}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})},$$

$$I_3(\boldsymbol{s}) = \frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \varepsilon_i}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})}.$$

Under assumptions (A1), (A2) and (A6), and given that the difference $m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}) - m_{\boldsymbol{\beta}_0}(\boldsymbol{s}) = O_p(n^{-1/2})$, it is obtained that

$$n|\boldsymbol{H}|^{1/2} \int I_1^2(\boldsymbol{s})w(\boldsymbol{s})d\boldsymbol{s} = n|\boldsymbol{H}|^{1/2} \int \left\{ \frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \left[m_{\boldsymbol{\beta}_0}(\boldsymbol{s}_i) - m_{\hat{\boldsymbol{\beta}}}(\boldsymbol{s}_i) \right]}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})} \right\}^2 w(\boldsymbol{s})d\boldsymbol{s}$$

$$= O_p(|\boldsymbol{H}|^{1/2}).$$

For the term $I_2(\boldsymbol{s})$, using the assumption (A2), it follows that

$$n|\boldsymbol{H}|^{1/2} \int I_2^2(\boldsymbol{s})w(\boldsymbol{s})d\boldsymbol{s} = n|\boldsymbol{H}|^{1/2} \int \left[\frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) n^{-1/2}|\boldsymbol{H}|^{-1/4}g(\boldsymbol{s}_i)}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})} \right]^2 w(\boldsymbol{s})d\boldsymbol{s}$$

$$= n|\boldsymbol{H}|^{1/2}n^{-1}|\boldsymbol{H}|^{-1/2} \int \left[K_{\boldsymbol{H}} * g(\boldsymbol{s}) \right]^2 w(\boldsymbol{s})d\boldsymbol{s}$$

$$= \int \left[K_{\boldsymbol{H}} * g(\boldsymbol{s}) \right]^2 w(\boldsymbol{s})d\boldsymbol{s}, \tag{12}$$

which corresponds to b_{1H} in Theorem 1. Finally, $I_3(s)$ (associated with the error component) can be decomposed as:

$$
n|\boldsymbol{H}|^{1/2} \int I_3^2(\boldsymbol{s}) w(\boldsymbol{s}) d\boldsymbol{s} = n|\boldsymbol{H}|^{1/2} \int \left[\frac{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \, \varepsilon_i}{\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})} \right]^2 w(\boldsymbol{s}) d\boldsymbol{s}
$$

$$
= n|\boldsymbol{H}|^{1/2} \int \frac{\sum_{i=1}^n K_{\boldsymbol{H}}^2(\boldsymbol{s}_i - \boldsymbol{s}) \, \varepsilon_i^2}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \right]^2} w(\boldsymbol{s}) d\boldsymbol{s}
$$

$$
+ n|\boldsymbol{H}|^{1/2} \int \frac{\sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) K_{\boldsymbol{H}}(\boldsymbol{s}_j - \boldsymbol{s}) \, \varepsilon_i \varepsilon_j}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \right]^2} w(\boldsymbol{s}) d\boldsymbol{s}
$$

$$
= I_{31} + I_{32}.
$$

Close expressions of I_{31} and I_{32} can be obtained computing the expectation and the variance of these terms. Under assumption (A6), it can be proved that

$$
\mathbb{E}(|\boldsymbol{H}|^{1/2} I_{31}) = \mathbb{E} \left\{ n|\boldsymbol{H}| \int \frac{\sum_{i=1}^n K_{\boldsymbol{H}}^2(\boldsymbol{s}_i - \boldsymbol{s}) \, \varepsilon_i^2}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \right]^2} w(\boldsymbol{s}) d\boldsymbol{s} \right\}
$$

$$
= \sigma^2 K^{(2)}(0) \int w(\boldsymbol{s}) d\boldsymbol{s} \cdot [1 + o(1)]. \tag{13}
$$

Similarly, using assumptions (A3), (A6) and (A7), it can be obtained that

$$
\mathrm{Var}(|\boldsymbol{H}|^{1/2} I_{31}) = \mathrm{Var} \left\{ n|\boldsymbol{H}| \int \frac{\sum_{i=1}^n K_{\boldsymbol{H}}^2(\boldsymbol{s}_i - \boldsymbol{s}) \, \varepsilon_i^2}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) \right]^2} w(\boldsymbol{s}) d\boldsymbol{s} \right\}
$$

$$
= 2n^2 |\boldsymbol{H}|^2 \sum_{i=1}^n \sum_{j=1}^n \int \int \frac{K_{\boldsymbol{H}}^2(\boldsymbol{s}_i - \boldsymbol{s}) K_{\boldsymbol{H}}^2(\boldsymbol{s}_j - \boldsymbol{t}) \, [\mathrm{Cov}(\varepsilon_i, \varepsilon_j)]^2}{\left[\sum_{i=1}^n K_{\boldsymbol{H}}^2(\boldsymbol{s}_i - \boldsymbol{s}) \right] \left[\sum_{i=1}^n K_{\boldsymbol{H}}^2(\boldsymbol{s}_i - \boldsymbol{t}) \right]} w(\boldsymbol{s}) w(\boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}
$$

$$
= 2\sigma^4 |\boldsymbol{H}| \int \int \int \int K^2(\boldsymbol{v}) K^2(\boldsymbol{z}) w^2(\boldsymbol{s}) \rho_n^2 [\boldsymbol{H}(\boldsymbol{v} - \boldsymbol{z} + \boldsymbol{u})] d\boldsymbol{v} d\boldsymbol{z} d\boldsymbol{s} d\boldsymbol{u} \cdot [1 + o(1)].
$$

Let

$$
j_n(\boldsymbol{v}, \boldsymbol{u}) = n|\boldsymbol{H}| \int K^2(\boldsymbol{v}) \rho_n^2 [\boldsymbol{H}(\boldsymbol{v} - \boldsymbol{z} + \boldsymbol{u})] d\boldsymbol{z}.
$$

Notice that, using assumption (A3),

$$
|j_n(\boldsymbol{v}, \boldsymbol{u})| \leq K_M^2 \left\{ n|\boldsymbol{H}| \int |\rho_n^2 [\boldsymbol{H}(\boldsymbol{v} - \boldsymbol{z} + \boldsymbol{u})]| d\boldsymbol{z} \right\}
$$

$$
\leq K_M^2 \left[n \int |\rho_n(\boldsymbol{t})| d\boldsymbol{t} \right]
$$

$$
\leq K_M^2 \rho_M,
$$

where $K_M = \max_{\boldsymbol{s}}(K(\boldsymbol{s}))$ and $\rho_M = \max_{\boldsymbol{s}}(\rho_n(\boldsymbol{s}))$, and using assumptions (A2), (A3), (A6) and (A8), one gets that

$$
\mathrm{Var}(I_{31}) = o(1). \tag{14}
$$

From (13) and (14) it follows that

$$I_{31} = \sigma^2 |\boldsymbol{H}|^{-1/2} K^{(2)}(\boldsymbol{0}) \int w(\boldsymbol{s})d\boldsymbol{s} \cdot [1 + o_p(1)]. \qquad (15)$$

Now, consider the term

$$I_{32} = n|\boldsymbol{H}|^{1/2} \int \frac{\sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) K_{\boldsymbol{H}}(\boldsymbol{s}_j - \boldsymbol{s}) \varepsilon_i \varepsilon_j}{[\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})]^2} w(\boldsymbol{s})d\boldsymbol{s}.$$

Let

$$\kappa_{ij} = n|\boldsymbol{H}|^{1/2} \int \frac{K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) K_{\boldsymbol{H}}(\boldsymbol{s}_j - \boldsymbol{s})}{[\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})]^2} w(\boldsymbol{s})d\boldsymbol{s}\varepsilon(\boldsymbol{s}_i)\varepsilon(\boldsymbol{s}_j).$$

Thus,

$$I_{32} = \sum_{i \neq j} \kappa_{ij},$$

and this can be seen as a U-statistic with degenerate kernel. To obtain the asymptotic normality of I_{32} we apply the central limit theorem for reduced U-statistics under dependence given by Kim et al. (2013).

For this term I_{32} we have

$$\mathbb{E}\left(|\boldsymbol{H}|^{1/2}I_{32}\right) = \mathbb{E}\left\{ n|\boldsymbol{H}| \int \frac{\sum_{i \neq j} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) K_{\boldsymbol{H}}(\boldsymbol{s}_j - \boldsymbol{s}) \varepsilon_i \varepsilon_j}{[\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})]^2} w(\boldsymbol{s})d\boldsymbol{s} \right\}$$

$$= n^{-1}|\boldsymbol{H}| \sum_{i \neq j} \mathrm{Cov}[\varepsilon(\boldsymbol{s}_i), \varepsilon(\boldsymbol{s}_j)] \int \frac{K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s}) K_{\boldsymbol{H}}(\boldsymbol{s}_j - \boldsymbol{s})}{[\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{s}_i - \boldsymbol{s})]^2} w(\boldsymbol{s})d\boldsymbol{s} \cdot [1 + o(1)]$$

$$= (n-1)|\boldsymbol{H}|\sigma^2 \int \int \int K(\boldsymbol{v}) K(\boldsymbol{z}) \rho_n[\boldsymbol{H}(\boldsymbol{v} - \boldsymbol{z})] w(\boldsymbol{s})d\boldsymbol{v}d\boldsymbol{z}d\boldsymbol{s} \cdot [1 + o(1)].$$

Under the assumptions (A4)–(A8), as shown by Liu (2001),

$$\lim_{n \to \infty} n|\boldsymbol{H}| \int K(\boldsymbol{v})K(\boldsymbol{z})\rho_n[\boldsymbol{H}(\boldsymbol{v} - \boldsymbol{z})]d\boldsymbol{v}d\boldsymbol{z} = K^{(2)}(0)\rho_c.$$

It follows that

$$\mathbb{E}(|\boldsymbol{H}|^{1/2}I_{32}) = \sigma^2 K^{(2)}(0)\rho_c \int w(\boldsymbol{s})d\boldsymbol{s} \cdot [1 + o(1)]. \qquad (16)$$

Similarly, it can be obtained that the asymptotic variance of I_{32} is

$$V = \sigma^4 K^{(4)}(0) \int w^2(\boldsymbol{s})d\boldsymbol{s} \left(1 + \rho_c + 2\rho_c^2\right). \qquad (17)$$

The term I_{32} converges in distribution to a normally distributed random variable with mean the second term of $b_{0\boldsymbol{H}}$ and variance V.

In virtue of the Cauchy–Bunyakovsky–Schwarz inequality, the cross terms in T_n resulting from the products of I_1, I_2 and I_3 are all of small order. Therefore, combining

the results given in equations (12) and (15), and the asymptotic normality of I_{32} (with its bias (16) and its variance (17)), it follows that

$$V^{-1/2}(T_n - b_{0\boldsymbol{H}} - b_{1\boldsymbol{H}}) \to_{\mathscr{L}} N(0,1), \text{ as } n \to \infty,$$

where

$$b_{0\boldsymbol{H}} = |\boldsymbol{H}|^{-1/2}\sigma^2 K^{(2)}(0) \int w(\boldsymbol{s})d\boldsymbol{s}\,(1+\rho_c),$$

$$b_{1\boldsymbol{H}} = \int [K_{\boldsymbol{H}} * g(\boldsymbol{s})]^2 w(\boldsymbol{s})d\boldsymbol{s},$$

$$V = \sigma^4 K^{(4)}(0) \int w^2(\boldsymbol{s})d\boldsymbol{s}\,\left(1+\rho_c+2\rho_c^2\right).$$