Testing for Normality in Geostatistics. A New Approach Based on the Mahalanobis Distance

A discussion on the use of univariate normality tests in this context

Ramón Giraldo[1] and Emilio Porcu[2]

Abstract

Simple kriging is a best linear predictor (BLP) and ordinary kriging is a best linear unbiased predictor (BLUP). When the underlying process is normal, simple kriging is not only a BLP but a best predictor (BT) as well, that is, under squared loss, this predictor coincides with the conditional expectation of the predictor given the information. In this scenario, ordinary kriging provides an approximation to the BP. For this reason, in applied geostatistics, it is important to test for normality. Given a realization of a spatial random process, the simple kriging predictor will be optimal if the random vector follows a multivariate normal distribution. Some classical tests, such as Shapiro-Wilk (SW), Shapiro-Francia (SF), or Anderson-Darling (AD) are frequently used to evaluate the normality assumption. Such approaches assume independence and hence are not effective for at least two reasons. On the one hand, observations in a geostatistical analysis are typically spatially correlated. On the other hand, kriging optimality as mentioned above is based on multivariate rather than univariate normality. In this work, we provide a simulation study to describe the negative effect of using normality univariate tests with geostatistical data. We also show how the Mahalanobis distance can be adapted to the geostatistical context to test for normality.

Keywords: Chi-square distribution, Multivariate normal distribution, Mahalanobis distance, Normality test, Random field, Monte Carlo simulation.

Resumen

En geoestadística, bajo estacionariedad, kriging simple (KS) es el mejor predictor lineal (MPL) y kriging ordinario (KO) es el mejor predictor lineal insesgado (MPLI). Cuando el proceso estocástico es Normal, KS no es solo un MPL sino un mejor predictor (MP), es decir que bajo la función de pérdida cuadrática, éste coincide con la esperanza condicional del predictor dada la información. En este escenario, el predictor KO sirve como aproximación del MP. Por esta razón, en geoestadística aplicada, es importante probar el supuesto de normalidad. Dada una realización de un proceso espacial, KS será un predictor óptimo si el vector aleatorio subyacente sigue una distribución normal multivariada. Algunas pruebas de normalidad clásicas como Shapiro-Wilk (SW), Shapiro-Francia (SF), o Anderson-Darling (AD) son usadas para evaluar este supuesto. Estas asumen independencia y por ello no son apropiadas en geoestadística (y en general en estadística espacial). Por un lado, las observaciones en geoestadística son espacialmente correlacionadas. Por otro lado la optimalidad del kriging es fundamentada en normalidad multivariada (no en normalidad univariada). En este trabajo se presenta un estudio de simulación para mostrar por qué es inapropiado el uso de pruebas univaridas de normalidad con datos geoestadísticos. También, como solución al problema anterior, se propone una adaptación de la prueba de Mahalanobis al contexto geoestadístico para hacer de manera correcta el test de normalidad en este ámbito.

Palabras clave: Distribución chi-cuadrado, Distribución normal multivariada, Distancia de Mahalanobis, Prueba de normalidad, Campo aleatorio, Simulación de Monte Carlo.

---

[1]Department of Statistics, Universidad Nacional de Colombia, Bogotá, Colombia. Email: rgiraldoh@unal.edu.co
[2]Department of Mathematics, Khalifa University, Abu Dhabi, United Arab Emirate

1 Introduction

Geostatistics plays an important role in applied fields as diverse as ecology [1], agronomy [2], mining [3], meteorology [4], environmental pollution [5], amongst others. Usually, a geostatistical analysis is carried out in three steps, namely: exploratory data analysis, spatial dependence estimation, and kriging prediction [6]. Classical results from decision theory [27] show that under squared loss, the best predictor (BP) is the conditional expectation of the predictor given the information. For normal processes, the BP is equivalent to the best linear predictor (BLP), called simple kriging in geostatistics [8]. When the mean is unknown and the random field is normal, ordinary kriging provides as an approximation to the BLP [9]. The solution, in this case, includes an additional constraint to ensure the predictor to be unbiased. Hence, ordinary kriging is a best linear unbiased predictor (BLUP). Such an explicit link between BP and normality calls for testing the normality assumption on the basis of a limited number of observations. The literature has typically relied on univariate normality tests: [10] have studied the spatial distribution of some chemical variables in a mine located in Ireland. The authors use a Kolmogorov-Smirnov test and Box-Cox transformations to achieve normality. [11] study the spatial variability of mechanical resistance to penetration in a ferralsol cropped with corn, in Ilha Solteira, Brazil. The normality hypothesis is tested by using a Lilliefors test. [12] use a SW normality test with data of penetration resistance of soil in a region of Poland. [13] assess the distribution pattern of organic carbon in a region of central Japan. [13] assess distribution pattern of organic carbon in a region of central Japan. They use a SW normality test in a previous step of the analysis. Other references about the use of univariate normality tests with spatially distributed data are [14], [15], [16], [17] and [18]. These are just a few examples on the use of classical normality tests with samples from spatial processes. These tests are defined on the basis of the assumptions of independence and identical distribution. This is clearly inadequate in geostatistics, where the observations are typically spatially correlated. Hence, these tests might be inappropriate to say the least. The literature about normality tests in the presence of spatial correlation is sparse, with the exception of [19], who provide a method for obtaining unbiased estimates of the probability density function by weighting the observations using block kriging, and [20], who use a generalized form of the bootstrap method to model the distribution of the statistic of the Kolmogorov-Smirnov test.

Using classical multivariate normality tests [21, 22] in geostatistics is limited because several samples of the random vector are required. In geostatistics, we generally have only one realization of the random field of interest (just one observation of the process is available). Consequently, such tests (in their usual way) cannot be assumed as a viable alternative under a geostatistical framework. In this work, we focus two aspects on. Initially, we show (through simulation) that using univariate normality tests (SW, SF, AD, etc) in geostatistics is not appropriate because these ones do not take into account the spatial dependence. We also show how the Mahalanobis distance [23] can be adapted to a geostatistical scenario to test for multivariate normality of $\mathbf{Y} = (Y(x_1), \cdots, Y(x_n))^T$ of the random process based on just one an observation $y(x) = (y(x_1), \ldots, y(x_n))^T$.

The paper is organized as follows. Section 2 presents an overview of kriging and optimal prediction under normality. In Section 3, we show the approach based on the Mahalanobis distance [23]. A simulation study is developed in Section 4. The paper ends with a short discussion and suggestions for further research.

2 Materials and Methods

In this section we show the basics of simple kriging, ordinary kriging and BP under multivariate normality.

2.1 Simple Kriging and BLP

This section is based on [9], [24], [25], and [26]. Let A be a countable subset of $\mathbb{R}^2$. Geostatistics concerns with the analysis of a real valued stochastic process $\{Y(x) : x \in A\}$, which is typically considered to be a partial realization of a stochastic process $\{Y(x) :$

$x \in \mathbb{R}^2$}. Let $\mathbf{Y} = (Y(x_1), \cdots, Y(x_n))^T$ be a random vector following any multivariate distribution $\mathbf{Y} \sim f(\boldsymbol{\mu}, \Sigma)$, with $\boldsymbol{\mu}$ denoting the mean vector and Σ the covariance matrix. Here T denotes transpose. Assume that $\mathbb{E}(Y(x_i)) = \mu$ and $\mathbb{V}(Y(x_i)) = \sigma^2$, for all $i = 1, \ldots, n$. Then $\mathbf{Y}$ can be written as

$$\mathbf{Y} = \begin{pmatrix} Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} + \begin{pmatrix} \varepsilon(x_1) \\ \vdots \\ \varepsilon(x_n) \end{pmatrix} \quad (1)$$
$$= \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \ x_1, \ldots, x_n \in A,$$

where we assume that

$$\mathbb{V}(\boldsymbol{\varepsilon}) = \Sigma = \begin{pmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma^2 \end{pmatrix},$$

with $\mathbb{V}(Y(x_i)) = \mathbb{V}(\varepsilon(x_i)) = \sigma^2, i = 1, \cdots, n$, and $\mathbb{C}(Y(x_i), Y(x_j)) = \mathbb{C}(\varepsilon(x_i), \varepsilon(x_j)) = \sigma_{ij}, i, j = 1, \ldots, n, i \neq j$. Suppose we want to predict $Y(x_0)$, $x_0 \in A$, based on $\mathbf{Y}$ where $\boldsymbol{\mu}$ and Σ are known. Simple kriging consider linear predictors of the form

$$\widehat{Y}(x_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i Y(x_i)$$
$$= \lambda_0 + \boldsymbol{\lambda}^T \mathbf{Y}, x_1, \ldots, x_n \in A,$$

which minimizes the mean-squared prediction error, denoted MSPE, and defined as

$$\mathbb{E}(\widehat{Y}(x_0) - Y(x_0))^2 = \mathbb{V}(\widehat{Y}(x_0) - Y(x_0))$$
$$+ (\mathbb{E}(\widehat{Y}(x_0) - Y(x_0))^2. \quad (2)$$

Both terms in Equation (2) are nonnegative and the MSPE will be minimized when each term is as small as possible. The second term in the right hand side of Equation (2) is minimized when

$$\mathbb{E}(\widehat{Y}(x_0)) = \mathbb{E}(Y(x_0)),$$

that is, when

$$\lambda_0 + \boldsymbol{\lambda}^T \boldsymbol{\mu} = \mu, \quad \text{with} \quad \lambda_0 = \mu - \boldsymbol{\lambda}^T \boldsymbol{\mu}.$$

On the other hand,

$$\mathbb{V}(\widehat{Y}(x_0) - Y(x_0)) = \boldsymbol{\lambda}^T \Sigma \boldsymbol{\lambda} - 2\mathbf{c}^T \boldsymbol{\lambda} + \sigma^2, \quad (3)$$

with $\mathbb{C}(\mathbf{Y}, Y(x_0)) = (\sigma_{10}, \ldots, \sigma_{n0})^T = \mathbf{c}$. Differentiating (3) with respect to $\boldsymbol{\lambda}$ and equating to zero gives $\boldsymbol{\lambda}^T = \mathbf{c}^T \Sigma^{-1}$. Thus, the BLP is

$$\widehat{Y}(x_0) = \mu + \sum_{i=1}^{n} \lambda_i (Y(x_i) - \mu)$$
$$= \mu + \boldsymbol{\lambda}^T (\mathbf{Y} - \boldsymbol{\mu}) \quad (4)$$
$$= \mu + \mathbf{c}^T \Sigma^{-1} (\mathbf{Y} - \boldsymbol{\mu}).$$

Replacing $\boldsymbol{\lambda}^T = \mathbf{c}^T \Sigma^{-1}$ in (3) yields the simple kriging variance

$$\sigma_{sk}^2 = \sigma^2 - \mathbf{c}^T \Sigma^{-1}. \quad (5)$$

2.2 Ordinary Kriging and BLUP

This section assumes the same conditions as in Section 2.1 but μ is now unknown. The ordinary kriging predictor of $Y(x_0)$ has the same expression as the simple kriging predictor:

$$\widehat{Y}(x_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i Y(x_i), \ x_1, \ldots, x_n \in A.$$

Again, the parameters are estimated by minimizing the MSPE as defined in Equation (2). Note that μ unknown implies λ_0 to be unknown as well. In this case, we guarantee that $\mathbb{E}(\widehat{Y}(x_0)) = \mathbb{E}(Y(x_0))$ by taking $\lambda_0 = 0$ and $\sum_{i=1}^{n} \lambda_i = 1$. The predictor is now restricted to these unbiasedness conditions. This is called best linear unbiased predictor (BLUP). The ordinary kriging predictor is consequently defined defined as

$$\widehat{Y}(x_0) = \sum_{i=1}^{n} \lambda_i Y(x_i)$$
$$= \boldsymbol{\lambda}^T \mathbf{Y}, \quad (6)$$

and the parameters are estimated by minimizing $\mathbb{E}(\widehat{Y}(x_0) - Y(x_0))^2$ subject to the unbiasedness constraint, that is, by considering the following optimization problem:

$$\min_{\boldsymbol{\lambda}} \mathbb{V}(\widehat{Y}(x_0) - Y(x_0)), \ \text{s.t.} \ \sum_{i=1}^{n} \lambda_i = 1. \quad (7)$$

To solve (7), we let m denote the Lagrange multiplier. Then, the optimization function is given by

$$\min_{\lambda_i, m} \mathbb{V}(\widehat{Y}(x_0)) + \mathbb{V}(Y_0) - 2\mathbb{C}(\widehat{Y}(x_0), Y(x_0))$$

$$- 2m(\sum_{i=1}^n \lambda_i - 1) = \min_{\lambda, m} \mathbb{V}(\lambda^T \mathbf{Y}) + \sigma^2$$

$$- 2\mathbb{C}(\lambda^T \mathbf{Y}, Y(x_0)) - 2m(\lambda^T \mathbf{1} - 1)$$

$$= \min_{\lambda, m} \lambda^T \Sigma \lambda + \sigma^2 - 2\lambda^T \mathbf{c} - 2m(\lambda^T \mathbf{1} - 1),$$

where the second line in the chain of equalities comes from the definition of $\widehat{Y}$, while the third line comes from the fact that covariance is a linear operator [23]. Differentiation with respect to λ and m provides the system

$$\Sigma \lambda - \mathbf{c} - m\mathbf{1} = 0, \quad \text{with} \quad \lambda^T \mathbf{1} - 1 = 0. \quad (8)$$

Solving with respect to λ in (8) gives

$$\lambda = \Sigma^{-1}(\mathbf{c} + m\mathbf{1}). \quad (9)$$

To obtain m, we plug λ into the unbiasedness condition in (8), and find

$$(\Sigma^{-1}(\mathbf{c} + m\mathbf{1}))^T \mathbf{1} = 1,$$

$$m = \frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}. \quad (10)$$

Plugging (10) into (9) shows that

$$\lambda = \Sigma^{-1}\left(\mathbf{c} + \frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}\mathbf{1}\right), \text{ and}$$

$$\lambda^T = \left(\mathbf{c} + \mathbf{1}\frac{1 - \mathbf{1}^T(\Sigma^{-1}\mathbf{c})}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}\right)^T \Sigma^{-1}. \quad (11)$$

According to the solution in (11), the kriging predictor $\widehat{Y}(x_0)$ as defined at (6) is given by

$$\widehat{Y}(x_0) = \sum_{i=1}^n \lambda_i Y(x_i) = \lambda^T \mathbf{Y}$$

$$= \left[\left(\mathbf{c} + \mathbf{1}\frac{1 - \mathbf{1}^T \Sigma^{-1}\mathbf{c}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}\right)^T \Sigma^{-1}\right]\mathbf{Y}$$

$$= \mathbf{c}^T \Sigma^{-1}\mathbf{Y} + \left[1 - (\mathbf{1}^T \Sigma^{-1}\mathbf{c})^T\right]\frac{\mathbf{1}^T \Sigma^{-1}\mathbf{Y}}{\mathbf{1}^T \Sigma^{-1}\mathbf{1}}$$

$$= \mathbf{c}^T \Sigma^{-1}\mathbf{Y} + \left[1 - (\mathbf{c}^T \Sigma^{-1}\mathbf{1})\right]\widehat{\mu}$$

$$= \widehat{\mu} + \mathbf{c}^T \Sigma^{-1}(\mathbf{Y} - \mathbf{1}\widehat{\mu}), \quad (12)$$

where $\widehat{\mu}$ is the generalized least squares estimator of μ defined in Equation (1). From the linear regression model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, with $\mathbb{V}(\boldsymbol{\varepsilon}) = \Sigma$, the parameters are estimated by [9]

$$\widehat{\boldsymbol{\beta}} = \left(X^T \Sigma^{-1} X\right)^{-1}\left(X^T \Sigma^{-1}\mathbf{Y}\right). \quad (13)$$

Taking $X = \mathbf{1}$ and $\widehat{\boldsymbol{\beta}} = \widehat{\mu}$ in Equation (13) we obtain, $\widehat{\mu} = \left(\mathbf{1}^T \Sigma^{-1}\mathbf{1}\right)^{-1}\left(\mathbf{1}^T \Sigma^{-1}\mathbf{Y}\right)$. This is the expression for $\widehat{\mu}$ in Equation (12). The variance of the predictor (12) is given by

$$\sigma^2_{\text{ok}} = \sigma^2 - \mathbf{c}^T \Sigma^1 \mathbf{c} + \frac{(1 - \mathbf{1}^T \Sigma^{-1}\mathbf{c})^2}{(\mathbf{1}^T \Sigma^{-1}\mathbf{1})}. \quad (14)$$

Comparing (5) and (14) we observe that $\sigma^2_{\text{sk}} < \sigma^2_{\text{ok}}$, since the last term on the right-hand side of (14) is positive. An unknown mean increases the MSPE. In practice, Σ and $\mathbf{c}$ are unknown and must be estimated. Then, a *plug-in* estimation of λ and m is usually carried out. Thus the ordinary kriging predictor is not a BLUP but an estimated BLUP (EBLUP, throughout).

2.3 Best Prediction

We provide a well know result about optimal prediction, establishing that under the squared loss function, the BP is the conditional expectation of the predictor given the information. For a formal proof the reader is refereed to [25] and [27].

Theorem 2.1. *Let* $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ *be a vector composed of a finite collection of n random variables. Denote the corresponding realizations as* $\mathbf{y} = (y_1, \ldots, y_n)^T$. *Let Y_0 be a random variable, and* $\widehat{Y}_0 = \varphi(\mathbf{Y})$ *be any function of $\mathbf{Y}$ for $\varphi : \mathbb{R}^n \to \mathbb{R}$. Then,* $\mathbb{E}[(Y_0 - \widehat{Y}_0)^2]$ *attains its minimum value when* $\widehat{Y}_0 = \mathbb{E}(Y_0 | \mathbf{Y})$.

Let $\{Y(x) : x \in R^2\}$ be a normal random field [24]. Then

$$\begin{pmatrix} Y(x_0) \\ Y(x_1) \\ \vdots \\ Y(x_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \mathbf{c}^T \\ \mathbf{c} & \Sigma \end{pmatrix}\right),$$

$$(15)$$

with the vector $\mathbf{c}$ and the matrix Σ as defined in Section 2.1. Standard properties of the multivariate normal distribution [23] show that, if

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

then $\mathbf{Y}^{(1)} | \mathbf{Y}^{(2)} \sim N(\boldsymbol{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Y}^{(2)} - \boldsymbol{\mu}^{(2)}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ [23]. This property

in concert with Equation (15) shows that

$$Y(x_0)|\mathbf{Y} \sim N(\mu + \mathbf{c}^T\Sigma^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \sigma^2 - \mathbf{c}^T\Sigma^{-1}\mathbf{c}).$$

Consequently, the BP under the squared loss function with the stationary normal model is

$$\hat{Y}(x_0) = \mathbb{E}(Y(x_0)|\mathbf{Y}) = \mu + \mathbf{c}^T\Sigma^{-1}(\mathbf{Y} - \boldsymbol{\mu}), \quad (16)$$

with variance

$$\mathbb{V}(Y(x_0)|\mathbf{Y}) = \sigma^2 - \mathbf{c}^T\Sigma^{-1}\mathbf{c}.$$

The predictors in Equations (4) and (16) are actually the same. Thus, simple kriging is an optimal predictor if the random field is normal. This result has become popular in applied geostatistics ([14], [12], [17], [18]). However, it has been sometimes misunderstood. Some comments are in order:

- Data recorded in geostatistics correspond to a realization of a stochastic process. These should not be regarded as one sample of a random variable. Then, testing for univariate normality in this field is meaningless. If $\{Y(x) : x \in R^2\}$ is a normal spatial random process then the random vector $\mathbf{Y} = (Y(x_1), \ldots, Y(x_n))^T \sim N(\boldsymbol{\mu}, \Sigma)$. Only when the process is stationary and uncorrelated (mean and variance being constant and $\Sigma = \sigma^2 I$) it has sense testing for univariate normality.

- Simple kriging, as defined in Section 2.1, is a BP whether the stochastic process of interest is normal, and it is a BLP regardless of whether the process is normal. In practice, using simple kriging is limited because μ and Σ are unknown.

- Ordinary kriging is not a BP, even when the stochastic process is normal. This predictor is always a BLUP (when μ is estimated and Σ is known) or an EBLUP (when μ and Σ are estimated).

- When the stochastic process is normal, maximum likelihood can be used for estimating μ, Σ, and $\mathbf{c}$ and consequently for carrying out prediction by using ordinary kriging.

Although common practice, the use of classic univariate normality tests does not make sense in geostatistics. It is required to take into account the spatial dependency structure. The following section proposes an adaptation of the Mahalanobis statistic for this purpose.

3 A New Test for Normality in Geostatistics Based on the Mahalanobis Distance

We now illustrate how a normality test based on the Mahalanobis distance [23], can be adapted to a geostatistical framework. Let $A = \{x_1, \ldots, x_n\}$ and $\mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma)$. It can be shown that

$$\mathbf{Q_1} = (\mathbf{Y} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2(n), \quad (17)$$

where $\chi^2(n)$ denotes the central χ^2 distribution with n degrees of freedom [28].

Assume that we have a random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_r$ from the n-dimensional random vector $\mathbf{Y}$, that is we have a random matrix defined as

$$\begin{pmatrix} \mathbf{Y}_1^T \\ \mathbf{Y}_2^T \\ \vdots \\ \mathbf{Y}_r^T \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{r1} & Y_{r2} & \cdots & Y_{rn} \end{pmatrix},$$

where $Y_{ij}, i = 1, \ldots, r$, is a random sample of size r from the random variable $Y_j, j = 1, \ldots, n$. It is known that asymptotically

$$\mathbf{Q_2} = (\mathbf{Y} - \bar{\mathbf{Y}})^T\hat{\Sigma}^{-1}(\mathbf{Y} - \bar{\mathbf{Y}}) \sim \chi^2(n), \quad (18)$$

where $\bar{\mathbf{Y}}$ is the sample mean vector and $\hat{\Sigma}$ is the sample covariance matrix [28].

Let $\mathbf{Y} = (Y(x_1), \ldots, Y(x_n))^T$ be a realization of size n from a stationary random process $\{Y(x) : x \in R^2\}$ with $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}$ and $\mathbb{V}(\mathbf{Y}) = \Sigma$. Suppose we have an observation $\mathbf{y} = (y(x_1), \ldots, y(x_n))^T$ of this process and we want to test the hypothesis

$$\text{H}_o : \mathbf{Y} \sim N(\boldsymbol{\mu}, \Sigma) \quad \text{against} \quad \text{H}_a : \mathbf{Y} \nsim N(\boldsymbol{\mu}, \Sigma).$$

Given $\boldsymbol{\mu}$ and Σ known (an unrealistic scenario), the statistic in Equation (17) is calculated as

$$\mathbf{Q_1^c} = (\mathbf{y} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu}). \quad (19)$$

The null hypothesis of normality is then rejected at level α if

$$\mathbf{Q_1^c} < \chi^2_{\frac{\alpha}{2}}(n) \text{ or } \mathbf{Q_1^c} > \chi^2_{1-\frac{\alpha}{2}}(n).$$

On the other hand, let $\bar{\mathbf{y}} = (\bar{y}, \dots, \bar{y})^T_{n \times 1}$ be a mean vector with $\bar{y} = \frac{\sum_{i=1}^n y(x_i)}{n}$ and

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}^2_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1n} \\ \hat{\sigma}_{21} & \hat{\sigma}^2_{22} & \cdots & \hat{\sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{n1} & \hat{\sigma}_{n2} & \cdots & \hat{\sigma}^2_{nn} \end{pmatrix}, \qquad (20)$$

be the estimated covariance matrix, where $\hat{\sigma}_{ij} = \hat{C}(Y(x_i), Y(x_j)), i,j = 1, \dots, n$. The elements of this matrix can be obtained by fitting to the experimental covariogram a covariance model (Exponential, Gaussian, spherical, etc) through ordinary or weighted least squares [26]. Then, based on $\bar{\mathbf{y}}$ and $\hat{\Sigma}$, we can calculate the statistic defined in Equation (18) as

$$\mathbf{Q_2^c} = (\mathbf{y} - \bar{\mathbf{y}})^T \hat{\Sigma}^{-1} (\mathbf{y} - \bar{\mathbf{y}}). \qquad (21)$$

Consequently, the null hypothesis of normality of the random process is rejected at level α if

$$\mathbf{Q_2^c} < \chi^2_{\frac{\alpha}{2}}(n) \text{ or } \mathbf{Q_2^c} > \chi^2_{1-\frac{\alpha}{2}}(n).$$

We study through Monte Carlo simulation the performance of some classical normality tests (SW, SF, and AD) as well as the tests based on the Mahalanobis distance. The goal is two-fold: to show that using univariate normality tests in a geostatistical context can be inappropriate, and to assess the behavior of the statistics $\mathbf{Q_1}$ and $\mathbf{Q_2}$ defined in Equations (17) and (18).

4 Simulation Study

4.1 Simulation Under Normality

We simulate normal random fields over regular grids with $n = 25, 49$ and 100 points on a square of 10×10 units (Figure 1). Also in order to establish the effect of the spatial distribution of the sampling points on the tests a size $n = 100$ clustered configuration is considered (Figure 1).
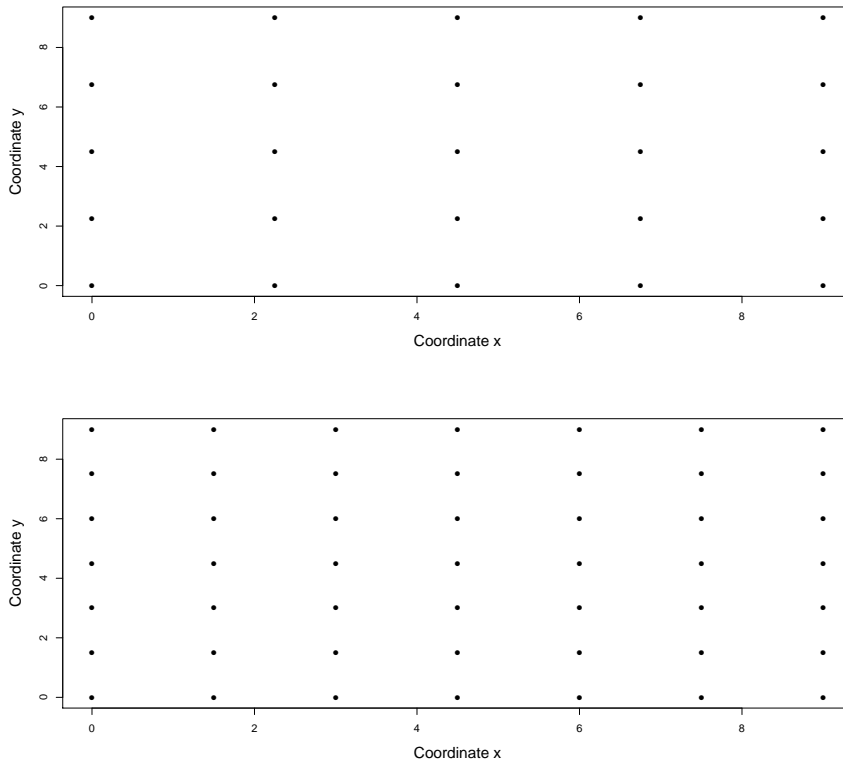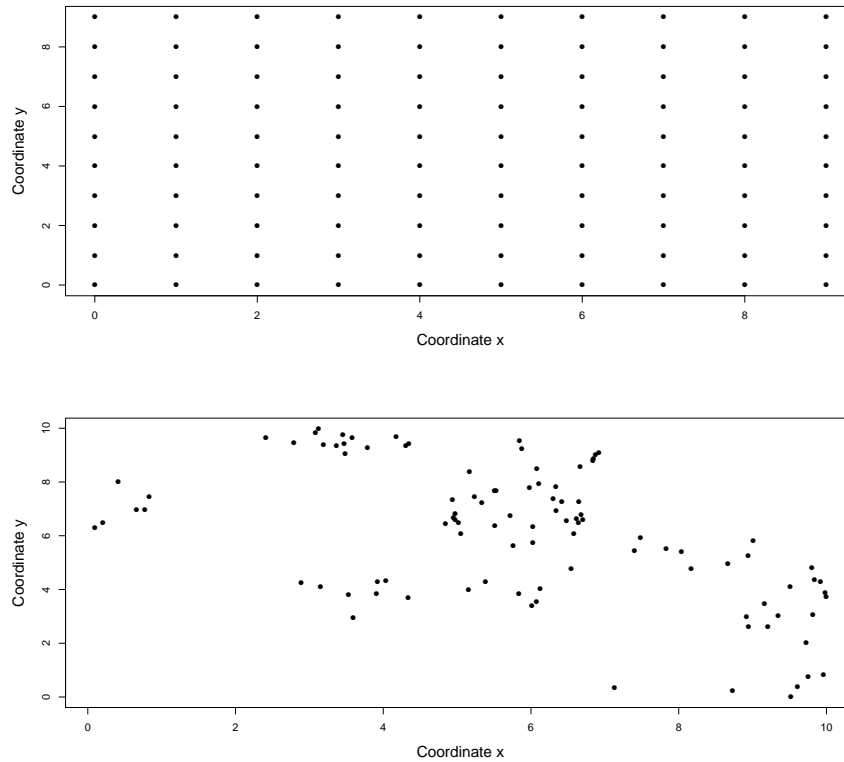
Figure 1. Spatial configurations used in the simulations. Regular grids size $n = 25$ (top), $n = 49$ (second panel) and $n = 100$ (third panel). Clustered grid $n = 100$ (bottom).

At each case, we simulate 1000 realizations of $\mathbf{Y} = (Y(x_1), \ldots, Y(x_n))^T \sim N(\boldsymbol{\mu}, \Sigma)$, with $\boldsymbol{\mu} = (100, \ldots, 100)^T_{(n \times 1)}$, and the covariance matrix given by

$$\Sigma = \begin{pmatrix} \sigma^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma^2 \end{pmatrix}, \qquad (22)$$

where $\sigma_{ij} = C(h) = \sigma^2 \exp(-\frac{h}{\phi}), \sigma^2, \phi > 0$, and $h = \|x_i - x_j\|$, that is, $C(Y(x_i), Y(x_j)), i, j = 1, \ldots, n$, is defined by an exponential model [24]. We consider two values for σ^2 (8 and 400) and three values for ϕ (3, 7 and 12), establishing thus two levels of variability (low and high) and three levels of spatial correlation (low, medium and high), respectively. The models considered are plotted in Figure 2. Combining n (25, 49, 100, 100), σ^2 (8 and 400) and ϕ (3, 7, 12) values, in total, we have 24 simulation scenarios. We use the software R [29] to obtain the simulations. The random fields are simulated by using the function grf of the library geoR [30]. The clustered configuration of points is defined by

using the function rMatClust of the library sp. The functions shapiro.test, sf.test and ad.test of the library nortest [32] are used to carry out the tests of SW, SF and AD, respectively. The tests based on the Mahalanobis distance ($\mathbf{Q_1}$ and $\mathbf{Q_2}$) are conducted by using the package base of R [29]. A number of 1000 simulations under each scenario (combinations of n, σ^2 and ϕ) are obtained and the rejection rate for each one of the five tests is estimated by

$$\hat{\alpha} = \frac{\sum_{i=1}^{1000} \delta_i}{1000},$$

with δ_i a dichotomous variable taking the value 1 if the null hypothesis is rejected and 0 otherwise. In all cases, the tests are carried out by using a significance level $\alpha = 5\%$, i.e, the test considered is appropriate if the estimation of α ($\hat{\alpha}$) obtained from the simulation is close to the nominal level 5%. The farther away from this value, the worse its performance. The results of the estimations are shown in Table 1.
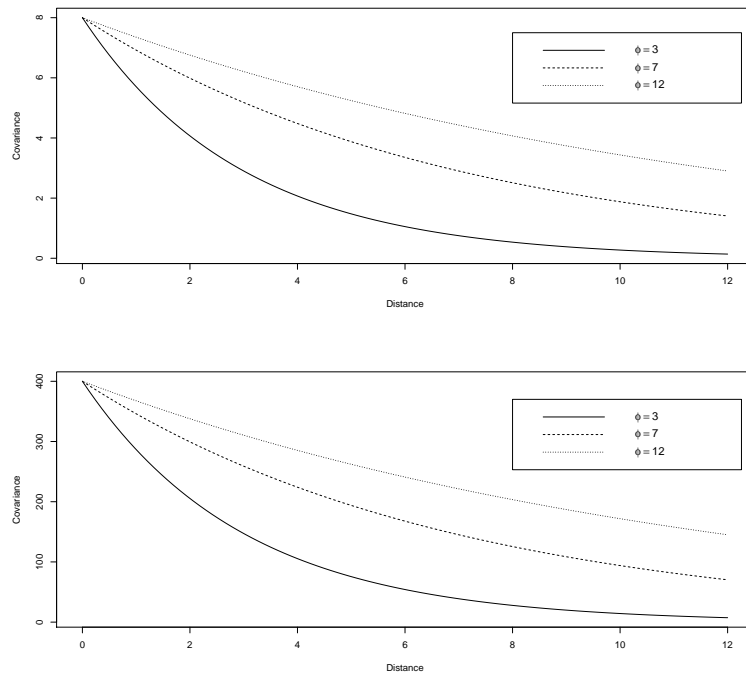
Figure 2. Covariance models with parameters $\sigma^2 = 8$ (top) and $\sigma^2 = 400$ (bottom).

Table 1. Estimated rejection rates ($\hat{\alpha}$) of the null hypothesis of normality obtained by using Monte Carlo simulation of normal random fields with different levels of variability (σ^2) and correlation (ϕ). In all cases is considered that the process follows an exponential correlation model. The rejection rates are calculated for three regular grids ($n = 25, 49$ and 100) and one clustered grid ($n = 100$, fourth line at each case).

Test	Grid	$\sigma^2 = 8,$ $\phi = 3$	$\sigma^2 = 8,$ $\phi = 7$	$\sigma^2 = 8,$ $\phi = 12$	$\sigma^2 = 400,$ $\phi = 3$	$\sigma^2 = 400,$ $\phi = 7$	$\sigma^2 = 400,$ $\phi = 12$
Shapiro-Wilk	$n = 25$	0.05	0.06	0.07	0.05	0.06	0.07
	$n = 49$	0.08	0.12	0.15	0.08	0.12	0.16
	$n = 100$	0.15	0.29	0.33	0.16	0.28	0.33
	$n = 100$	0.43	0.52	0.60	0.42	0.53	0.58
Shapiro-Francia	$n = 25$	0.05	0.06	0.06	0.05	0.05	0.06
	$n = 49$	0.06	0.09	0.12	0.06	0.09	0.12
	$n = 100$	0.12	0.23	0.28	0.13	0.23	0.28
	$n = 100$	0.37	0.46	0.54	0.37	0.47	0.52
Anderson-Darling	$n = 25$	0.05	0.06	0.07	0.05	0.06	0.07
	$n = 49$	0.08	0.12	0.15	0.08	0.13	0.16
	$n = 100$	0.16	0.28	0.33	0.17	0.28	0.34
	$n = 100$	0.43	0.52	0.60	0.42	0.53	0.57
$\mathbf{Q_1}$	$n = 25$	0.05	0.05	0.05	0.05	0.05	0.05
	$n = 49$	0.05	0.05	0.05	0.05	0.05	0.05
	$n = 100$	0.05	0.05	0.05	0.05	0.05	0.05
	$n = 100$	0.06	0.04	0.06	0.06	0.05	0.05
$\mathbf{Q_2}$	$n = 25$	0.02	0.07	0.09	0.03	0.06	0.09
	$n = 49$	0.01	0.05	0.06	0.02	0.05	0.06
	$n = 100$	0.01	0.02	0.05	0.01	0.02	0.05
	$n = 100$	0.01	0.02	0.06	0.01	0.03	0.05

Concerning the univariate tests (SW, SF and AD), several points are identified from Table 1. The estimations of the rejection rates have a very similar behavior (although the values with SF are slightly lower). In general the results are unsatisfactory. It is clear that these tests do not have a good performance because the $\hat{\alpha}$ values obtained are greater than the nominal $\alpha = 5\%$ defined in the simulations, except in those cases where there is both low spatial dependence ($\phi = 3$) and small sample size ($n = 25$). It can also be observed that the stronger the spatial dependence, the bigger the difference between $\hat{\alpha}$ and α (the higher the spatial correlation, the worse the performance of these tests). A comparison of the $\hat{\alpha}$ values reported for the three statistics (SW, SF, AD) in the two 100-point spatial configurations (regular and clustered, Table 1) indicates that the magnitude of the error (overestimation of α) increases if a clustered sampling is considered. All the above results are consistent: it is natural that the classical tests are not satisfactory, because they are based on the assumption of independence and the simulations are obtained from normal random fields with an underlying correlation structure.

As for the tests based on the statistics $\mathbf{Q_1}$ and $\mathbf{Q_2}$, we can observe in Table 1 that in all cases the estimated rejection probabilities ($\hat{\alpha}$) (after rounding to two decimal places) are close to the significance level $\alpha = 5\%$ considered (particularly when $n \geq 49$), i.e. these tests are unbiased [33]. The rejection rates obtained with $\mathbf{Q_1}$ indicate that under ideal situations (when the parameters are known), the Mahalanobis distance would perform satisfactorily (as expected in this case all the $\hat{\alpha}$ values are practically equal to α because the parameters $\boldsymbol{\mu}$ and Σ in Equation (17) should not be estimated). Likewise, the rejection rates estimated with $\mathbf{Q_2}$, although slightly different from the 5%, are also appropriate (in all cases less than 9%). The relationship between n and $\hat{\alpha}$ corresponding to the statistics $\mathbf{Q_2}$ (for all correlation and variability levels considered) is illustrated in Figure 3. Only the estimations obtained with the regular grids are considered (for the clustered distribution we only have the estimations with $n = 100$ and these are practically the same as those obtained with the regular grid of the same size). We note according to the interpolations lines in this

Figure, that the optimal n is around 50 when $\phi = 7$, while this value is close to 100 when $\phi = 12$, i.e., the greater the spatial dependence, the bigger the sample size required to get a good approximation to the nominal level $\alpha = 5\%$ defined in the simulations.

Note from Table 1 that in general (for all the test considered) when n and ϕ are fixed the estimations of the rejection probabilities obtained under the two levels of variability ($\sigma^2 = 8$ and 400) are very similar, i.e, $\hat{\alpha}$ values in columns 3 and 6, 4 and 7, and 5 and 8 are close. In other words, the variability of the random process has no influence on the result of the normality test. The tests depends only on the number of observations (n) of the random process and the level of spatial correlation (ϕ). This pattern is apparent in Figure 3 (to the case of $\mathbf{Q_2}$), where it can be seen that the variation of $\hat{\alpha}$ (with respect to n and ϕ) is practically the same in both panels ($\sigma^2 = 8$, above, and $\sigma^2 = 400$, below).

According to the above comments, the classical tests of normality (SW, SF, AD) should not be used in geostatistics. On the other hand, the test based on the $\mathbf{Q_1}$ statistic, although satisfactory, is not applicable in practice because the vector of means and the covariance matrix are unknown. An appropriate test, according to the estimates given in Table 1, is that based on the $\mathbf{Q_2}$ statistic. In summary, the estimations based on $\mathbf{Q_2}$ given in Table 1 reveal that this statistic can be used in applied geostatistics for carrying out the normality test and also indicate that its performance depends on the sample size.

4.2 Simulation Under Log-Normality

We also evaluate the power of the test, denoted with π throughout. For this purpose, we simulate data from log-normal stationary processes. Assuming $\{Y(x), x \in \mathbb{R}^2\}$ is a normal stationary stochastic process with mean μ and variance σ^2, a log-normal process can be obtained from $Y(x)$ by means of the transformation $T(x) = \exp(Y(x))$ [24]. It is known that mean and variance of $T(x)$ are given by [26]

$$\mathbb{E}(T(x)) = \exp\left(\mu + \frac{\sigma^2}{2}\right), \text{ and}$$

$$\mathbb{V}(T(x)) = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right). \quad (23)$$
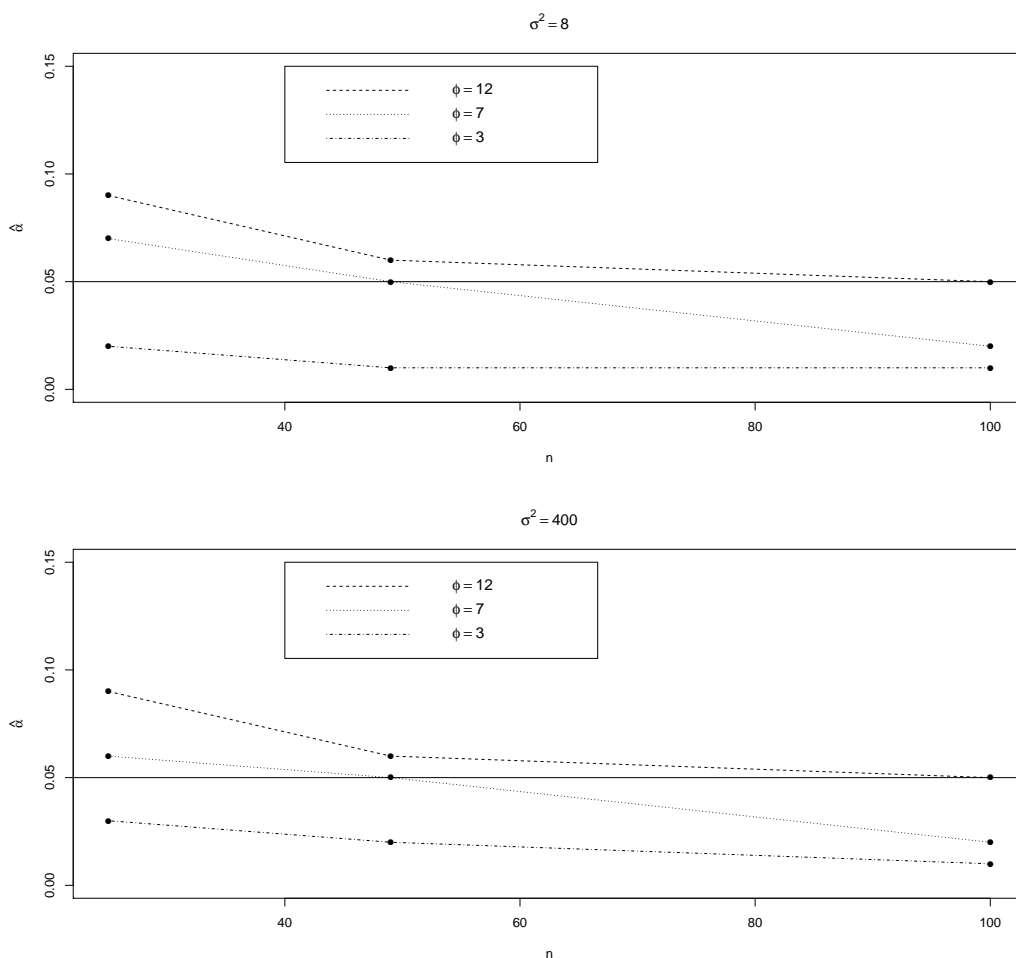
Figure 3. Optimal number of sites (points) required to have a good estimation of α when is used the statistic $\mathbf{Q_2}$. Lines calculated only with the estimations of the regular grids.

We use the expressions in (23) to obtain realizations of log-normal random fields starting from simulations of normal processes with $\mu = 0$ and $\sigma = 1$. The parameters μ and σ are changed with respect to those used in Section 4.1 for simplicity on the calculations. It is clear that these values have no effect on the estimation of the rejection probabilities $\hat{\pi}$. With these specifications of μ and σ we have

$$\mathbb{E}(T(x)) = \exp\left(\mu + \frac{\sigma^2}{2}\right) = 1.65, \qquad (24)$$

and

$$\begin{aligned} \mathbb{V}(T(x)) &= \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right) \\ &= \exp(1)\left(\exp(1) - 1\right) \\ &= 4.67 = \sigma_2^2. \end{aligned} \qquad (25)$$

The covariance matrix Σ (see Equation 22) required to calculate the statistic $\mathbf{Q_1}$ (Equation 17) for the log-normal processes is based on Equations (24) and (25). The elements of the matrix are now given by

$$\sigma_{ij} = \sigma_2^2 \exp\left(-\frac{h}{\phi}\right), \text{with } \phi > 0; \text{ and } h = \|x_i - x_j\|.$$

The matrix $\hat{\Sigma}$ used for obtaining $\mathbf{Q_2}$ is calculated at each case by fitting covariance exponential models by ordinary least squares to the simulated processes (analogously to the procedure used in Section 4.1). Three sample sizes $n = (25, 49, 100)$ and five correlation coefficients $\phi = (10, 20, 30, 40, 50)$ are used to evaluate empirically the power of the tests. A size $n = 100$ clustered configuration (as defined in Section 4.1) is also considered to evaluate the effect of spatial distribution on the estimation of π. The results are shown in Table 2.

Table 2. Estimated power ($\hat{\pi}$) of the normality hypothesis obtained by using Monte Carlo simulation of log-normal random fields with different levels of correlation (ϕ) and sample size (n). In all cases it is considered that the process has an exponential spatial correlation structure. The power is calculated for three regular grids ($n = 25, 49$ and 100) and one clustered grid ($n = 100$, fourth line at each case).

Test	Grid	$\phi = 10$	$\phi = 20$	$\phi = 30$	$\phi = 50$	$\phi = 70$
Shapiro-Wilk	$n = 25$	0.77	0.58	0.47	0.34	0.27
	$n = 49$	0.94	0.77	0.68	0.56	0.47
	$n = 100$	0.99	0.93	0.84	0.75	0.68
	$n = 100$	0.95	0.88	0.83	0.74	0.72
Shapiro-Francia	$n = 25$	0.74	0.55	0.44	0.31	0.25
	$n = 49$	0.92	0.74	0.65	0.52	0.43
	$n = 100$	0.99	0.91	0.81	0.71	0.64
	$n = 100$	0.93	0.85	0.79	0.72	0.67
Anderson-Darling	$n = 25$	0.70	0.52	0.43	0.30	0.24
	$n = 49$	0.90	0.70	0.61	0.50	0.42
	$n = 100$	0.98	0.88	0.78	0.68	0.62
	$n = 100$	0.92	0.84	0.78	0.73	0.69
$\mathbf{Q_1}$	$n = 25$	0.77	0.80	0.82	0.82	0.82
	$n = 49$	0.84	0.85	0.86	0.87	0.88
	$n = 100$	0.88	0.90	0.92	0.92	0.92
	$n = 100$	0.89	0.90	0.90	0.91	0.92
$\mathbf{Q_2}$	$n = 25$	0.75	0.75	0.75	0.75	0.75
	$n = 49$	0.81	0.83	0.83	0.84	0.85
	$n = 100$	0.86	0.88	0.88	0.88	0.89
	$n = 100$	0.88	0.89	0.89	0.90	0.90

Two points can be highlighted from the $\hat{\pi}$ values. First, the estimations based on the statistics $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are in general much bigger than derived with the classical tests (except when $\phi = 10$ or both $\phi = 20$ and $n = 100$) i.e., the tests based on the Mahalanobis distance are more powerful. On the other hand, there are marked differences between trends of change when the level of correlation increases ($\phi \geq 30$). The three classical tests (SW, SF, AD) show the same pattern. There is an inverse relationship between $\hat{\pi}$ and ϕ, that is, the stronger the spatial dependence, the lower the estimation of the rejection probability ($\hat{\pi}$). On the contrary, in the case of the estimations based on $\mathbf{Q_1}$ and $\mathbf{Q_2}$ the relationship is monotonous increasing and $\hat{\pi}$ is around 90% (when $\phi = 70$ and $n = 100$ (in both spatial configurations)).

Summarizing, the results in Tables 1 and 2 allow us to establish empirically that the classical normality tests such as SW, SF and AD are not appropriate to evaluate this assumption in geostatistics because tend to both overestimate the significance level ($\hat{\alpha}$ values for these tests in Table 1 are much greater than the nominal level $\alpha = 5\%$

considered, particularly when the sample size n and the correlation level ϕ increase) and underestimate the power ($\hat{\pi}$ values in Table 2 for these tests are much lower that obtained with $\mathbf{Q_1}$). Results also show that the Mahalanobis distance-based approaches ($\mathbf{Q_1}$ and $\mathbf{Q_2}$) have a good performance ($\hat{\alpha}$ values (Table 1) are close to the nominal level $\alpha = 5\%$ and the power values $\hat{\pi}$ (Table 2) increase as expected when the sample size n and the correlation level ϕ increase). Consequently the approaches proposed and particularly the statistics $\mathbf{Q_2}$ can be applied in practical problems of geostatistics for testing for normality when the process under study is stationary (when simple and ordinary kriging is going to be used to carry out spatial prediction).

Kriging prediction can be applied independently of the distribution that follows the process. However, under normality the prediction is optimal (simple kriging is a BP). In addition to the combination of the method-of-moments and least squares, maximum likelihood can be used to estimate the parameters of the covariance model. This is a straightforward procedure widely used when the random process is

normal. For these reasons (optimality and simplicity of estimation) testing for normality is a relevant step in any geostatistical analysis. Classical normality tests based on SW, SF or AD are not suitable in this scenario. The statistics proposed in this work are a good alternative to fulfill this stage of the analysis. Particularly in practical geostatistics, the statistic $\mathbf{Q_2}$ can be used in a preliminary step of the analysis for testing for normality.

5   Conclusion and Further Research

In this paper, we introduced an approach to test for normality in geostatistics. Specifically we showed how the Mahalanobis distance, traditionally used in multivariate analysis, can be adapted to a geostatistical context to test for normality. Simulation results indicated firstly that using univariate tests such as Shapiro-Wilk, Shapiro-Francia or Anderson-Darling in geostatistics is inadequate and on the other hand, that the strategy proposed is attractive from a practical point of view because it can be calculated based on just one realization of the underlying random process. There is room for research in the context of this paper. A generalization to the non-stationary case is required. Also the test could be extended to a multivariate geostatistical scenario.

References

[1] G. Robertson. "Geostatistics in ecology: interpolating with known variance", *Ecology*, vol. 68, no. 3, pp 744-748, 1987.

[2] T. Hooks, D. Marx, S. Kachman, J. Pedersen, R. Eigenberg, R. "Analysis of covariance with spatially correlated secondary variables", *Revista Colombiana de Estadística*, vol. 31, no. 1, pp. 95-109, 2008.

[3] G. Severino, M. Scarfato, G. Toraldo. "Mining geostatistics to quantify the spatial variability of certain soil flow properties", *Procedia Computer Science*, vol. 98, 419-424, 2016.

[4] I. Gundogdu. "Usage of multivariate geostatistics in interpolation processes for meteorological precipitation maps", *Theoretical and Applied Climatology*, vol. 127, no. 1-2, 81-86, 2017.

[5] R. Giraldo, L. Herrera, V. Leiva. "Cokriging prediction using as secondary variable a functional random field with application in environmental pollution", *Mathematics*, vol. 8, no. 8, 1305, 2020.

[6] M. Oliver, R. Webster. *Basic steps in geostatistics: The variogram and kriging*. Springer. 2015.

[7] E. Lehmann, G. Casella. *Theory of Point Estimation*. Springer-Verlag, 1998.

[8] J. Chiles, P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, 1999.

[9] O. Schabenberger, C. Gotway. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall, 2005.

[10] D. McGratha, C. Zhangb, O. Cartona. "Geostatistical analysis and hazard assessment on soil lead in silvermines area, Ireland", *Environmental Pollution*, vol. 127, 239-248, 2004.

[11] V. Júnior, M. Carvalho, J. Dafonte, O. Freddi, E. Vidal, O. Ingaramoc. "Spatial variability of soil water content and mechanical resistance of Brazilian ferralsol ", *Soil & Tillage Research*, vol. 85, 166-177, 2006.

[12] A. Dexter, A. Czyz, O. Gate. "A method for prediction of soil penetration resistance ", *Soil & Tillage Research*, vol. 93, 412-419, 2007.

[13] K. Kamarudin, M. Tomita M, K. Kondo, S. Abe. "Geostatistical estimation of surface soil carbon stock in Mt. Wakakusa grassland of Japan", *Landscape and Ecological Engineering*, vol. 15, no.2, 215-221, 2019.

[14] J. Iqbal, J. Thomasson, J. Jenkins, P. Owens, Whisler, F. "Spatial variability analysis of soil physical properties of alluvial soils", *Soil Science Society of America Journal*, vol. 69, 1338-1350, 2005.

[15] L. Pozdnyakova, D. Giménez, P. Oudemans. "Spatial analysis of cranberry yield at three scales", *Agronomy Journal*, vol. 97, 49-57, 2005.

[16] M. Carrara, A. Castrignano, A. Comparetti, P. Febo, S. Orlando. "Mapping of Penetrometer Resistance in Relation to Tractor Traffic Using Multivariate Geostatistics", *Geoderma*, vol. 142, 294-307, 2007.

[17] J. Lima, S. Silva. "Multivariate analysis and geostatistics of the fertility of a humic rhodic hapludox under coffee cultivation", *Revista Brasileira de Ciencia do Solo*, vol. 36, no. 2, 467-474, 2012.

[18] E. Barca, D. De Benedetto, A. Stellaccic. "Contribution of EMI and GPR proximal sensing data in soil water content assessment by using linear mixed effects models and geostatistical approaches ", *Geoderma*, vol. 343, no. 1, 280-293, 2019.

[19] E. Pardo-Igúzquiza, P. Dowd. "Normality tests for spatially correlated data", *Mathematical Geology*, vol. 36, no. 6, 659-681, 2004.

[20] R. Olea, V. Pavlosky. "Kolmogorov-Smirnov test for spatially correlated data", *Stochastic Environmental Research and Risk Assessment*, vol. 23, no.6, 749-757, 2008.

[21] K. Mardia. "Measures of multivariate skewness and kurtosis with applications", *Biometrika*. vol. 57, no. 3, 519-530, 1970.

[22] G.Szekely, M. Rizzo. "A new test for multivariate normality", *Journal of Multivariate Analysis*. vol. 93, 58-80, 2005.

[23] T. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 1984.

[24] N. Cressie. *Statistic for Spatial Data*. John Wiley & Sons, 1993.

[25] P. Diggle, P. Ribeiro. *Model-Based Geostatistics*. 14th Simposio Nacional de Probabilidade e Estatística, Associacao Brasileira de Estatística, 2000.

[26] P. Diggle, P. Ribeiro. *Model-Based Geostatistics*. Springer, 2007.

[27] E. Lehmann, G. Casella. *Theory of Point Estimation*. Springer-Verlag, 1998.

[28] I. Cardoso De Oliveira, D. Ferreira. "Multivariate extension of Chi-squared univariate normality test", *Journal of Statistical Computation and Simulation*. vol. 80, no. 5, 513-526, 2010.

[29] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. 2019.

[30] P. Ribeiro, P. Diggle. "geoR: A package for geostatistical analysis", *Journal of Statistical Computation and Simulation*. vol. 1, no. 2, 14-18, 2007.

[31] E. Pebesma, R. Bivand. "S classes and methods for spatial data: The sp package", *R news*, vol. 5, no. 2, 9-13, 2005.

[32] J. Gross, U. Ligges. nortest: Tests for Normality. R package version 1.0-4, URL https://CRAN.R-project.org/package=nortest. 2015.

[33] Dudewicz E, Mishra S. *Modern Mathematical Statistics*. John Wiley & Sons, 1988.

[34] H. R. Vega, E. Manzanares, V.M. Hernández, G. A. Mercado, E. Gallego y A. Lorente, "Características dosimétricas de fuentes isotópicas de neutrones", *Revista Mexicana de Física*, vol. 51, no. 5, pp. 494-501, 2005.

[35] H. R. Vega y C. Torres, "Low energy neutrons from a 239PuBe isotopic neutron source inserted in moderating media", *Revista Mexicana de Física*, vol. 48, no. 5, pp. 405-412, 2002.

[36] N. C. Tam, J. Bagi y L. Lakosi, "Determining Pu isotopic composition and Pu content of PuBe sources by neutron coincidence technique", *Nuclear Instruments and Methods in Physics Research*, vol. 262, pp. 75-80, 2007.

[37] I. ElAgib, J. Csikai, J. Jordanova y L. OlaAh, "Leakage neutron spectra from spherical samples with a Pu-Be source", *Applied Radiation and Isotopes*, vol. 51, pp. 329-333, 1999.

[38] J. Bagi, N. C. Tam y L. Lakosi, "Assessment of the Pu content of Pu-Be neutron sources", *Nuclear Instruments and Methods in Physics Research*, vol. 222, pp. 242-248, 2004.

[39] R. H. Zachary, "Neutron flux and energy characterization of a plutonium-beryllium isotopic neutron source by Monte Carlo simulation with verification by neutron activation analysis", Las Vegas, USA: University of Nevada, 2010.

[40] N. A. Carrillo y H. R. Vega, "Cálculo de los espectros de neutrones de una fuente isotópica moderada", Presentado en 5as Jornadas de Investigación de la UAZ, CB/UEN-10/042, 2001.

[41] B. Pelowitz, "MCNPX User's Manual version 2.5.0 Los Alamos National Laboratory Report LA-UR-02-2607", Presentado en Computation Topical Meeting, Avignon, France, 2005.