

Empirical variogram for achieving the best valid variogram

Esam Mahdi^{1,a}, Ali H. Abuzaid^b, Abdu M. A. Atta^a

^aDepartment of Mathematics, Statistics, and Physics, Qatar University, Qatar;

^bDepartment of Mathematics, Al-Azhar University, Gaza

Abstract

Modeling the statistical autocorrelations in spatial data is often achieved through the estimation of the variograms, where the selection of the appropriate valid variogram model, especially for small samples, is crucial for achieving precise spatial prediction results from kriging interpolations. To estimate such a variogram, we traditionally start by computing the empirical variogram (traditional Matheron or robust Cressie-Hawkins or kernel-based nonparametric approaches). In this article, we conduct numerical studies comparing the performance of these empirical variograms. In most situations, the nonparametric empirical variable nearest-neighbor (VNN) showed better performance than its competitors (Matheron, Cressie-Hawkins, and Nadaraya-Watson). The analysis of the spatial groundwater dataset used in this article suggests that the wave variogram model, with hole effect structure, fitted to the empirical VNN variogram is the most appropriate choice. This selected variogram is used with the ordinary kriging model to produce the predicted pollution map of the nitrate concentrations in groundwater dataset.

Keywords: spatial dependence, variogram, nonparametric estimation, kriging

1. Introduction

In many situations we have data measured at spatial locations that can be analyzed by geostatistical methods. Such methods are concerned in modeling the spatial trend and spatial correlation where the traditional statistical methods cannot reflect the spatial features and dependencies in such data. Modelling the statistical autocorrelation in spatial data is often achieved through the estimation of the variogram, instead of a correlogram or covariogram, where the variogram plays an important role in predicting the value of the target variable allocated at an unsampled distinct location (Matheron, 1962, 1963a, 1963b; Barry and Pace, 1997; Cressie, 1993, Chapter 3).

In geostatistics, it is common to assume that the spatial random process $\{Z(s) : s \in D \subset \mathbb{R}^d\}$, where D is a subset of d -dimension Euclidean space, to be an isotropic and intrinsically stationary random function; that is,

$$E[Z(s_i) - Z(s_j)] = 0, \quad \forall s_i, s_j \in D, \quad (1.1)$$

$$\text{Var}[Z(s_i) - Z(s_j)] = 2\gamma(\mathbf{h}), \quad \forall s_i, s_j \in D \quad (1.2)$$

or equivalently (under the constant mean assumption), $2\gamma(\mathbf{h}) = E[(Z(s_i) - Z(s_j))^2]$, where $2\gamma(\mathbf{h})$ and $\gamma(\mathbf{h})$ referred to *Variogram* and *Semivariogram* respectively, and $\mathbf{h} \in \mathbb{R}^d$ is the lag interval that

¹ Corresponding author: Department of Mathematics, Statistics, and Physics, Qatar University, University Street, Doha, P.O. Box: 2713, Qatar. E-mail: emahdi@qu.edu.qa.

represents the distance between the two points s_i and s_j . In this article, the word variogram is used synonymously with semivariogram.

One crucial assumption of the variogram is conditional negative-definite, defined as

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(s_i - s_j) \leq 0, \quad (1.3)$$

for any finite number of spatial locations s_1, \dots, s_n and real numbers a_1, \dots, a_n satisfying $\sum_{i=1}^n a_i = 0$ (Armstrong and Jabin, 1981; Dunn, 1983).

The conditional negative-definite variogram has the spectral representation

$$2\gamma(\mathbf{h}) = \int_0^\infty (1 - \Omega_d(\mathbf{u}\mathbf{h})) dF(\mathbf{u}), \quad (1.4)$$

where

$$\Omega_d(\mathbf{x}) = \left(\frac{2}{\mathbf{x}}\right)^{\frac{d-2}{2}} \Gamma\left(\frac{d}{2}\right) K_{\frac{d-2}{2}}(\mathbf{x}), \quad (1.5)$$

$K_\kappa(\cdot)$ is the Bessel function of the first kind of order κ and $F(\cdot)$ is a non-decreasing function on $(0, \infty)$, satisfying $\int_0^\infty \mathbf{u}^2 dF(\mathbf{u}) / (1 + \mathbf{u}^2) < \infty$ (Yaglom, 1987, Section 25).

Therefore, it is important to use a valid estimated variogram that satisfies the conditional negative-definite assumption given in (1.3) and (1.4) so that it can be used in spatial prediction models; otherwise, we may get negative mean squared prediction errors (Jin and Kelly, 2017).

In practice, the first step in modeling and estimating the valid variogram is to compute the traditional empirical variogram or the kernel-based nonparametric version of the traditional empirical variogram (Section 2).

In many situations, the empirical variogram estimator is not guaranteed to be conditionally negative-definite (Armstrong and Jabin, 1981; Cressie, 1993, p.69). In this respect, the empirical variogram is often approximated by a theoretical one that is guaranteed to be conditionally negative definite (Cressie, 1993). This theoretical variogram can be approximated by either a parametric or a nonparametric valid model. Comparative studies of both approaches, conducted by Menezes *et al.* (2005) and Yu *et al.* (2007), showed superior results of the nonparametric estimators.

In this article, we evaluate the performance of the empirical variogram based on the traditional Matheron, Cressie-Hawkins, and the kernel-based nonparametric approaches. In most cases, our simulation results indicate that the kernel-based nonparametric empirical *variable nearest-neighbor* (VNN) has a better performance than its competitors: the traditional Matheron, Cressie-Hawkins, and the kernel-based *Nadaraya-Watson* (NW). We construct interpolated pollution maps of the nitrate concentrations in the drinking water wells of the Gaza Strip in Palestine. We find that the wave variogram fitted to the nonparametric empirical VNN variogram is the most appropriate model to describe the spatial dependency in the nitrate dataset, so we use it in producing the final interpolated kriging pollution maps.

In the next section, brief reviews of modeling and estimating the variograms are given. Simulation studies in Section 3 comparing the performance of the traditional empirical variograms (Matheron and Cressie-Hawkins) to two kernel-based nonparametric variograms (VNN and NW) are presented. The environmental spatial dataset (the nitrate concentrations in the groundwater of Gaza Strip) being analyzed is described in Section 4. Section 5 summarizes the conclusion results with brief recommendations that can be used as a general guideline for decision-makers, water-resource managers, and private users in Gaza Strip for preparing strategic plans for protecting water supplies.

2. Estimation and modeling variogram

Let $\{Z(s) : s \in D \subset \mathbb{R}^d\}$ be a random process defined on a subset D of d -dimension Euclidean space which can be modeled as follows

$$Z(s) = \mu(s) + \varepsilon(s), \tag{2.1}$$

where $\mu(s)$ denotes the deterministic trend which is usually modelled as a linear form $\sum_{j=0}^p \beta_j X_j(s)$ where $\{X_j(s); j = 0, \dots, p, \text{ with } X_0(s) = 1\}$ is a collection of nonrandom explanatory variables that may or may not depend on spatial locations and β_0, \dots, β_p are unknown regression coefficients; $\{\varepsilon(\cdot)\}$ is a zero-mean isotropic and intrinsically stationary random process with stationary autocovariance function

$$\text{Cov}(\varepsilon(s + \mathbf{h}), \varepsilon(s)) = C(\mathbf{h}), \quad \forall s \in D \text{ and } \mathbf{h} > \mathbf{0}. \tag{2.2}$$

Suppose that we want to predict the value of the target variable Z at an unobserved location $s_0, \hat{Z}(s_0)$, based on the values of the observed data of size n . This would be done by kriging interpolation methods that give the best linear unbiased prediction (Isaaks and Srivastava, 1989, p.289). Universal kriging predictor takes the form of a linear combination of the data as follows

$$\hat{Z}(s_0) = \sum_{j=0}^p \hat{\beta}_j X_j(s_0) + \sum_{i=1}^n w_i(s_0)(Z(s_i) - \hat{\mu}(s_i)), \tag{2.3}$$

where $X_0(s_0) = 1$; $\hat{Z}(s_0)$ is the estimated random variable at unobserved sites s_0 ; $s_i, i = 1, \dots, n$ is the i^{th} spatial location; $\hat{\beta}_j$ is the generalized-least-square estimator of β_j ; and the weights $\{w_i(s_0), i = 1, \dots, n\}$. The kriging predictors are constructed by minimizing $\text{Var}(\hat{z}(s) - z(s))$ subject to $E[\hat{Z}(s)] = E[Z(s)] = \mu(s)$. The ordinary kriging is a special case of the universal kriging for which p in (2.3) equals zero. Note that the general prediction formula of the inverse distance weighting is the same as the one given by the ordinary kriging, where the weights depend solely on the observations within the nearest-neighbor locations.

The assumption of a constant mean is not realistic in most spatial data; therefore, we have to detrend the surface and then calculate the sample variogram. After the trend removal, we consider $Z(s)$ to be the second order stationary. Under the stationary assumptions, the bounded variogram is related to the autocovariance function by the equation $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$ so that kriging interpolation relies on a valid variogram instead of a covariogram.

In this article, we render the residuals stationary and use $Z(s)$ to compute the variogram for further analysis. The traditional empirical variogram estimator as introduced by Matheron (1962) based on the method-of-moments is

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(i,j) \in N(\mathbf{h})} (Z(s_i) - Z(s_j))^2, \tag{2.4}$$

where $|N(\mathbf{h})|$ stands for the number of distinct pairs in $N(\mathbf{h})$, which is the set formed by those locations that are a distance \mathbf{h} apart. A tolerance region for the distance \mathbf{h} is generally used (Cressie, 1993). Although, there is no specific rule for determining the size of the lag \mathbf{h} , choosing the lag classes well is important so that the researchers can use their knowledge of the phenomenon they are analyzing (somewhat *ad hoc* procedures) (García-Soidán *et al.*, 2004).

The Matheron's estimator is unbiased; however, it has some drawbacks such as being badly affected by outliers. Under the Gaussianity assumption, Cressie and Hawkins (1980) introduced a stable robust to the outliers estimate of the variogram

$$2\hat{\gamma}(\mathbf{h}) = \frac{\left\{ \frac{1}{|N(\mathbf{h})|} \sum_{(i,j) \in N(\mathbf{h})} |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|^{\frac{1}{2}} \right\}^4}{\left(0.457 + \frac{0.494}{|N(\mathbf{h})|}\right)}, \quad (2.5)$$

Genton (1998) indicated that the Cressie and Hawkins (1980) does not set boundaries for the effect of outliers and a single large outlier could make the estimate meaningless.

Genton (1998) introduced a robust scale estimator based on the M-estimators of scale as proposed by Croux and Rousseeuw (1992), Rousseeuw and Croux (1993). The resulting estimator of Genton (1998) is

$$2\hat{\gamma}(\mathbf{h}) = \left[2.219 \left\{ |Z(\mathbf{s}_i) - Z(\mathbf{s}_j)|; i < j \right\}_{\left(\frac{H}{2}\right)} \right]^2, \quad (2.6)$$

where H is the integer part of $n/2 + 1$, and the term $\{.\}_{\left(\frac{H}{2}\right)}$ denotes the value of the $\left(\frac{H}{2}\right)$ ordered term in braces.

These empirical estimators have some drawbacks in situations where the variogram curve is not a smooth function due to non-increasing behaviors; therefore, it does not capture all spatial information presented in the pattern. However, the kernel-based approach provides a nonparametric estimator of the variogram which is smooth enough to capture most of the spatial information presented in the pattern for which the other approach might not reveal (Diggle, 1985; Diggle *et al.*, 1987; Berman and Diggle, 1989; Hall *et al.*, 1994; Stein, 1999; Lahiri *et al.*, 1999; Menezes *et al.*, 2005; Yu *et al.*, 2007; Huang *et al.*, 2011; García-Soidán and Menezes, 2012).

The standard nonparametric empirical variogram of Matheron (1962) based on the Nadaraya-Watson's kernel estimator as proposed by Hall *et al.* (1994), García-Soidán *et al.* (2003, 2004) is

$$2\hat{\gamma}(\mathbf{h}) = \frac{\sum_i \sum_j K\left(\frac{\mathbf{h} - \|\mathbf{s}_i - \mathbf{s}_j\|}{b}\right) (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2}{\sum_i \sum_j K\left(\frac{\mathbf{h} - \|\mathbf{s}_i - \mathbf{s}_j\|}{b}\right)}, \quad (2.7)$$

where b is the smoothing parameter which presents the kernel bandwidth and $K(\cdot)$ is a symmetric kernel with bounded second-order moments. It is generally agreed that the exact shape of the kernel function, $K(\cdot)$, is not as important as the smoothing parameter and the choice of the bandwidth is based on the usual trade-off between bias and variance. In this respect, the optimal choice of the smoothing parameter given by minimizing the *asymptotic mean integrated squared error* (AMISE) is

$$\hat{b} = \left(\frac{n^{-1}R(K)}{m_2(K)^2 R(\gamma''(\mathbf{h}))} \right)^{\frac{1}{5}}, \quad (2.8)$$

where

$$\text{AMISE} = (n\mathbf{h})^{-1}R(K) + \frac{1}{4}\mathbf{h}^2 m_2(K)^2 R(\gamma''(\mathbf{h})), \quad (2.9)$$

with $R(K) = \int_{-\infty}^{\infty} K(u)^2 du$; $m_2(K) = \int_{-\infty}^{\infty} u^2 K(u) du$.

Yu *et al.* (2007) proposed two kernel estimators by replacing b in Hall *et al.* (1994) estimator by a constant (δ) and a non-constant ($\delta_0(s_{ij})$) nearest-neighbor parameters, where $s_{ij} = \|s_i - s_j\|$. Yu *et al.* (2007) indicated that, in most of the situations, the estimator based on the non-constant nearest-neighbor parameters is more general and better than the other one. The *constant nearest-neighbor parameter estimator* (CNN) is defined as:

$$2\hat{\gamma}(\mathbf{h}) = \frac{\sum_i \sum_j I(|\mathbf{h} - s_{ij}| < \delta) (Z(s_i) - Z(s_j))^2}{\sum_i \sum_j I(|\mathbf{h} - s_{ij}| < \delta)}, \tag{2.10}$$

where $I(\cdot)$ is the indicator function, whereas the variable nearest-neighbor estimator is given by:

$$2\hat{\gamma}(\mathbf{h}) = \frac{\sum_i \sum_j \frac{1}{\delta_0(s_{ij})} K\left(\frac{\mathbf{h}-s_{ij}}{\delta_0(s_{ij})}\right) (Z(s_i) - Z(s_j))^2}{\sum_i \sum_j \frac{1}{\delta_0(s_{ij})} K\left(\frac{\mathbf{h}-s_{ij}}{\delta_0(s_{ij})}\right)}. \tag{2.11}$$

The optimal choice of the nearest-neighbor smoothing parameter is derived by minimizing the AMISE

$$\text{AMISE} = \int_0^\infty (B^2(\mathbf{h}) + V(\mathbf{h})) d\mathbf{h}, \tag{2.12}$$

with $B^2(\mathbf{h}) = (1/4)\gamma''(\mathbf{h})^2 m_2(K)^2 (\delta\delta_0(\mathbf{h}))^4$ and $V(\mathbf{h}) = 2\hat{\gamma}(\mathbf{h})^2 2R(K) / \{n(n-1)\delta\delta_0(\mathbf{h})\}$.

Yu *et al.* (2007) showed that the optimal choice of the constant nearest-neighbor smoothing parameters derived by minimizing takes the form

$$\hat{\delta} = \left(\frac{2R(K) \int_0^a 2\hat{\gamma}(\mathbf{h})^2 d\mathbf{h}}{n(n-1)m_2(K)^2 \int_0^a \gamma''(\mathbf{h})^2 d\mathbf{h}} \right)^{\frac{1}{5}}, \tag{2.13}$$

where $a > 0$, in practice, can be taken as the estimated range or a percentage of the variogram, whereas the optimal choice of the variable nearest-neighbor smoothing parameters derived by minimizing take the form

$$\hat{\delta} = \xi \left(\frac{2R(K)}{n(n-1)m_2(K)^2} \right)^{\frac{1}{5}}, \tag{2.14}$$

and

$$\hat{\delta}_0(\mathbf{h}) = \begin{cases} \xi \left(\frac{2\hat{\gamma}(\mathbf{h})}{|\gamma''(\mathbf{h})|} \right)^{\frac{2}{5}}, & \text{if } \gamma''(\mathbf{h}) \neq 0, \\ \xi^*, & \text{otherwise,} \end{cases} \tag{2.15}$$

with ξ and ξ^* being two arbitrary positive constants related to the degree of the spatial smoothing parameters. Yu *et al.* (2007) suggested to consider of $\xi = 1, 2$ and ξ^* to be greater than or equal to $2\hat{\gamma}(\mathbf{h})^2$. Yu *et al.* (2007, Theorem 3) showed that the AMISE based on the optimal choice of $\hat{\delta}$ in (2.14) and $\hat{\delta}_0(\mathbf{h})$ in (2.15) is less than the optimal one given in (2.13). It is worth noting that the indicator kernel function in $d = 2$, the case of most interest in practice, indicates that $\hat{\delta}$ in (2.13) essentially equals \hat{b} given in (2.8). Note also that the estimators in Equations (2.8), (2.13), and (2.15) depend on the unknown true variogram. In practice, for \mathbb{R}^2 , it is common to use the Epanechnikov kernel

$$K(u) = 0.75b^{-1} \left(1 - \left(\frac{u}{b} \right)^2 \right) I(|u| \leq b), \tag{2.16}$$

where Diggle (1985) suggests to consider the estimated bandwidth value $\hat{b} = 0.68n^{-1/5}$ for a unit square.

2.1. Valid models

The drawback of the aforementioned empirical variograms (kernel and traditional) is that they may fail to satisfy the conditionally negative-definite property which may lead to non-valid variograms that we cannot use directly for spatial prediction. In this respect, the empirical variogram is often approximated by a theoretical one that is guaranteed to be conditionally negative definite (Cressie, 1993). Estimation of the distribution function of the variogram at a specific location can be addressed in two ways: the parametric way and the nonparametric way. Both approaches are strongly defended by their authors. The parametric method assumes that the spatial data are drawn from a pre-specified parametric distribution under some assumptions.

Let $r = \| \mathbf{h} \|$ so that $\hat{\gamma}_i = \hat{\gamma}(r_i)$, $i = 1, 2, \dots, n$ are the estimated variogram values (e.g., empirical variogram in (2.4), the first nonparametric approach to select a valid model was introduced by Shapiro and Botha (1991). They used the *weighted least squares* (WLS) criterion based on minimizing the weighted sum squares

$$\sum_{i=1}^n w_i [\hat{\gamma}(r_i) - \gamma(r_i)]^2. \quad (2.17)$$

The weights w_i are given by $w_i = N(r_i)\gamma(r_i)^{-2}$, where $N(r_i)$ is the number of contributing pairs used to estimate the variogram at the i^{th} lag, which satisfies the conditions of the spectral representation in (1.4) and (1.5) under some nonlinear constraints (Cressie, 1985). It should be noted that García-Soidán *et al.* (2003) adopted the local linear regression techniques to estimate the nonparametric empirical variogram given in (2.7) where they applied the Shapiro and Botha (1991) test to estimate the local linear empirical estimator by a valid variogram.

For the parametric approach, the common valid variogram models are the Linear, Circular, Exponential, Spherical, Gaussian, Matérn, and Wave models defined as follows, (the parameters $\boldsymbol{\theta} = (c_0, c_1, r)$ are the nugget effect (Although the definition of the variogram indicates that $\gamma(\mathbf{0}) = 0$, the estimator $\hat{\gamma}(\mathbf{0}) \neq 0$ due to the high variability between the observations taken at two very close points that leads to increase the size of the discontinuity at zero; in addition, the discontinuity of the variogram at zero is called a *nugget effect.*), the partial sill (It is very often that the variogram stops increasing beyond a certain distance and becomes more or less stable around a limit value called a *sill* value $= c_0 + c_1$), and the range (The correlation between two variables tends to be equal to zero when the distance \mathbf{h} becomes too large. The distance r beyond which $C(\mathbf{h})$ considered to be zero is called the *range.*) respectively (Chilés and Delfiner, 1999; Cressie, 1993):

- **The spatial bounded linear variogram model**

$$\gamma(h; \boldsymbol{\theta}) = \begin{cases} c_0 + c_1 \left(\frac{h}{r}\right), & \text{if } 0 < h \leq r, \\ c_0 + c_1, & \text{if } h > r, \\ 0, & \text{if } h = 0, \end{cases} \quad (2.18)$$

where this model is valid in one-dimensional case, i.e., $d = 1$.

- **The spatial circular variogram model**

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} c_0 + c_1 \left(1 - \frac{1}{\pi} \cos^{-1} \left(\frac{h}{r}\right) + \frac{2h}{\pi} \sqrt{1 - \frac{h^2}{r^2}}\right), & \text{if } \mathbf{h} \leq r, \\ c_0 + c_1, & \text{otherwise.} \end{cases} \quad (2.19)$$

The circular model reaches the sill at $\|\mathbf{h}\| = r$, where the tangent at $\|\mathbf{h}\| = 0$ would cross the sill $c = c_0 + c_1$ at a distance of $(\pi/4)r$.

- **The spatial exponential variogram model**

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = c_0 + c_1 \left(1 - \exp\left(-3\frac{\|\mathbf{h}\|}{r}\right)\right), \quad \text{for } \mathbf{h} \neq 0. \quad (2.20)$$

The exponential model reaches 95% of the sill at $\|\mathbf{h}\| = r$ asymptotically, where the tangent at $\|\mathbf{h}\| = 0$ would cross the sill at a distance of $(1/3)r$.

- **The spatial spherical variogram model**

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \begin{cases} c_0 + c_1 \left(1.5\left(\frac{\|\mathbf{h}\|}{r}\right) - 0.5\left(\frac{\|\mathbf{h}\|}{r}\right)^3\right), & \text{if } \|\mathbf{h}\| < r, \\ c_0 + c_1, & \text{otherwise.} \end{cases} \quad (2.21)$$

The spherical model reaches the sill at $\|\mathbf{h}\| = r$, where the tangent at $\|\mathbf{h}\| = 0$ would cross the sill at a distance of $(2/3)r$.

- **The spatial Gaussian variogram model**

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = c_0 + c_1 \left(1 - \exp\left(-3\frac{\|\mathbf{h}\|^2}{r^2}\right)\right). \quad (2.22)$$

The Gaussian model reaches 95% of the sill at $\|\mathbf{h}\| = r$ asymptotically, where the tangent at $\|\mathbf{h}\| = 0$ would cross the sill $c = c_0 + c_1$ at a distance of $(1/\sqrt{3})r$.

- **The Matérn variogram model**

$$\gamma(\mathbf{h}; \boldsymbol{\theta}; \kappa) = c_0 + c_1 \left[1 - \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{\|\mathbf{h}\|}{r}\right)^{\kappa} K_{\kappa}\left(\frac{\|\mathbf{h}\|}{r}\right)\right], \quad (2.23)$$

where $K_{\kappa}(\cdot)$ is the modified Bessel function of second kind of order κ , and κ is a parameter controlling the smoothness of the realized random field (Matérn, 1960). Note that the spatial case of Matérn model $\kappa = 0.5$ gives the exponential model and $\kappa = \infty$ is the Gaussian model.

- **The spatial wave (hole-effect) variogram model**

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = c_0 + c_1 \left(1 - \pi^{-1} \left(\frac{r}{\|\mathbf{h}\|}\right) \sin\left(\frac{\pi\|\mathbf{h}\|}{r}\right)\right), \quad \text{for } \mathbf{h} \neq 0. \quad (2.24)$$

The wave model is widely used when there is some periodicity in the data resulting in a hole effect due to positive and negative correlations among different distant regions. The range in this model is the shortest distance that occurs on the initial rise in the variogram function for which the variogram equals the sill.

Gorsich and Genton (2000) noticed that most of these valid parametric variogram models appear very similar, especially for large samples; however, their derivatives with respect to the lags are not. Gorsich and Genton (2000) showed that the derivative of the nonparametric variogram estimator can be used as a good tool to aid in the selection of the underlying variogram model. They proposed a method for the selection of a valid parametric model via the derivation of a nonparametric variogram estimate to avoid or simplify the task of checking different valid parametric variogram models.

2.2. Best valid variogram

A common approach to achieve a valid variogram estimator is accomplished through the following three steps:

- Step 1: Estimate an empirical parametric (classical Matheron or robust Cressie-Hawkins) or kernel-based nonparametric variogram (Nadaraya-Watson or variable nearest-neighbor). Typically, this variogram is non-valid and not useful in kriging interpolation.
- Step 2: Select a reasonable best theoretical model from the family of valid parametric or nonparametric variograms that can fit the empirical model obtained from Step 1. This model is usually selected by *fitting-by-eye* and *ad hoc* methods.
- Step 3: Estimate the parameters of the fitted theoretical model. The fitting procedures have either been by-eye, least squares, weighted least squares, or restricted maximum likelihood estimation.

We propose an alternative procedure for selecting the best theoretical variogram model. The proposed algorithm slightly modifies the aforementioned steps as:

- Step 1: Estimate the empirical parametric and nonparametric variograms appearing in the literature (Matheron, Cressie-Hawkins, Nadaraya-Watson, and variable nearest-neighbor).
- Step 2: Fit some reasonable theoretical models to each one of the estimated variograms. In this step, the parameters of the fitted model can be approximated by the WLS, *ordinary least squares* (OLS), or *residual maximum likelihood* (REML) method.
- Step 3: Use kriging (ordinary kriging or regression-kriging) with each of the estimated theoretical models, and then evaluate the prediction efficiency for each case. The accuracy of the prediction can be evaluated by using the *adjusted coefficient of multiple determination* (R_a^2):

$$R_a^2 = 1 - \left(\frac{n-1}{n-p} \right) (1 - R^2), \quad (2.25)$$

where R^2 indicates the amount of variance explained by the model and p denotes the number of variables used in the regression model.

An alternative method, that we recommend, is to use the *K-fold cross-validation* ($CV_{(K)}$) technique based on the *mean squared prediction error* (MSPE) (e.g., Hastie *et al.*, 2009, Chapter 7):

$$CV_{(K)} = \left(\frac{1}{K} \right) \sum_{k=1}^K MSPE_k, \quad (2.26)$$

where

$$MSPE_k = \left(\frac{1}{m_k} \right) \sum_{i=1}^{m_k} (z_i - \hat{z}_i^{(-k)})^2, \quad (2.27)$$

where for $k = 1, \dots, K$ folds, you fit the model with the k^{th} fold removed, and obtain the predictions $\hat{z}_i^{(-k)}$ for $i = 1, \dots, m_k$, where m_k is the number of observations in the k^{th} fold.

As a rule of thumb, $K = 5$ or $K = 10$ is a good choice, but one can also use the leave-one-out-cross-validation by choosing the number of folds to be $K = m$.

- **Step 4:** Choose the model that provides the lowest $CV_{(K)}$ where the estimated *mean squared standardized prediction error* (MSSPE) should be as close to 1 as possible:

$$\text{MSSPE}_k = \left(\frac{1}{m_k} \right) \sum_{i=1}^{m_k} \frac{(z_i - \hat{z}_i^{(-k)})^2}{\sigma_i^{2(-k)}}, \quad (2.28)$$

where $\sigma_i^{2(-k)}$, $i = 1, \dots, m_k$ are the kriging variances for $k = 1, \dots, K$ folds.

Once, you select the best model, reuse it with the full dataset.

The proposed algorithm may be expensive due to the need for computational efficiency when selecting a valid model; however, with advances in modern computing power has made this easy to accomplish.

3. Simulation studies

In this section, we conduct two different simulation studies to investigate the proposed algorithm for achieving the best valid variogram. The simulations were all run using the **R** packages **geoR** (Ribeiro and Diggle, 2001), **npsp** (Fernández-Casal, 2016), **sp** (Pebesma and Bivand, 2005), and **gstat** (Gräler *et al.*, 2016). For the first one, we simulate Gaussian spatial data, each of length $n = 109$ with a mean of zero. The reason for selecting this sample size is to imitate the real spatial dataset, which has 109 observations (Section 4). This also reflects the fact that many applications including environmental and epidemiology real data are relatively small. The coordinates of the real spatial dataset are used as a polygon data to simulate the conditional irregular grids of the spatial locations in our simulated data. For the second, we consider the case of regular grid by simulating several Gaussian dataset, each of length $n = 200$ with a mean of zero and spatial locations $s_i = (x_i, y_i)$, $i = 1, 2, \dots, n$, using a uniform distribution on $[0, 1] \times [0, 1]$.

For both simulation schemes, we consider the aforementioned six theoretical variograms: Circular, Exponential, Spherical, Gaussian, Matérn, and Wave models to specify the spatial dependency. The parameters of these models were selected in a way that the curves were comparable to the corresponding range.

In the first simulation study, we fix the values of the nugget, c_0 , and the partial sill, c_1 , to be 1,600 and 2,000, respectively. The range parameter is chosen depending on the theoretical model: circular, $r = 4,000$; exponential, $r = 1,330$; spherical, $r = 4,000$; Gaussian, $r = 2,300$; Matérn, $r = 2,000$; and wave, $r = 2,830$. The order of the Bessel function, κ , of the Matérn model equals 0.7.

In the other simulation scheme, we consider $\kappa = 0.7$ and fix the values of the nugget and the partial sill to be 0.2 and 5, respectively. The values of the range parameters corresponding to the circular, exponential, spherical, Gaussian, Matérn, and wave variogram are 0.5, 0.167, 0.5, 0.289, 0.250, and 0.354, respectively. Some of these parameters values were selected from similar previous simulation studies (e.g., Menezes *et al.*, 2005, 2008).

With these selections, we have sill values of 3,600 and 5.2 in the first and second cases, respectively, and for each case, we have a range that reaches the corresponding sill or asymptotically approaches 95% of the sill when the range is not finite. For both cases, we consider 1,000 replications where we focus our attention on the isotropic case as the analysis of the real spatial dataset shows no clear signs of anisotropy (see the rose diagram displayed in Figure 5).

First, we evaluate the empirical parametric and nonparametric variograms by comparing the esti-

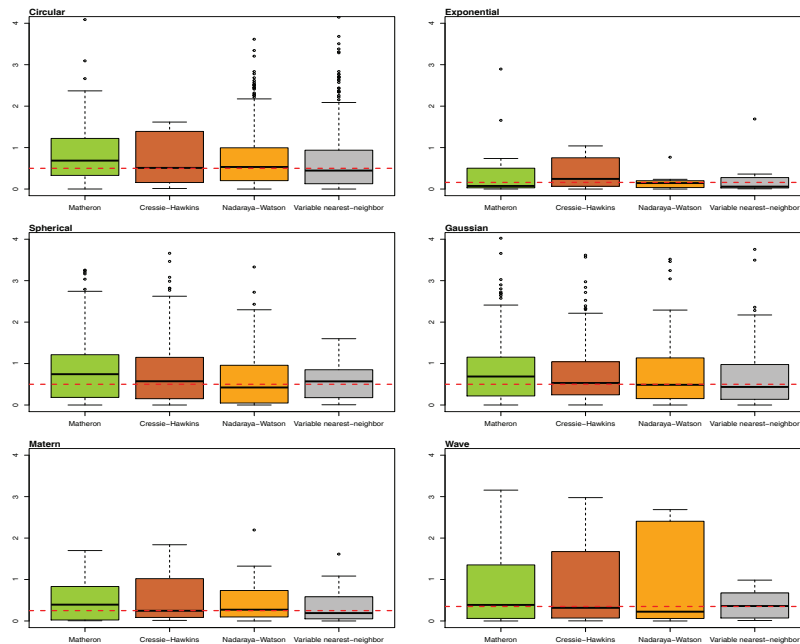


Figure 1: Box-plots of the evaluated ISE from four empirical parametric and nonparametric variograms are based on fitting the true theoretical model to a generated irregular spatial data with length 109 according to the first simulation scheme with 1,000 replications. The dashed line denotes the true median parameter value.

mated *integrated squared error* (ISE) given by

$$\text{ISE} = \int_{h_0}^{h_1} [\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h})]^2 d\mathbf{h}, \quad (3.1)$$

where $\hat{\gamma}(\mathbf{h})$ and $\gamma(\mathbf{h})$ are the empirical estimator and the theoretical curve, respectively, and the integral is calculated over an interval of selected distances $[h_0, h_1]$. The ISE provides a useful measure of model accuracy which measures the deviation between each of the empirical estimators and the theoretical variogram, and the smaller the values of the ISE the better the selected model. The ISE can be used as a sufficient criterion for model diagnostics in spatial analysis and cross-validation is unnecessary (Yu *et al.*, 2007; Cressie, 1993). For each generated data set, we obtain the estimates of the empirical parametric variograms given in (2.4) and (2.5) by considering the lags of 20 and 30 pairs for the first and second simulation methods, respectively, whereas the empirical variograms given in (2.7) and (2.11) are estimated using the optimal choices of the smoothing parameters that minimizes the AMISE given in (2.9) and (2.12), respectively. For each simulation method, we approximate the true marginal distribution by fitting the true theoretical model based on the weighted least squares method and then we use the trapezoidal rule to estimate the ISE values between this model and each of the empirical models.

Figures 1–2 display the box-plots describing the distribution of the ISE values based on the median and interquartile range. The dashed red line denotes the true median parameter value. In addition, the median values of the empirical variogram estimators are close to the theoretical values and the distribution of the ISE of each variogram is slightly right skewed suggesting that the majority of the

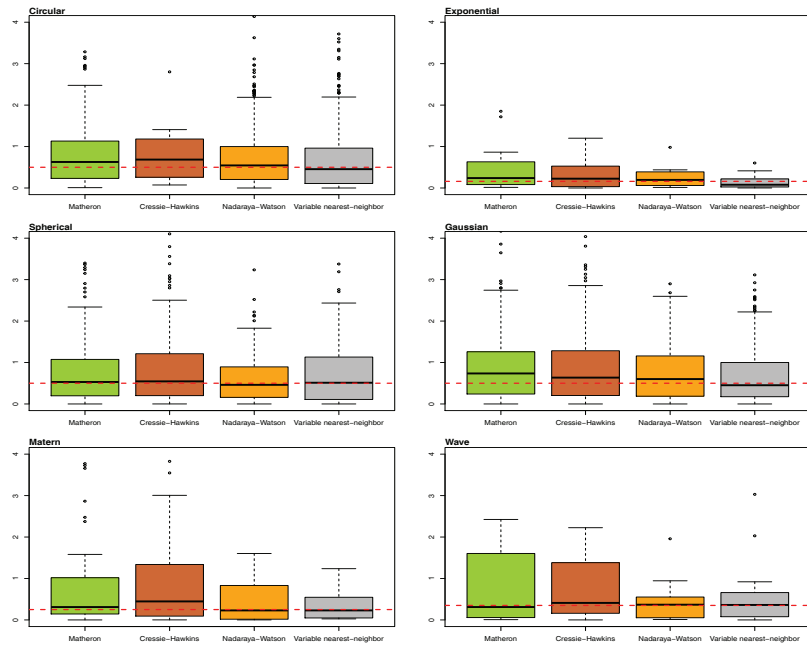


Figure 2: *Box-plots of the evaluated ISE from four empirical parametric and nonparametric variograms, based on fitting the true theoretical model to a generated regular spatial data with length 200 according to the second simulation scheme with 1,000 replications. The dashed line denotes the true median parameter value.*

ISE values are small. In most cases, the figures show that the best is achieved by the variable nearest-neighbor followed by Nadaraya-Watson kernel models in the presence of the exponential, Gaussian, Spherical, Matérn, and wave models. It is also evident that the lowest ISE values are associated with the wave curve and the largest ones are associated with the circular curve, where the robust parametric Cressie-Hawkins variogram is the preferred choice in the circular model. In general, based on the simulation results, we conclude that the traditional Matheron and Cressie-Hawkins estimators may have some drawbacks in situations where the variogram curve is not a smooth function due to non-increasing behaviors; therefore, it does not capture all spatial information presented in the pattern. However, the kernel-based approach provides a nonparametric estimator of the variogram which is smooth enough to capture most of the spatial information presented in the pattern that the other approach might not reveal. Next, we fit each of the six parametric models given by (2.19)–(2.24) to the four empirical parametric and nonparametric variograms. We use the weighted least squares method to estimate the parameters of the fitted model and then use it with simple kriging. For each case, we randomly split the generated data into $K = 10$ folds as test data and consider the remaining $K - 1$ folds as training data and estimate the $MSPE_k$, $MSSPE$, and $CV_{(K)}$ values.

Table 1 shows the estimated $CV_{(K)}$ and $MSSPE$ (in parenthesis) values, where the $CV_{(K)}$ with the lowest value is put in boldface to assist the reader. When $MSSPE < 1$, we underestimate the true estimation; when $MSSPE > 1$, we overestimate; when $MSSPE \approx 1$, the actual estimated error equals on average to the error predicted by the model. It is worth noting that we did not include the nonparametric model given by Shapiro and Botha (1991) in our numerical study as the simulation results indicate that the performance of each of the empirical nonparametric estimators was superior

Table 1: $CV_{(K)}$ (and associated MSSPE values in parenthesis) based on the kriging at 10-Fold validations when each of the theoretical models given by (2.19), (2.20), (2.21), (2.22), (2.23), (2.24), and (2.17) is fitted (using WLS method) to the four empirical parametric and nonparametric variograms for simulated irregular and regular spatial data with lengths $n = 109$ and $n = 200$ according to two the simulation scheme with *circular*, *exponential*, *spherical*, *Gaussian*, *Matérn*, and *wave* models.

Fitted model	Empirical model, $n = 109$				Empirical model, $n = 200$			
	Matheron	Cressie-Hawkins	NW	VNN	Matheron	Cressie-Hawkins	NW	VNN
Circular								
Circular	1.11(0.487)	1.02 (0.545)	1.72(0.289)	1.63(0.201)	1.01(0.481)	1.09 (0.526)	1.51(0.309)	1.60(0.300)
Exponential	5.43(0.132)	6.02(0.121)	6.19(0.101)	5.98(0.128)	5.04(0.099)	5.00(0.164)	6.01(0.141)	5.76(0.111)
Spherical	3.56(0.286)	2.93(0.297)	2.06(0.267)	2.00(0.284)	3.08(0.302)	2.43(0.298)	2.29(0.289)	1.89(0.275)
Gaussian	6.65(0.281)	5.14(0.331)	3.93(0.291)	3.44(0.288)	4.44(0.303)	5.82(0.299)	3.33(0.278)	3.72(0.225)
Matérn	2.96(0.362)	2.93(0.390)	2.66(0.316)	2.64(0.329)	2.71(0.354)	2.73(0.390)	2.29(0.310)	2.19(0.311)
Wave	2.74(0.158)	2.51(0.200)	2.33(0.187)	2.20(0.192)	2.01(0.212)	2.22(0.229)	1.99(0.205)	1.83(0.213)
Exponential								
Circular	4.27(0.487)	4.10(0.419)	3.59(0.496)	3.82(0.484)	4.09(0.455)	4.11(0.491)	3.62(0.501)	3.59(0.513)
Exponential	1.99(0.612)	2.00(0.618)	0.98(0.797)	0.80 (0.812)	1.51(0.712)	1.40(0.727)	0.77(0.848)	0.58 (0.879)
Spherical	2.11(0.656)	2.00(0.683)	1.88(0.700)	1.87(0.726)	1.39(0.667)	1.49(0.680)	1.06(0.769)	1.00(0.801)
Gaussian	2.82(0.619)	2.08(0.679)	2.53(0.695)	2.02(0.718)	2.52(0.635)	2.32(0.664)	2.12(0.783)	2.29(0.790)
Matérn	2.48(0.660)	2.26(0.678)	1.55(0.693)	1.49(0.710)	2.27(0.681)	2.19(0.693)	2.03(0.712)	1.93(0.743)
Wave	4.04(0.518)	3.89(0.546)	2.89(0.568)	2.67(0.582)	3.88(0.569)	3.69(0.598)	2.94(0.637)	2.37(0.655)
Spherical								
Circular	2.77(0.505)	2.70(0.570)	2.62(0.629)	2.59(0.701)	2.72(0.591)	2.65(0.628)	2.19(0.729)	2.17(0.787)
Exponential	2.64(0.577)	2.58(0.583)	2.46(0.601)	2.38(0.627)	2.61(0.616)	2.59(0.698)	2.44(0.719)	2.11(0.728)
Spherical	1.08(0.761)	0.99(0.770)	0.81(0.876)	0.82 (0.901)	1.03(0.810)	1.00(0.824)	0.77(0.885)	0.74 (0.942)
Gaussian	2.64(0.600)	2.44(0.628)	2.39(0.678)	2.34(0.690)	2.60(0.660)	2.53(0.703)	2.39(0.712)	2.27(0.735)
Matérn	2.22(0.606)	2.19(0.610)	2.13(0.718)	2.11(0.727)	2.00(0.671)	2.04(0.693)	1.95(0.723)	1.97(0.777)
Wave	1.86(0.688)	1.83(0.702)	1.72(0.715)	1.76(0.721)	1.65(0.714)	1.71(0.735)	1.50(0.775)	1.49(0.810)
Gaussian								
Circular	3.67(0.755)	3.00(0.730)	2.59(0.798)	2.61(0.800)	3.09(1.126)	3.11(1.204)	2.61(1.128)	2.54(1.081)
Exponential	1.87(0.729)	1.78(0.766)	1.31(0.826)	1.22(0.886)	1.62(1.328)	1.52(1.275)	1.20(0.874)	1.19(0.893)
Spherical	1.22(0.459)	1.25(0.528)	0.88(0.709)	0.80(0.693)	1.19(0.442)	1.09(0.592)	0.76(0.683)	0.75(0.708)
Gaussian	1.39(1.231)	1.20(1.301)	0.78 (.901)	0.99(0.923)	1.00(0.875)	1.00(0.890)	0.66 (0.952)	0.87(0.900)
Matérn	1.78(1.213)	1.66(1.190)	1.00(0.960)	0.99(0.947)	1.35(1.200)	1.38(1.198)	1.03(1.102)	0.95(1.002)
Wave	3.00(0.699)	2.01(0.730)	1.99(0.780)	1.63(0.820)	2.87(0.773)	2.66(0.767)	1.71(0.824)	1.43(0.829)
Matérn								
Circular	2.55(0.802)	2.60(0.791)	2.51(0.854)	2.46(0.873)	2.50(1.220)	2.50(1.375)	2.12(1.109)	2.04(1.213)
Exponential	1.76(0.676)	1.44(0.703)	1.41(0.782)	1.38(0.788)	1.42(0.801)	1.31(0.825)	1.22(0.867)	0.95(0.881)
Spherical	3.57(0.470)	3.65(0.448)	3.32(0.497)	3.27(0.503)	3.45(0.539)	3.41(0.556)	3.20(0.605)	3.13(0.646)
Gaussian	2.44(0.336)	2.41(0.323)	2.35(0.395)	2.30(0.398)	2.30(0.441)	2.19(0.453)	1.99(0.493)	2.00(0.497)
Matérn	1.00(1.284)	1.01(1.190)	0.85(0.991)	0.69 (1.012)	0.97(1.300)	0.95(1.167)	0.86(1.117)	0.66 (1.082)
Wave	1.55(0.760)	1.46(0.832)	1.30(0.803)	1.30(0.900)	1.47(0.794)	1.43(0.802)	1.26(0.901)	1.20(0.910)
Wave								
Circular	1.60(0.770)	1.87(0.729)	1.53(0.815)	1.63(0.826)	1.92(0.764)	1.37(0.826)	1.28(0.844)	1.16(0.876)
Exponential	3.99(0.459)	4.08(0.525)	3.42(0.584)	3.33(0.594)	2.91(0.506)	3.01(0.476)	2.72(0.619)	2.50(0.637)
Spherical	3.76(0.508)	3.54(0.553)	3.17(0.598)	3.13(0.608)	3.52(0.558)	3.43(0.617)	3.29(0.622)	3.21(0.650)
Gaussian	3.03(0.533)	3.19(0.524)	2.86(0.621)	2.65(0.628)	2.89(0.575)	2.78(0.596)	2.39(0.626)	2.21(0.640)
Matérn	2.55(0.663)	2.41(0.688)	2.16(0.707)	2.15(0.702)	2.49(0.716)	2.46(0.722)	2.10(0.748)	1.97(0.752)
Wave	0.98(1.208)	1.01(1.196)	0.45(1.116)	0.42 (1.009)	0.86(1.194)	0.84(1.123)	0.38(1.010)	0.33 (1.002)

when compared to empirical parametric estimators. In the case of generating data using a circular variogram, the empirical parametric Cressie-Hawkins variogram is slightly better with smaller values of $CV_{(K)}$ than the Matheron (1.02 versus 1.11) in achieving the best valid model (true circular model). The robust empirical parametric Cressie-Hawkins gives the lowest values of mean prediction errors

for the case of spatial dependency with a circular model. Similarly, we noticed that the Nadaraya-Watson kernel model achieved a higher accuracy of prediction when the true theoretical Gaussian models fit a variogram from a generated Gaussian model compared with the others. For all other cases, the variable nearest-neighbor variogram presents the best performance to achieve the best valid model whose MSSPE values are closest to 1 with the lowest mean prediction errors.

4. Nitrate concentrations in groundwater: a case study in Gaza Strip

Gaza Strip is a narrow sliver of land with a total area of 365 square kilometers located in arid and semi-arid areas at longitudes $34^{\circ}20'$ East and latitudes $31^{\circ}25'$ North. There are 25 municipalities distributed in the five governorates of the Gaza Strip: Northern Gaza, Gaza, Deir el-Balah, Khan Yunis, and Rafah governorates. Approximately 49% of the land in the Gaza strip are agricultural areas, 28.8% are sand dunes, 3.3% are wooded, and 18.9% are urban areas. As of the year 2014, the Gaza Strip has a population density of 4,583 inhabitants per a square kilometer and the estimated natural increasing rate was 4.24 (Palestinian Central Bureau of Statistics, 2014). This indicates that the Gaza Strip is one of the world's most densely populated areas, which causes many problems in water resources and groundwater quality Shomer *et al.* (2004).

There are more than 4000 wells within the Gaza Strip. More than 97% of these are privately owned, where they have been exploited irrationally for agricultural purposes. Only 109 wells are operated by municipalities where they are mainly used for domestic supply of drinking water. The irrigation and drinking water quality in Gaza Strip is not optimal, where chemical analyses of irrigation wells indicate high salinity and sodium adsorption ratio, SAR, ratio (Al-Najar, 2011). In this context, in August 2012, the United Nations issued a warning report in which they warned that the water aquifer in the Gaza Strip may become unusable by the year 2016 with an irreversible damage by the year 2020 (United Nations, 2012). Nitrate concentration in groundwater generally increases with higher nitrogen input and higher aquifer vulnerability (Nolan *et al.*, 1998). Drinking water in excess of admissible nitrate limit can result in severe health problems, especially for infants less than two months of age (Spalding and Exner, 1993). Consequently, the European Union has established a drinking-water standard of 50 milligrams per liter (mg/L) nitrate (or approximately 3.91 mg/L for log-transformation scale) as nitrogen (European Union, 2006), whereas the United States Environmental Protection Agency, EPA, considered 10 (mg/L) nitrate as the maximum potable as nitrogen (or approximately 2.30 mg/L for log-transformation scale) United States Environmental Protection Agency (1995).

In this section, we analyze the spatial dependency in the NO_3 groundwater concentrations dataset obtained from the Ministry of Planning in the Gaza Strip, Palestine in 2012. We hope that the results of this research can help decision-makers, water-resource managers and private users to protect water supplies.

Figure 3 below shows the distribution of the groundwater wells in the study area in the Gaza Strip. The target variable is the nitrate, NO_3 concentration in drinking water wells, where we have 109 observations irregularly measured on spatial locations $77585.36 \leq X \leq 106475.1$ and $78491.00 \leq Y \leq 107217.9$.

Figure 4 shows the sample locations of the concentration levels of the NO_3 measured in milligrams per liter (mg/L). Table 2 gives a statistical summary for nitrate concentration levels in groundwater measured in the Gaza Strip. Based on the groundwater quality regulated by United States Environmental Protection Agency (1995), the results in Table 2 show that all wells in the Gaza Strip suffer from groundwater contamination by nitrates (more than 10 mg/L). In addition, more than 86% of the

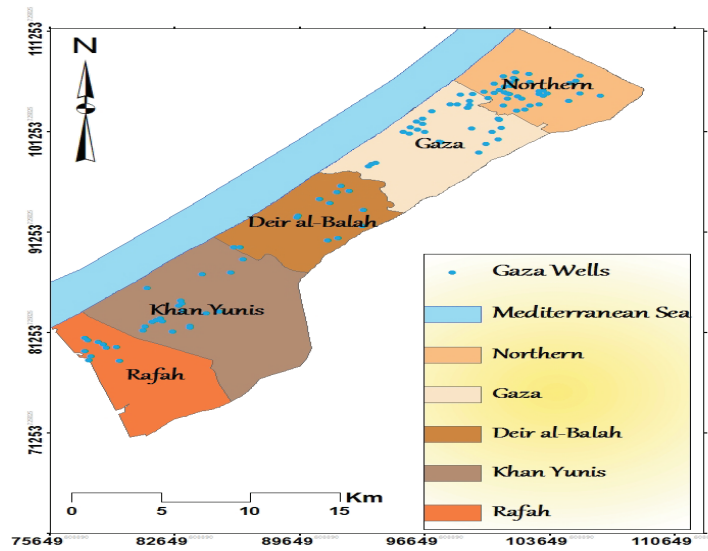


Figure 3: Distribution of groundwater wells in Gaza Strip (365 km²).

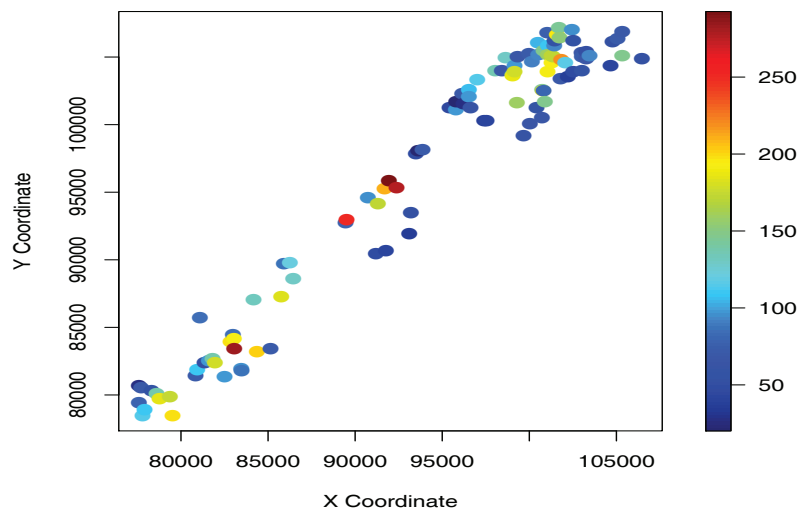


Figure 4: The sample locations (in Gaza Strip) for the NO₃-nitrate concentration data. The bullets indicate the nitrate concentration measurements in milligrams per liter (mg/L).

wells with NO₃ concentration exceed the maximum potable concentration of 50 mg/L (approximately 3.91 mg/L for log-transformation of NO₃), as classified by European Union (2006), which can potentially cause fatal conditions in infants (Spalding and Exner, 1993). The results suggest that the groundwater quality is severely deteriorated in the south region (Khan Yunis and Rafah governorates) of the Gaza Strip. The apparent variance-mean relationship (Table 2) suggests that the observed data departs from Gaussian distribution. In this respect, we use the log-transformation in order to alleviate

Table 2: Summary statistics for nitrate concentration (and log-transformed) levels in groundwater measured in Gaza Strip

Statistics		Northern	Gaza	Deir el-Balah	Khan Yunis	Rafah	Total
Number	wells	29	35	16	19	10	109
	> 50 mg NO ₃ /L	27	29	11	18	9	94
Mean	untransformed	105.2	96.6	119.3	131.3	115.3	110
	transformed	(4.5)	(4.4)	(4.5)	(4.8)	(4.6)	(4.5)
Median	untransformed	76.5	93.6	82.9	110.1	112.1	94.7
	transformed	(4.3)	(4.5)	(4.4)	(4.7)	(4.7)	(4.6)
Standard deviation	untransformed	54.17	47.96	91.8	60.23	57.90	60.80
	transformed	(0.50)	(0.56)	(0.80)	(0.45)	(0.62)	(0.58)
Minimum	untransformed	41.7	20.1	21.7	50.0	27.2	20.1
	transformed	(3.7)	(3.0)	(3.1)	(3.9)	(3.3)	(3.0)
Maximum	untransformed	219.8	196.8	292.6	284.1	198.6	292.6
	transformed	(5.4)	(5.3)	(5.7)	(5.6)	(5.3)	(5.7)

the departure from the non-symmetric distribution.

We use the Moran's I index to test the presence of global spatial autocorrelation in the values of the target variable based on the locations where they were measured (Moran, 1950). The Moran's I test statistic for a sample of size n is given by

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{k=1}^n (z_k - \bar{z})^2}, \quad (4.1)$$

where z_i is the i^{th} observation, \bar{z} is the sample mean of the target variable, and w_{ij} is the spatial weight of the link between i and j . The p -value based on Moran's I test is 1.24×10^{-07} , and it indicates strong spatial correlations among the NO₃ values (Moran, 1950). Therefore, it is reasonable to model the spatial correlation by estimating a valid variogram that can be used for kriging.

To model the valid variogram, we first use the rose diagram to visualize the empirical parametric and nonparametric variogram values in all directions of anisotropy. Figure 5 displays the rose diagrams (Isaaks and Srivastava, 1989; Jammalamadaka and Sengupta, 2001; Waller and Crawford, 2004) of the nonparametric Nadaraya-Watson variogram for the NO₃ data for the five governorates together with the corresponding rose diagram of all governorates. The rose maps clearly show consistent pattern around a regular disk with very minor directions to the south west region. We have found a similar pattern based on the parametric variograms (results are available upon request); consequently, we use the isotropic variogram in our analysis due to space constraints.

We then estimate the empirical parametric variogram given by (2.4) and (2.5). In addition, we also estimate the nonparametric NW and VNN empirical variograms using the optimal choices of the smoothing parameters in Equations (2.14), and (2.15) as suggested by Yu *et al.* (2007) along with an Epanechnikov kernel model, $\xi = 1$ and $\xi^* = 2\hat{\gamma}(\mathbf{h})^2$.

Then we fit the theoretical circular, exponential, spherical, Gaussian, Matérn, and wave models to each empirical model (Figures 6–9).

After that, we apply different ordinary kriging models based on estimated valid variograms to empirical variograms as well as use cross-validation. We use 10% of the data as a test data and consider the remainder 90% as a training data to select the best kriging model. We choose the final model that provides the lowest value of the $CV_{(K)}$. The results in Table 3 suggest that the wave (hole-effect) variogram model (with nugget 498.01, partial sill 1875.23, and range 1192.15) fitted to the empirical variable nearest-neighbor variogram is the best for modeling the spatial structures of

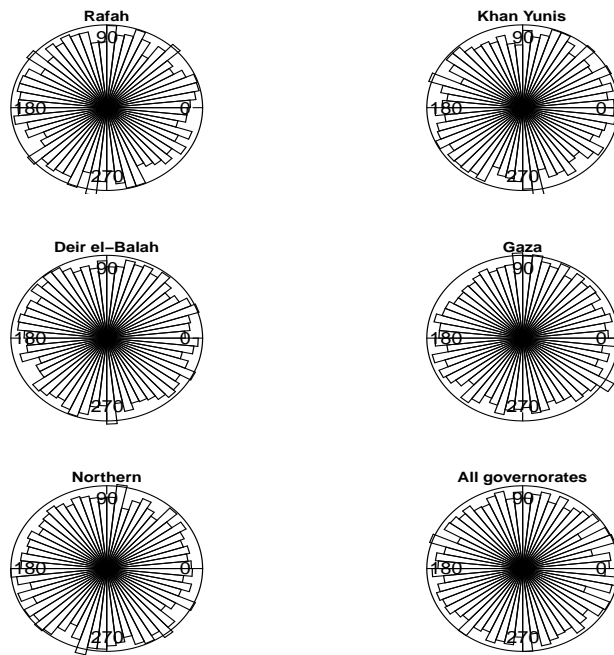


Figure 5: Rose diagrams of the nonparametric Nadaraya-Watson variogram the NO_3 data.

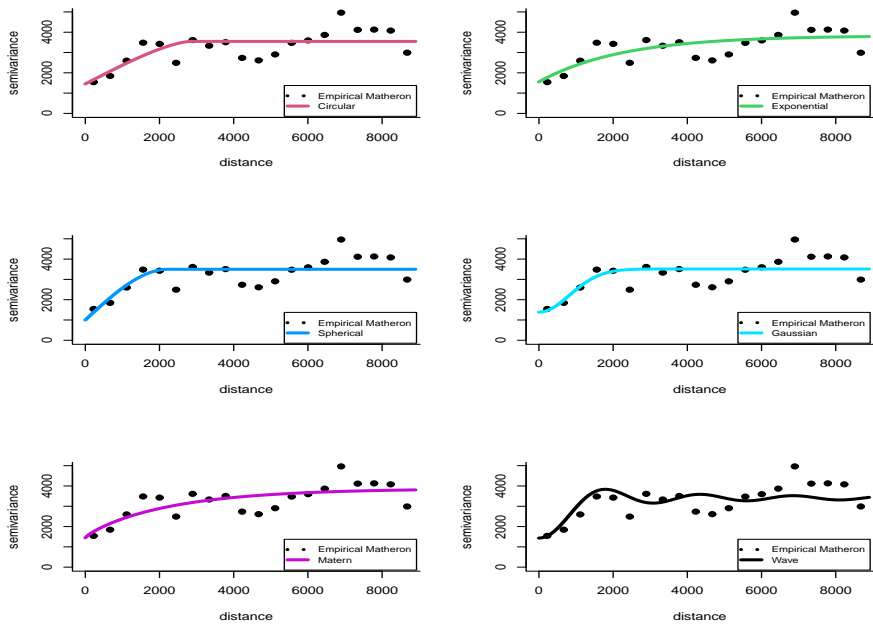


Figure 6: The circular, exponential, spherical, Gaussian, Matérn, and wave parametric variograms are fitted to the empirical Matheron's variogram for the NO_3 concentrations.

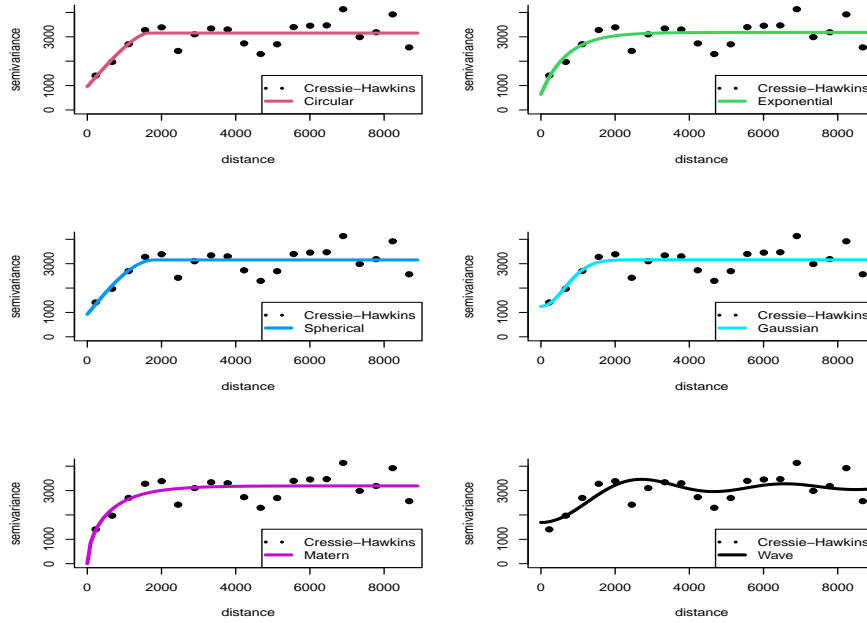


Figure 7: The circular, exponential, spherical, Gaussian, Matérn, and wave parametric variograms are fitted to the empirical Cressie-Hawkins' modulus variogram for the NO_3 concentrations.

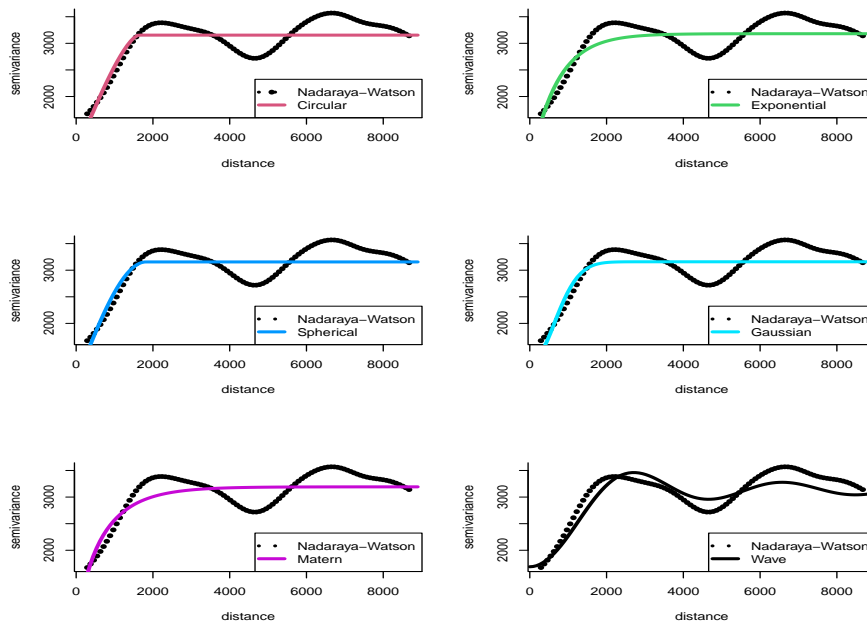


Figure 8: The circular, exponential, spherical, Gaussian, Matérn, and wave parametric variograms are fitted to the Nadaraya-Watson's nonparametric variogram for the NO_3 concentrations.

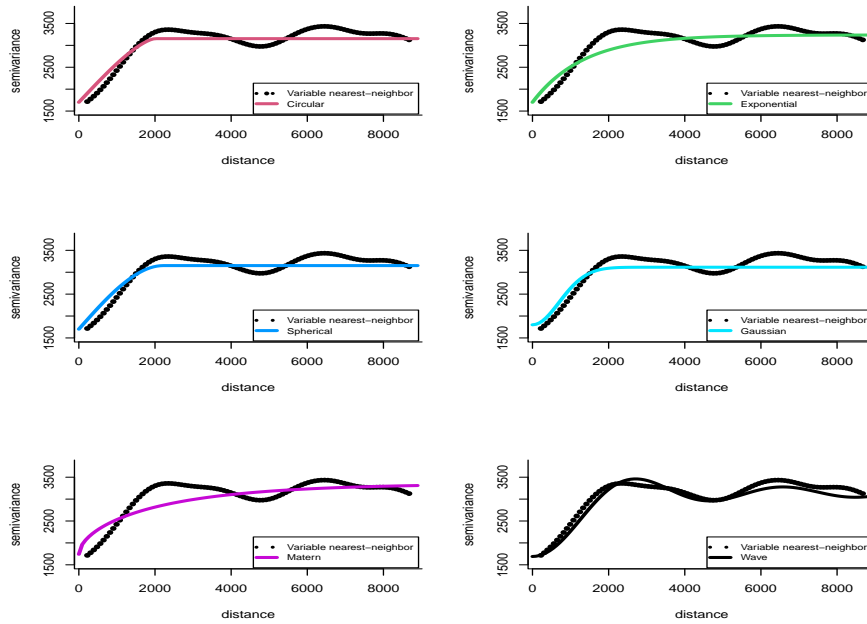


Figure 9: The circular, exponential, spherical, Gaussian, Matérn, and wave parametric variograms are fitted to the variable nearest-neighbor nonparametric variogram for the NO_3 concentrations.

Table 3: Cross validation results of fitting variogram models for nitrate concentrations

Models	Prediction errors							
	Root mean squared				Root mean squared standardized			
	Matheron	Cressie-Hawkins	NW	VNN	Matheron	Cressie-Hawkins	NW	VNN
Circular	0.471	0.483	0.485	0.491	0.998	0.993	0.976	0.970
Exponential	0.471	0.475	0.486	0.490	0.995	0.980	0.977	0.969
Spherical	0.476	0.479	0.487	0.485	1.003	0.984	0.978	0.970
Gaussian	0.498	0.504	0.509	0.517	1.036	1.022	1.020	1.019
Matérn ($\kappa = 0.7$)	0.476	0.480	0.492	0.502	1.001	0.987	0.987	0.986
Wave	0.413	0.424	0.429	0.409	1.001	1.007	1.006	1.003

the log-transformation of the NO_3 data. The resulted interpolation map is shown in the left panel of Figure 11. The map shows eight levels of contamination risk of groundwater from very low nitrogen input (green areas on the map), to low nitrogen input (yellow areas), to median nitrogen input (brown areas), to high nitrogen input (gray areas on the map). The right panel of Figure 11 shows the kriging prediction error variances map produced by kriging interpolation, which clearly demonstrates that locations near existing groundwater wells have lower kriging standard error values, whereas higher kriging standard error values can be found in areas having less or no wells. The p -value of the Moran's I test is 0.24 indicating that we have no evidence of spatial correlation in the kriging residuals.

5. Concluding remarks

The selection of the appropriate valid variogram model (especially for small samples) is crucial to obtain accurate prediction results from kriging interpolations; otherwise, negative mean squared pre-

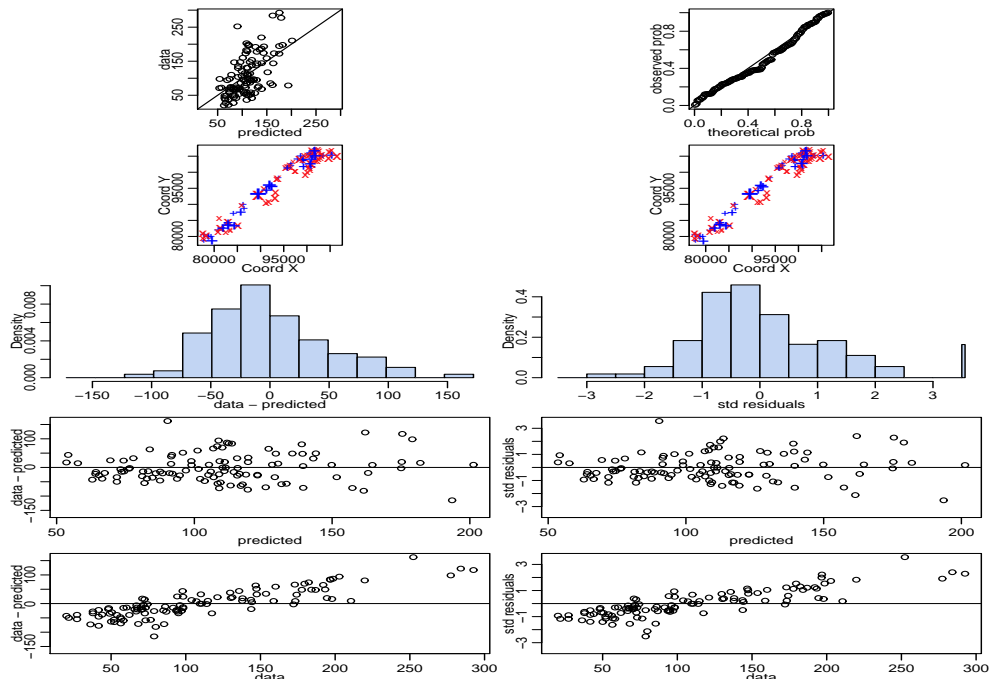


Figure 10: Cross-validation of kriging results with wave model.

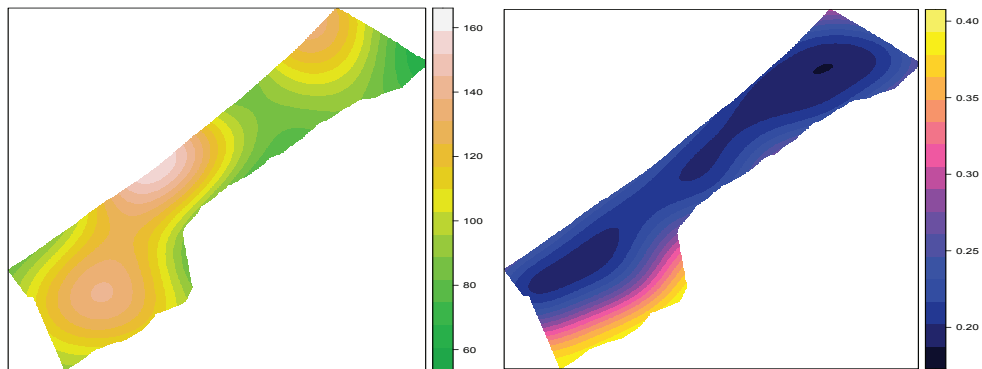


Figure 11: The panel on the left hand side shows the predicted pollution maps of the nitrate concentrations based on the ordinary kriging interpolation with a wave variogram. The panel on the right hand side shows the kriging prediction error map.

diction errors may be obtained. To estimate such a variogram, we often first estimate the empirical variogram: the traditional variogram (Matheron or robust Cressie-Hawkins); or the kernel-based variogram (Nadaraya-Watson or variable nearest-neighbor). In many situations, the empirical variogram estimator is not guaranteed to be valid. In practice, we estimate the empirical variogram by a theoretical model selected from the family of valid variograms. The results obtained in this study suggest following the procedure algorithm to achieve the best valid variogram. We apply this approach on

the nitrate concentrations in the groundwater dataset and then find that the wave (hole-effect) variogram model fitted to the empirical variable nearest-neighbor, VNN, variogram provides the lowest values of the $CV_{(K)}$. The interpolation map indicates that the quality of the groundwater in the Gaza Strip has a serious issue due to high levels of nitrate concentrations. Decision-makers, water-resource managers, and private users in the Gaza Strip are strongly encouraged to implement an immediate strategic solution to protect the groundwater wells of the Gaza Strip, especially in the south region.

References

- Al-Najar H (2011). The integration of FAO-CropWat model and GIS techniques for estimating irrigation water requirement and its application in the Gaza strip, *Natural Resources*, **2**, 146–154.
- Armstrong M and Jabin R (1981). Variogram models must be positive-definite, *Journal of the International Association of Mathematical Geology*, **13**, 5–459.
- Barry RP and Pace RK (1997). Kriging with large data sets using sparse matrix techniques, *Communications in Statistics - Simulation and Computation*, **26**, 619–629.
- Berman M and Diggle P (1989). Estimating weighted integrals of the second-Order Intensity of a spatial point process, *Journal of the Royal Statistical Society: Series B*, **51**, 81–92.
- Chilés JP and Delfiner P (1999). *Geostatistics: Modelling Spatial Uncertainty*, John Wiley & Sons, Wiley-Interscience.
- Cressie N (1985). Fitting variogram models by weighted least squares, *Mathematical Geology*, **17**, 563–586.
- Cressie N (1993). *Statistics for Spatial Data*, John Wiley & Sons, New York.
- Cressie N and Hawkins DM (1980). Robust estimation of the variogram, *Journal of the International Association of Mathematical Geology*, **12**, 115–125.
- Croux C and Rousseeuw PJ (1992). Time-efficient algorithms for two highly robust estimators of scale, *Computational Statistics*, **1**, 411–428.
- Diggle PJ (1985). A kernel method for smoothing point process data, *Applied Statistics*, **34**, 138–147.
- Diggle PJ, Gates DJ, and Stibbard A (1987). A nonparametric estimator for pairwise-interaction point processes, *Biometrika*, **74**, 763–770.
- Dunn MR (1983). A simple sufficient condition for a variogram model to yield positive variances under restrictions, *Mathematical Geology*, **15**, 553–564.
- European Union (December 2006). Directive 2006/118/EC of the European parliament and of the council on the protection of groundwater against pollution and deterioration, *Official Journal of the European Union*, **L 372**, 19–31.
- Fernández-Casal R (2016). *npsp: Nonparametric Spatial (Geo)statistics*, R package version 0.5-3.
- García-Soidán PH, Febrero-Bande M, and González-Manteiga W (2004). Nonparametric kernel estimation of an isotropic variogram, *Journal of Statistical Planning and Inference*, **121**, 65–92.
- García-Soidán PH, González-Manteiga W, and Febrero-Bande M (2003). Local linear regression estimation of the variogram, *Statistics and Probability Letters*, **64**, 169–179.
- García-Soidán PH and Menezes R (2012). Estimation of the spatial distribution through the kernel indicator variogram, *Environmetrics*, **23**, 535–548.
- Genton MG (1998). Highly robust variogram estimation, *Mathematical Geology*, **30**, 213–221.
- Gorsich DJ and Genton MG (2000). Variogram model selection via nonparametric derivative estimation, *International Association for Mathematical Geology*, **32**, 249–270.
- Gräler B, Pebesma E, and Heuvelink G (2016). Spatio-Temporal Interpolation using gstat, *The R Journal*, **8**, 204–218.

- Hall P, Fisher NI, and Hoffmann B (1994). On the nonparametric estimation of covariance functions, *The Annals of Statistics*, **22**, 2115–2134.
- Hastie T, Tibshirani R, and Friedman J (2009). *The Elements of Statistical Learning*, Springer, New York.
- Huang C, Hsing T, and Cressie N (2011). Nonparametric estimation of the variogram and its spectrum, *Biometrika*, **98**, 775–789.
- Isaaks EH and Srivastava RM (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York.
- Jammalamadaka SR and Sengupta A (2001). *Topics in Circular Statistics*, World Scientific, Singapore.
- Jin R and Kelly GE (2017). A comparison of sampling grids, cut-off distance and type of residuals in parametric variogram estimation, *Communications in Statistics - Simulation and Computation*, **46**, 1781–1795.
- Lahiri SN, Kaiser MS, Cressie N, and Hsu NJ (1999). Prediction of spatial cumulative distribution functions using subsampling, *Journal of the American Statistical Association*, **94**, 86–97.
- Matérn B (1960; reprinted 1986). *Spatial Variation* (2nd ed), Springer-Verlag, Berlin.
- Matheron G (1962). *Traité de géostatistique appliquée, tome i, mémoires du Bureau de recherches géologiques et minières*, Editions Technip, **14**, Paris.
- Matheron G (1963a). *Traité de géostatistique appliquée, tome ii, le krigeage. mémoires du bureau de recherches géologiques et minières*, Editions Bureau de Recherches Géologiques et Minières, **24**, Paris.
- Matheron G (1963b). Principles of geostatistics, *Economic Geology*, **58**, 1246–1266.
- Menezes R, García-Soidán PH, and Febrero-Bande M (2005). A comparison of approaches for valid variogram achievement, *Computational Statistics*, **20**, 623–642.
- Menezes R, García-Soidán PH, and Febrero-Bande M (2008). A Kernel variogram estimator for clustered data, *Scandinavian Journal of Statistics*, **35**, 18–37.
- Moran PAP (1950). Notes on continuous stochastic phenomena, *Biometrika*, **37**, 17–23.
- Nolan BT, Ruddy BC, Hitt KJ, and Helsel DR (1998). A national look at nitrate contamination of ground water, *Water Conditioning and Purification*, **39**, 76–79.
- Palestinian Central Bureau of Statistics (2014). Palestinian Central Bureau of Statistics reports. Available from: <http://www.pcbs.gov.ps/default.aspx>
- Pebesma EJ and Bivand RS (2005). Classes and methods for spatial data in R, *R News*, **5**, 9–13.
- Ribeiro Jr PJ and Diggle PJ (2001). geoR: a package for geostatistical analysis, *R-NEWS*, **1**, 15–18.
- Rousseeuw PJ and Croux C (1993). Alternatives to the median absolute deviation, *Journal of the American Statistical Association*, **88**, 1273–1283.
- Shapiro A and Botha JD (1991). Variogram fitting with a general class of conditionally nonnegative definite functions, *Computational Statistics and Data Analysis*, **11**, 11–96.
- Shomer B, Müller G, and Yahya A (2004). Potential use of treated wastewater and sludge in agricultural sector of the Gaza strip, *Clean Technologies and Environmental Policy*, **6**, 128–137.
- Spalding RF and Exner ME (1993). Occurrence of nitrate in groundwater – a review, *Journal of Environmental Quality*, **22**, 392–402.
- Stein ML (1999). *Interpolation of Spatial Data - Some Theory for Kriging*, Springer Verlag, New York.
- United Nations (August 2012). Gaza in 2020: a liveable place?, *United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA)*, 1–24.
- United States Environmental Protection Agency (1995). *Drinking water regulations and health advi-*

- sories*, Office of Water, Washington.
- Waller LA and Crawford CAG (2004). *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, Hoboken, NJ.
- Yaglom A (1987). *Correlation Theory of Stationary and Related Random Functions*, Springer Verlag, New York.
- Yu K, Mateu J, and Porcu E (2007). A kernel-based method for nonparametric estimation of variograms, *Statistica Neerlandica*, **61**, 173–197.

Received May 16, 2020; Revised July 12, 2020; Accepted August 17, 2020