WILEY | Hindawi

Research Article

More Efficient Prediction for Ordinary Kriging to Solve a Problem in the Structure of Some Random Fields

Mohammad Mehdi Saber [ID] [1] **and Ramy Abdelhamid Aldallal**[2]

[1]*Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran*
[2]*College of Business Administration in Hotat Bani Tamim, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia*

Correspondence should be addressed to Mohammad Mehdi Saber; mmsaber@eghlid.ac.ir

Recently, some specific random fields have been defined based on multivariate distributions. This paper will show that almost all these random fields have a deficiency in spatial autocorrelation structure. The paper recommends a method for coping with this problem. Another application of these random fields is spatial data prediction, and the Kriging estimator is the most widely used method that does not require defining the mentioned random fields. Although it is an unbiased estimator with a minimum mean-squared error, it does not necessarily have a minimum mean-squared error in the class of all linear estimators. In this work, a biased estimator is introduced with less mean-squared error than the Kriging estimator under some conditions. Asymptotic behavior of its basic component will be investigated too.

1. Introduction

Random fields (RFs) as a statistical model can be applied in many real-life events, such as biological sequences, management of soil resources in agriculture and forestry, text and image processing, designing environmental monitoring networks, artificial intelligence, and road and tunnel planning.

An RF is usually constructed based on a multivariate distribution's finite-dimensional distributions. For example, the skew-normal (SN) RF based on multivariate SN distribution [1] has been defined by Kim and Mallick [2]. The closed skew-normal (CSN) RF based on multivariate CSN distribution [GF_2004] has been established by Allard and Naveau [3]. Similar work has been performed by Karimi et al [4], Hosseini et al. [5], and Karimi and Mohammadzadeh [6]. Zareifard and Jafari Khaledi [7] defined a second-order stationary RF named unified skew-normal (SUN) RF by using multivariate SUN distribution [8]. The generalized Skew-Normal (GSN) RF introduced by Mahmoudian [9] has also been constructed from multivariate GSN distribution as shown by Sahu et al. [10], while the generalized asymmetric Laplace (GAL) RF is defined by multivariate GAL distribution as shown by Kozubowski et al. [11] and was introduced by Saber et al. [12]. Some other RFs may be defined by multivariate skew-t distribution and its general form or multivariate extended skew-t (EST) distribution as shown by Arellano-Valle and Genton [13].

For the spatial structure of the model, Wall [14] studied the spatial structure implied by the CAR and SAR models. According to Saber et al. [12] and Mahmoudian [15], almost all mentioned RFs in the previous paragraph are not well defined by the Kolmogorov existence theorem. Saber et al. [12] showed that these RFs do not consider proper spatial autocorrelation structure (PSAS). In the present article, we will show that almost all aforementioned skew RFs have this deficiency. In fact, they do not consider PSAS.

A final application of any random field is spatial data prediction, which has been applied in some works of authors such as Basu and Reinsel [16], Saber and Nematollahi [17], and Saber [18]. Many researchers have studied the prediction of spatial data using mentioned unsuitable RFs. To solve the previously discussed problem, some authors such as

Mojiri et al. [19] have considered a univariate distribution for detection of errors of the process, which results in a new problem in modeling spatial correlation. Some others such as Hosseini and Karimi [21] have used an approximate skew-normal RF. Although this work has some benefits for reaching a high percentage of PSAS, it has some fundamental problems, and it cannot solve the two mentioned problems.

Kriging is a spatial prediction methodology and is the most widely used method of spatial data prediction that has increased during recent years. Because this method does not require defining the previously mentioned random fields that have problem in PSAS, Cressie [22] presented an unbiased linear estimator for Kriging and has a minimum mean-squared error (MSE), although it is not the best estimator in the larger class of all linear estimators. On the other hand, according to Moyeed and Papritz [23] linear Kriging is as good as any nonlinear method, especially for symmetric data. However, for skewed data, some of the nonlinear methods perform better in estimating prediction uncertainty. The focus of this article is on linear Kriging as we suggested a biased estimator for Kriging where the only interest is to minimize the MSE. Our recommended estimator has less MSE than the Cressie estimator under some conditions. Therefore, it is appropriate to use it for skewed data when we have a problem in defining a valid RF and for ordinary Kriging that does not have precision for nonlinear methods.

The paper is organized as follows: In Section 2, some requirements of the previously mentioned multivariate distributions are reviewed to show defined RFs that do not consider the PSAS property. After that, an estimator for the mean of the process and then a predictor with a minimum MSE in all the linear predictors for ordinary Kriging are presented in Section 3. Finally, Section 4 is devoted to investigating some asymptotic behavior for the recommended estimator.

2. A Problem in Some RFs

The first part of this section reviews the variance matrix of multivariate SN, CSN, EST, GSN, SUN, and GAL distributions.

Consider a p-dimensional continuous random vector $\mathbf{X} = (X_1, \ldots, X_p)'$ that has an SN distribution, denoted by $\mathbf{X} \sim SN_p(\mu, \Sigma, \lambda)$. Then, we have

$$\text{Var}(\mathbf{X}) = \Sigma - \frac{2}{\pi(1 + \boldsymbol{\lambda}^T \boldsymbol{\lambda})} \Sigma^{1/2} \boldsymbol{\lambda}\boldsymbol{\lambda}^T \Sigma^{1/2}, \quad (1)$$

and consider another p-dimensional continuous random vector $\mathbf{X} = (X_1, \ldots, X_p)'$ but with a CSN distribution, denoted by $\mathbf{X} \sim CSN_{p,q}(\mu, \Sigma, \Gamma, \xi, \Delta)$. Then, we have

$$\text{Var}(\mathbf{X}) = \Sigma + \Sigma\Gamma^T \boldsymbol{\Lambda} \Gamma\Sigma - \Sigma\Gamma^T \boldsymbol{\Psi}\boldsymbol{\Psi}^T \Gamma\Sigma, \quad (2)$$

where Λ and Ψ are complicated matrices based on $\Phi_q(\mathbf{0}; \xi, \Delta + \Gamma\Sigma\Gamma')$. The detail about these matrices is not essential for us, although they are present in the study conducted by González-Farías et al. [24].

The variance matrix of a multivariate EST distribution is denoted by $\mathbf{X} \sim EST_p(\mu, \Sigma, \lambda, \nu, \tau)$, which is given as follows:

$$\text{Var}(\mathbf{X}) = \alpha_*^2 \left(\boldsymbol{\omega}^{-1} \Sigma \boldsymbol{\omega}^{-1} - \boldsymbol{\delta}\boldsymbol{\delta}^T \right) + \sigma_*^2 \boldsymbol{\delta}\boldsymbol{\delta}^T. \quad (3)$$

However, the variance matrix of a multivariate GSN distribution is denoted by $\mathbf{X} \sim GSN_p(\mu, \Sigma, \lambda)$, which is given as follows:

$$\text{Var}(\mathbf{X}) = \Sigma + \left(1 - \frac{2}{\pi} \right) \text{diag}(\boldsymbol{\lambda}^2). \quad (4)$$

Arellano-Valle and Azzalini [25] recently provided a very complicated and cumbersome computation of the variance matrix of multivariate SUN. Let $\mathbf{X} \sim SUN_{d,m}(\xi, \Sigma, \Delta, \tau, \Gamma)$. Then we have,

$$\text{Var}(\mathbf{X}) = \Sigma + \boldsymbol{\omega}\Delta\mathbf{H}\Delta^T \boldsymbol{\omega}. \quad (5)$$

Finally, the variance matrix of the GAL distribution continuous p-dimensional random vector $\mathbf{X}$ denoted by $\mathbf{X} \sim GAL_p(\mu, \Sigma, q)$ is given as follows:

$$\text{Var}(\mathbf{X}) = q\left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T \right). \quad (6)$$

After that, we state the following remark.

Remark 1. Let $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ be the observations from an RF $\{Z(\mathbf{s}): \mathbf{s} \in \mathbf{D} \subseteq R^d\}$ at n locations $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ in the area $\mathbf{D}$. It is assumed that $\text{Cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = q(|\mathbf{s}_i - \mathbf{s}_j|)$ where $|\mathbf{h}|$ is the norm of the vector $\mathbf{h}$ and q is a real value function. In a matrix form, $\text{Var}(\mathbf{Z}) = \mathbf{C}$ where $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))$ and $\mathbf{C} = [C_{i,j}]_{i,j=1}^n$ with a notation $C_{i,j} = q(|\mathbf{s}_i - \mathbf{s}_j|)$. Some RFs have been defined so that they satisfy the condition $\text{Var}(\mathbf{Z}) = \mathbf{C}$, while in some other cases, the covariance matrix $\mathbf{C}$ has been used somehow in which $\text{Var}(\mathbf{Z}) \neq \mathbf{C}$.

In continuation of this section, we investigate the condition where $\text{Var}(\mathbf{Z}) = \mathbf{C}$ is named as PSAS in some well-known RFs.

Gaussian RF: $\mathbf{Z} \sim N_n(\mu, \Sigma)$ and $\text{Var}(\mathbf{Z}) = \Sigma$. Therefore, we choose $\Sigma = \mathbf{C}$ to have a spatial correlation between variables.

t RF: $\mathbf{Z} \sim t_n(\mu, \Sigma, d)$ $(d \geq 3)$ and $\text{Var}(\mathbf{Z}) = d/d - 2\Sigma$. It seems to be an appropriate choice to reach the PSAS when $\Sigma = d - 2/d\mathbf{C}$.

SN RF: $\mathbf{Z} \sim SN_n(\mu, \Sigma, \lambda)$ and $ar(\mathbf{Z}) = \Sigma - 2/\pi(1 + \lambda^T\lambda)\Sigma^{1/2}\lambda\lambda^T\Sigma^{1/2}$ where it is convenient to substitute $\Sigma - 2/\pi(1 + \lambda^T\lambda)\Sigma^{1/2}\lambda\lambda^T\Sigma^{1/2} = C$ to get to the PSAS.

CSN RF: $\mathbf{Z} \sim CSN_{n,q}(\mu, \Sigma, \Gamma, \xi, \Delta)$ and $ar(\mathbf{Z}) = \Sigma + \Sigma\Gamma^T\Lambda\Gamma\Sigma - \Sigma\Gamma^T\Psi\Psi^T\Gamma\Sigma$. Choosing $\Sigma + \Sigma\Gamma^T\Lambda\Gamma\Sigma - \Sigma\Gamma^T\Psi\Psi^T\Gamma\Sigma = \mathbf{C}$ provides an adequate choice to reach the PSAS.

EST RF: $\mathbf{Z} \sim EST_n(\mu, \Sigma, \lambda, \nu, \tau)$ and $ar(\mathbf{Z}) = \alpha_*^2(\omega^{-1}\Sigma\omega^{-1} - \delta\delta^T) + \sigma_*^2\delta\delta^T$. By substituting $\alpha_*^2(\omega^{-1}\Sigma\omega^{-1} - \delta\delta^T) + \sigma_*^2\delta\delta^T = \mathbf{C}$, we will obtain the PSAS.

SUN RF: $\mathbf{Z} \sim SUN_{n,m}(\xi, \Sigma, \Delta, \tau, \Gamma)$ and $ar(\mathbf{Z}) = \Sigma + \omega\Delta\mathbf{H}\Delta^T\omega$. Making $\Sigma + \omega\Delta\mathbf{H}\Delta^T\omega = \mathbf{C}$ will fulfill the PSAS.

GSN RF: $\mathbf{Z} \sim GSN_n(\mu, \Sigma, \lambda 1)$ and $ar(\mathbf{Z}) = \Sigma + \lambda(1 - 2/\pi)I$. It provides a suitable choice to choose $\Sigma + \lambda(1 - 2/\pi)I = C$ and then consider λ as a nugget and $\Sigma + \lambda(1 - 2/\pi)I = C$ to accomplish the PSAS.

GAL RF: $\mathbf{Z} \sim \mathrm{GAL}_n(\mu, \Sigma, q)$ and $ar(\mathbf{Z}) = q(\Sigma + \mu\mu^T)$. After finding the PSAS, we will choose $q(\Sigma + \mu^2\, \mathbf{1}\, \mathbf{1}^T) = \mathbf{C}$, which results as $\mu = \mathbf{X}\beta$ and $\Sigma = 1/q\mathbf{C} - \mathbf{X}\beta\beta^T\mathbf{X}^T$ where $\mathbf{X}$ is the matrix of covariates and β is the regression coefficient.

As previously shown, some of the RFs satisfy the PSAS, while others do not. In the following remark, we denote this matter concisely.

Remark 2. The Gaussian RF, *t* RF, GSN RF, and GAL RF satisfy the PSAS. Other RFs containing the SN RF, CSN RF, EST RF, and SUN RF do not satisfy the PSAS. In some of the previous works that applied a second group of RFs, it was supposed that $\Sigma = \mathbf{C}$, although it does not lead to $\mathrm{Var}(\mathbf{Z}) = \mathbf{C}$ for achieving the PSAS.

Imposing the equality $\mathrm{Var}(\mathbf{Z}) = \mathbf{C}$, without any other constraints, may result in a nonpositive definite for the matrix Σ. There is a method from which we can add a small value to the diagonal of Σ, but it does not work in every case. So, there is still a problem in getting the PSAS and positive definiteness for the matrix. Σ

3. Linear Kriging as an Alternative for the Nonlinear Predictors Conducted by RFs

Let RF $\{Z(\mathbf{s}): \mathbf{s} \in \mathbf{D} \subseteq R^d\}$ satisfies the following model:

$$Z(\mathbf{s}) = \mu + Y(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{7}$$

where μ is the mean value, $Y(\cdot)$ is the zero-mean intermediate-scale variation, and a white noise error denoted by $\varepsilon(\cdot)$ with a variance σ_ε^2.

Kriging belongs to the family of linear least square estimation algorithms. Given the observations $Z = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))$, a common task is to predict $Z(\mathbf{s}_0)$ at an unobserved location $\mathbf{s}_0$ and to calculate the prediction error variance at each such location. The Kriging estimator is a linear predictor for $\widehat{Z}(\mathbf{s}_0)$, which is as follows:

$$\widehat{Z}(\mathbf{s}_o) = \sum_{i=1}^{n} l_i Z(\mathbf{s}_i), \tag{8}$$

where $l_i\, s$ (the Kriging weights) are some scalars that may be found under two conditions:

(a) $E(\widehat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)) = 0$ (unbiased condition)

(b) $\mathrm{Argmin}\{\mathrm{MSE}(\widehat{Z}(\mathbf{s}_0)) = \mathrm{E}(\widehat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0))^2\}$

Simple Kriging (known mean) and ordinary Kriging (unknown but constant mean) are two well-known divisions of Kriging based on the mean value specification. In fact, in simple Kriging μ is assumed to be known, but it is unknown in the case of ordinary Kriging, which occurs more in the application. As in the previous section, let $\mathbf{C} = (\mathrm{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)))_{i,j}$ denotes the covariance matrix of the observations and let $\gamma = (\mathrm{cov}(Z(\mathbf{s}_0), Z(\mathbf{s}_1)), \ldots, \mathrm{cov}(Z(\mathbf{s}_0), Z(\mathbf{s}_n)))$ denotes the vector containing the covariances between observed and unobserved location $\mathbf{s}_0$.

According to Cressie [22], the simple Kriging estimator and its error variance are given, respectively, by

$$\widehat{Z}_{\mathrm{simple}}(\mathbf{s}_0) = \gamma' \mathbf{C}^{-1}\mathbf{Z}, \tag{9}$$

and

$$\mathrm{Var}\big(\widehat{Z}_{\mathrm{simple}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\big) = \sigma_z^2(\mathbf{s}_0) - \gamma' \mathbf{C}^{-1}\gamma. \tag{10}$$

Also, the ordinary Kriging and its error variance are given, respectively, by

$$\widehat{Z}(\mathbf{s}_0) = \left(\gamma + 1 \frac{\left(1 - \mathbf{1}' \mathbf{C}^{-1}\gamma\right)}{\mathbf{1}' \mathbf{C}^{-1}\mathbf{1}}\right)' \mathbf{C}^{-1}\mathbf{Z}, \tag{11}$$

and

$$\sigma_r^2(\mathbf{s}_0) = V\big(\widehat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\big)$$
$$= \sigma_z^2(\mathbf{s}_0) - \gamma' \mathbf{C}^{-1}\gamma + \frac{\left(1 - \mathbf{1}' \mathbf{C}^{-1}\gamma\right)^2}{\mathbf{1}' \mathbf{C}^{-1}\mathbf{1}}, \tag{12}$$

where $\mathbf{1}' = (1, \ldots, 1)$ is the row vector of size n in which its components are $1\,s$.

3.1. New Estimator. The Kriging estimator for ordinary Kriging is the best estimator regarding unbiasedness and the minimum MSE. However, a biased estimator with a less MSE is preferable in some circumstances. An estimator with this condition exists, and we denote our motivation for recommending this estimator in the following theorem.

Theorem 1. *An unbiased linear estimator for the mean of the process given in (7) exists and has a minimum MSE among all unbiased linear estimators for the mean. This estimator is in the form of*

$$\widehat{\mu} = \frac{\mathbf{1}' \mathbf{C}^{-1}\mathbf{Z}}{\mathbf{1}' \mathbf{C}^{-1}\mathbf{1}}, \tag{13}$$

and its variance is given by

$$\mathrm{Var}(\widehat{\mu}) = \frac{1}{\mathbf{1}' \mathbf{C}^{-1}\mathbf{1}}. \tag{14}$$

Proof. A linear estimator for μ is in the form of $\sum_{i=1}^{n} a_i Z(\mathbf{s}_i) = \mathbf{a}' \mathbf{Z}$. Unbiasedness implies that $\mathbf{a}' \mathbf{1} = \sum_{i=1}^{n} a_i = 1$. Hence, under this condition, $\mathrm{Var}(\mathbf{a}' \mathbf{Z}) = \mathbf{a}' \mathbf{C}\mathbf{a}$ must be minimized. By using the Lagrange equation $\Gamma = \mathbf{a}' \mathbf{C}\mathbf{a} + \lambda(\mathbf{a}' \mathbf{1} - 1)$ and by vector differentiating, one can solve this problem to reach $\widehat{a} = \mathbf{C}^{-1}\mathbf{1}/\mathbf{1}' \mathbf{C}^{-1}\mathbf{1}$, and its results are shown in (13). Now, we have $\mathrm{Var}(\mathbf{1}' \mathbf{C}^{-1}/ \mathbf{1}' \mathbf{C}^{-1}\mathbf{1}\mathbf{Z}) = \mathbf{1}' \mathbf{C}^{-1}\mathbf{C}\mathbf{C}^{-1}\mathbf{1}/ (\mathbf{1}' \mathbf{C}^{-1}\mathbf{1})^2 = \mathbf{1}' \mathbf{C}^{-1}\mathbf{1}/(\mathbf{1}' \mathbf{C}^{-1}\mathbf{1})^2 = 1/\mathbf{1}' \mathbf{C}^{-1}\mathbf{1}$, which completes the proof.

Now, a simple Kriging method can be applied instead of an ordinary Kriging method. By substituting (13) in (7), we can reach $Z(\mathbf{s}) - \widehat{\mu} = Y(\mathbf{s}) + \varepsilon(\mathbf{s})$ and $Z(\mathbf{s}) - \mathbf{1}' \mathbf{C}^{-1}\mathbf{Z}/\mathbf{1}' \mathbf{C}^{-1}\mathbf{1} = Y(\mathbf{s}) + \varepsilon(\mathbf{s})$. From (9), we can conclude that

$$\widehat{Z}_{\text{new}}(\mathbf{s}_0) = \gamma' \mathbf{C}^{-1}\left(\mathbf{Z} - \frac{1'\mathbf{C}^{-1}\mathbf{Z}}{1'\mathbf{C}^{-1}1}\right)$$

$$= \left(\gamma - \frac{1}{1'\mathbf{C}^{-1}1}\right)'\mathbf{C}^{-1}\mathbf{Z}, \tag{15}$$

which is the new estimator for interpolation. $\widehat{Z}_{\text{new}}(\mathbf{s}_0)$ is a special case of $\widehat{Z}(\mathbf{s}_0)$ when $\gamma' \mathbf{C}^{-1}1 = 2$. □

$$E\left(\widehat{Z}_{\text{new}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) = \left(\gamma - \frac{1}{1'\mathbf{C}^{-1}1}\right)'\mathbf{C}^{-1}\mu 1 - \mu$$

$$= \mu\left(\gamma'\mathbf{C}^{-1}1 - 2\right),$$

$$\text{Var}\left(\widehat{Z}_{\text{new}}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) = \text{Var}\left(\widehat{Z}_{\text{new}}(\mathbf{s}_0)\right) + \text{Var}\left(Z(\mathbf{s}_0)\right) - 2\text{cov}\left(\widehat{Z}_{\text{new}}(\mathbf{s}_0), Z(\mathbf{s}_0)\right) \tag{17}$$

$$= \left(\gamma - \frac{1}{1'\mathbf{C}^{-1}1}\right)'\mathbf{C}^{-1}\left(\gamma - \frac{1}{1'\mathbf{C}^{-1}1}\right) + \sigma_z^2(\mathbf{s}_0) - 2\left(\gamma - \frac{1}{1'\mathbf{C}^{-1}1}\right)'\mathbf{C}^{-1}\gamma$$

$$= \sigma_z^2(\mathbf{s}_0) - \gamma'\mathbf{C}^{-1}\gamma + \frac{1}{1'\mathbf{C}^{-1}1}.$$

Because of this fact, we deduce that $\widehat{Z}_{\text{new}}(\mathbf{s}_0)$ is a biased estimator. Now, to compare with the ordinary Kriging estimator given by (11), variance is not a useful criterion. The MSE is better when solving this problem and not only in this case but also while comparing between estimators even though at least one of them is biased. Finally, Theorem 2 shows that $\widehat{Z}_{\text{new}}(\mathbf{s}_0)$ is better than the ordinary Kriging estimator (11) regarding the MSE under some conditions. □

Theorem 2. *Under one of the following two conditions (i) and (ii), $\widehat{MSE}\,\widehat{Z}_{new}(\mathbf{s}_0)$ will be less than $\widehat{MSE}(\widehat{Z}(\mathbf{s}_0))$.*

(i) $(1'\mathbf{C}^{-1}\mathbf{Z})^2 > 1'\mathbf{C}^{-1}1$ and $\gamma'\mathbf{C}^{-1}1$ are between 2 and $2(1'\mathbf{C}^{-1}\mathbf{Z})^2/(1'\mathbf{C}^{-1}\mathbf{Z})^2 - 1'\mathbf{C}^{-1}1$

(ii) $(1'\mathbf{C}^{-1}\mathbf{Z})^2 < 1'\mathbf{C}^{-1}1$ and $\gamma'\mathbf{C}^{-1}1 > 2$ or $\gamma'\mathbf{C}^{-1}1 < 2(1'\mathbf{C}^{-1}\mathbf{Z})^2/(1'\mathbf{C}^{-1}\mathbf{Z})^2 - 1'\mathbf{C}^{-1}1)$

Proof. From Corollary 1, we have $\text{MSE}(\widehat{Z}_{\text{new}}(\mathbf{s}_0)) = (\mu(\gamma'\mathbf{C}^{-1}1 - 2))^2 + \sigma_z^2(\mathbf{s}_0) - \gamma'\mathbf{C}^{-1}\gamma + 1/1'\mathbf{C}^{-1}1$. On the other hand, Equation (12) results in $\text{MSE}(\widehat{Z}(\mathbf{s}_0)) = \sigma_z^2(\mathbf{s}_0) - \gamma'\mathbf{C}^{-1}\gamma + (1 - \gamma'\mathbf{C}^{-1}1)^2/1'\mathbf{C}^{-1}1$. Then, by defining $c = \gamma'\mathbf{C}^{-1}1$ and $a = 1'\mathbf{C}^{-1}1$, one can show that $d = \text{MSE}(\widehat{Z}_{\text{new}}(\mathbf{s}_0)) - \text{MSE}(\widehat{Z}(\mathbf{s}_0)) = 2c - c^2/a + \mu^2(2-c)^2$

From the positive definiteness of $\mathbf{C}$, it is found that $a > 0$. By multiplying the above equality by a, we get $ad = 2c - c^2 + a\mu^2(2-c)^2$. Replacing μ by $\widehat{\mu}$ leads to

Corollary 1. *The biasedness and prediction error variance of the estimator $\widehat{Z}_{new}$ is given by*

$$E\left(\widehat{Z}_{\text{new}}(\mathbf{s}_0) - Z(s_0)\right) = \mu\left(\gamma'\mathbf{C}^{-1}1 - 2\right),$$

$$\text{Var}\left(\widehat{Z}_{\text{new}}(\mathbf{s}_0) - Z(s_0) = \sigma_z^2(\mathbf{s}_0)\right) - \gamma'\mathbf{C}^{-1}\gamma + \frac{1}{1'\mathbf{C}^{-1}1}. \tag{16}$$

Proof. The proof comes from Theorem 1, and a direct computation is presented in the following equations:

$$a\widehat{d} = c^2(e-a) + c(2a - 4e) + 4e, \tag{18}$$

where $e = (1'\mathbf{C}^{-1}\mathbf{Z})^2$. The quadratic form (18) has two roots $c_1 = 2$ and $c_2 = 2e/e - a$ with respect to c. Under condition (i) or (ii), Equation (18) will be negative; hence, $\widehat{d} \le 0$ under condition (i) or (ii) as well. This completes the proof. □

4. Asymptotic Behavior of the Mean Estimator

Since the estimator of μ (13) in Theorem 1 has an essential role, its asymptotic behavior should be interesting for this study. First, we denote the following lemma, which is required for the continuation of this section.

Lemma 1. *For two sequences of unbiased estimators $\widehat{\theta}_n^{(1)}$ and $\widehat{\theta}_n^{(2)}$ we can say that $\widehat{\theta}_n^{(2)} \longrightarrow^L 2\theta$, $\widehat{\theta}_n^{(1)} \longrightarrow^L 2\theta$, if $\text{Var}(\widehat{\theta}_n^{(2)}) \ge \text{Var}(\widehat{\theta}_n^{(1)})$. Here, $\longrightarrow^L 2$ states the mean square convergence.*

Theorem 3. *Let a random process $Z(s)$ exists on a lattice Λ in R^k with coordinates, $\mathbf{s}_1, \ldots, \mathbf{s}_n$, $\text{Var}(Z(\mathbf{s}_i)) \le c < \infty$ for all $i = 1, \ldots, n$ and $\text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \longrightarrow 0$ as $|\mathbf{s}_i - \mathbf{s}_j|$ tends to infinity, then $\overline{Z}_n \longrightarrow^L 2\mu$ and $\widehat{\mu} \longrightarrow^L 2\mu$.*

Proof. Without losing generality, assume Z_i denotes $Z(\mathbf{s}_i)$. First, we know that $E(\overline{Z}_n) = 1/n(\sum_{i=1}^n E(Z_i)) = \mu$. Now, we can compute the variance of $\overline{Z}_n$

$$\mathrm{Var}\left(\overline{Z}_n\right) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Var}(Z_i)\mathrm{Var}(Z_j)\rho(Z_i, Z_j)$$

$$\leq \frac{c}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\rho(Z_i, Z_j)\right|$$

$$= \frac{c}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\rho(Z_i, Z_j)\right| I_{\left\{\left|\mathbf{s}_i - \mathbf{s}_j\right| \geq N_\varepsilon\right\}} \quad (19)$$

$$+ \frac{c}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left|\rho(Z_i, Z_j)\right| I_{\left\{\left|\mathbf{s}_i - \mathbf{s}_j\right| \geq N_\varepsilon\right\}}.$$

The last term has n^2 components. The assumption which $\rho(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) \longrightarrow 0$, as $|\mathbf{s}_i - \mathbf{s}_j|$ tends to infinity gives that for all $\varepsilon > 0$, there exists an integer N_ε enough large such that $|\rho(Z(\mathbf{s}_i), Z(\mathbf{s}_j))| \leq \varepsilon$ for all $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$ and that $|\mathbf{s}_i - \mathbf{s}_j| \geq N_\varepsilon$. For a fixed $\mathbf{s}_i$, the number of $\mathbf{s}_j$ so that $|\mathbf{s}_i - \mathbf{s}_j| \leq N_\varepsilon$ is less than $(2N_\varepsilon)^k$. Therefore, the total number of points with condition $|\mathbf{s}_i - \mathbf{s}_j| \leq N_\varepsilon$ is at last $n(2N_\varepsilon)^k$ which leads to $\mathrm{Var}(\overline{Z}_n) \leq c/n^2 (n(2N_\varepsilon)^k + (n^2 - n(2N_\varepsilon)^k)\varepsilon) = c(1/n(2N_\varepsilon)^k + (1 - 1/n(2N_\varepsilon)^k)\varepsilon)$ which clearly proves $\overline{Z}_n \longrightarrow^L 2\mu$. By Theorem 1, $\widehat{\mu}$ is unbiased for μ and $\mathrm{Var}(\widehat{\mu}) < \mathrm{Var}(\overline{Z}_n)$. Finally, by Lemma 1 we get that $\widehat{\mu} \longrightarrow^L 2\mu$. □

Corollary 2. *It is an immediate consequence of Theorem 4 that under its conditions $1'\mathbf{C}^{-1}1$ tends to infinity as n tends to infinity.*

Theorem 4. *Under conditions of Theorem 4 and at least one of the followings (i) or (ii) and $\widehat{Z}_n(\mathbf{s}_0) \longrightarrow^L 2\mu$.*

(i) $\gamma'\mathbf{C}^{-1}\gamma \longrightarrow 0$ as n tends to infinity.
(ii) $1'\mathbf{C}^{-1}1\gamma'\mathbf{C}^{-1}\gamma - (\gamma'\mathbf{C}^{-1}1)^2$ is bounded.

Proof. Regarding Lemma 1 and the unbiasedness of $\widehat{Z}_n(\mathbf{s}_0)$, we see that $\mathrm{var}(\widehat{Z}_n(\mathbf{s}_0)) \longrightarrow 0$ if it satisfies condition (i) or (ii). Since $\widehat{\mu}_n$ is the linear unbiased estimator for μ with the minimum variance and $\widehat{Z}_n(\mathbf{s}_0)$ is the linear unbiased estimator for μ, we obtain that

$$0 \leq \mathrm{var}(\widehat{\mu}_n) - \mathrm{var}(\widehat{Z}_n(\mathbf{s}_0)) = k_n. \quad (20)$$

Corollary 1 and equation (20) result in $var(\widehat{Z}_n(\mathbf{s}_0)) \longrightarrow 0$ if and only if $k_n \longrightarrow 0$. On the other hand, a direct computation shows that $k_n = 1'\mathbf{C}^{-1}1\,\gamma'\mathbf{C}^{-1}\gamma - (\gamma'\mathbf{C}^{-1}1)^2/1'\mathbf{C}^{-1}1$ and under condition (ii), it obviously tends to 0. Another form of k_n is $k_n = \gamma'\mathbf{C}^{-1}\gamma - (\gamma'\mathbf{C}^{-1}1)^2/1'\mathbf{C}^{-1}1$, with the fact that $(\gamma'\mathbf{C}^{-1}1)^2/1'\mathbf{C}^{-1}1 > 0$ and if (i) satisfies, it will tend to 0. □

5. Conclusion

In this paper, we showed some previously defined RFs that have been extensively applied to modeling time series, spatial, and spatiotemporal data. However, they have a few limitations that they cannot consider the PSAS, which has a crucial and basic role in modeling the previously mentioned data. For that, we presented some alternative RFs without the mentioned problem, and having a more general solution is still needed. The paper recommended a method for coping with this problem by finding a new predictor in ordinary Kriging. The customary predictor for ordinary Kriging has a minimum MSE in the class of all linear unbiased predictors. However, it does not necessarily have a minimum MSE in the class of all linear predictors. Therefore, we obtained a biased predictor with a less MSE than the Kriging predictor. Asymptotic behavior of this predictor alongside the past Kriging predictor was provided.

We studied six well-known multivariate distributions in this article. Some other multivariate distributions studied by Kotz et al. [26] such as the truncated multivariate normal distribution, truncated multivariate t distribution, Linnik's distribution, gamma distribution, and logistic distribution can be surveyed under this paper's point of view.

Despite providing some solutions for the problem of lacking PSAS, it seems that the best method for modeling skew spatial data is defining a consistent skew RF. This matter may be performed in the future.

Data Availability

There are no data related to this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

The idea of the paper was conceived and planned by Mohammad Mehdi Saber. Ramy Abdelhamid Aldallal was involved in writing the manuscript. All authors provided critical feedback and helped shape the manuscript's research, analysis, and construction.

References

[1] A. Azzalini, "A class of distributions which includes the normal ones," *Scandinavian Journal of Statistics*, vol. 12, pp. 171–178, 1985.

[2] H.-M. Kim and B. Mallick, "A Bayesian prediction using the skew Gaussian distribution," *Journal of Statistical Planning and Inference*, vol. 120, no. 1-2, pp. 85–101, 2004.

[3] D. Allard and P. Naveau, "A new spatial skew-normal random field model," *Communications in Statistics - Theory and Methods*, vol. 36, no. 9, pp. 1821–1834, 2007.

[4] O. Karimi, H. Omre, and M. Mohammadzadeh, "Bayesian closed-skew Gaussian inversion of seismic AVO data for elastic material properties," *Geophysics*, vol. 75, pp. 1–11, 2010.

[5] F. Hosseini, J. Eidsvik, and M. Mohammadzadeh, "Approximate Bayesian inference in spatial GLMM with skew normal latent variables," *Computational Statistics & Data Analysis*, vol. 55, no. 4, pp. 1791–1806, 2011.

[6] O. Karimi and M. Mohammadzadeh, "Bayesian spatial prediction for discrete closed skew Gaussian random field," *Mathematical Geosciences*, vol. 43, no. 5, pp. 565–582, 2011.

[7] H. Zareifard and M. J. Jafari Khaledi, "Non-Gaussian modeling of spatial data using scale mixing of a unified skew Gaussian process," *Journal of Multivariate Analysis*, vol. 114, pp. 16–28, 2013.

[8] R. B. Arellano-Valle and A. Azzalini, "On the unification of families of skew-normal distributions," *Scandinavian Journal of Statistics*, vol. 33, no. 3, pp. 159–188, 2006.

[9] B. Mahmoudian, "A skewed and heavy-tailed latent random field model for spatial extremes," *Journal of Computational & Graphical Statistics*, vol. 26, no. 3, pp. 658–670, 2017.

[10] S. K. Sahu, D. K. Dey, and M. D. Branco, "A new class of multivariate skew distributions with applications to Bayesian regression models," *Canadian Journal of Statistics*, vol. 31, no. 2, pp. 129–150, 2003.

[11] T. J. Kozubowski, K. Podgórski, and I. Rychlik, "Multivariate generalized Laplace distribution and related random fields," *Journal of Multivariate Analysis*, vol. 113, pp. 59–72, 2013.

[12] M. M. Saber, A. R. Nematollahi, and M. Mohammadzadeh, "Generalized asymmetric Laplace random fields: existence and application," *Journal of Data Science*, vol. 18, pp. 51–68, 2018.

[13] R. B. Arellano-Valle and M. G. Genton, "Multivariate extended skew-t distributions and related families," *Metron*, vol. 68, no. 3, pp. 201–234, 2010.

[14] M. Wall, "A close look at the spatial structure implied by the CAR and SAR models," *Journal of Statistical Planning and Inference*, vol. 121, no. 2, pp. 311–324, 2004.

[15] B. Mahmoudian, "On the existence of some skew-Gaussian random field models," *Statistics & Probability Letters*, vol. 137, pp. 331–335, 2018.

[16] S. Basu and G. C. Reinsel, "Properties of the spatial unilateral first-order ARMA model," *Advances in Applied Probability*, vol. 25, no. 3, 1993.

[17] M. M. Saber and A. R. Nematollahi, "Comparison of spatial interpolation methods in the first order stationary multiplicative spatial autoregressive models," *Communications in Statistics - Theory and Methods*, vol. 46, no. 18, pp. 9230–9246, 2017.

[18] M. M. Saber, "Performance of extrapolation based on Pitman's measure of closeness in spatial regression models with extended skew t innovations," *Communications in Statistics - Theory and Methods*, vol. 48, no. 2, pp. 282–299, 2019.

[19] A. Mojiri, Y. Waghei, H. R. Sani, and G. ., R. Borzadaran, "Comparison of predictions by kriging and spatial autoregressive models," *Communications in Statistics - Simulation and Computation*, vol. 47, no. 6, pp. 1785–1795, 2018.

[20] A. Mojiri, Y. Waghei, H. R. Nili Sani, and G. R. Mohtashami Borzadaran, "Non-stationary spatial autoregressive modeling for the prediction of lattice data," *Communications in Statistics - Simulation and Computation, forthcoming*, 2021.

[21] F. Hosseini and O. Karimi, "Approximate pairwise likelihood inference in SGLM models with skew normal latent variables," *Journal of Computational and Applied Mathematics*, vol. 398, 2021.

[22] N. A. C. Cressie, *Statistics for Spatial Data*, Wiley, New York, NY, USA, 1993.

[23] R. A. Moyeed and A. Papritz, "An empirical comparison of kriging methods for nonlinear spatial point prediction," *Mathematical Geology*, vol. 34, no. 4, pp. 365–386, 2002.

[24] G. Gonzales-Farias, G Dominguez-Molin, and A. K. Gupta, M. G. Genton, The closed skew normal distribution," in *Skew-Elliptical Distribution and Their Applications: A Journey beyond Normality*, pp. 25–42, Chapman & Hall, Boca Rayton, FL, USA, 2004.

[25] R. B. Arellano-Valle and A. Azzalini, "Some properties of the unified skew-normal distribution," *Mathematics arXiv: Statistics Theory*, vol. 63, pp. 461–487, 2020.

[26] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions. Models and Applications*, Wiley, vol. 1, New York, NY, USA, , 2000.