

# Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function

Yuan Yan | Marc G. Genton

CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

**Correspondence**

Marc G. Genton, CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.

Email: marc.genton@kaust.edu.sa

**Funding information**

King Abdullah University of Science and Technology (KAUST)

Gaussian likelihood inference has been studied and used extensively in both statistical theory and applications due to its simplicity. However, in practice, the assumption of Gaussianity is rarely met in the analysis of spatial data. In this paper, we study the effect of non-Gaussianity on Gaussian likelihood inference for the parameters of the Matérn covariance model. By using Monte Carlo simulations, we generate spatial data from a Tukey  $g$ -and- $h$  random field, a flexible trans-Gaussian random field, with the Matérn covariance function, where  $g$  controls skewness and  $h$  controls tail heaviness. We use maximum likelihood based on the multivariate Gaussian distribution to estimate the parameters of the Matérn covariance function. We illustrate the effects of non-Gaussianity of the data on the estimated covariance function by means of functional boxplots. Thanks to our tailored simulation design, a comparison of the maximum likelihood estimator under both the increasing and fixed domain asymptotics for spatial data is performed. We find that the maximum likelihood estimator based on Gaussian likelihood is overall satisfying and preferable than the non-distribution-based weighted least squares estimator for data from the Tukey  $g$ -and- $h$  random field. We also present the result for Gaussian kriging based on Matérn covariance estimates with data from the Tukey  $g$ -and- $h$  random field and observe an overall satisfactory performance.

**KEYWORDS**

Gaussian likelihood, heavy tails, kriging, log-Gaussian random field, Matérn covariance function, non-Gaussian random field, skewness, spatial statistics, Tukey  $g$ -and- $h$  random field

## 1 | INTRODUCTION

In the analysis of spatial data, estimation of the covariance function, which captures the spatial dependence structure that cannot be solely explained by a global trend, is a crucial step in statistical inference. Knowing the covariance structure not only helps in the estimation of the global trend but also allows the prediction of the spatial process at new locations. A parametric model is usually assumed for the covariance function due to the positive-definiteness requirement. Sometimes, interpretation of covariance function parameters is equally important as estimating the trend and making predictions. The maximum likelihood (ML) method based on Gaussian distributions has been studied and used extensively for covariance function estimation in environmental problems due to its simplicity and many attractive properties. However, the assumption of Gaussianity rarely holds in practice when spatial data are analyzed. It is therefore important to understand the impact of a misspecified distribution assumption on ML inference with Gaussian likelihood.

Recently, models for non-Gaussian random fields have received growing attention for real data analyses, for example, precipitation and wind speed data. The main approaches to constructing non-Gaussian random fields include trans-Gaussian random

fields (Cressie, 1993), skew-Gaussian processes (Zhang & El-Shaarawi, 2010), scale-mixing Gaussian random fields (Fonseca & Steel, 2011), log-skew-elliptical random fields (Marchenko & Genton, 2010), spatial copula models (Gräler, 2014; Krupskii, Huser, & Genton, 2017), and non-Gaussian Matérn fields based on stochastic partial differential equation (Wallin & Bolin, 2015). These models are usually extensions of their univariate and multivariate counterparts. Jones (2015) gave a summary and comparison of various existing methods to build univariate non-Gaussian distributions with shape parameters; see Genton and Zhang (2012), Xu and Genton (2017), and references therein for a more detailed review of non-Gaussian random fields.

Consider a random field  $\{Y(\mathbf{s}), \mathbf{s} \in S \subset \mathbb{R}^q\}$ . To simplify the model, we assume that  $Y(\mathbf{s})$  has zero mean and a stationary and isotropic covariance function. By ignoring the mean structure, we avoid the problem of bias when estimating the regression coefficients of the global trend and parameters of the covariance function simultaneously using the ML method. The restricted (or residual) ML (REML) estimation was proposed to reduce the bias. There are several papers on the effects of the number of regression coefficients to be estimated and on comparisons of the REML and ML methods (Irvine, Gitelman, & Hoeting, 2007; Zimmerman & Zimmerman, 1991). With our focus on the effect of misspecified Gaussianity, here, we consider only parameters of the covariance function.

Denote by  $C(d; \boldsymbol{\theta})$  a parametric isotropic covariance function for observations that are  $d$  units apart, where  $\boldsymbol{\theta} \in \mathbb{R}^p$  is a vector of parameters for the covariance function. Our goal is to estimate  $\boldsymbol{\theta}$  from spatial observations generated by the underlying random field at  $n$  locations,  $\mathbf{Y} = \{Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)\}^\top$ . The Matérn covariance model is a flexible and widely used class of valid isotropic covariance functions; see Guttorp and Gneiting (2006) for a review of the development of the Matérn family. The Matérn covariance function takes the following form:

$$C_{\mathcal{M}}(d; \boldsymbol{\theta}) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{d}{\phi}\right)^\nu K_\nu\left(\frac{d}{\phi}\right),$$

where  $\boldsymbol{\theta} = (\sigma^2, \phi, \nu)^\top$ ,  $\sigma^2$  is the variance,  $\phi > 0$  is the range parameter,  $\nu > 0$  controls the smoothness of the spatial process, and  $K_\nu$  is the modified Bessel function of the second kind of order  $\nu$ . The exponential covariance,  $C_{\text{exp}}(d; \sigma^2, \phi) = \sigma^2 \exp(-d/\phi)$ , is a special case of the Matérn family with  $\nu = 0.5$ .

Under the assumption of a Gaussian random field of zero mean,  $\mathbf{Y}$  follows a zero-mean multivariate normal distribution with a log-likelihood function:

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \log |\Sigma(\boldsymbol{\theta})| - \frac{1}{2} \mathbf{Y}^\top \Sigma(\boldsymbol{\theta})^{-1} \mathbf{Y} - \frac{n}{2} \log(2\pi), \quad (1)$$

where  $\Sigma(\boldsymbol{\theta})_{ij} = C(\|\mathbf{s}_i - \mathbf{s}_j\|; \boldsymbol{\theta})$ ,  $i, j = 1, \dots, n$ . Taking the derivative of Equation 1 with respect to  $\theta_i$ , we get the following score function:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i} = -\frac{1}{2} \text{tr} \left\{ \Sigma(\boldsymbol{\theta})^{-1} \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_i} \right\} + \frac{1}{2} \mathbf{Y}^\top \Sigma(\boldsymbol{\theta})^{-1} \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_i} \Sigma(\boldsymbol{\theta})^{-1} \mathbf{Y}. \quad (2)$$

Then, the ML estimator (MLE),  $\hat{\boldsymbol{\theta}}$ , of  $\boldsymbol{\theta}$  is computed by maximizing  $l(\boldsymbol{\theta})$ , which is equivalent to solving  $\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i} = 0$ , for  $i = 1, \dots, p$ .

We are interested in the performance of the parameter estimation for the covariance model and prediction at unknown locations based on the Gaussian likelihood when the true underlying random field is non-Gaussian. We focus our attention on one flexible class of trans-Gaussian random fields proposed by Xu and Genton (2017), the Tukey  $g$ -and- $h$  (TGH) random field, in which  $g \in \mathbb{R}$  controls skewness and  $h \geq 0$  controls tail heaviness.

We investigate the impact of a misspecified Gaussian assumption on the MLE by means of Monte Carlo simulations. We generate spatial data from a TGH random field with the Matérn covariance function. We acknowledge that generating data from a non-Gaussian random field with a given covariance structure is difficult, and we circumvent this difficulty by utilizing the explicit form of the transformed covariance function of the TGH random field. We then compute the MLE based on Equation 1 to estimate the parameters of the Matérn covariance function. We illustrate the effects of non-Gaussianity on the estimated covariance function parameters with boxplots and the estimated covariance function as a whole with functional boxplots (Sun & Genton, 2011). Thanks to our tailored simulation design, a comparison of the MLE under both the increasing and fixed domain asymptotics for spatial data is performed. In addition, we compute the weighted least squares (WLS) estimator and compare its performance with the MLE. We also run simulations to assess the performance of the Gaussian kriging prediction with the MLE plugged in for the covariance function.

The remainder of this paper is organized as follows. Section 2 introduces the TGH random field and briefly describes the procedure to generate spatial data from a TGH random field with a given covariance function. Section 3 describes the simulation design and presents results from estimations based on Gaussian ML and WLS methods, as well as from predictions. We end this paper with a discussion in Section 4.

## 2 | TGH RANDOM FIELDS WITH MATÉRN COVARIANCE

Although Gaussian random fields are among the most popular models used to fit spatial observations, data collected in practice, for example, environmental data, oftentimes do not fit the Gaussian model very well. Indeed, marginal features, such as skewness and elongation, are not captured by the Gaussian distribution. Parametric models for non-Gaussian random fields have been proposed to fit spatial data that are skewed and/or have heavy tails. Non-Gaussian random fields can be built from trans-Gaussian random fields. Commonly used transformations include lognormal (De Oliveira, 2006), square-root (Johns, Nychka, Kittel, & Daly, 2003), Box–Cox (De Oliveira, Kedem, & Short, 1997), and power transformations (Allcroft & Glasbey, 2003). Recently, a new family of trans-Gaussian random fields, the TGH random field, was proposed by Xu and Genton (2017). The TGH random fields accommodate different levels of skewness and tail heaviness in marginal distributions. It is based on the bijective TGH transformation (Tukey, 1977):

$$\tau_{g,h}(z) = \begin{cases} g^{-1} \{ \exp(gz) - 1 \} \exp(hz^2/2), & g \neq 0, h \geq 0, z \in \mathbb{R}, \\ z \exp(hz^2/2), & g = 0, h \geq 0, z \in \mathbb{R}. \end{cases}$$

The standard TGH random field is formed by applying the TGH transformation to a standard Gaussian random field,  $Z(\mathbf{s})$ , which is a stationary Gaussian random field with mean 0, variance 1, and correlation function  $\rho_Z(\mathbf{s}_1, \mathbf{s}_2)$ :  $T(\mathbf{s}) = \tau_{g,h}\{Z(\mathbf{s})\}$ .

The mean  $\mu_T$  and variance  $\sigma_T^2$  of the TGH random field are:

$$\mu_T = E\{T(\mathbf{s})\} = \frac{1}{g\sqrt{1-h}} \left[ \exp \left\{ \frac{g^2}{2(1-h)} \right\} - 1 \right], \quad 0 \leq h < 1, \quad (3)$$

$$\sigma_T^2 = \text{var}\{T(\mathbf{s})\} = C_T(0) = \frac{1}{g^2\sqrt{1-2h}} \left[ \exp \left( \frac{2g^2}{1-2h} \right) - 2 \exp \left\{ \frac{g^2}{2(1-2h)} \right\} + 1 \right] - \mu_T^2, \quad 0 \leq h < \frac{1}{2}. \quad (4)$$

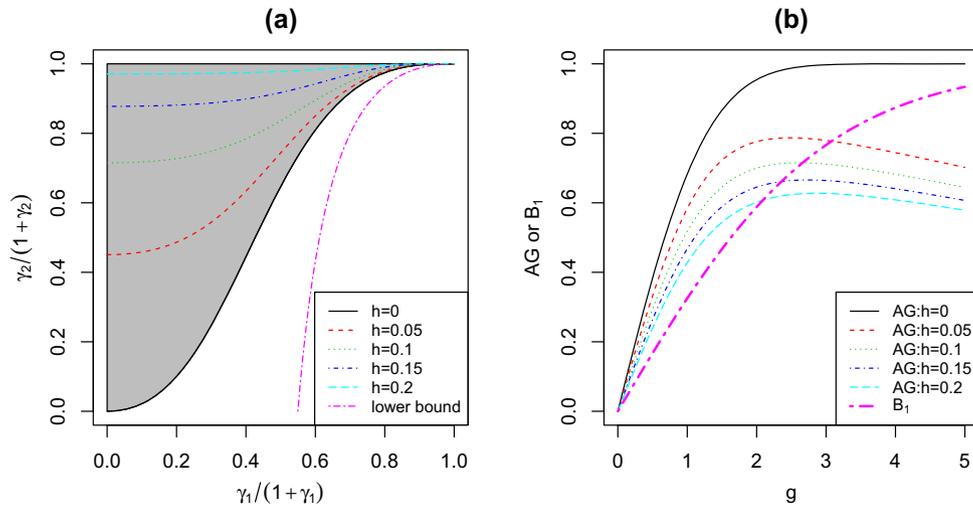
Essentially, the parameter  $g$  controls skewness, whereas  $h$  controls tail heaviness. In particular, for  $h = 0, g > 0$ , the skewness and kurtosis of the TGH random field are the same as for the lognormal distribution with  $\mu = 0$  and  $\sigma = g$ . We express skewness ( $\gamma_1$ ) and excess kurtosis ( $\gamma_2$ ) analytically for the special cases when  $g$  or  $h$  equals to 0:

$$\gamma_1\{T(\mathbf{s})\} = \begin{cases} \text{sgn}(g)\{\exp(g^2) + 2\} \sqrt{\exp(g^2) - 1}, & h = 0, \\ 0, & g = 0, 0 \leq h < \frac{1}{3}, \end{cases}$$

$$\gamma_2\{T(\mathbf{s})\} = \begin{cases} \exp(4g^2) + 2 \exp(3g^2) + 3 \exp(2g^2) - 6, & h = 0, \\ 3 \frac{(1-2h)^2}{(1-4h)^{5/2}} - 3 \approx 12h + 66h^2, & g = 0, 0 \leq h < \frac{1}{4}. \end{cases}$$

For  $g \neq 0$  and  $h \neq 0$ , the explicit forms of  $\gamma_1$  and  $\gamma_2$  become extremely complicated, and it is easier to illustrate the skewness–kurtosis relation graphically. We only plot the case when  $g \geq 0$ , which results in positive skewness, because the figures in the following are symmetric about the  $y$ -axis for  $g < 0$ . As the skewness and excess kurtosis could range from 0 to infinity for distributions with positive skewness and tails heavier than the normal distribution, we rescale both the skewness and excess kurtosis coefficients to the range  $[0,1)$  through the rescaling function  $f(x) = x/(1+x)$  as in Jones and Pewsey (2009). The attainable regions in the rescaled skewness–kurtosis plane for the TGH distribution is shown (shaded area) in Figure 1(a) overlapped with curves that represent the scaled skewness–kurtosis relationship for different values of  $h$  when  $g$  spans the real line. The magenta line is the theoretical lower bound for non-Gaussian distributions.

Because higher moments do not exist when  $h$  is large for the TGH distribution, and the classical moment-based measures of skewness is sometimes not a good measure of skewness level for the density shape, we discuss alternative measures of skewness associated with the  $g$  parameter. Arnold and Groeneveld (1995) proposed an interpretable measure of skewness with respect to the mode, which is defined as one minus two times the cumulated mass to the left of the mode, that is,  $AG = 1 - 2F(\text{mode})$ , where  $F$  is the cumulative distribution function. Figure 1(b) shows the relationship between  $g$  and AG skewness for different values of  $h$  for the TGH distribution. We can see that, as  $h$  increases, the attainable level of skewness measured by AG decreases. Bowley's coefficient is a quantile-based measure of skewness, which is defined by  $B_1 = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$ , where  $Q_k$  is the  $k$ th quartile of the distribution. For the TGH distribution, Bowley's coefficient depends solely on  $g$ , which shows that  $g$  indeed controls the level of skewness in the quantile sense regardless of the kurtosis. Figure 1(b) also shows Bowley's coefficient as a monotonic increasing function of  $g$ .



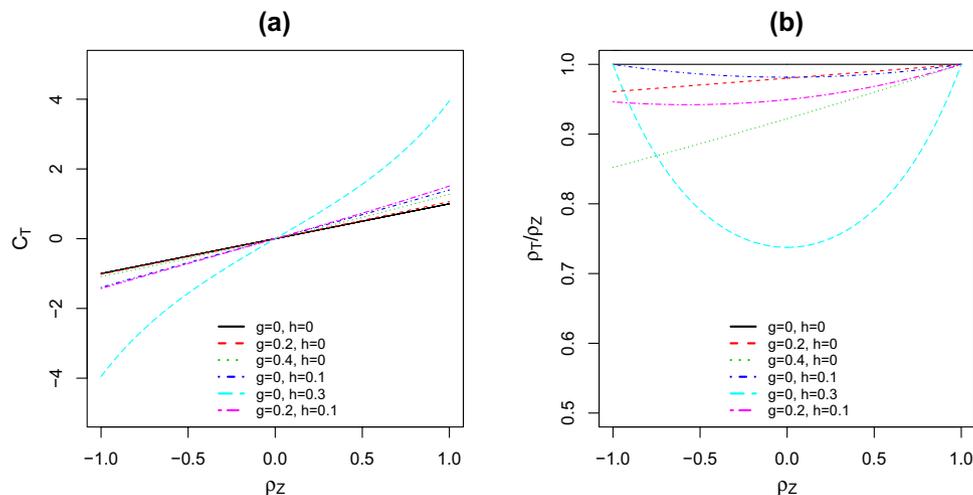
**FIGURE 1** (a) Attainable regions (shaded area) in the rescaled skewness–kurtosis plane for the Tukey  $g$ -and- $h$  distribution with curves representing the scaled skewness–kurtosis relationship for different values of  $h$  when  $g$  spans the real line. The magenta line is the theoretical lower bound. (b) Relationship between  $g$  and Arnold–Groeneveld skewness for different values of  $h$  and relationship between  $g$  and Bowley’s coefficient for the Tukey  $g$ -and- $h$  distribution

The covariance and correlation functions  $C_T(d)$  and  $\rho_T(d)$  of the transformed random field  $T(s)$  can be found explicitly, given the correlation function of the standard Gaussian random field  $\rho_Z(d)$  (for simplicity, we consider only isotropic covariance functions). For  $h < 1/2$ ,

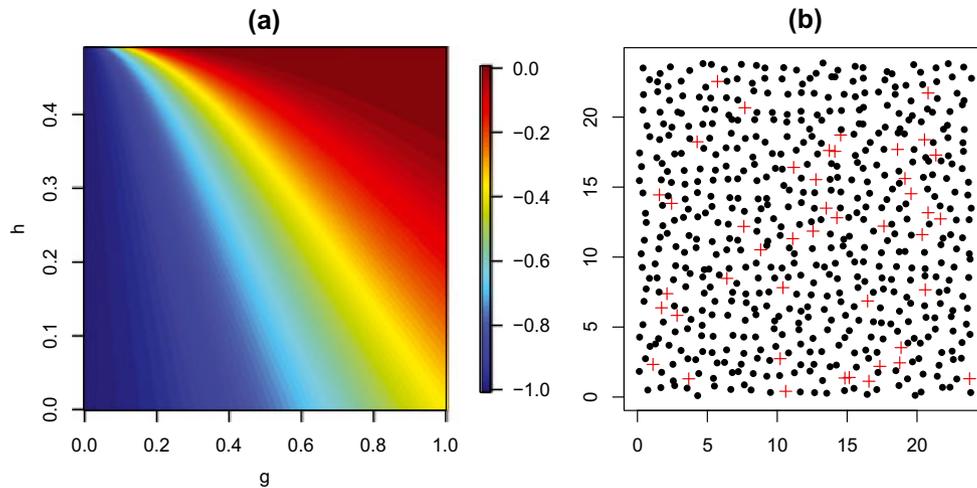
$$C_T(d) = \frac{\exp\left[\frac{1+\rho_Z(d)}{1-h\{1+\rho_Z(d)\}}g^2\right] - 2\exp\left[\frac{1-h\{1-\rho_Z^2(d)\}}{(1-h)^2-h^2\rho_Z^2(d)}\frac{g^2}{2}\right] + 1}{g^2\sqrt{(1-h)^2 - \rho_Z^2(d)h^2}} - \mu_T^2, \tag{5}$$

$$\rho_T(d) = \frac{C_T(d)}{\sigma_T^2}. \tag{6}$$

It is easy to show that the mapping from  $\rho_Z$  to  $C_T$  in Equation 5 or from  $\rho_Z$  to  $\rho_T$  in Equation 6 is monotonically increasing for any fixed  $g$  and  $h$ . Figure 2(a) displays the relationship between  $C_T$  and  $\rho_Z$  for six pairs of different values of  $g$  and  $h$  that are used in our simulation study. We notice that the transformed random field always has a variance larger than or equal to 1 and  $h$  has a strong impact on the magnitude of  $C_T$ . Figure 2(b) displays the ratio of  $\rho_T$  to  $\rho_Z$  as a function of  $\rho_Z$  as it ranges from  $-1$  to  $1$  for the same pairs of values for  $g$  and  $h$ . We can see that  $|\rho_T(d)| \leq |\rho_Z(d)|$  for a fixed  $d$ , which means that the transformation weakens the correlation.



**FIGURE 2** Tukey  $g$ -and- $h$  random field: (a) relationship between  $C_T$  and  $\rho_Z$ ; (b) the  $\rho_T/\rho_Z$  ratio as a function of  $\rho_Z$



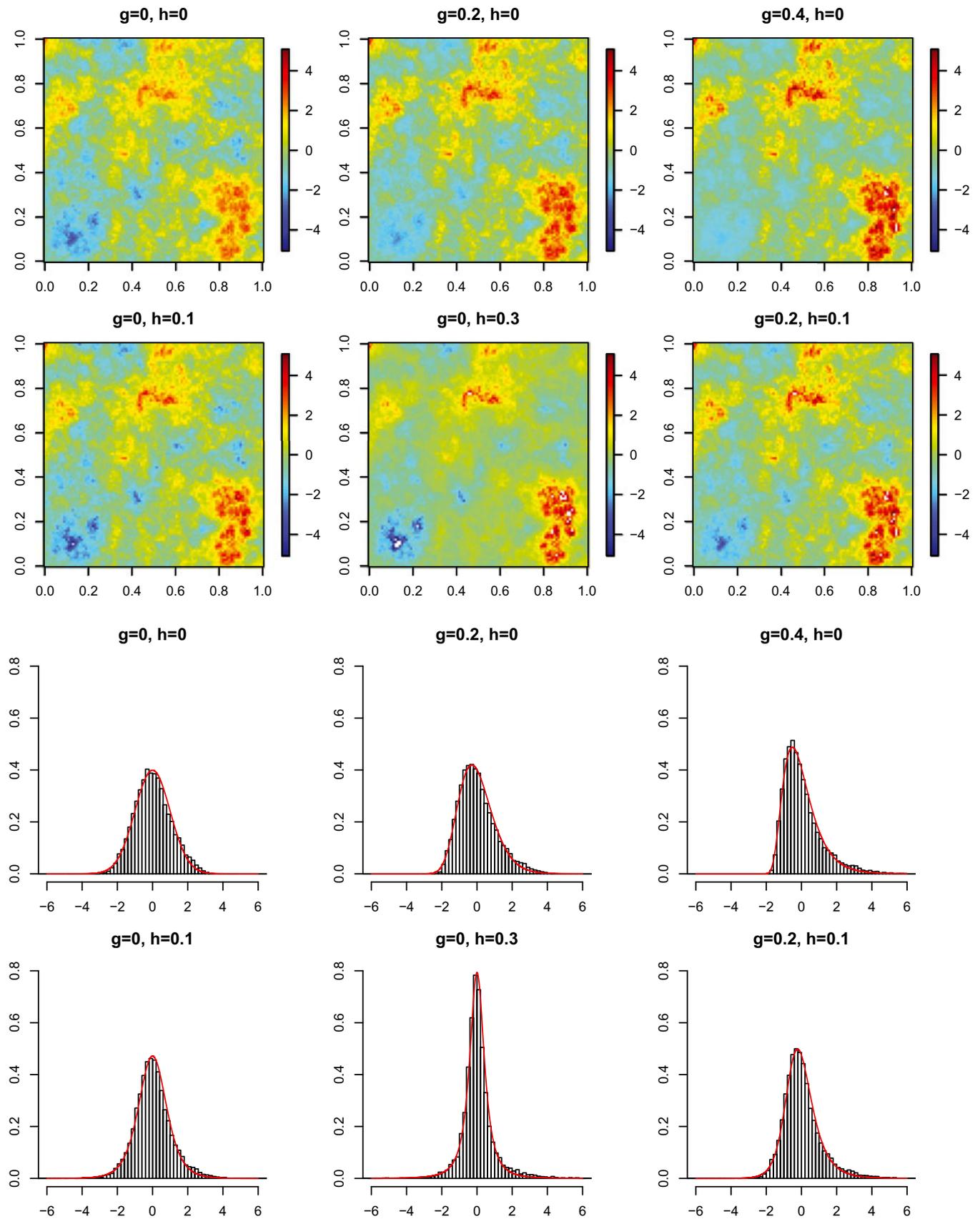
**FIGURE 3** (a)  $\rho^*$  (lower bound of the correlation coefficient after the Tukey  $g$ -and- $h$  transformation) as a function of  $g$  and  $h$ . (b) Realization of 576 perturbed grid points, with 532 (black dots) for estimation and 44 (red plus signs) for prediction validation

Because of the one-to-one relationship of  $\rho_Z$  to  $\rho_T$ , given a desired correlation model for  $T(\mathbf{s})$  and certain values of  $g$  and  $h$ , we can numerically find the correlation function  $\rho_Z(d)$  of the underlying standard Gaussian random field at any distance  $d$ . However, the correlation  $\rho_T(d)$  obtained by the TGH transformation has a lower bound,  $-1 \leq \rho^* < 0$ , where  $\rho^*$  corresponds to setting  $\rho_Z = -1$  in Equation 6 and  $\rho^*$  depends on values of  $g$  and  $h$ . Figure 3(a) displays the colormap for  $\rho^*$  as a function of  $g$  and  $h$ . There is thus a restriction on the correlation model for the TGH random fields, so that it can be inverted. Under the assumption that the covariance model is isotropic, there is already a lower bound for a valid isotropic covariance function in  $\mathbb{R}^q$  because of Schoenberg's theorem. For example, when  $q = 2$ , the lower bound is about  $-0.403$ , and for an isotropic covariance function to be valid in all dimensions, it has to be positive (Stein, 1999, p. 45). The Matérn covariance model is valid in all dimensions. There should therefore be no problem with solving for  $\rho_Z(d)$  under the restriction of  $\rho^*$ . However, there is an additional restriction that the inverted  $\rho_Z(d)$  should be a positive definite function. This condition needs to be checked case by case, and we suspect it depends on the values of  $g$  and  $h$ . In our simulation study with moderate values for  $g$  and  $h$ , we do not meet any inversion problem.

It is also worth mentioning that, as a result of this one-to-one relationship between  $\rho_Z$  and  $\rho_T$ , we have  $\rho_T = 0$  if and only if  $\rho_Z = 0$ . This further suggests that, for pairs of observations drawn from a TGH random field, not being correlated implies independence. This can be explained by the fact that the Gaussian copula structure remains the same after a monotonic transformation.

Because the mean and variance of the standard TGH random field are determined by  $g$  and  $h$ , we need a more general TGH random field,  $\tilde{T}(\mathbf{s})$ , defined by  $\tilde{T}(\mathbf{s}) = \xi + \omega T(\mathbf{s})$ , where  $\xi \in \mathbb{R}$  is a location parameter and  $\omega > 0$  is a scale parameter. We want to generate data at  $n$  spatial locations,  $\mathbf{s}_1, \dots, \mathbf{s}_n$ , from a TGH random field with mean zero and a given Matérn covariance function. Figure 3(b) shows a realization of  $n = 576$  perturbed grid points for the experiment design described in detail later in Section 3.2. For given values of  $g$  and  $h$ , we first find the mean  $\mu_T$  and variance  $\sigma_T^2$  of the standard TGH random field by Equations 3 and 4, respectively. Then, given a desired covariance function  $C(d)$ , with  $C(0) = \sigma_1^2$ , we solve for  $\rho_Z(d_{ij})$  with  $C_T(d_{ij}) = \frac{\sigma_T^2}{\sigma_1^2} C(d_{ij})$  in Equation 5 for every pairwise distance,  $d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|, i, j = 1, \dots, n$ . Next, we generate data,  $\mathbf{Z} = \{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}^\top$ , from the standard Gaussian random field with the correlation matrix,  $\rho_Z(d_{ij})$ , we just determined. Then, by applying the TGH transformation (2) to  $\mathbf{Z}$ , we get  $\mathbf{T} = \tau_{g,h}(\mathbf{Z}) = \{T(\mathbf{s}_1), \dots, T(\mathbf{s}_n)\}^\top$ , which is from a TGH random field with mean  $\mu_T$  and covariance function  $C_T(d) = \frac{\sigma_T^2}{\sigma_1^2} C(d)$ . Finally, we subtract the mean and rescale the data by taking  $\tilde{T}(\mathbf{s}_i) = \frac{\sigma_1}{\sigma_T} \{T(\mathbf{s}_i) - \mu_T\}, i = 1, \dots, n$ . In this way,  $\tilde{\mathbf{T}} = \{\tilde{T}(\mathbf{s}_1), \dots, \tilde{T}(\mathbf{s}_n)\}^\top$  is a realization from a TGH random field with mean 0 and the given isotropic covariance function  $C(d)$ .

Figure 4 shows realizations in the unit square of TGH random fields,  $\tilde{\mathbf{T}}$ , with exponential covariance with  $\sigma^2 = 1, \phi = 0.1$  for different values of  $g$  and  $h$  generated by the procedure described earlier and the corresponding histogram overlaid with the true density function. This covariance model has an effective range of about 0.3, the distance after which pairs of observations have a correlation less than 0.05. We can clearly see that a larger  $h$  results in heavy-tailed data and a larger  $g$  leads to asymmetric data (right skewed when  $g > 0$ ). Each realization contains 10,000 observations. The slight discrepancy of the histogram from the



**FIGURE 4** Top: Realizations in the unit square of Tukey  $g$ -and- $h$  random fields with an exponential covariance function, that is, Matérn with  $\theta = (1, 0.1, 0.5)^T$ , for different values of  $g$  and  $h$ . Bottom: Corresponding histograms overlaid with theoretical density functions (red curves). Each panel contains 10,000 observations

true density function is due to the fact that observations are not independent. Also worth noticing, with the help of the histogram plots, is that, with large values of  $h$  and when there are more extreme values, the non-extreme values tend to be closer to 0.

### 3 | MONTE CARLO SIMULATION STUDY

#### 3.1 | Motivation

The ML and least squares (LS) methods are described as two canonical methods for parameter estimation in textbooks on spatial statistics (e.g., Gelfand, Diggle, Guttorp, & Fuentes, 2010). The MLE has many appealing properties and has been used widely by the statistical community. However, for the ML method, the distribution of the spatial process has to be assumed in advance so that the likelihood function can be written explicitly. On the other hand, LS methods are simple and visually appealing and avoid distribution assumptions, but they are not optimal in a statistical sense. LS methods are especially popular among the geostatistical community. They are sometimes suspected to outperform MLEs when the distribution of the spatial process is misspecified, but no simulation study has been carried out to verify that yet.

In spatial statistics, there are at least two frameworks of asymptotics. The more familiar one is increasing-domain asymptotics, where as sample size grows, the minimum distance between data locations is bounded away from zero and observations are made from an unbounded spatial domain. The other framework is the so-called infill or fixed-domain asymptotics in which data are observed more densely inside a bounded region.

Mardia and Marshall (1984) established consistency and asymptotic normality of the MLE under the assumption of a Gaussian random field with increasing-domain asymptotics. Fixed domain asymptotic results for the MLE with parameters in a Matérn covariance model under Gaussian assumption were obtained by Ying (1991) and Zhang (2004). They noted that, under fixed-domain asymptotics, not all parameters in a Matérn covariance model can be estimated consistently by the ML method.

Among the LS methods, the WLS method is a compromise between the strong assumption of independent and identically distributed data for the ordinary LS and intensive computational effort for the generalized LS. In using the WLS method to fit a parametric covariance model, different choices of weights were proposed and comparisons of different LS methods were made; see Cressie (1985), Zimmerman and Zimmerman (1991), and Müller (1999). Consistency and asymptotic normality results for the LS methods were developed by Lahiri, Lee, and Cressie (2002) under certain regularity conditions in different asymptotic frameworks.

We run Monte Carlo simulations to investigate several issues related to estimation of the Matérn covariance parameters. First, and the main concern of this paper, is the issue of how a misspecified Gaussian likelihood influences the MLE, given data that are generated from a zero-mean TGH random field with different values for  $g$  and  $h$ . We want to know if the WLS method alleviates the effect because it is not based on any distributional assumption. We also want to see the difference between fixed-domain and increasing-domain asymptotics for the MLE under misspecification. As a byproduct, we observe the estimation performance when different numbers of parameters are estimated. Additionally, we also show results for the case when a nugget effect exists as well as results for prediction.

#### 3.2 | Simulation design

In this study, we first simulate spatial observations from TGH random fields with the Matérn covariance function by the procedure described in Section 2. The spatial data are generated on perturbed grid points as used by Kaufman, Schervish, and Nychka (2008). Using the perturbed grid points helps to avoid numerical problems when observation locations are very close. The perturbed grid points are generated by the following procedure: First, we generate  $n_x \times n_x$  regular grid points with unit lengths; then, we add random noise uniformly distributed on  $[-0.4, 0.4]$  to each coordinate. In this way, the minimum distance between two points will be greater than 0.2. We use part of the observations for estimation and the other part for validation of the prediction result. Next, on the basis of the estimation part of the simulated data, we use the ML method as well as the WLS method to estimate the parameters in the Matérn covariance model. Finally, we compute Gaussian kriging predictors for the prediction locations by plugging in the MLE for covariance parameters.

To incorporate the two frameworks of asymptotics into a single simulation study, we notice that the joint distribution of random variables generated from a TGH random field with the Matérn covariance model are the same if the range parameter  $\phi$  and  $n_x$  vary proportionally. For example, the same realizations as in Figure 4 can be generated in  $[0, 10] \times [0, 10]$  with  $\phi = 1$ . Thus, we can always scale the spatial observations into a unit square with a scaled range parameter to investigate the fixed-domain asymptotic properties. A similar idea was described by Zhang and Zimmerman (2005). Table 1 includes different

**TABLE 1** Equivalent effective range when observations generated by the TGH random field with the Matérn covariance model are scaled to a unit square with the range parameter  $\phi$  in the perturbed grid points for different values of  $n_x$

$\nu=1$	$\phi=0.6$	$\phi=1.2$	$\phi=2.4$
$\nu=0.5$	$\phi=0.8$	$\phi=1.6$	$\phi=3.2$
$n_x = 12, n = 100$	0.2	0.4	0.8
$n_x = 24, n = 532$	0.1	0.2	0.4

sets of values for  $n_x$ ,  $\phi$ , and  $\nu$  selected in our simulation study for the Matérn covariance model. We consider  $\nu = 0.5$  and 1 in the Matérn model, corresponding to the exponential and Whittle covariance functions, respectively. The values in the table are the corresponding effective range if we scale the spatial observations to a unit square. The scaled effective range of 0.1, 0.2, 0.4, and 0.8 represents weak, mild, mild, and strong spatial dependence, respectively. In this way, by examining results vertically as  $n_x$  grows from 12 to 24, we have an increasing-domain framework. Results for fixed domain can be obtained by checking the results diagonally, as the arrows in Table 1 demonstrate. In all settings, we fix  $\sigma^2 = 1$ .

For each value of  $n_x$ , we randomly select 44 locations among the perturbed grid points for prediction and the remaining  $n = n_x^2 - 44$  for estimation. Figure 3(b) shows one realization of the 576 perturbed grid points when  $n_x = 24$ , with 532 locations for estimation and 44 locations for prediction.

We consider different values for  $g$  and  $h$  in the TGH random fields. For  $g = 0$  and  $h > 0$ , the TGH random field has a Pareto-like marginal distribution. With a fixed  $g = 0$  and varying  $h = 0, 0.1, 0.3$ , we can examine the effect of using Gaussian likelihood on the estimation when the data have heavy tails. For  $h = 0$  and  $g > 0$ , the TGH random field becomes a shifted log-Gaussian random field. With  $h = 0, g = 0, 0.2, 0.4$ , the influence on the estimation when the data are asymmetric can be checked. We also consider a combination of both skewness and elongation by setting  $g = 0.2, h = 0.1$ .

We consider three different settings in which different numbers of parameters are estimated. For convenience, we denote the three settings by [1], [2], and [3] as follows:

- [1]  $\theta = \phi$ : to estimate only the range parameter;
- [2]  $\theta = (\sigma^2, \phi)^\top$ : to estimate the variance and range parameters jointly;
- [3]  $\theta = (\sigma^2, \phi, \nu)^\top$ : to estimate the variance, range, and smoothness parameters together.

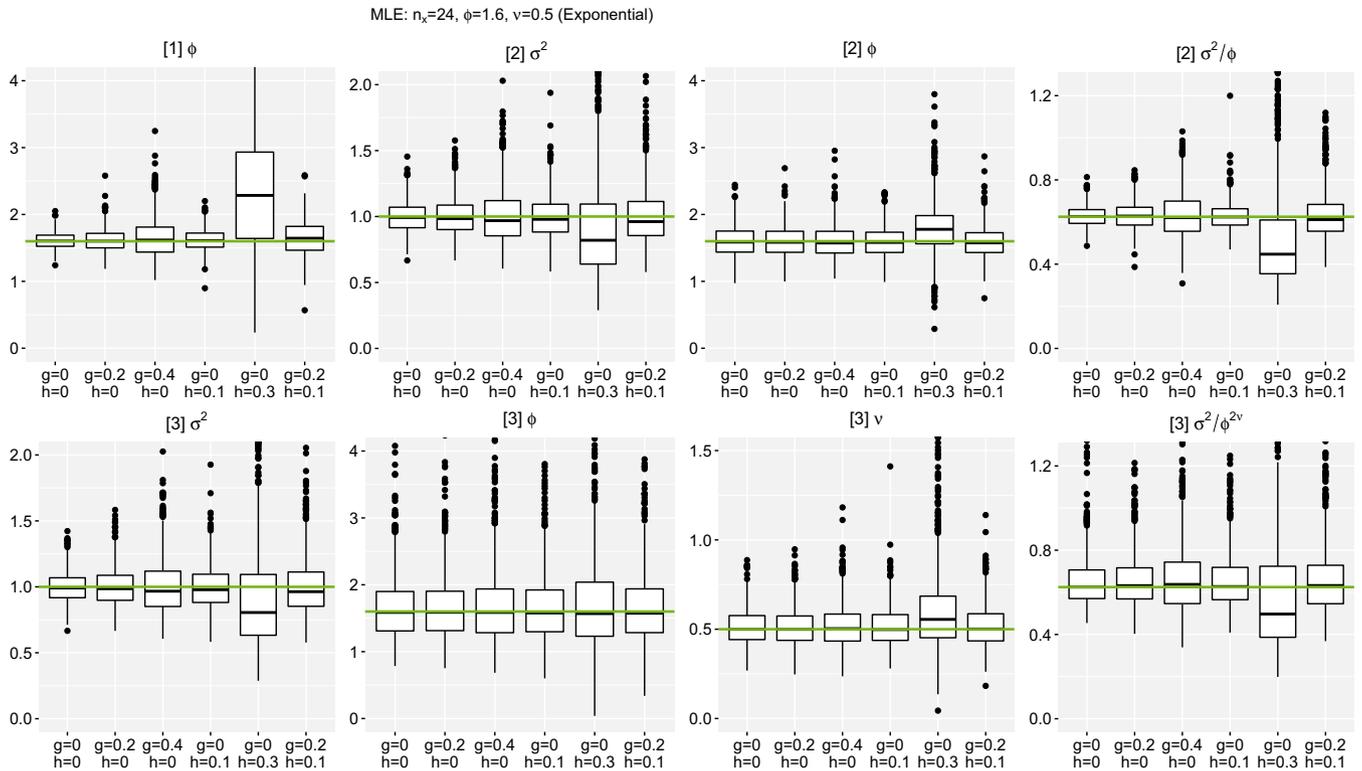
The data simulation, parameter estimation, and prediction steps are implemented in R (R Development Core Team, 2016). We run 1,000 replications for each setting to assess the estimation results.

To display the estimation results, previous authors used tables and/or boxplots to summarize values of the estimated parameters individually. In this paper, in addition to boxplots, we adopt an informative exploratory method to illustrate the estimation of the covariance function per se by the functional boxplot (Sun & Genton, 2011). This method is an extension of the boxplot to functional data, where the basic unit of information is a function instead of a single number. With a center-outward order based on the modified band depth (López-Pintado & Romo, 2009), the median, 50% central region, and outliers of the functional data can be found. In this way, we can look at the estimation of the covariance function as a whole instead of examining the estimation of each parameter separately. The effect of non-Gaussianity on the estimation illustrated by the boxplots and functional boxplots are consistent for both the exponential and Whittle covariance functions and among different range parameters. Unless otherwise specified, the boxplot and functional boxplot results shown in the next subsections are all for  $n_x = 24, \sigma^2 = 1, \phi = 1.6, \nu = 0.5$ , which is an exponential covariance function with an effective range of 0.2 when scaled to the unit square. Under this mild spatial dependence, Irvine et al. (2007) and Zimmerman and Zimmerman (1991) observed that the ML method outperforms the REML approach and the method of moments.

### 3.3 | Gaussian ML estimation

The boxplots in Figure 5 show estimation results of the MLE for all the three different settings [1], [2], and [3]. Ying (1991) proved that, for the exponential-without-nugget model, consistency results only hold for  $\sigma^2/\phi$  but not for individual parameters under infill asymptotics. Zhang (2004) extended this result to the Matérn model, stating that, when estimating three parameters together, only  $\sigma^2/\phi^{2\nu}$  can be estimated consistently under infill asymptotics. Therefore, we also display boxplots of this ratio value,  $\sigma^2/\phi$ , for setting [2], and  $\sigma^2/\phi^{2\nu}$ , for setting [3]. For the functional boxplot, only the results for setting [2] are shown in Figure 6 because similar results are obtained in settings [1] and [3].

From the boxplots in Figure 5 and the functional boxplots in Figure 6, we can see that the performance of the MLE is satisfactory, especially when the underlying distribution is truly Gaussian, and even when all three parameters are estimated

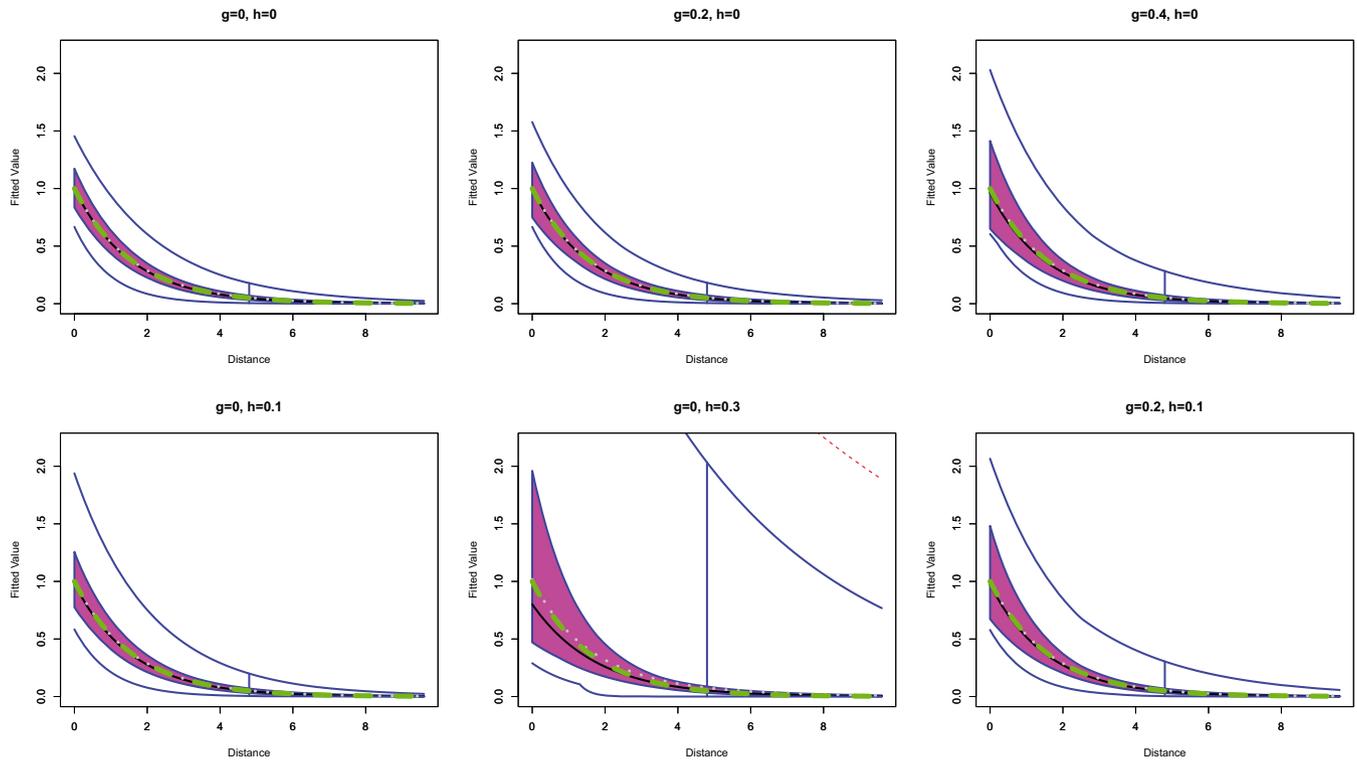


**FIGURE 5** Boxplots of the MLE with Gaussian likelihood for the Matérn covariance parameters  $\sigma^2$ ,  $\phi$ ,  $\nu$  as well as the ratio  $\sigma^2/\phi^{2\nu}$  in the three estimation settings [1], [2], and [3], when data are generated from a Tukey  $g$ -and- $h$  random field with different values of  $g$  and  $h$

together because the sample size is quite adequate ( $n = 532$ ). We can also observe from the functional boxplots that, as the value of  $g$  or  $h$  becomes larger, the estimated covariance function has larger variability and, as  $h$  becomes larger, the estimation is biased and tends to underestimate the covariance function.

We also obtain results under weak and strong spatial correlations (plots not shown). We observe that the quality of estimation worsens when spatial dependence strengthens. With strong spatial dependence, the ML method tends to underestimate the covariance function even with Gaussian random fields. This underestimation with strong spatial dependence is also mentioned by Irvine et al. (2007). When spatial correlation is strong, there is a lack of information from the spatial observations because the influence of  $\sigma^2$  on fluctuations of the random field is small when the longest distance between observations is not large compared to the effective range. This adds to the difficulty of parameter estimation. On the other hand, when spatial correlation is weak, observations that are separated by distances larger than the effective range can be made. These almost independent observations ameliorate parameter estimation.

Figure 7 compares the two asymptotic frameworks for cases of the arrows shown in Table 1. The top row is for  $n_x = 12$  and the middle and bottom rows are both for  $n_x = 24$ , but the sample size grows under increasing domain (middle) and fixed domain (bottom). The boxplots are for different values of  $g$  and  $h$  under setting [2], and the functional boxplots are for  $g = 0.2, h = 0.1$ . Notice that, even though the true range parameters are different in the bottom row from those in the top and middle rows, the y-axis in the boxplots and the x-axis in the functional boxplots are adjusted proportionally, so that results are comparable by ignoring the axis labels. We can see that, for different values of  $g$  and  $h$ , the variance of the estimation for  $\sigma^2$  and  $\phi$  is larger for the fixed domain than for the increasing domain. Also, the estimated covariance function has larger variance under fixed domain, which is obvious by comparing the functional boxplots. However, for the ratio  $\sigma^2/\phi$ , which can be estimated consistently under both asymptotic frameworks, the boxplots show no significant difference for the two asymptotic frameworks. These findings were observed by Zhang (2004) in the Gaussian case. The essential difference between fixed-domain and increasing-domain asymptotics for estimation of the individual parameters is the strength of spatial correlation within the observation area. Under increasing domain, spatial correlation relative to the observation area decreases as the sample size grows. In Figure 7, the difference between the estimation under increasing domain and fixed domain is equivalent to when the scaled effective ranges are 0.1 and 0.2, as shown in Table 1.



**FIGURE 6** Functional boxplots of estimated covariance function by the maximum likelihood method in setting [2] for different values of  $g$  and  $h$  (green dashed lines show the true covariance function, and gray dotted lines show the pointwise mean of the estimated covariance function)

### 3.4 | WLS estimation

The LS method usually uses the variogram rather than the covariance function. With a parametric variogram model,  $\gamma(d; \theta)$ , the WLS estimator is found from

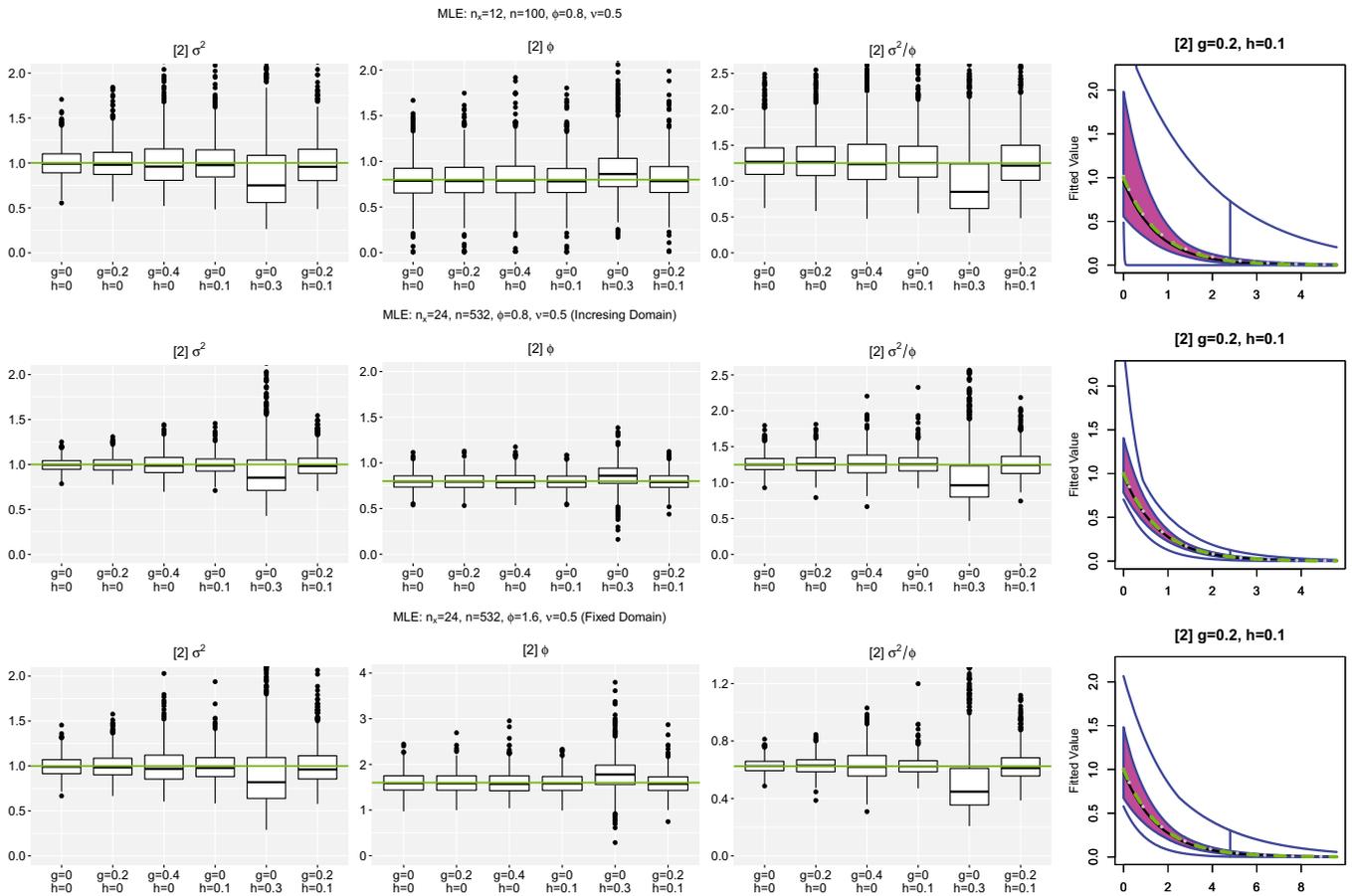
$$\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^K w_k \{\hat{\gamma}_k - \gamma(d_k; \theta)\}^2,$$

where  $\hat{\gamma}_k$  is the empirical variogram at the  $k$ th bin,  $d_k$  is the center of the  $k$ th bin,  $K$  is the total number of bins, and  $w_k$  is the weight.

The performance of the WLS estimator depends heavily on how the empirical variogram is defined, especially the maximum distance and the number of bins. The weight function provides another source of variation for the WLS method. Different weights for the WLS estimate of the variogram have been suggested by Cressie (1985), Zhang, van Eijkeren, and Heemink (1995), and Genton (1998). McBratney and Webster (1986), Gotway (1991), and Zhang et al. (1995) compared these weights via numerical studies.

The weight suggested by Cressie (1985) is a popular choice:  $w_k = \frac{N_k}{\gamma^2(d_k; \theta)}$ , where  $N_k$  is the number of pairs falling into the  $k$ th bin. However, this weight is derived with a Gaussian model. To avoid distribution assumption, in our simulation study, we use the weights  $w_k = N_k$ . Our simulation shows that, with the Cressie weights, the estimation for  $\phi$  is biased even when data are truly Gaussian. By using the weights  $w_k = N_k$ , the estimation bias is much smaller, but the variance is larger than by using the Cressie weights. The empirical variogram is estimated by the classical method-of-moments estimator in which the maximum distance equals half of the greatest distance from the sample and  $K = 13$ .

The boxplots in Figure 8 and the functional boxplots in Figure 9 show that the WLS estimators (under our choice of specific weight and tuning parameters) are quite biased. Compared with the MLE, WLS estimators have larger variances. As values of  $g$  and  $h$  increase, the WLS estimations are affected similarly to the MLE. This shows that, at least with data from the TGH random fields, the WLS estimator does not benefit from distributional assumptions not being made.



**FIGURE 7** Comparison of the two asymptotic frameworks: From the top row to the middle/bottom row, sample size grows under increasing/fixed domain with boxplots of the MLE for the Matérn covariance parameters  $\sigma^2$ ,  $\phi$  and the ratio  $\sigma^2/\phi$  in setting [2] and functional boxplots for the case  $g = 0.2$  and  $h = 0.1$

### 3.5 | Estimation with nugget effect

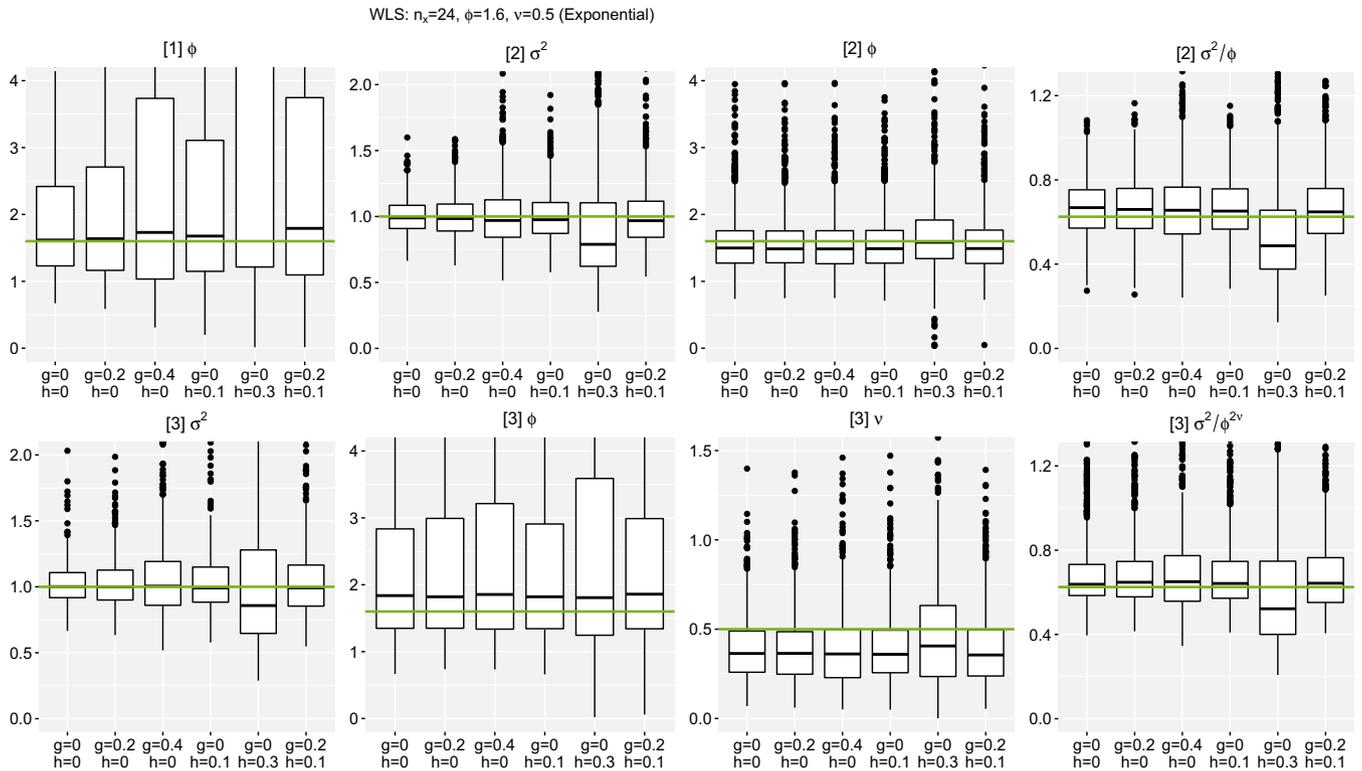
The nugget effect captures measurement error in repeated observations at any single site or it captures the microscale variability occurring at such a small scale that cannot be distinguished from the effect of measurement error. With nugget effect  $\epsilon^2$ , the covariance function becomes discontinuous at 0 and the Matérn-with-nugget covariance function is

$$C_{\mathcal{M}}(d; \theta) = \begin{cases} \sigma^2 + \epsilon^2, & d = 0, \\ \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{d}{\phi}\right)^{\nu} K_{\nu}\left(\frac{d}{\phi}\right), & d > 0, \end{cases}$$

where  $\theta = (\sigma^2, \phi, \nu, \epsilon^2)^T$ .

Irvine et al. (2007) performed a thorough study of the effect of different sampling schemes and different levels of spatial correlation (measured in terms of the nugget-to-sill ratio,  $\epsilon^2/(\epsilon^2 + \sigma^2)$ , range parameter, or some combination of these) on the ML and REML estimates.

Here, we examine the estimation of the extra nugget effect parameter,  $\epsilon^2$ , by a simulation study and present results for the true nugget parameter  $\epsilon^2 = 0.1$  with the other parameters remaining the same. The overall estimation performance is similar to the estimation performance without the nugget effect. We notice that, even with an additional parameter to be estimated, the ML method performs quite well. Figure 10(b) shows a realization of the TGH random field with the Matérn-with-nugget covariance function with  $g = 0.2, h = 0.1$ . Figure 10(a) shows boxplots of the MLE in the Matérn-with-nugget model for all three settings. Figure 10(c) shows the functional boxplot of the estimated covariance function by the ML method under setting [2] with  $g = 0.2, h = 0.1$ . We can see that the MLE performs very well except for the case when  $h$  is very large.



**FIGURE 8** Boxplots of the WLS estimator for the Matérn covariance parameters  $\sigma^2$ ,  $\phi$ ,  $v$  in the three estimation settings [1], [2], and [3], when data are generated from a Tukey  $g$ -and- $h$  random field with different values of  $g$  and  $h$

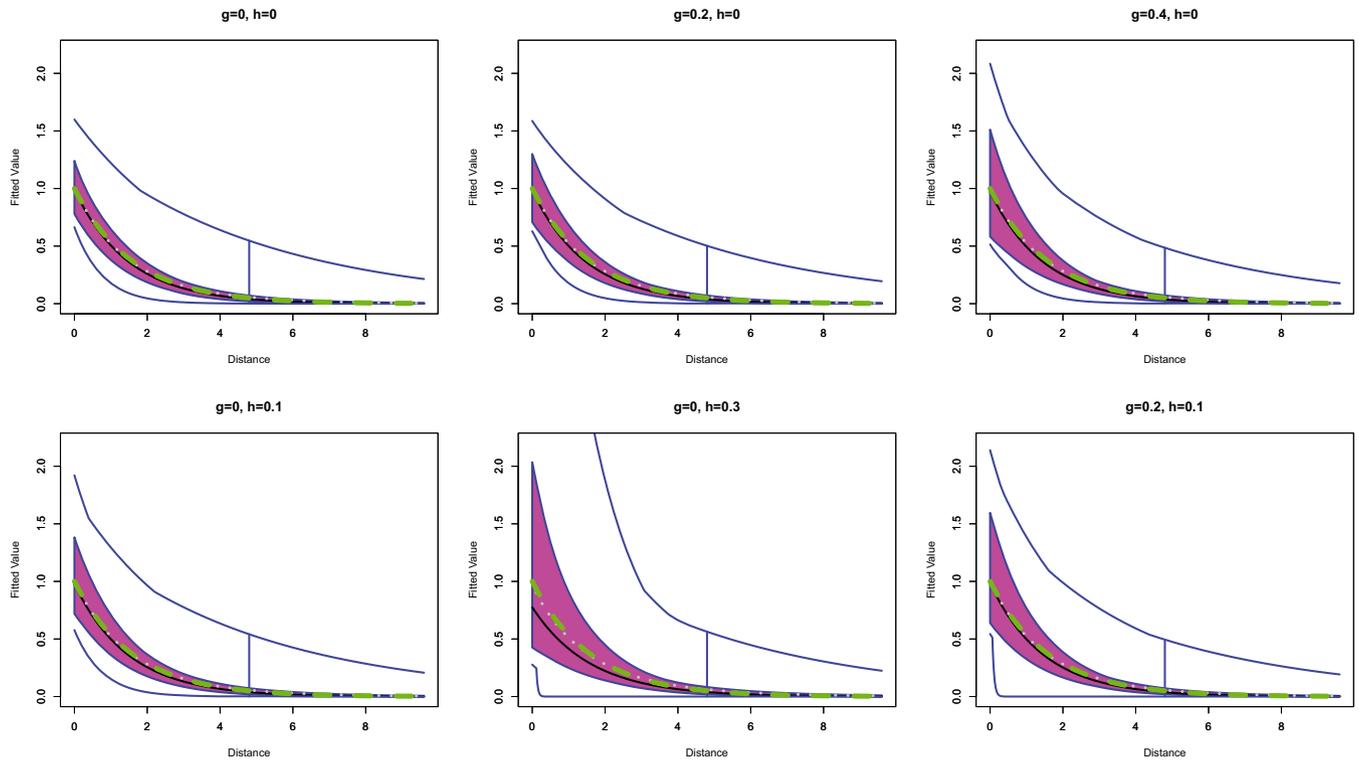
### 3.6 | Simulation study for prediction

Prediction is the final stage and often the ultimate objective of statistical analysis for a spatial data set. A class of linear interpolation methods called *kriging* is used to predict the values of  $Y(\mathbf{s}^*)$  at new locations  $\mathbf{s}^*$ . The kriging predictor is a linear combination of the observed data and the best linear unbiased predictor when the underlying process is Gaussian. In this section, we present simulation results of the performance of the Gaussian kriging predictor given data drawn from the TGH random field and where the MLEs of the covariance parameters based on Gaussian likelihood are plugged in.

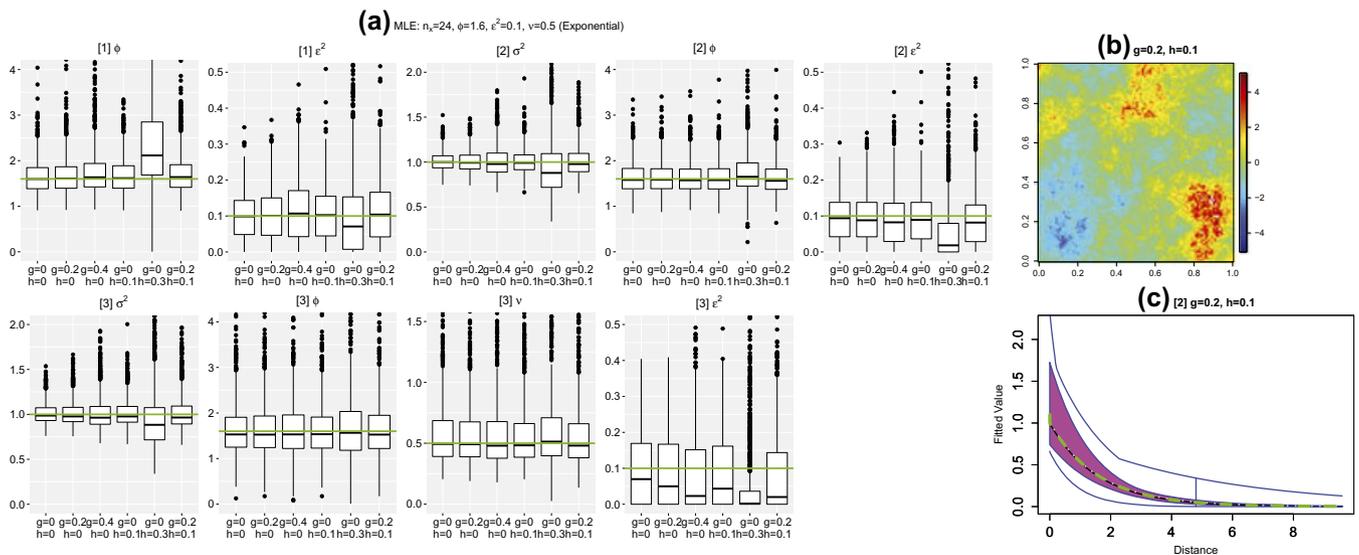
We illustrate results only for estimation under setting [2] (estimate  $\sigma^2$  and  $\phi$  together) and with true parameters  $n_x = 24$ ,  $\sigma^2 = 1$ ,  $\phi = 1.6$ ,  $v = 0.5$ , because we obtained similar results for all settings.

We evaluate the kriging performance by computing the mean square prediction error,  $MSPE^j = \frac{1}{44} \sum_{i=1}^{44} \left\{ \hat{y}(\mathbf{s}_i^j) - y(\mathbf{s}_i^j) \right\}^2$ , for each replication  $j = 1, \dots, 1000$ . Figure 11(a) shows the boxplots of MSPE for kriging predictions in our simulation study. The boxplots of MSPE show an interesting phenomenon that the median of MSPE is the smallest when  $g = 0, h = 0.3$ , where the estimation behavior is the worst as shown in Section 3.2. On the other hand, the variance of the MSPE is extremely large with  $g = 0, h = 0.3$ . This can be explained by the fact that, for the TGH random fields, if we keep the variance unchanged, as  $h$  grows larger, the marginal density becomes concentrated on values closer to 0, which can be seen from the histogram of the realization for  $g = 0, h = 0.3$  in Figure 4. Therefore, observations from the TGH random fields are mostly composed of very small values, except for a few extreme points. This motivates us to adjust the MSPE criteria to define the mean relative prediction error as  $MRPE^j = \frac{1}{44} \sum_{i=1}^{44} \left\{ \frac{\hat{y}(\mathbf{s}_i^j) - y(\mathbf{s}_i^j)}{y(\mathbf{s}_i^j)} \right\}^2$ , for  $j = 1, \dots, 1000$ . Figure 11(b) shows that, by accounting for this relative effect, the Gaussian kriging predictor behaves best with data generated from a Gaussian random field. Overall, the kriging predictor performs quite well under different values of  $g$  and  $h$ ; that is, it works well with data with different levels of skewness and heavy tails even when the covariance parameters have been estimated from a misspecified Gaussian likelihood.

We also notice that, as spatial dependence grows stronger, the kriging predictor performs better, even though the MLE behaves worse as shown in Section 3.1. This can be explained by the fact that strong spatial dependence provides a lot of information for spatial interpolation.



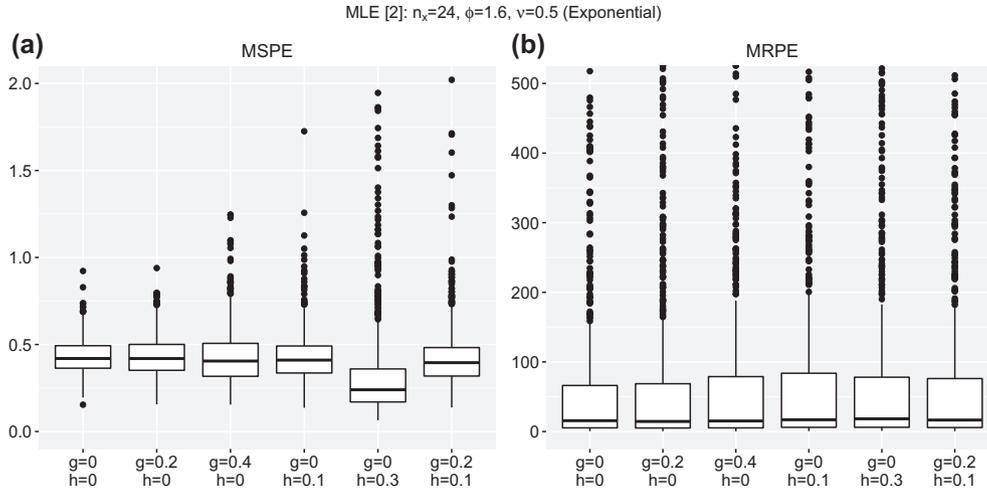
**FIGURE 9** Functional boxplots of estimated covariance functions by the WLS method in setting [2] for different values of  $g$  and  $h$  (green dashed lines show the true covariance function, and gray dotted lines show the pointwise mean of the estimated covariance function)



**FIGURE 10** Exponential covariance with the nugget effect  $\epsilon^2 = 0.1$ . (a) Boxplots for MLE; (b) realization for  $g = 0.2, h = 0.1$ ; (c) functional boxplot for MLE under setting [2] with  $g = 0.2, h = 0.1$

## 4 | DISCUSSION

In this paper, we described a procedure to generate data from a trans-Gaussian random field with a given covariance matrix based on the TGH random field. We made use of this procedure to run simulation studies to examine the performances of estimation and prediction based on Gaussian likelihood when data are not truly Gaussian but from the TGH random field with a given Matérn covariance function.



**FIGURE 11** (a) Mean square prediction error (MSPE) and (b) mean relative prediction error (MRPE) for the Gaussian kriging predictor with data generated from the TGH random fields with different values of  $g$  and  $h$

The simulation study on estimation shows the superiority of the Gaussian ML method over the WLS method when the true spatial process is a TGH random field with tails not too heavy. Consistency and asymptotic normality for the MLE based on Gaussian likelihood when the underlying random field is in fact non-Gaussian with a given Matérn covariance function can be derived from the theory of unbiased estimating equations. Let  $\mathbf{G}(\mathbf{Y}; \boldsymbol{\theta})$  be a vector-valued estimating equation. If  $E\{\mathbf{G}(\mathbf{Y}; \boldsymbol{\theta})\} = \mathbf{0}$  for all  $\boldsymbol{\theta}$ , it is said to be an unbiased estimating equation. The solution to an unbiased estimating equation under increasing-domain asymptotics is consistent and asymptotically normal. We show that the derivative of the log-likelihood function (2),  $\mathbf{G}_i = \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i}$ , is an unbiased estimating equation when observations  $\mathbf{Y}$  are from any random field with a given Matérn covariance function:

$$\begin{aligned} E\{\mathbf{G}_i(\mathbf{Y}; \boldsymbol{\theta})\} &= E\left\{\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i}\right\} = -\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right\} + \frac{1}{2}E\left[\text{tr}\left\{\mathbf{Y}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{Y}\right\}\right] \\ &= -\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right\} + \frac{1}{2}\text{tr}\left[E\left\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{Y}\mathbf{Y}^\top\right\}\right] \\ &= -\frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\right\} + \frac{1}{2}\text{tr}\left\{\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\frac{\partial \boldsymbol{\Sigma}(\boldsymbol{\theta})}{\partial \theta_i}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{\Sigma}(\boldsymbol{\theta})\right\} = 0, \end{aligned}$$

because  $E(\mathbf{Y}\mathbf{Y}^\top) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$  is independent of the underlying process when the covariance structure is given. This is a general result for any random field with a given covariance structure and is not confined only to the TGH random fields. It thus follows that the MLE with misspecified Gaussian likelihood is still consistent and asymptotically normal given data from a non-Gaussian random field with a given Matérn covariance function.

The asymptotic result confirms that the MLE still possesses preferable properties even when the likelihood is misspecified. Simulation results on the Gaussian kriging predictor show satisfactory performance based on Gaussian likelihood inference even when data are not from a Gaussian random field. Nevertheless, in this paper, we only used the TGH random fields to model data from non-Gaussian random fields in our simulation study. Other models of non-Gaussian random fields, such as those mentioned in the introduction, may be used in similar investigations for comparisons.

## ACKNOWLEDGEMENT

This research was supported by the King Abdullah University of Science and Technology (KAUST).

## REFERENCES

- Allcroft, D. J., & Glasbey, C. A. (2003). A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. *Journal of the Royal Statistical Society, Series C*, 52, 487–498.
- Arnold, B. C., & Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician*, 49, 34–38.
- Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17, 563–586.
- Cressie, N. (1993). *Statistics for spatial data*. New York, NY: Wiley.

- De Oliveira, V. (2006). On optimal point and block prediction in log-Gaussian random fields. *Scandinavian Journal of Statistics*, 33, 523–540.
- De Oliveira, V., Kedem, B., & Short, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, 92, 1422–1433.
- Fonseca, T. C. O., & Steel, M. F. J. (2011). Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika*, 98, 761–774.
- Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. Boca Raton, FL: Chapman & Hall/CRC.
- Genton, M. G. (1998). Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Mathematical Geology*, 30, 323–346.
- Genton, M. G., & Zhang, H. (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics*, 3(2), 171–179.
- Gotway, C. A. (1991). Fitting semivariogram models by weighted least squares. *Computers & Geosciences*, 17, 171–172.
- Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 10, 87–102.
- Guttorp, P., & Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93(4), 989–995.
- Irvine, K. M., Gitelman, A. I., & Hoeting, J. A. (2007). Spatial designs and properties of spatial correlation: Effects on covariance estimation. *Journal of Agricultural, Biological, and Environmental Statistics*, 12(4), 450–469.
- Johns, C. J., Nychka, D., Kittel, G. F., & Daly, C. (2003). Infilling sparse records of spatial fields. *Journal of the American Statistical Association*, 98, 796–806.
- Jones, M. C. (2015). On families of distributions with shape parameters. *International Statistical Review*, 83(2), 175–192.
- Jones, M. C., & Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96(4), 761–780.
- Kaufman, C. G., Schervish, M. J., & Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484), 1545–1555.
- Krupskii, P., Huser, R., & Genton, M. G. (2017). Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, 112. to appear.
- Lahiri, S. N., Lee, Y., & Cressie, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning & Inference*, 103, 65–85.
- López-Pintado, S., & Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104, 718–734.
- Marchenko, Y. V., & Genton, M. G. (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics*, 21, 318–340.
- Mardia, K. V., & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71, 135–146.
- McBratney, A. B., & Webster, R. (1986). Choosing functions and fitting them to sampling estimates. *Journal of Soil Science*, 37, 617–639.
- Müller, W. G. (1999). Least-squares fitting from the variogram cloud. *Statistics & Probability Letters*, 43, 93–98.
- R Development Core Team (2016). R: A language and environment for statistical computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. New York: Springer.
- Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20, 316–334.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
- Wallin, J., & Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, 42, 872–890.
- Xu, G., & Genton, M. G. (2017). Tukey *g*-and-*h* random fields. *Journal of the American Statistical Association*, 112. to appear.
- Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis*, 36, 280–296.
- Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465), 250–261.
- Zhang, H., & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, 21, 33–47.
- Zhang, H., & Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92(4), 921–936.
- Zhang, X., van Eijkeren, J., & Heemink, A. (1995). On the weighted least-squares method for fitting a semivariogram model. *Computers & Geosciences*, 21, 605–608.
- Zimmerman, D. L., & Zimmerman, M. B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, 33, 77–91.

**How to cite this article:** Yan Y, Genton MG. Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function. *Environmetrics*. 2018;29:e2458. <https://doi.org/10.1002/env.2458>