Tukey g-and-h Random Fields

Ganggang Xu & Marc G. Genton

Tukey g-and-h Random Fields

Ganggang Xu[1] and Marc G. Genton[2]

Abstract

We propose a new class of trans-Gaussian random fields named Tukey g-and-h (TGH) random fields to model non-Gaussian spatial data. The proposed TGH random fields have extremely flexible marginal distributions, possibly skewed and/or heavy-tailed, and, therefore, have a wide range of applications. The special formulation of the TGH random field enables an automatic search for the most suitable transformation for the dataset of interest while estimating model parameters. Asymptotic properties of the maximum likelihood estimator and the probabilistic properties of the TGH random fields are investigated. An efficient estimation procedure, based on maximum approximated likelihood, is proposed and an extreme spatial outlier detection algorithm is formulated. Kriging and probabilistic prediction with TGH random fields are developed along with prediction confidence intervals. The predictive performance of TGH random fields is demonstrated through extensive simulation studies and an application to a dataset of total precipitation in the south east of the United States.

Some key words: Continuous Rank Probability Score; Heavy tails; Kriging; Log-Gaussian random field; Non-Gaussian random field; PIT; Probabilistic prediction; Skewness; Spatial outliers; Spatial statistics; Tukey g-and-h distribution.

Short title: Tukey g-and-h Random Fields

[1]Department of Mathematical Sciences, Binghamton University, Binghamton, NY 13902, USA.
E-mail: gang@math.binghamton.edu
[2]CEMSE Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: marc.genton@kaust.edu.sa

1   Introduction

Gaussian random fields are among the most popular tools for analyzing spatial data because they can be simply characterized by a mean structure and a valid covariance function. Various parametric and nonparametric covariance functions have been studied in the literature and have proven useful in practice, which further enhances the modeling power of Gaussian random fields. Unfortunately, Gaussianity is a strong assumption that is rarely met in reality. Data collected from a wide range of applications often display strong skewness and heavy tails in their distributions; for example, wind speed data (Zhu and Genton, 2012), temperature data (North et al., 2011), and precipitation data (Marchenko and Genton, 2010), to mention but a few.

As a motivating example, we consider total precipitation data (in centimeters) in November 1994 over the southeastern United States. Longitude in the region under study ranges from -91.23 to -75.55 and latitude ranges from 25.02 to 37.38. The data are available at *http://www.image.ucar.edu/Data/US.monthly.met*. Although there are 991 observations recorded in this region, as illustrated in Figure 3(a), the locations with observed values still appear quite sparse on the map. It is of great interest to create a precipitation map with a much finer resolution based on the observed data from monitoring stations, which can be useful for revealing short and long-term climate patterns. There are ongoing projects to create such climate maps for precipitation and temperature patterns on a regular basis, see, for example, the PRISM climate group (*http://www.prism.oregonstate.edu*). To this end, it is critical to build geostatistical models that can explain spatial or spatio-temporal variations in precipitation well. However, precipitation data usually do not have a marginal normal distribution (Allcroft and Glasbey, 2003; Johns et al., 2003;) and special care must be taken to accommodate the skewness and potential outliers in the data.

Various approaches have been proposed to model non-Gaussian geostatistical data, such as

skew-Gaussian processes (Zhang and El-Shaarawi, 2010; Genton and Zhang, 2012; Kim and Mallick, 2012; Rimstad and Omre, 2014), scale mixing Gaussian random fields (Palacios and Steel, 2006; Fonseca and Steel, 2011), log-skew-elliptical random fields (Marchenko and Genton, 2010), T-distributed random fields (Røislien and Omre, 2006), trans-Gaussian random fields (Cressie, 1993; De Oliveira et al., 1997; Allcroft and Glasbey, 2003; Butler and Glasbey, 2008), spatial copula models (Gräler, 2014) and non-Gaussian Matérn fields (Wallin and Bolin, 2015). Of these methods, one particularly appealing approach is the trans-Gaussian random field obtained by applying some non-linear transformations to the original data. Typically, $Y(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^q$, $q \geq 1$, is said to be a trans-Gaussian random field if there exists a transformation, $\psi(\cdot)$, so that

$$\psi\{Y(\mathbf{s})\} = \xi + \mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + V(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{1}$$

where $\xi \in \mathbb{R}$ is a location parameter, $\mathbf{X}(\mathbf{s}) \in \mathbb{R}^p$ is a vector of some observed covariates at location $\mathbf{s}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of regression parameters, $V(\mathbf{s})$ is a Gaussian random field with mean 0 and some covariance function, and $\varepsilon(\mathbf{s})$ is a Gaussian white noise process independent of $V(\mathbf{s})$ with mean 0 and variance σ_e^2. Common choices of transformation $\psi(\cdot)$ include log-normal (De Oliveira, 2006), square-root (Johns et al., 2003), Box-Cox (De Oliveira et al., 1997), and power transformations (Allcroft and Glasbey, 2003). In principle, statistical analyses can be carried out on the transformed data, $\psi\{Y(\mathbf{s})\}$, using any techniques available for Gaussian random fields, which adds significant flexibility to the trans-Gaussian random fields when modeling real spatial data. However, it can be difficult to find an adequate transformation $\psi(\cdot)$, if not impossible. Furthermore, for a given $\psi(\cdot)$, some appealing properties of the latent Gaussian random field, $V(\mathbf{s})$, may not be inherited by the transformed random field, $Y(\mathbf{s})$. For example, Wallin and Bolin (2015) pointed out that the transformation $\psi(\cdot)$ in (1) may induce dependence between the mean structure and the covariance function of $Y(\mathbf{s})$. As a result, even if $V(\mathbf{s})$ is second-order stationary, the covariance function of $Y(\mathbf{s})$

may still be nonstationary. Wallin and Bolin (2015) proposed non-Gaussian Matérn fields derived from stochastic partial differential equations to model non-Gaussian spatial data. Although this provides an interesting alternative, their approach is mathematically involved and its statistical properties are much less understood than the trans-Gaussian random field; see Wallin and Bolin (2015) for a detailed discussion.

Another challenge of modeling non-Gaussian spatial data lies often in the presence of potential outliers. Although some moderate outliers can be accommodated by using a random field with heavy-tailed marginal distributions, such as the Student-t distribution, there is no guarantee that the suggested model will be flexible enough to cope with more extreme outliers, which in turn may have significant impacts on the estimation of model parameters. Consequently, if one wishes to conduct a probabilistic prediction/forecast of the distribution of some uncertain quantity at a spatial location (Gneiting et al., 2007; Gneiting and Katzfuss, 2014), imprecise parameter estimates may lead to appreciable deviations from the truth. Therefore, not only it is of great importance to construct non-Gaussian random fields with more flexible marginal distributions, it is also important to develop methods that can effectively identify extreme outliers that cannot be accommodated well by the suggested model. Identifying spatial outliers in a systematic way can be challenging because unlike the independent case, extreme outliers in a spatial random field are not necessarily the largest or the smallest observations from that field. But rather, because of spatial dependence, they are more likely to be some observations that appear to be significantly different from their local neighbors. We will show how to identify extreme spatial outliers using our proposed method.

In this paper, we propose a new class of trans-Gaussian random fields named the Tukey g-and-h (TGH) random fields, which have extremely flexible marginal distributions. The proposed model is parameterized in a way such that all parameters in the transformation can be estimated together with the covariance function of the latent Gaussian random field. This enables us to search for the

most suitable transformation for the observed data in a very large family of transformations. In this sense, the TGH random field is more flexible than existing trans-Gaussian random fields based on pre-given transformations. We will show that the TGH random field enjoys appealing statistical properties and can be used to effectively identify extreme spatial outliers, an issue that has not been well studied for existing trans-Gaussian random fields.

The rest of our paper is organized as follows. The probabilistic properties of TGH random fields, such as second-order moments, are investigated in Section 2, whereas their estimation, based on maximum approximated likelihood, is described in Section 3. Point prediction, also called kriging, and probabilistic prediction with TGH random fields is considered in Section 4, along with prediction confidence intervals. The results of Monte Carlo simulation studies of the estimation and predictive performance with TGH random fields are reported in Section 5. An application of our methodology to the aforementioned spatial precipitation dataset is presented in Section 6. The article ends with a discussion in Section 7. Derivations of our theoretical results are collected in the supplementary materials.

2 Tukey g-and-h Random Fields

2.1 Definitions

The g-and-h distribution was introduced by Tukey (1977) to model distributions that are severely skewed and subject to large outliers. Tukey's g-and-h transformation is

$$\tau_{g,h}(z) = g^{-1}\{\exp(gz) - 1\}\exp(hz^2/2), \tag{2}$$

which is a strictly monotone function of z when $h \geq 0$ and $g \in \mathbb{R}$. Here and in the sequel, for all quantities involving g, their values for the case with $g = 0$ are defined as their limits attained when $g \to 0$. The random variable $Y = \tau_{g,h}\{Z\}$ is said to have a Tukey g-and-h distribution if Z follows

a standard normal, $N(0, 1)$, distribution. The first parameter, g, in (2) controls the skewness of Y's distribution, where $g > 0$ yields a right-skewed distribution and $g < 0$ makes the distribution left-skewed. The second parameter, h, governs the tail behavior of Y's distribution, with a larger value of h indicating a heavier tail. Because of its flexible shapes, the Tukey g-and-h distribution provides a powerful tool to model non-normal data; see, for example, Field (2004) and He and Raghunathan (2012).

In this paper, we apply the Tukey g-and-h transformation to a Gaussian random field. More specifically, let $Z(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^q$, $q \geq 1$, be a standard Gaussian random field, that is, $\mathrm{E}\{Z(\mathbf{s})\} = 0$ and $\mathrm{var}\{Z(\mathbf{s})\} = 1$, with some correlation function $\mathrm{corr}\{Z(\mathbf{s}_1), Z(\mathbf{s}_2)\} = \rho_Z(\mathbf{s}_1, \mathbf{s}_2)$. A standard *Tukey g-and-h* (TGH) random field, $T(\mathbf{s})$, is defined by

$$T(\mathbf{s}) = \tau_{g,h}\{Z(\mathbf{s})\}. \tag{3}$$

Then, a more general Tukey g-and-h random field, $Y(\mathbf{s})$, can be defined as

$$Y(\mathbf{s}) = \xi + \mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + \omega T(\mathbf{s}), \tag{4}$$

where $\xi \in \mathbb{R}$ is a location parameter, $\omega > 0$ is a scale parameter, and $\mathbf{X}(\mathbf{s}), \boldsymbol{\beta} \in \mathbb{R}^p$ are vectors of some observed covariates at location $\mathbf{s}$ and their regression coefficients. This formulation of $Y(\mathbf{s})$ is close in spirit to the definition of finite dimensional multivariate Tukey g-and-h distributions in Field and Genton (2006) and He and Raghunathan (2012). The TGH random field, $Y(\mathbf{s})$, includes a large family of trans-Gaussian random fields with extremely flexible marginal distributions; for example, when $g = h = 0$, $Y(\mathbf{s})$ reduces to a Gaussian random field. For $h = 0$ and $g > 0$, $Y(\mathbf{s})$ is essentially a shifted log-Gaussian random field and for $g = 0$ and $h > 0$, $Y(\mathbf{s})$ becomes a random field with a Pareto-like marginal distribution. In fact, the Tukey g-and-h distribution can adequately approximate many distributions including the Student's t, exponential, Cauchy, Weibull and logistic distributions (Martinez and Iglewicz, 1984; Hoaglin, 1985). In Figure 1, we give a few

examples of distributions with different levels of skewness and tail heavyness that can be well approximated by the Tukey *g*-and-*h* distribution; see Genton (2004) and Azzalini and Capitanio (2014)'s books for more information on the skew-normal and skew-*t* distributions.

2.2   Second-order moments of TGH random fields

The following lemma gives the mean, variance, and covariance functions of a standard Tukey *g*-and-*h* random field, $T(\mathbf{s})$.

Lemma 1 *If $h < 1$, then the standard TGH random field $T(\mathbf{s})$ defined in (3) has a mean*

$$E\{T(\mathbf{s})\} = \frac{1}{g\sqrt{1-h}}\left[\exp\left\{\frac{g^2}{2(1-h)}\right\} - 1\right], \tag{5}$$

and if $h < 1/2$, then $T(\mathbf{s})$ has a covariance function $C_T(\mathbf{s}_1, \mathbf{s}_2) = cov\{T(\mathbf{s}_1), T(\mathbf{s}_2)\}$ as

$$C_T(\mathbf{s}_1, \mathbf{s}_2) = \frac{\exp\left[\frac{1+\rho_Z(\mathbf{s}_1,\mathbf{s}_2)}{1-h\{1+\rho_Z(\mathbf{s}_1,\mathbf{s}_2)\}}g^2\right] - 2\exp\left[\frac{1-h\{1-\rho_Z^2(\mathbf{s}_1,\mathbf{s}_2)\}}{(1-h)^2-h^2\rho_Z^2(\mathbf{s}_1,\mathbf{s}_2)}\frac{g^2}{2}\right] + 1}{g^2\sqrt{(1-h)^2 - \rho_Z^2(\mathbf{s}_1,\mathbf{s}_2)h^2}} - [E\{T(\mathbf{s})\}]^2. \tag{6}$$

The proof is given in the supplementary materials. Following Lemma 1, for the general TGH random field, $Y(\mathbf{s})$, defined in (4), we have

$$E\{Y(\mathbf{s})\} = \xi + \mathbf{X}^{\mathrm{T}}(\mathbf{s})\boldsymbol{\beta} + \omega E\{T(\mathbf{s})\}, \quad cov\{Y(\mathbf{s}_1), Y(\mathbf{s}_2)\} = \omega^2 cov\{T(\mathbf{s}_1), T(\mathbf{s}_2)\}.$$

As in any other trans-Gaussian random field, we can introduce complicated mean and covariance structures for $Y(\mathbf{s})$. Any stationary or non-stationary covariance structure available for Gaussian random fields can be applied to the TGH random field. Additional non-stationarity can be added into the formulation of (4) by allowing the scale parameter, ω, to depend on the location, $\mathbf{s}$. For the ease of presentation, we shall keep ω as a constant in this paper. The following theorem is a direct consequence of Lemma 1.

Theorem 1 *Suppose that the Gaussian random field $Z(\mathbf{s})$ in (3) is second-order stationary and that $h < 1/2$ in the standard TGH random field $T(\mathbf{s})$. Then we have: (a) $T(\mathbf{s})$ is also second-order stationary; (b) $T(\mathbf{s})$ is mean-square continuous if and only if $Z(\mathbf{s})$ is mean-square continuous; (c) $T(\mathbf{s})$ is m-times mean-square differentiable if $Z(\mathbf{s})$ is m-times mean-square differentiable.*

The proof is given in the supplementary materials.

Theorem 1 states that unlike many other trans-Gaussian random fields, nice properties such as stationarity, mean-square continuity, and degrees of mean-square differentiability can be inherited by the TGH random field from the latent Gaussian random field, $Z(\mathbf{s})$. The key difference between the TGH random field (4) and the traditional trans-Gaussian random field (1) is that the parameterizations of model (4) allow us to separate the mean structure from the transformation function $\tau_{g,h}(\cdot)$ and assume that the latent Gaussian random field $Z(\mathbf{s})$ always has a mean 0. This leads to Theorem 1. Of course, we can do this because model (4) is flexible enough for us to estimate $\xi, \omega, g, h, \boldsymbol{\beta}$ together with the parameters in the correlation function simultaneously.

Figure 2 presents realizations of the standard Gaussian random field $Z(\mathbf{s})$ with Matérn spatial correlation function (17) and its corresponding standard TGH random fields with various values of g and h. The blank areas indicate extreme values produced by the transformation that cannot fit into the color scheme. Figure 2(a) appears to be much smoother than Figure 2(c), indicating longer and stronger spatial dependence. Another observation is that although increasing either values of g or h leads to large observations, they seem to work in different fashions. Between (a)-(b), the overall image patterns appear to be similar except that almost every local area becomes more extreme because h can produce both positive and negative outliers. On the contrary, a positive g produces mainly positive large values while shrinking smaller values of $Z(\mathbf{s})$ toward 0 and thus changes the overall image pattern.

2.3 Connection with log-Gaussian random fields

Log-Gaussian random fields are popular for spatial data analysis in areas such as ecology and meteorology, where the data collected often have right-skewed distributions with possible outliers (De Oliveira, 2006). In what follows, we show that log-Gaussian random fields can be viewed as a special type of TGH random field defined in (4) with $h = 0$ and appropriately chosen ξ, ω and g. Following the setting of De Oliveira (2006), we say that $Y^*(\mathbf{s})$ is a log-Gaussian random field if $Z^*(\mathbf{s}) = \log\{Y^*(\mathbf{s})\}$ is a Gaussian random field with $\mathrm{E}\{Z^*(\mathbf{s})\} = \mu_{Z^*}$ and $\mathrm{cov}\{Z^*(\mathbf{s}_1), Z^*(\mathbf{s}_2)\} = \sigma_{Z^*}^2 \rho_{Z^*}(\mathbf{s}_1, \mathbf{s}_2)$ for some $\mu_{Z^*} \in \mathbb{R}$, $\sigma_{Z^*}^2 > 0$ and some correlation function $\rho_{Z^*}(\mathbf{s}_1, \mathbf{s}_2)$.

Lemma 2 *For the TGH random field, $Y(\mathbf{s})$, defined in (4), by letting $\boldsymbol{\beta} = 0$, $h = 0$, and setting the constraint $\xi = \omega/g$ and $g > 0$ for parameters ξ, ω and g, $Y(\mathbf{s})$ becomes a log-Gaussian random field with $Z(\mathbf{s}) = \exp(\omega) + gZ(\mathbf{s})$.*

The proof of Lemma 2 is trivial and is thus omitted. From Lemma 2, we see that the TGH random field is more flexible than the log-Gaussian random field in modeling spatial data. For example, while the log-Gaussian random field is usually used to model right-skewed data, the TGH random field can also model left-skewed data with $g < 0$.

3 Estimation of TGH Random Fields

3.1 Asymptotic properties of the MLE

Denote $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^{\mathrm{T}}, \xi, \omega, g, h)^{\mathrm{T}}$ and let $\boldsymbol{\theta}_2$ be the parameter vector of $\rho_Z(\mathbf{s}_1, \mathbf{s}_2)$ in (4). Consider a dataset $\mathcal{D}_n = \{(y(\mathbf{s}_1), \mathbf{x}(\mathbf{s}_1)), \dots, (y(\mathbf{s}_n), \mathbf{x}(\mathbf{s}_n))\}$ collected from the TGH random field, $Y(\mathbf{s})$, at locations $\mathbf{s}_1, \dots, \mathbf{s}_n$. The maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}})^{\mathrm{T}}$ is defined as

the maximizer of the log-likelihood function

$$
\begin{aligned}
L_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathcal{D}_n) \propto {}& -\frac{1}{2} \mathbf{Z}_{\theta_1}^{\mathrm{T}} (\mathbf{R}_{\theta_2}^{-1} + h \mathbf{I}_n) \mathbf{Z}_{\theta_1} - \frac{1}{2} |\mathbf{R}_{\theta_2}| \\
& - \sum_{i=1}^{n} \log \left[\exp(g z_{\theta_1, \mathbf{s}_i}) + g^{-1} \{ \exp(g z_{\theta_1, \mathbf{s}_i}) - 1 \} h z_{\theta_1, \mathbf{s}_i} \right] - n \log \omega,
\end{aligned}
\tag{7}
$$

where $z_{\theta_1, \mathbf{s}_i} = \tau_{g,h}^{-1} \left\{ \frac{y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^{\mathrm{T}} \beta - \xi}{\omega} \right\}$, $\mathbf{Z}_{\theta_1} = (z_{\theta_1, \mathbf{s}_1}, \ldots, z_{\theta_1, \mathbf{s}_n})^{\mathrm{T}}$, and $\mathbf{R}_{\theta_2}$ is the $n \times n$ correlation matrix whose (i, j)th element is $\rho_Z(\mathbf{s}_i, \mathbf{s}_j)$.

To study the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$, we follow the work of Sweeting (1980). Suppose Θ is an open subset of $\mathbb{R}^p$ and assume that $L_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathcal{D}_n)$ is twice continuously differentiable for any $\boldsymbol{\theta} \in \Theta$. Let $\mathbf{H}_n(\boldsymbol{\theta})$ be the random matrix consisting of sub-matrices $\mathbf{H}_{\theta_i \theta_j, n} = -\frac{\partial L_n(\theta_1, \theta_2 | \mathcal{D}_n)}{\partial \theta_i \partial \theta_j^{\mathrm{T}}}$, $i, j = 1, 2$. Following Sweeting (1980), we use $\to^u$ and $\Rightarrow^u$ to indicate uniform convergence and uniform weak convergence in compact subsets of Θ, respectively. Define the norm of a matrix $\mathbf{A}$ as $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}^{\mathrm{T}} \mathbf{A})}$ and a matrix sequence $\mathbf{A}_1, \ldots, \mathbf{A}_n$ converges to $\mathbf{A}$ if and only if $\|\mathbf{A}_n - \mathbf{A}\| \to 0$ as $n \to \infty$. Assume that $\mathbf{B}_{i,n}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}(\mathbf{H}_{\theta_i \theta_i, n})$, $i = 1, 2$, exist and are positive definite for all $\boldsymbol{\theta} \in \Theta$. The following conditions are sufficient to ensure the asymptotic normality of the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$:

C1 **(Information Growth)**: The non-random matrices $\mathbf{B}_{1,n}(\boldsymbol{\theta})$ and $\mathbf{B}_{2,n}(\boldsymbol{\theta})$ are both continuous in $\boldsymbol{\theta}$ and as $n \to \infty$, $\|\mathbf{B}_{i,n}^{-1}(\boldsymbol{\theta})\| \to^u 0$, $i = 1, 2$, for all $\boldsymbol{\theta} \in \Theta$.

C2 **(Convergence)**: There exists a (random) matrix $\mathbf{W}(\boldsymbol{\theta})$, which is positive definite with probability 1, such that for any $\boldsymbol{\theta} \in \Theta$, as $n \to \infty$:

$$
\mathbf{W}_n(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{B}_{1,n}^{-1/2}(\boldsymbol{\theta}) & 0 \\ 0 & \mathbf{B}_{2,n}^{-1/2}(\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} \mathbf{H}_{\theta_1 \theta_1, n} & \mathbf{H}_{\theta_1 \theta_2, n} \\ \mathbf{H}_{\theta_2 \theta_1, n} & \mathbf{H}_{\theta_2 \theta_2, n} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{1,n}^{-1/2}(\boldsymbol{\theta}) & 0 \\ 0 & \mathbf{B}_{2,n}^{-1/2}(\boldsymbol{\theta}) \end{pmatrix} \Rightarrow^u \mathbf{W}(\boldsymbol{\theta}).
$$

C3 **(Continuity)**: Let $\mathbf{A}_n(\boldsymbol{\theta}) = \mathrm{diag}\{\mathbf{B}_{1,n}^{1/2}(\boldsymbol{\theta}), \mathbf{B}_{2,n}^{1/2}(\boldsymbol{\theta})\}$. We assume that for any $c > 0$, (i) $\sup_{\theta'} \|\mathbf{A}_n^{-1}(\boldsymbol{\theta}) \mathbf{A}_n(\boldsymbol{\theta}') - \mathbf{I}_p\| \to^u 0$, where the supremum is taken over the set $\{\boldsymbol{\theta}' : \|\mathbf{A}_n(\boldsymbol{\theta})(\boldsymbol{\theta}' - \boldsymbol{\theta})\| \le c\}$; (ii) define a $p \times p$ matrix $\Gamma = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_p')$ with $\boldsymbol{\theta}_k' \in \Theta$ and let $\mathbf{H}_n(\Gamma)$ be the

matrix whose kth row is the kth row of $\mathbf{H}_n(\theta'_k)$, $k = 1, \ldots, p$. Then, $\sup_\Gamma \|\mathbf{A}_n^{-1}(\theta)\{\mathbf{H}_n(\Gamma) - \mathbf{H}_n(\theta)\}\mathbf{A}_n^{-1}(\theta) - \mathbf{I}_p\| \to^u 0$ in probability, where the supremum is taken over the set $\{\Gamma : \|\mathbf{A}_n(\theta)(\theta'_k - \theta)\| \le c, k = 1, \ldots, p\}$.

Theorem 2 *Under conditions C1-C3, the MLE $\widehat{\theta}_n$ is consistent for θ and*

$$\mathbf{A}_n(\theta)(\widehat{\theta}_n - \theta) \Rightarrow^u N_p(\mathbf{0}, \mathbf{W}^{-1}(\theta)),$$

where $\mathbf{A}_n(\theta) = \mathrm{diag}\{\mathbf{B}_{1,n}^{1/2}(\theta), \mathbf{B}_{2,n}^{1/2}(\theta)\}$ and $\|\mathbf{A}_n^{-1}(\theta)\| \to^u 0$ as $n \to \infty$.

Theorem 2 follows readily from Theorems 1-2 of Sweeting (1980) and thus the proof is omitted. Conditions C1-C3 are standard conditions used to derive asymptotic normality of a general maximum likelihood estimator and have been used by many authors, for example, see Sweeting (1980), Mardia and Marshall (1984), and Cressie and Lahiri (1993). Condition C3 essentially imposes some mild smoothness conditions on the likelihood function and is generally reasonable for a spatial model with a covariance function that is twice differentiable with respect to its parameters. Condition C1 is critical to ensure the consistency of $\widehat{\theta}_n$. Tailored to the likelihood function (7), Theorem 3 of Mardia and Marshall (1984) shows that, under the increasing domain asymptotic framework, when the data are observed in a spatial domain $\mathbf{s} \in \mathbf{D}_n \subseteq \mathbb{R}^d$ that is increasing with the sample size n, one has that $\|\mathbf{B}_{2,n}^{-1}(\theta)\| \to^u 0$ and $\mathbf{B}_{2,n}^{-1/2}(\theta)\mathbf{H}_{\theta_2\theta_2,n}\mathbf{B}_{2,n}^{-1/2}(\theta) \Rightarrow^u \mathbf{I}_{p_2}$. In addition, under the increasing domain framework, condition (i) of Theorem 2 in Mardia and Marshall (1984) requires that $\lambda_{\max}\{\mathbf{R}_{\theta_2}\} < C$ for some constant $C > 0$, which essentially controls the overall strength of the spatial dependence. It is reasonable to expect that $\|\mathbf{B}_{1,n}^{-1}(\theta)\| \to^u 0$ and $\mathbf{B}_{1,n}^{-1/2}(\theta)\mathbf{H}_{\theta_1\theta_1,n}\mathbf{B}_{1,n}^{-1/2}(\theta) \Rightarrow^u \mathbf{I}_{p_1}$ as long as the overall spatial dependence is not too strong, with an extreme case $\mathbf{R}_{\theta_2} = \mathbf{I}_n$ being supported by Xu and Genton (2015). Our simulation study in Subsection 5.1 also confirms the consistency of the MLE $\widehat{\theta}_n$.

Conditions C1-C2 are suitable for increasing domain asymptotic framework but may break down in the infill asymptotic framework, when the spatial domain $\mathbf{D} \subseteq \mathbb{R}^d$ is fixed as the sample size n increases. Let $\mathbb{P}_k$, $k = 1, 2$, be two probability measures defined on the same measurable space $(\Omega, \mathfrak{F})$. If $\mathbb{P}_1(A) = 0$ for any $A \in \mathfrak{F}$ such that $\mathbb{P}_2(A) = 0$, $\mathbb{P}_1$ is said to be absolutely continuous with respect to $\mathbb{P}_2$, denoted as $\mathbb{P}_1 \ll \mathbb{P}_2$. Next, $\mathbb{P}_1$ is said to be equivalent to $\mathbb{P}_2$, denoted as $\mathbb{P}_1 \equiv \mathbb{P}_2$, if $\mathbb{P}_1 \ll \mathbb{P}_2$ and $\mathbb{P}_2 \ll \mathbb{P}_1$. It is well known that for a family of equivalent measures $\{\mathbb{P}_\theta : \boldsymbol{\theta} \in \Theta\}$, regardless of what is observed, any estimator $\widehat{\boldsymbol{\theta}}_n$ cannot be weakly consistent for all $\boldsymbol{\theta} \in \Theta$; see Stein (1999) and Zhang (2004) for more detailed discussions. Using this argument, the following theorem indicates that, under the infill asymptotic framework, the parameter $\boldsymbol{\theta}$ for the TGH random field may not be consistently estimable.

Theorem 3 *Let $\mathbf{D}$ be a bounded subset of $\mathbb{R}^d$ for $d = 1, 2, 3$ and $\mathbb{P}_{Y,k}$, $k = 1, 2$, be two probability measures such that under $\mathbb{P}_{Y,k}$, $Y(\mathbf{s})$, $\mathbf{s} \in \mathbf{D}$, is a TGH random field (4) with $\boldsymbol{\beta}_k = \mathbf{0}$ and a Matérn correlation function (17). Assume that the parameter vector associated with $\mathbb{P}_{Y,k}$ is of the form $\boldsymbol{\theta}^{(k)} = (\xi, \omega_k, g_k, h_k, \phi_k, \nu)^\mathrm{T}$, $h_k > 0$, $k = 1, 2$. If $g_1/\omega_1 = g_2/\omega_2$, $h_1/\omega_1^2 = h_2/\omega_2^2$ and $\omega_1^2/\phi_1^{2\nu} = \omega_2^2/\phi_2^{2\nu}$, then $\mathbb{P}_{Y,1} \equiv \mathbb{P}_{Y,2}$ on the path of $Y(\mathbf{s})$, $\mathbf{s} \in \mathbf{D}$.*

The proof is given in the supplementary materials.

In practice, for a given n, both infill and increasing domain asymptotics may be appropriate. The key difference is whether the condition C1 can be met, see Zhang and Zimmerman (2005). Our results show that the TGH random field may be more suitable for applications where the increasing domain framework is appropriate. More general infill asymptotic properties of TGH random fields is an interesting future research topic.

3.2 Efficient computation of the MLE

Although the maximum likelihood estimator $\widehat{\theta}_n$ has appealing asymptotic properties, its computation is challenging. The main reason is that the inverse function of $\tau_{g,h}(z)$ in (2), denoted by $\tau_{g,h}^{-1}(\cdot)$, does not have a closed form. As a result, a direct maximization of $L_n(\theta_1, \theta_2 | \mathcal{D}_n)$ in (7) is not feasible. For independent data, one popular strategy to bypass this issue is to estimate ξ, ω, g, and h by matching a set of sample quantiles with their population counterparts; see Xu and Genton (2015) and references therein for a complete review. However, for spatial data, how to define quantiles is still an open problem and thus the quantile-matching method is not applicable. As an alternative, Xu and Genton (2015) proposed to use the estimator obtained by maximizing an approximated likelihood function, which was shown to be as efficient as the maximum likelihood estimator. The following estimation procedure can be viewed as an extension of Xu and Genton (2015) from independent to spatially dependent data.

For a fixed sample size n, we first choose a positive number b_n and then introduce K_n equally spaced knots over the interval $[-b_n, b_n]$, denoted as $-b_n = t_1 < t_2 < \cdots < t_{K_n} = b_n$. Correspondingly, $T_{k,\theta_1} = \xi + \omega\tau_{g,h}(t_k)$, $k = 1, \ldots, K_n$ define K_n knots in the transformed scale. If $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^\mathrm{T}\boldsymbol{\beta} \in [T_{1,\theta_1}, T_{K_n,\theta_1}]$, there must exist a k such that $T_{k,\theta_1} \le y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^\mathrm{T}\boldsymbol{\beta} < T_{k+1,\theta_1}$. Since $\tau_{g,h}(\cdot)$ is a monotone function, the $z_{\theta_1,\mathbf{s}_i}$ associated with $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^\mathrm{T}\boldsymbol{\beta}$ must lie between the knots t_k and t_{k+1}. As a result, we can define the following approximation to $z_{\theta_1,\mathbf{s}_i}$ as

$$\tilde{z}_{\theta_1,\mathbf{s}_i} = t_k + \frac{y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^\mathrm{T}\boldsymbol{\beta} - T_{k,\theta_1}}{T_{k+1,\theta_1} - T_{k,\theta_1}}(t_{k+1} - t_k) \quad \text{if} \quad T_{k,\theta_1} \le y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^\mathrm{T}\boldsymbol{\beta} < T_{k+1,\theta_1}, \tag{8}$$

$i = 1, \ldots, n$. Since $\tau_{g,h}(z)$ is continuous with a bounded derivative for any $z \in [-b_n, b_n]$, $z_{\theta_1,\mathbf{s}_i}$ can be well approximated by $\tilde{z}_{\theta_1,\mathbf{s}_i}$ provided that K_n is sufficiently large. Furthermore, if we can choose an appropriate b_n such that all data points in $\mathcal{D}_n$ meet the condition that $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^\mathrm{T}\boldsymbol{\beta} \in [T_{1,\theta_1}, T_{K_n,\theta_1}]$, the likelihood function (7) can also be well approximated by replacing $z_{\theta_1,\mathbf{s}_i}$ with

$\tilde{z}_{\theta_1,\mathbf{s}_i}$ for $i = 1,\ldots,n$. Such a choice of b_n is equivalent to finding a b_n such that all $z_{\theta_1,\mathbf{s}_i}$'s fall into the interval $[-b_n, b_n]$. When $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_{10}$ with $\boldsymbol{\theta}_{10}$ being the true values of $\boldsymbol{\theta}_1$, the $z_{\theta_1,\mathbf{s}_i}$'s are realizations from the standard Gaussian random field $Z(\mathbf{s})$ by definition. From a practical point of view, almost all observations from a $N(0, 1)$ distribution lie between $[-4, 4]$ and hence b_n needs not be an extremely large value. In fact, we found that $b_n = 10$ is sufficiently large for most applications, as long as $\boldsymbol{\theta}_1$ is not too far away from its true value $\boldsymbol{\theta}_{10}$. More rigorous discussions can be found in Xu and Genton (2015). After choosing appropriate values for K_n and b_n, for any given $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$, if all $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^T\boldsymbol{\beta} \in [T_{1,\theta_1}, T_{K_n,\theta_1}]$, $i = 1,\ldots,n$, we define the approximated likelihood function as

$$
\begin{aligned}
\tilde{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathcal{D}_n) \propto &-\frac{1}{2}\tilde{\mathbf{Z}}_{\theta_1}^T(\mathbf{R}_{\theta_2}^{-1} + h\mathbf{I}_n)\tilde{\mathbf{Z}}_{\theta_1} - \frac{1}{2}|\mathbf{R}_{\theta_2}| \\
&- \sum_{i=1}^{n} \log\left[\exp(g\tilde{z}_{\theta_1,\mathbf{s}_i}) + g^{-1}\{\exp(g\tilde{z}_{\theta_1,\mathbf{s}_i}) - 1\}h\tilde{z}_{\theta_1,\mathbf{s}_i}\right] - n\log\omega,
\end{aligned}
\tag{9}
$$

where $\tilde{\mathbf{Z}}_{\theta_1} = (\tilde{z}_{\theta_1,\mathbf{s}_1}, \ldots, \tilde{z}_{\theta_1,\mathbf{s}_n})^T$ with $\tilde{z}_{\theta_1,\mathbf{s}_i}$'s obtained using (8) for all $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^T\boldsymbol{\beta}$'s. If there exist any $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^T\boldsymbol{\beta} \notin [T_{1,\theta_1}, T_{K_n,\theta_1}]$, we set $\tilde{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathcal{D}_n) = -\infty$.

Using this $\tilde{L}_n(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathcal{D}_n)$, it is important to pick a sensible starting value $\boldsymbol{\theta}_{2,0} = (\boldsymbol{\beta}_0^T, \xi_0, g_0, h_0)^T$ such that all observed $y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^T\boldsymbol{\beta}_0$'s fall into the interval $[\xi_0 + \omega_0\tau_{g_0,h_0}(-b_n), \xi_0 + \omega_0\tau_{g_0,h_0}(b_n)]$. Once a sensible initial $\boldsymbol{\theta}_{1,0}$ is chosen, Algorithm I described below will start searching for local maxima in a neighborhood of $\boldsymbol{\theta}_{1,0}$. Like in many other maximum likelihood estimation procedures, one can try multiple initial starting values to search for a global maximum.

ALGORITHM I: THE MAXIMUM APPROXIMATED LIKELIHOOD ESTIMATOR

Step I: Set $K_n = \max(1000, n)$ and $b_n = 10$;

Step II: obtain initial values for $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^{\mathrm{T}}, \xi, \omega, g, h)^{\mathrm{T}}$:

    (a) Find the ordinary least square fit $\hat{\boldsymbol{\beta}}_0 = \{\mathbf{X}^{\mathrm{T}}\mathbf{X}\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ where $\mathbf{X}$ is the covariate design matrix and $\mathbf{Y}$ the vector of observations;

    (b) Use the residuals $\hat{\varepsilon}_i(\mathbf{s}_i) = y(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)^{\mathrm{T}}\hat{\boldsymbol{\beta}}_0$ to find initial estimators $(\hat{\xi}_0, \hat{\omega}_0, \hat{g}_0, \hat{h}_0)$ using Algorithms in Xu and Genton (2015).

Step III: Suppose we have obtained an estimator $\hat{\boldsymbol{\theta}}^k = (\hat{\boldsymbol{\theta}}_1^{k\mathrm{T}}, \hat{\boldsymbol{\theta}}_2^{k\mathrm{T}})^{\mathrm{T}}$, $k = 0, 1, 2, \ldots$;

    (a) Fix $\boldsymbol{\theta}_1 = \hat{\boldsymbol{\theta}}_1^k$, obtain $\hat{\boldsymbol{\theta}}_2^{k+1}$ by maximizing (9) with respect to $\boldsymbol{\theta}_2$;

    (b) Fix $\boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2^{k+1}$, obtain $\hat{\boldsymbol{\theta}}_1^{k+1}$ by maximizing (9) with respect to $\boldsymbol{\theta}_1$;

Repeat Step III until $\hat{\boldsymbol{\theta}}^{k+1} = (\hat{\boldsymbol{\theta}}_1^{(k+1)\mathrm{T}}, \hat{\boldsymbol{\theta}}_2^{(k+1)\mathrm{T}})^{\mathrm{T}}$ converges.

The choice $K_n = \max(1000, n)$ is recommended by Xu and Genton (2015) and yields sufficiently good estimation accuracy in all our simulation studies. In practice, one can increase K_n as long as the computational cost is acceptable. In all our simulation studies, Algorithm I converges quickly. To increase the chance of finding the global maxima of the approximated likelihood function, it is highly recommended to use multiple starting values for parameters in $\boldsymbol{\theta}_2$.

With regard to the computational cost of Algorithm I, Step I costs $O(K_n + n)$ floating operations (Xu and Genton, 2015) provided that $\mathbf{R}_{\theta_2}^{-1}$ and $|\mathbf{R}_{\theta_2}|$ were returned from Step II, which cost $O(n^3)$ floating operations to compute. Hence, the overall computational complexity of Algorithm I is of the order $O(K_n + n^3)$, which is only feasible for small to moderate n. If n is large to massive, tools such as Gaussian predictive processes (Banerjee et al., 2008), fixed rank kriging (Cressie and Johannesson, 2008), or Gaussian Markov Random Fields (Rue and Held, 2005; Xu et al., 2015), can be used for the efficient computation of the TGH random field; see also the review by Sun et al. (2012).

3.3 Detection of extreme spatial outliers

In practice, every geostatistical model has its limitations and it is always possible that some extreme observations from a spatial random field cannot be modeled well. In such cases, even a small number of extreme outliers may have a significant impact on model estimation. In particular, the parameter h in the definition of the TGH random field (4) is introduced to accommodate possible outliers in the random field and therefore its estimation is more sensitive to extreme outliers. Detection of extreme spatial outliers can be challenging because unlike in the independent case, the extreme outliers in a spatial random field are not necessarily the largest or the smallest observations from that field. Instead, because of the spatial dependence, they are more likely to be observations that appear to be significantly different from their local neighbors.

Next, we show how to use the TGH random field to identify extreme spatial outliers. Suppose we have obtained an initial estimator of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ using Algorithm I with all available data and denote them by $\hat{\boldsymbol{\theta}}_1^{int}$ and $\hat{\boldsymbol{\theta}}_2^{int}$ where "int" stands for "initial". Define

$$\widehat{\mathbf{Z}}^{int} = \widehat{\mathbf{L}} \tilde{\mathbf{Z}}_{\hat{\theta}_1^{int}}, \tag{10}$$

where $\widehat{\mathbf{L}} = \mathbf{R}_{\hat{\theta}_2^{int}}^{-1/2}$, $\tilde{\mathbf{Z}}_{\hat{\theta}_1^{int}}$ and $\mathbf{R}_{\hat{\theta}_2^{int}}$ are as defined in (9) with $\boldsymbol{\theta}_j = \hat{\boldsymbol{\theta}}_j^{int}$ for $j = 1, 2$. The $\widehat{\mathbf{L}}$ matrix can be computed through an eigen decomposition of $\mathbf{R}_{\hat{\theta}_2^{int}}^{-1} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{\mathrm{T}}$ with the diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$ and then let $\widehat{\mathbf{L}} = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}^{\mathrm{T}}$ where $\boldsymbol{\Lambda}^{1/2} = \mathrm{diag}\{\lambda_1^{1/2}, \ldots, \lambda_n^{1/2}\}$. If the TGH random field (4) provides an adequate fit for the dataset $\mathcal{D}_n$, the elements in the $n \times 1$ vector $\widehat{\mathbf{Z}}^{int}$ are independent $N(0, 1)$ random variables. Unusually large values in $\widehat{\mathbf{Z}}^{int}$ often indicate potential existence of extreme outliers in the spatial random field. Denote the jth element of $\widehat{\mathbf{Z}}^{int}$ as $\hat{z}_j^{int}$ and $\widehat{\mathbf{L}}_j^{\mathrm{T}} = (\hat{l}_{j,1}, \ldots, \hat{l}_{j,n})$ as the jth row of the matrix $\widehat{\mathbf{L}}$. Then we have the relationship $\hat{z}_j^{int} = \widehat{\mathbf{L}}_j^{\mathrm{T}}\tilde{\mathbf{Z}}_{\hat{\theta}_1^{int}}$. If $|\hat{z}_j^{int}|$ is suspiciously large, for example, larger than 3, we can trace the major source of this abnormality back to those components in $\tilde{\mathbf{Z}}_{\hat{\theta}_1^{int}}$. Then we shall remove the most extreme outlier from the entire random field and refit the TGH random field to compute new values of $\widehat{\mathbf{Z}}^{int}$. We will keeping

repeating this process until the resulting $\widehat{\mathbf{Z}}^{int}$ appears to be close enough to $N(0, 1)$ distribution. We summarize the extreme spatial outlier detection procedure in the following algorithm.

ALGORITHM II: EXTREME SPATIAL OUTLIER DETECTION

Step I: Find the initial estimates $\hat{\boldsymbol{\theta}}_1^{int}$ and $\hat{\boldsymbol{\theta}}_2^{int}$ using Algorithm I with all available data;
Step II: Find $\mathbf{R}_{\hat{\theta}_2^{int}}$ and $\tilde{\mathbf{Z}}_{\hat{\theta}_1^{int}}$ in (9) by plugging in $\boldsymbol{\theta}_j = \hat{\boldsymbol{\theta}}_j^{int}$ for $j = 1, 2$;
Step III: Perform the eigen-decomposition of $\mathbf{R}_{\hat{\theta}_2^{int}}^{-1} = \mathbf{P}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{P}$ and $\widehat{\mathbf{L}} = \mathbf{P}^{\mathrm{T}}\boldsymbol{\Lambda}^{1/2}\mathbf{P}$;
Step IV: Find $\widehat{\mathbf{Z}}^{int} = \widehat{\mathbf{L}}\tilde{\mathbf{Z}}_{\hat{\theta}_1^{int}}$, compute the p-value of Shapiro-Wilk test on $\widehat{\mathbf{Z}}^{int}$;
  (a) If p-value> 0.10, stop the algorithm; otherwise, continue as follows;
  (b) If the set $\mathcal{J} = \{j : |z_j^{int}| > \eta\}$ for some $\eta > 0$ (we recommend using $\eta = 3$) is empty, stop
  the algorithm; otherwise, continue as follows;
  (c) Let $j^o = \arg\max_{l=1,\dots,n;} |z_j^{int}|$ and remove the largest contributor to $z_{j^0}^{int}$, that is, remove the
  point $k = \arg\max_{l=1,\dots,n;}\{|\hat{l}_{j^o,l}\tilde{z}_{\hat{\theta}_1^{int},l}|\}$.
Repeat Steps I-IV until algorithm stops.

As an illustration, we apply Algorithm II to the precipitation dataset described in Section 6. Figure 3(c) gives the QQ-plot of the initial fitted value of $\widehat{\mathbf{Z}}^{int}$ in (10) obtained by fitting the TGH random field (20) with all $n = 991$ observed data points. Obviously, the potential inadequate fit to the data is indicated by the existence of very large and small values in $\widehat{\mathbf{Z}}^{int}$. By applying the Algorithm II with $\eta = 3$, we successfully identify 13 extreme spatial outliers in the original dataset as indicated in Figure 3(a). To see whether these 13 points are truly extreme spatial outliers, we give the boxplots of observations of identified locations together with their nearest 20 neighbors in Figure 3(b). We can see that all 13 points can be considered as extreme outliers in their local spatial neighborhoods, which demonstrates the effectiveness of Algorithm II in detecting extreme spatial outliers.

After identifying these 13 extreme spatial outliers, we manually remove them and refit the TGH random field (20) with the remaining $n = 978$ data points. Denote the resulting estimator of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ as $\widehat{\boldsymbol{\theta}}_1^{psc}$ and $\widehat{\boldsymbol{\theta}}_2^{psc}$, where the superscript "psc" stands for the "post-screen" estimator.

As illustrated in Figure 3(d), the components in $\widehat{\mathbf{Z}}^{psc}$ obtained by replacing $\widehat{\boldsymbol{\theta}}_1^{int}$ and $\widehat{\boldsymbol{\theta}}_2^{int}$ with $\widehat{\boldsymbol{\theta}}_1^{psc}$ and $\widehat{\boldsymbol{\theta}}_2^{psc}$ in (10) appear to be much closer to a $N(0, 1)$ distribution. The vector $\mathbf{Z}^{psc}$ was computed based only on the remaining 978 data points. This is a strong indication of an adequate fit of the TGH random field to the remaining precipitation data. In Table 3 of Section 6, we summarize the parameter estimates for the TGH random field before and after removing the 13 extreme spatial outliers, where we can see that the estimate for the parameter h decreases from 0.095 to 0.045. This is expected considering that the role of h is to model outliers in the TGH random field, making it sensitive to extreme spatial outliers.

4   Spatial Prediction with TGH Random Fields

One of the primary goals of geostatistical modeling is to make predictions at spatial locations without observations. There have been two major approaches to make spatial predictions: (i) point prediction or kriging (Cressie, 1993); (ii) probabilistic prediction or forecast (Gneiting et al., 2007; Gneiting and Katzfuss, 2014). The goal of kriging is to predict the actual value of $Y(\mathbf{s}_0)$ at a new location $\mathbf{s}_0$ based on observed data $\mathcal{D}_n$. On the contrary, probabilistic prediction aims at providing a complete predictive distribution for $Y(\mathbf{s}_0)$ conditional on $\mathcal{D}_n$. Either approach requires partial or full knowledge of the conditional distribution of $Y(\mathbf{s}_0)$ given $\mathcal{D}_n$, which we derive in this section for the TGH random field. First, we give a definition of the multivariate g-and-h distribution. For simplicity, from now on, we use $\tau_{g,h}(\mathbf{Z})$ (or $\tau_{g,h}^{-1}(\mathbf{Z})$) to denote componentwise (or componentwise inverse) g-and-h transformation of elements in a vector $\mathbf{Z}$.

A random vector $\mathbf{T} = (T_1, \ldots, T_n)^{\mathrm{T}}$ is said to have a multivariate g-and-h distribution if $\mathbf{T} = \tau_{g,h}(\mathbf{Z})$ with $\mathbf{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$ having a multivariate normal distribution $N_n(\boldsymbol{\mu}_n, \Sigma_n)$ with a mean vector $\boldsymbol{\mu}_n$ and a covariance matrix Σ_n. We denote it by $\mathbf{T} \sim GH_n(\boldsymbol{\mu}_n, \Sigma_n, g, h)$. With this definition,

the following Lemma 3 states that if two random vectors $\mathbf{T}_1$ and $\mathbf{T}_2$ have a joint multivariate g-and-h distribution, then the marginal distributions of $\mathbf{T}_1$, $\mathbf{T}_2$ and the conditional distribution of $\mathbf{T}_1$ given $\mathbf{T}_2$ are all still within this family of distributions.

Lemma 3 *Suppose that* $(\mathbf{T}_1^\mathrm{T}, \mathbf{T}_2^\mathrm{T})^\mathrm{T} \sim GH_{n_1+n_2}\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, g, h\right]$ *for some* $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ *and* $h \geq 0$. *Then we have*

$$\mathbf{T}_1 \sim GH_{n_1}(\boldsymbol{\mu}_1, \Sigma_{11}, g, h), \quad \mathbf{T}_2 \sim GH_{n_2}(\boldsymbol{\mu}_2, \Sigma_{22}, g, h), \quad and$$

$$\mathbf{T}_1|\mathbf{T}_2 \sim GH_{n_1}\left[\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\{\tau_{g,h}^{-1}(\mathbf{T}_2) - \boldsymbol{\mu}_2\}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, g, h\right].$$

The proof is given in the supplementary materials.

4.1 Kriging with TGH random fields

The goal of kriging is to find an optimal point estimator of $Y(\mathbf{s}_0)$ by minimizing some accuracy measures. For TGH random fields, we consider two such measures: the absolute loss and the squared loss, under which the optimal predictors for $Y(\mathbf{s}_0)$ are:

absolute loss: $\quad \widehat{Y}_1^{opt}(\mathbf{s}_0) = \arg\min_c \mathrm{E}[\{|Y(\mathbf{s}_0) - c|\}|\mathcal{D}_n] = \mathrm{med}\{Y(\mathbf{s}_0)|\mathcal{D}_n\},$

squared loss: $\quad \widehat{Y}_2^{opt}(\mathbf{s}_0) = \arg\min_c \mathrm{E}[\{Y(\mathbf{s}_0) - c\}^2|\mathcal{D}_n] = \mathrm{E}\{Y(\mathbf{s}_0)|\mathcal{D}_n\},$

where $\mathrm{E}\{Y(\mathbf{s}_0)|\mathcal{D}_n\}$ and $\mathrm{med}\{Y(\mathbf{s}_0)|\mathcal{D}_n\}$ stand for the conditional mean and the conditional median of $Y(\mathbf{s}_0)$ given the data $\mathcal{D}_n$. For a general trans-Gaussian random field $Y(\mathbf{s}) = \psi\{Z(\mathbf{s})\}$, the exact form of $\mathrm{E}\{Y(\mathbf{s}_0)|\mathcal{D}_n\}$ is usually not available except for some special $\psi(\cdot)$. The following theorem provides closed-form solutions for $\widehat{Y}_1^{opt}(\mathbf{s}_0)$ and $\widehat{Y}_2^{opt}(\mathbf{s}_0)$.

Theorem 4 *Given a dataset* $\mathcal{D}_n = \{(y(\mathbf{s}_1), \mathbf{x}(\mathbf{s}_1)), \ldots, (y(\mathbf{s}_n), \mathbf{x}(\mathbf{s}_n))\}$ *generated from the TGH random field,* $Y(\mathbf{s})$, *defined in (4) with* $0 \leq h < 1$, *let* $\mathbf{R}_{\theta_2}$ *and* $\mathbf{Z}_{\theta_1}$ *be as defined in (7) and* $\mathbf{r}_{\theta_2}$ *be the*

$n \times 1$ vector whose ith element is $\rho_Z(\mathbf{s}_i, \mathbf{s}_0)$ for a new location $\mathbf{s}_0$. Then the conditional distribution of $T(\mathbf{s}_0)$ given the data $\mathcal{D}_n$ is

$$T(\mathbf{s}_0)|\mathcal{D}_n \sim GH_1(\tilde{\mu}, \tilde{\sigma}^2, g, h), \tag{11}$$

where $\tilde{\mu} = \mathbf{r}_{\theta_2}^{\mathrm{T}} \mathbf{R}_{\theta_2}^{-1} \mathbf{Z}_{\theta_1}$ and $\tilde{\sigma}^2 = 1 - \mathbf{r}_{\theta_2}^{\mathrm{T}} \mathbf{R}_{\theta_2}^{-1} \mathbf{r}_{\theta_2}$. As a result, the optimal predictors for $Y(\mathbf{s}_0)$ given the dataset $\mathcal{D}_n$ using the absolute loss and the squared loss are:

$$\widehat{Y}_1^{opt}(\mathbf{s}_0) = \xi + \mathbf{X}(\mathbf{s}_0)^{\mathrm{T}}\boldsymbol{\beta} + \omega\tau_{g,h}(\tilde{\mu}), \tag{12}$$

$$\widehat{Y}_2^{opt}(\mathbf{s}_0) = \xi + \mathbf{X}(\mathbf{s}_0)^{\mathrm{T}}\boldsymbol{\beta} + \frac{\omega}{g\sqrt{1-h\tilde{\sigma}^2}} \exp\left\{\frac{h\tilde{\mu}^2}{2(1-h\tilde{\sigma}^2)}\right\}\left[\exp\left\{\frac{g^2\tilde{\sigma}^2 + 2g\tilde{\mu}}{2(1-h\tilde{\sigma}^2)}\right\} - 1\right]. \tag{13}$$

The proof is given in the supplementary materials.

Although both $\widehat{Y}_1^{opt}(\mathbf{s}_0)$ and $\widehat{Y}_2^{opt}(\mathbf{s}_0)$ can be used in practice, we shall focus on $Y_1^{opt}(\mathbf{s}_0)$ because it is more robust to the skewness and potential outliers.

4.2 Probabilistic prediction with TGH random fields

Unlike kriging, probabilistic prediction aims at predicting the whole conditional distribution of some uncertain quantity at a location without observations (Gneiting et al., 2007; Gneiting and Katzfuss, 2014), which is more informative and better at capturing the uncertainty in prediction. Tailored to our case, we wish to use the conditional distribution of $Y(\mathbf{s}_0)|\mathcal{D}_n$ based on the TGH random field, denoted by $\mathcal{F}_{\mathbf{s}_0}$, to predict the true conditional distribution of $Y(\mathbf{s}_0)|\mathcal{D}_n$, denoted as $\mathcal{G}_{\mathbf{s}_0}$. By Theorem 4, the closed form of $\mathcal{F}_{\mathbf{s}_0}$ can be easily obtained through the distribution (11). The general practice of probabilistic prediction involves two steps: (i) calibration; and (ii) assessment of the sharpness of the predictive distribution (Gneiting et al., 2007). The calibration step is to provide some guidance on how close are $\mathcal{F}_{\mathbf{s}_0}$ and $\mathcal{G}_{\mathbf{s}_0}$. Let $F_{\mathbf{s}_0}(\cdot)$ and $G_{\mathbf{s}_0}(\cdot)$ be the cumulative distribution functions of $\mathcal{F}_{\mathbf{s}_0}$ and $\mathcal{G}_{\mathbf{s}_0}$, respectively. An important tool to assess the calibration is the probability integral transform (PIT, Dawid, 1984; Diebold et al., 1998), whose value is defined

as $p_{\mathbf{s}_0} = F_{\mathbf{s}_0}\{y(\mathbf{s}_0)\}$, where $y(\mathbf{s}_0)$ is an observed value from the distribution $G_{\mathbf{s}_0}$. If the predictive distribution $F_{\mathbf{s}_0} = G_{\mathbf{s}_0}$ and $F_{\mathbf{s}_0}$ is continuous, then $p_{\mathbf{s}_0}$ should have a uniform distribution. In practice, a histogram of PIT values at different locations of the random field can be created and an approximately uniform PIT histogram indicates calibration (Gneiting et al., 2006). After the predictive distribution passes the calibration, its sharpness can be assessed by the average width of confidence intervals, as discussed in Subsection 4.3.

Numerical assessments of probabilistic predictions are usually done based on some proper scoring rules such as the Hyvärinen score (Hyvärinen, 2005) and the continuous ranked probability score (CRPS, Gneiting et al., 2007). In particular,

$$CRPS(F, y) = \int_{-\infty}^{\infty} \{F(x) - I(y \le x)\}^2 \, dx = E(|Y - y|) - \frac{1}{2}E(|Y - Y^*|), \tag{14}$$

where $F(\cdot)$ is a cumulative distribution function, $I(\cdot)$ is an indicator function, Y and Y^* are independent random variables with the same cumulative distribution function $F(y)$ and a finite first moment (Gneiting and Katzfuss, 2014). The CRPS enjoys many appealing properties but may be difficult to derive in closed-form for a general distribution. However, using Theorem 4, we can easily derive the closed form of CRPS for the predictive distribution based on the TGH random field.

Lemma 4 *The continuous ranked probability score for the predictive cumulative distribution function $F_{\mathbf{s}_0}(\cdot)$ given in Theorem 4 for the TGH random field is*

$$
\begin{aligned}
CRPS\{F_{\mathbf{s}_0}, y(\mathbf{s}_0)\} &= \left\{y(\mathbf{s}_0) - \xi - \mathbf{x}^{\mathrm{T}}(\mathbf{s}_0)\beta\right\}\left[2\Phi\left\{\frac{z(\mathbf{s}_0) - \tilde{\mu}}{\tilde{\sigma}}\right\} - 1\right] \\
&\quad + \frac{2\omega \exp\left(\frac{h\tilde{\mu}^2}{2p^*}\right)}{g\sqrt{p^*}}\left(\Phi\left[\frac{\sqrt{p^*}}{\tilde{\sigma}}\left\{z(\mathbf{s}_0) - \frac{\tilde{\mu}}{p^*}\right\}\right] + \Phi\left(\frac{h\tilde{\mu}\tilde{\sigma}}{q^*}\right) - 1\right) \\
&\quad - \frac{2\omega \exp\left(\frac{g^2\tilde{\sigma}^2 + 2\tilde{\mu}g + h\tilde{\mu}^2}{2p^*}\right)}{g\sqrt{p^*}}\left(\Phi\left[\frac{\sqrt{p^*}}{\tilde{\sigma}}\left\{z(\mathbf{s}_0) - \frac{\tilde{\mu} + g\tilde{\sigma}^2}{p^*}\right\}\right] + \Phi\left(\frac{h\tilde{\mu}\tilde{\sigma} + g\tilde{\sigma}}{q^*}\right) - 1\right),
\end{aligned}
\tag{15}
$$

with $\tilde{\mu}$ and $\tilde{\sigma}$ as defined in (11),$0 \le h < 1$ and $z(\mathbf{s}_0) = \tau_{g,h}^{-1}\left\{\frac{y(\mathbf{s}_0) - \xi - \mathbf{x}^{\mathrm{T}}(\mathbf{s}_0)\beta}{\omega}\right\}$, $p^* = 1 - h\tilde{\sigma}^2$ and

$$q^* = \sqrt{2 - 3h\tilde{\sigma}^2 + h^2\tilde{\sigma}^4}.$$

The proof is given in the supplementary materials.

4.3 Prediction confidence intervals with TGH random fields

Whether it is to quantify the uncertainty of kriging or to assess the sharpness of the predictive distribution from the probabilistic prediction, a valid prediction confidence interval plays an important role. By Theorem 4, the most straightforward $(1 - \alpha)100\%$ prediction confidence interval can be defined as $[\xi + \mathbf{x}^\mathrm{T}(\mathbf{s}_0)\boldsymbol{\beta} + \omega\tau_{g,h}(\tilde{\mu} - z_{1-\alpha/2}\tilde{\sigma}),\ \xi + \mathbf{x}^\mathrm{T}(\mathbf{s}_0)\boldsymbol{\beta} + \omega\tau_{g,h}(\tilde{\mu} + z_{1-\alpha/2}\tilde{\sigma})]$, where $\tilde{\mu}$ and $\tilde{\sigma}^2$ are as given in Theorem 4 and z_α is the αth quantile of the standard normal distribution. However, this interval can be unnecessarily wide when the predictive distribution $\mathcal{F}_{\mathbf{s}_0}$ is severely skewed (i.e., $|g|$ is large). To resolve this issue, we follow the work of De Oliveira and Rui (2009) and propose the following shortest prediction interval for the TGH random field

$$\left[\xi + \mathbf{x}^\mathrm{T}(\mathbf{s}_0)\boldsymbol{\beta} + \omega\tau_{g,h}(\tilde{\mu} - z_{1-\gamma^{opt}}\tilde{\sigma}),\ \xi + \mathbf{x}^\mathrm{T}(\mathbf{s}_0)\boldsymbol{\beta} + \omega\tau_{g,h}(\tilde{\mu} + z_{1-\alpha+\gamma^{opt}}\tilde{\sigma})\right], \tag{16}$$

where $\gamma^{opt} \in [0, \alpha]$ is chosen by minimizing the length of the interval

$$\gamma^{opt} = \arg\min_{\gamma \in [0,\alpha]} \left\{\tau_{g,h}(\tilde{\mu} + z_{1-\alpha+\gamma^{opt}}\tilde{\sigma}) - \tau_{g,h}(\tilde{\mu} - z_{1-\gamma^{opt}}\tilde{\sigma})\right\}.$$

In practice, we need to plug-in the estimated values for parameters in the TGH random field and then obtain γ^{opt} numerically. As De Oliveira and Rui (2009) pointed out for log-Gaussian random fields, such a plug-in strategy may lead to undercoverage and needs to be adjusted when the sample size is small. We believe this might also be the case for the TGH random fields. How to make such adjustments is an interesting research topic.

5 Monte Carlo Simulation Study

In this section, we use Monte Carlo simulations to evaluate the performance of the proposed TGH random field. In all simulation studies, we assume that the latent Gaussian random field $Z(\mathbf{s})$ has the Matérn correlation function

$$\rho_Z(\mathbf{s}_1, \mathbf{s}_2) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(4\frac{\|\mathbf{s}_1 - \mathbf{s}_2\|}{\phi}\right)^{\nu} K_{\nu}\left(4\frac{\|\mathbf{s}_1 - \mathbf{s}_2\|}{\phi}\right), \tag{17}$$

where $\|\mathbf{s}_1 - \mathbf{s}_2\|$ is the distance between locations $\mathbf{s}_1$ and $\mathbf{s}_2$, ν is the smoothness parameter, ϕ is the range parameter, $\Gamma(\cdot)$ is the gamma function, and $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind. We adopt this special parameterization for the range parameter ϕ such that $\rho_Z(\mathbf{s}_1, \mathbf{s}_2) \approx 0.05$ for $\|\mathbf{s}_1 - \mathbf{s}_2\| = \phi$ when $\phi = 40$ and $\nu = 1$. For all simulations, we fix parameters $\xi = 0$, $\omega = 2$, $\phi = 40$, $\nu = 1$ for the TGH random field. In addition, one covariate $X \sim N(0, 1)$ was introduced in (4) as the regression random variable, whose coefficient was set to $\beta = 2$.

5.1 Evaluation of estimation accuracy with TGH random fields

In this simulation study, the data were generated using the TGH random field (4) with multiple values of g and h. The main goal is to investigate whether Algorithm I can produce consistent estimators for all parameters in model (4). To be consistent with the conditions of Theorem 2, we choose the spatial locations in an increasing domain as follows: $n = c_n^2$ locations were drawn uniformly over the region $[0, 10c_n] \times [0, 10c_n]$, with $c_n = 10, 15, 20$. Furthermore, for each simulation run, we applied the Algorithm II to the simulated data set and removed the detected outliers first. Then the remaining data were used to estimate the parameters. The empirical bias and root mean-squared error (RMSE) of the proposed estimators were computed based on 500 simulation runs, which are summarized in Table 1. In particular, $\hat{\beta}_0$ represents the least squares estimator for β by ignoring spatial dependence in (4) while $\hat{\beta}$ is the estimator of β using Algorithm I. As we can

see from Table 1, $\hat{\beta}$ is much more efficient than $\hat{\beta}_0$, which could lead to better spatial predictions because the mean structure of model (4) is estimated more precisely. Overall, from Table 1, we can see that the proposed estimators obtained using Algorithm I are unbiased and consistent for all parameters when the sample size is large enough. For the spatial case $g = h = 0$, we also report the estimation accuracies of the maximum likelihood estimator using the Gaussian random field (GRF), which appears to be quite close to those of the estimators obtained using Algorithm I. This partially illustrates the effectiveness of the estimation procedure outlined in Algorithm I.

5.2   Evaluation of kriging with TGH random fields

In this subsection, we compare the kriging performance of the TGH random field with that of the Gaussian random field. When generating data, we fix $g = 0.5$ and change h from 0 to 0.4 to illustrate the impacts of outliers on prediction accuracy. The data were generated on a 15×15 regular lattice over the region $[0, 100] \times [0, 100]$, as illustrated in Figure 4(a). Observations from 125 locations were used as the training data, and the remaining observations were held out as the validation data. The same set of spatial locations were used for estimation and prediction in each of 500 simulation runs. Therefore, each location in the validation set has 500 predictions. Three kriging approaches are considered: the TGH random field (gh-krig), the TGH random field by fixing $h = 0$ (g-krig), and the Gaussian random field (Gau-krig). For each prediction location, the kriging accuracy was measured by the following criterion

$$MAD(\mathbf{s}_i^p) = \mathrm{med}\left\{|\widehat{y}(\mathbf{s}_i^p)^b - y(\mathbf{s}_i^p)^b|, b = 1, \ldots, 500\right\}, \quad i = 1, \ldots, 100, \tag{18}$$

where $y(\mathbf{s}_i^p)^b$ and $\widehat{y}(\mathbf{s}_i^p)^b$ are the observed and predicted values at a prediction location $\mathbf{s}_i^p$ for the bth simulation run, respectively. In Figure 4(b), we summarize the predictive performance of these three kriging methods for $h = 0, 0.2, 0.4$ at 100 prediction locations. As we can see from

Figure 4(b), when $h = 0$, gh-krig and g-krig yield almost identical results, and both are better than Gau-krig. As h increases, g-krig is not flexible enough to accommodate additional outliers introduced by large values of h and thus yields suboptimal prediction results compared to those of gh-krig. Nevertheless, g-krig still manages to control the effect of the skewness in the random field to some extent and was able to outperform Gau-krig.

5.3 Evaluation of probabilistic prediction with TGH random fields

The same simulation setup was used as in the previous subsection. For each prediction location, the accuracy of the probabilistic prediction was measured by

$$mCRPS(\mathbf{s}_i^p) = \text{med}\{CRPS(\widehat{F}_{\mathbf{s}_i^p}^b, y(\mathbf{s}_i^p)^b), \quad b = 1, \ldots, 500\}, \quad i = 1, \ldots, 100, \tag{19}$$

where $\widehat{F}_{\mathbf{s}_i^p}^b$ is the estimated predictive distribution at the prediction location $\mathbf{s}_i^p$ for the bth simulation run and $y(\mathbf{s}_i^p)^b$ is the observed value. For $h = 0, 0.2, 0.4$, boxplots of $mCPRS(\mathbf{s}_i^p)$'s in Figure 4(c) are used to illustrate the performance of predictive distributions using the general TGH random field (gh-pred), the simplified TGH random field with $h = 0$ (g-pred), and the Gaussian random field (Gau-pred). We can see that as the value of h increases, the benefit of using the general TGH random field becomes more significant. In Figure 4(d-f), we also plot the PIT histograms of the three predictive distributions for the case with $g = 0.5$ and $h = 0.2$. In this case, the difference in the mCRPS between gh-pred and and g-pred is much smaller than the difference between their PIT histograms, indicating that simple comparison of mCRPS may be misleading for assessing the performance of probabilistic prediction.

5.4 Evaluation of prediction confidence intervals with TGH random fields

To study the sharpness of the three predictive distributions, we compute the average length of the 50% and 90% confidence intervals. For the gh-pred and g-pred methods, the confidence interval

defined in (16) is adopted while for the Gaussian random field, the usual symmetric prediction interval is used. The results are summarized in Table 2. As expected, the *gh*-pred gives the shortest confidence intervals on average.

6   Application to Precipitation Data

In this section, we apply the proposed TGH random field to the precipitation dataset introduced in Section 1; see Figure 3(a). Denote by $V(\mathbf{s})$ a stationary Gaussian random field with mean 0, variance σ_v^2, and a Matérn correlation function as defined in (17). Let $\varepsilon(s)$ be a Gaussian white noise process independent of $V(\mathbf{s})$ with a variance σ_e^2. Three covariates were considered for the mean structure: the longitude ($Lon(\mathbf{s})$), the latitude ($Lat(\mathbf{s})$), and the elevation ($Elev(\mathbf{s})$), which were standardized by removing the means and then dividing them by the standard deviations in the region under study.

The first model we use to model this dataset is a TGH random field

$$Y(\mathbf{s}) = \xi + \beta_1 Lon(\mathbf{s}) + \beta_2 Lat(\mathbf{s}) + \beta_3 Elev(\mathbf{s}) + \omega\tau_{g,h}\{V(\mathbf{s}) + \varepsilon(\mathbf{s})\}, \tag{20}$$

which is a special case of (4) with $Z(\mathbf{s}) = V(\mathbf{s}) + \varepsilon(\mathbf{s})$ after imposing the constraint $\sigma_v^2 + \sigma_e^2 = 1$. For comparison, we also consider trans-Gaussian random fields of the form

$$\psi\{Y(\mathbf{s})\} = \xi + \beta_1 Lon(\mathbf{s}) + \beta_2 Lat(\mathbf{s}) + \beta_3 Elev(\mathbf{s}) + V(\mathbf{s}) + \varepsilon(\mathbf{s}). \tag{21}$$

Two transformations are studied: $\psi(y) = \sqrt{y}$ and $\psi(y) = y$, corresponding to the Root-Gaussian random field (RGRF) and the Gaussian random field (GRF).

The parameters of models (20) and (21) were estimated in two ways: the first approach uses all $n = 991$ observations for estimation (Without Screen) and the second approach first removed the 13 extreme outliers identified in Subsection 3.3 and then used the remaining data to estimate

all parameters (With Screen). Results are summarized in Table 3. In addition, we use the lengths of the 90% bootstrap confidence intervals (bLen) based on 500 parametric bootstrap samples generated using estimated models to quantify the uncertainties of point estimators. It is also worth pointing out that estimated regression coefficients for Root-Gaussian random field have different interpretations than those for other fields and should not be compared directly with them.

We want to further evaluate the predictive performances of all fitted models. To compute the PIT and CRPS values, we use the following re-sampling approach: (a) randomly choose 80% of the data to estimate three random fields; (b) use the estimated random fields to compute PIT and CRPS values at the remaining 20% spatial locations. Repeat (a) and (b) for 500 times and record all PIT and CRPS values. We then first use the PIT histograms to calibrate the predictive distributions made by these three models, as illustrated in Figure 5. Obviously, the PIT histogram of predictive distributions made by the TGH random field estimated after screening out all extreme spatial outliers using Algorithm II appears more uniform than for all other models. In contrast, we observe systematic deviations from uniformity in Figure 5(b)-(c), suggesting departures of the predictive distributions based on the Root-Gaussian and Gaussian random fields from calibration. Another noticeable feature is that by removing 13 extreme outliers, the PIT histograms of the Root-Gaussian random field and even the Gaussian random fields become better. This indicates that the proposed Algorithm II for identifying extreme spatial outliers not only can improve the probabilistic predictive ability of the TGH random field, but also can be beneficial for other models.

Table 4 summarizes numerical assessments of prediction performances of the three models using the re-sampling approach mentioned above. The MAD and mCRPS are as defined in (18) and (19), respectively. The CP and mLen represent the empirical coverage probabilities and median interval length of the prediction confidence intervals, respectively, for all spatial locations. We can see that while the MAD and mCRPS do not show much differences between models, the TGH

random field fitted with screen gives the overall shortest confidence intervals with satisfactory coverage probabilities. Considering that the TGH random field fitted with screen also gives the best PIT histogram, we argue that it provides the best fit for this precipitation data among all candidates.

Finally, as mentioned earlier, one important goal in spatial modeling of precipitation data is to create a high resolution precipitation map in a spatial region using the observed data. In Figure 6(a)-(b), we plot the precipitation map produced by the fitted TGH random field (with screen) and the fitted Gaussian random field. Figure 6(c) illustrates the length differences of the 90% prediction confidence intervals produced using the TGH and Gaussian random field (Gaussian−TGH). We observe that the TGH random field produces much shorter prediction confidence intervals than the Gaussian random field in most areas of the map, except for those areas with large predicted values.

7 Discussion

We have introduced a new class of trans-Gaussian random fields named Tukey *g*-and-*h* random fields to model non-Gaussian spatial data. The proposed TGH random fields have extremely flexible marginal distributions and can, therefore, be applied to a wide range of applications. Unlike most of the existing trans-Gaussian random fields, the special formulation of TGH random fields enables us to automatically search for the best transformation among a large class of candidates for the dataset in hand while estimating model parameters. We have investigated the statistical properties of the TGH random field, proposed an efficient estimation approach and developed an extreme spatial outlier detection procedure based on the TGH random field. The estimation and predictive performances of the TGH random field were evaluated through extensive simulation studies and an application to a precipitation dataset, all of which demonstrated the effectiveness of the proposed

model.

One limitation of the current work lies in that if the measurement error is of interest, it can only be included in the Tukey g-and-h transformation function, as we did in the model (20) of the rainfall application. While this may make sense for some applications, an interesting alternative is to consider a TGH random field with additive measurement errors of the form $Y(\mathbf{s}) = \xi + \mathbf{X}(\mathbf{s})^{\mathrm{T}}\boldsymbol{\beta} + \omega\tau_{g,h}\{Z(\mathbf{s})\} + \varepsilon(\mathbf{s})$, where $\varepsilon(\mathbf{s})$ is a Gaussian measurement error process. The current Algorithm I cannot be directly used to estimate such a model because the joint finite-dimensional distribution of the random field $\tau_{g,h}\{Z(\mathbf{s})\} + \varepsilon(\mathbf{s})$ is a convolution of a multivariate Tukey g-and-h distribution given in Lemma 3 and a multivariate normal distribution, which is quite complicated. As a result, the likelihood function does not have a closed form as given in (7) and thus the Algorithm I is no longer applicable. We believe this issue can be resolved under a suitable Bayesian framework and this will be studied in future work.

References

Azzalini, A. and Capitanio, A. (2014), *The Skew-Normal and Related Families*, Cambridge University Press.

Allcroft, D. J., and 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.

Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008), "Gaussian predictive process models for large spatial datasets," *Journal of the Royal Statistical Society, Series B*, 70, 825-848.

Butler, A., and Glasbey, C. A. (2008), "A latent Gaussian model for compositional data with zeros," *Journal of the Royal Statistical Society, Series C*, 57, 505-520.

Cressie, N. (1993), *Statistics for Spatial Data*, New York, NY: Wiley,

Cressie, N., and Lahiri, S.N. (1993), "The asymptotic distribution of REML estimators," *Journal of Multivariate Analysis*, 45, 217-233.

Cressie, N., and Johannesson, G. (2008), "Fixed rank kriging for very large spatial processes," *Journal of the Royal Statistical Society, Series B*, 70, 209-226.

Dawid, A. P. (1984), "Statistical theory: The frequential approach," *Journal of the Royal Statistical Society, Series A*, 147, 278-292.

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., and Short, D. A. (1997), "Bayesian prediction of transformed Gaussian random fields," *Journal of the American Statistical Association*, 92, 1422-1433.

De Oliveira, V., and Rui, C. (2009), "On shortest prediction intervals in log–Gaussian random fields," *Computational Statistics & Data Analysis*, 53, 4345-4357.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), "Statistical Theory: The Prequential Approach," *Evaluating Density Forecasts With Applications to Financial Risk Management*, 39, 863-883.

Field, C. (2004), "Using the *gh* distribution to model extreme wind speeds," *Journal of Statistical Planning and Inference*, 122, 15-22.

Field, C., and Genton, M. G. (2006), "The multivariate g-and-h distribution," *Technometrics*, 48, 104-111.

Fonseca, T. C. O., and Steel, M. F. J. (2011), "Non-Gaussian spatiotemporal modelling through scale mixing," *Biometrika*, 98, 761-774.

Genton, M. G. (2004), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall/CRC, Boca Raton, FL.

Genton, M. G., and Zhang, H. (2012), "Identifiability problems in some non-Gaussian spatial random fields," *Chilean Journal of Statistics*, 3, 171-179.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007), "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society, Series B*, 69, 243-268.

Gneiting, T., and Katzfuss, M. (2014), "Probabilistic forecasting," *Annual Review of Statistics and Its Application*, 1, 125-151.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006), "Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method," *Journal of the American Statistical Association*, 101, 968-979.

Gräler, B. (2014), "Modelling skewed spatial random fields through the spatial vine copula," *Spatial Statistics*, 10, 87-102.

He, Y., and Raghunathan, T. E. (2012), "Multiple imputation using multivariate gh transformations," *Journal of Applied Statistics*, 39, 2177-2198.

Hoaglin, D. C. (1985), "Summarizing shape numerically: The g-and-h distributions," in *Data Analysis for Tables, Trends and Shapes: Robust and Exploratory Techniques*, D. C. Hoaglin, F. Mosteller, J. W. Tukey (eds.), New York: Wiley.

Hyvärinen, A. (2005), "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, 6, 695-709.

Johns, C. J., Nychka, D., Kittel, G. F., and Daly, C. (2003), "Infilling sparse records of spatial fields," *Journal of the American Statistical Association*, 98, 796-806.

Kim, H. M., and Mallick, B. K. (2012), "A Bayesian prediction using the skew Gaussian distribution," *Journal of Statistical Planning and Inference*, 120, 85-101.

Mardia, K. V., and Marshall, R. J. (1984), "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, 71, 135-146.

Marchenko, Y. V., and Genton, M. G. (2010), "Multivariate log-skew-elliptical distributions with applications to precipitation data," *Environmetrics*, 21, 318-340.

Martinez, J., and Iglewicz, B. (1984), "Some properties of the Tukey *g* and *h* family of distributions," *Communications in Statistics, Theory and Methods*, 13, 353-369.

North, G. R., Wang, J., and Genton, M. G. (2011), "Correlation models for temperature fields," *Journal of Climate*, 24, 5850-5862.

Palacios, M. B., and Steel, M. F. J. (2006), "Non-Gaussian Bayesian geostatistical modeling," *Journal of the American Statistical Association*, 101, 604-618.

Rimstad, K., and Omre, H. (2014), "Skew-Gaussian random fields," *Spatial Statistics*, 10, 43-62.

Røislien, J., and Omre, H. (2006), "T-distributed random fields: a parametric model for heavy-tailed well-log data," *Mathematical Geology*, 38, 821-849.

Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications.* Chapman & Hall/CRC.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

Sun, Y., Li, B., and Genton, M. G. (2012), "Geostatistics for large datasets," in *Space-Time Processes and Challenges Related to Environmental Problems*, E. Porcu, J. M. Montero, M. Schlather (eds), Springer, Vol. 207, Chapter 3, 55-77.

Sweeting, T. J. (1980), "Uniform asymptotic normality of the maximum likelihood estimator," *Annals of Statistics*, 8, 1375-1381.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.

Wallin, J., and Bolin, D. (2015), "Geostatistical modelling using non-Gaussian Matérn fields," *Scandinavian Journal of Statistics*, 42, 872-890.

Xu, G., and Genton, M. G. (2015), "Efficient maximum approximated likelihood inference for Tukey's *g*-and-*h* distribution," *Computational Statistics & Data Analysis*, 91, 78-91.

Xu, G., Liang, F., and Genton, M. G. (2015), "A Bayesian spatio-temporal geostatistical model with an auxiliary lattice for large datasets," *Statistica Sinica*, 25, 61-79.

Zhang, H. (2004), "Inconsistent estimation and asymptotically equal interpolations in model-based Geostatistics," *Journal of the American Statistical Association*, 99, 250-261.

Zhang, H., and El-Shaarawi, A. (2010), "On spatial skew-Gaussian processes and applications," *Environmetrics*, 21, 33-47.

Zhang, H., and Zimmerman, D. L. (2005), "Towards reconciling two asymptotic frameworks in spatial statistics," *Biometrika*, 92, 921-936.

Zhu, X., and Genton, M. G. (2012), "Short-term wind speed forecasting for power system operations," *International Statistical Review*, 80, 2-23.

Table 1: Empirical biases and RMSE (in parentheses) of the parameter estimators of the TGH random field using Algorithm I for various values of (g, h) and of parameters defined in the text.

(g, h)	n	$\hat{g}$	$\hat{h}$	$\hat{\xi}$	$\hat{\omega}$	$\hat{\phi}$	$\hat{\nu}$	$\hat{\beta}$	$\hat{\beta}_0$
$g = 0$	10^2	0.000	0.008	−0.03	−0.11	−3.37	0.45	0.000	0.01
$h = 0$		(0.08)	(0.03)	(0.45)	(0.25)	(9.66)	(1.36)	(0.09)	(0.18)
	15^2	−0.001	0.004	0.01	−0.06	−1.82	0.16	−0.000	0.004
		(0.05)	(0.01)	(0.33)	(0.16)	(6.48)	(0.52)	(0.06)	(0.13)
	20^2	0.001	0.004	−0.01	−0.03	−0.68	0.07	−0.001	−0.006
		(0.04)	(0.01)	(0.25)	(0.12)	(5.04)	(0.19)	(0.04)	(0.10)
$g = 0$	10^2	n/a	n/a	−0.03	−0.07	−3.04	0.43	0.000	0.01
$h = 0$		n/a	n/a	(0.43)	(0.23)	(9.81)	(1.35)	(0.09)	(0.18)
(GRF	15^2	n/a	n/a	0.01	−0.03	−1.51	0.15	−0.000	0.004
Estim.)		n/a	n/a	(0.32)	(0.15)	(6.62)	(0.52)	(0.06)	(0.13)
	20^2	n/a	n/a	−0.01	−0.01	−0.44	0.06	−0.002	−0.006
		n/a	n/a	(0.25)	(0.11)	(5.07)	(0.19)	(0.04)	(0.10)
$g = 0.5$	10^2	−0.03	−0.03	0.01	−0.02	−3.25	0.41	0.001	0.02
$h = 0.1$		(0.13)	(0.07)	(0.46)	(0.37)	(9.61)	(1.30)	(0.09)	(0.29)
	15^2	−0.02	−0.01	0.02	−0.004	−1.86	0.16	0.001	0.01
		(0.10)	(0.05)	(0.32)	(0.25)	(6.73)	(0.56)	(0.06)	(0.20)
	20^2	−0.006	−0.007	0.005	0.000	−0.82	0.07	−0.001	−0.01
		(0.07)	(0.03)	(0.26)	(0.19)	(5.08)	(0.20)	(0.04)	(0.15)
$g = 0.5$	10^2	−0.03	−0.04	0.003	−0.01	−3.06	0.43	0.004	0.02
$h = 0.3$		(0.22)	(0.11)	(0.47)	(0.40)	(9.69)	(1.32)	(0.11)	(0.59)
	15^2	−0.01	−0.02	0.01	−0.01	−1.83	0.16	0.002	−0.004
		(0.16)	(0.08)	(0.33)	(0.25)	(6.67)	(0.55)	(0.06)	(0.47)
	20^2	−0.01	−0.01	0.000	0.004	−0.78	0.07	−0.001	−0.02
		(0.13)	(0.06)	(0.27)	(0.20)	(5.11)	(0.20)	(0.05)	(0.31)

Table 2: Coverage probabilities (CP) and average lengths (Length) of prediction CIs.

(g, h)	Method	CP(50%)	Length	CP(90%)	Length
$g = -0.5$	gh-pred	47.9%	2.11	88.5%	6.14
$h = 0.2$	g-pred	54.7%	2.55	89.6%	6.39
	Gau-pred	45.9%	3.51	81.5%	8.56
$g = 0.5$	gh-pred	48.0%	2.15	88.5%	6.22
$h = 0.2$	g-pred	55.1%	2.58	89.6%	6.49
	Gau-pred	47.4%	3.81	82.0%	9.27
$g = -0.5$	gh-pred	47.9%	2.51	88.7%	8.62
$h = 0.4$	g-pred	64.3%	4.31	91.1%	10.79
	Gau-pred	61.7%	7.70	89.1%	18.80
$g = 0.5$	gh-pred	47.9%	2.48	88.5%	8.37
$h = 0.4$	g-pred	63.3%	4.17	90.9%	10.42
	Gau-pred	61.5%	7.58	89.2%	18.35

Table 3: Parameter estimates of the precipitation data

	Model	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\xi}$	$\hat{\omega}$	$\hat{g}$	$\hat{h}$	$\hat{\phi}$	$\hat{\nu}$	$\hat{\sigma}_v^2$	$\hat{\sigma}_e^2$
Without	TGH	0.55	1.41	0.38	9.53	3.98	0.31	0.095	8.16	0.73	0.81	0.19
Screen	(bLen)	1.10	1.10	0.55	3.61	1.23	0.23	0.07	3.93	1.07	0.16	0.16
	RGRF	0.16	0.07	0.09	3.18	*n/a*	*n/a*	*n/a*	13.11	0.93	1.19	0.13
	(bLen)	0.98	0.75	0.09	1.60	*n/a*	*n/a*	*n/a*	10.54	1.03	1.31	0.03
	GRF	0.73	0.29	0.55	10.98	*n/a*	*n/a*	*n/a*	10.55	1.20	45.61	4.76
	(bLen)	5.86	4.24	0.56	7.91	*n/a*	*n/a*	*n/a*	8.30	1.60	48.26	1.16
With	TGH	−0.14	1.53	0.31	9.56	4.15	0.29	0.045	8.41	0.65	0.86	0.14
Screen	(bLen)	1.47	1.26	0.49	4.01	1.47	0.16	0.06	4.09	0.78	0.14	0.14
	RGRF	0.19	0.07	0.09	3.19	*n/a*	*n/a*	*n/a*	15.48	0.78	1.25	0.09
	(bLen)	0.98	0.94	0.11	1.82	*n/a*	*n/a*	*n/a*	10.02	0.73	0.83	0.03
	GRF	0.90	0.37	0.54	11.13	*n/a*	*n/a*	*n/a*	12.44	1.03	51.21	3.42
	(bLen)	6.92	5.38	0.58	10.17	*n/a*	*n/a*	*n/a*	8.87	0.98	50.00	0.83

Table 4: Predictive performances of models for the precipitation data

	Model	*MAD*	*mCRPS*	50%CI		90%CI	
				CP	mLen	CP	mLen
Without Screen	TGH	1.32	0.82	54.5%	2.65	91.1%	6.74
	RGRF	1.29	0.85	59.0%	3.17	92.0%	7.74
	GRF	1.30	0.84	59.8%	3.28	92.4%	8.00
With Screen	TGH	1.11	0.76	52.7%	2.47	89.8%	6.14
	RGRF	1.20	0.77	54.9%	2.73	90.0%	6.65
	GRF	1.23	0.77	56.8%	2.87	90.6%	6.99
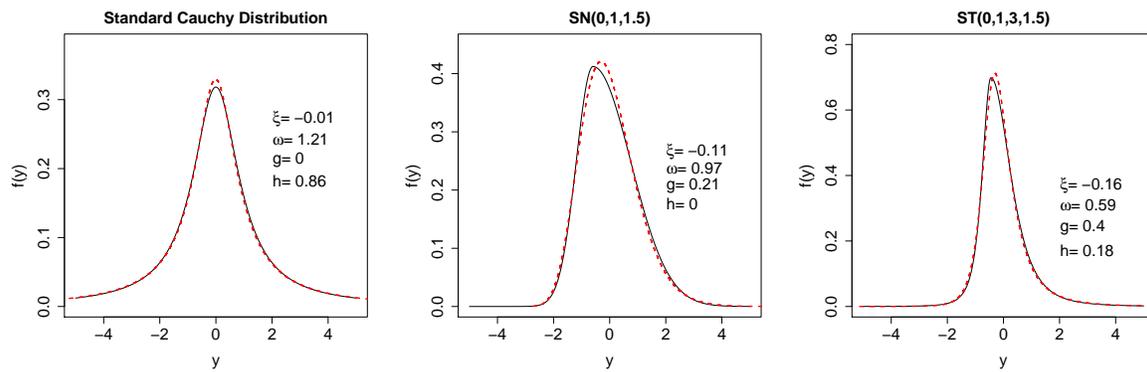
Figure 1: TGH distribution (red dash line) approximations to several distributions (solid line): (left) standard cauchy; (middle) skew-normal with location 0, scale 1 and skewness 1.5; (right) skew-*t* with location 0, scale 1, degrees-of-freedom 3 and skewness 1.5.
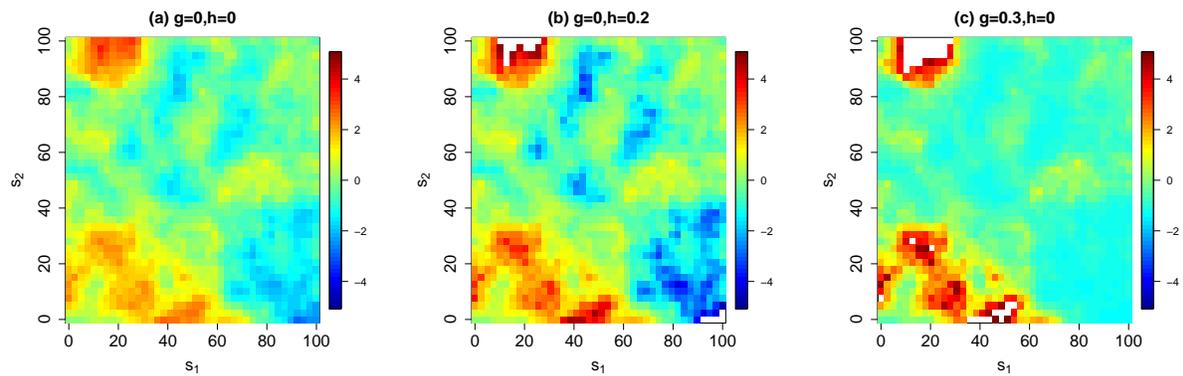
Figure 2: Realizations of standard TGH random fields (after removing the mean) with Matérn spatial correlation function (17).
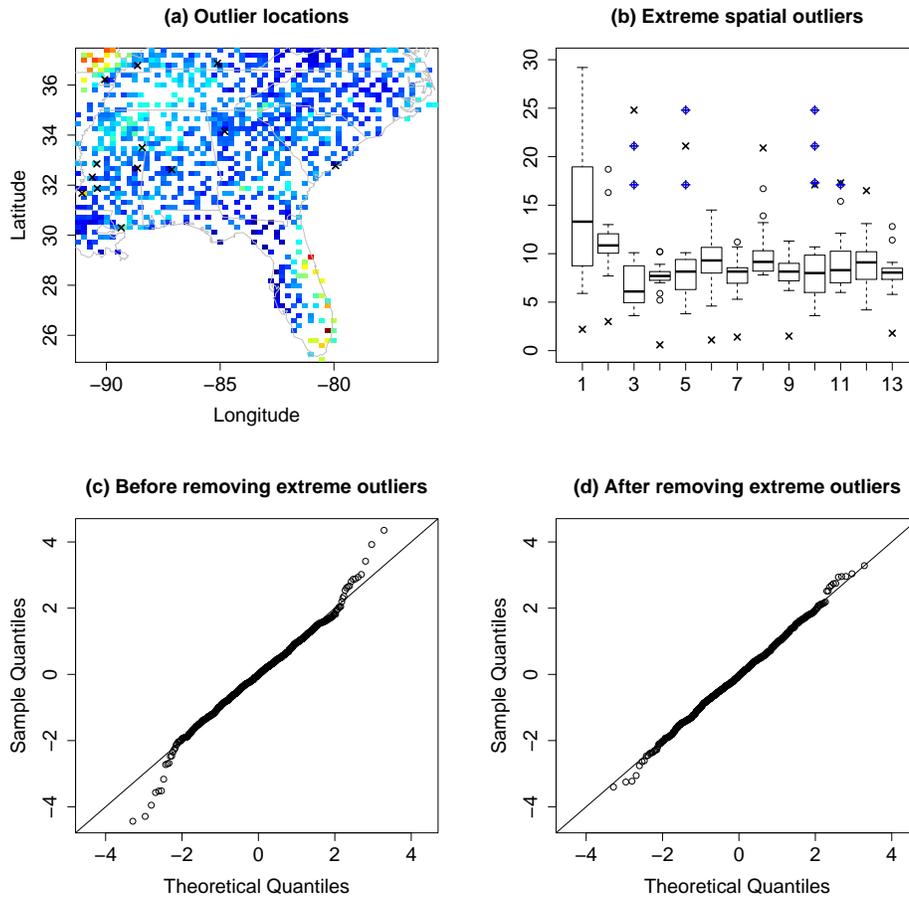
Figure 3: (a) Locations ("×") of 13 extreme outliers; (b) Boxplots of 13 extreme identified outliers ("×") and their nearest 20 neighbors, ("+" indicate other identified outliers in its neighborhoods); (c)-(d) Q-Q plot of components in $\widehat{\mathbf{Z}}^{int}$ in (10) and $\widehat{\mathbf{Z}}^{psc}$.
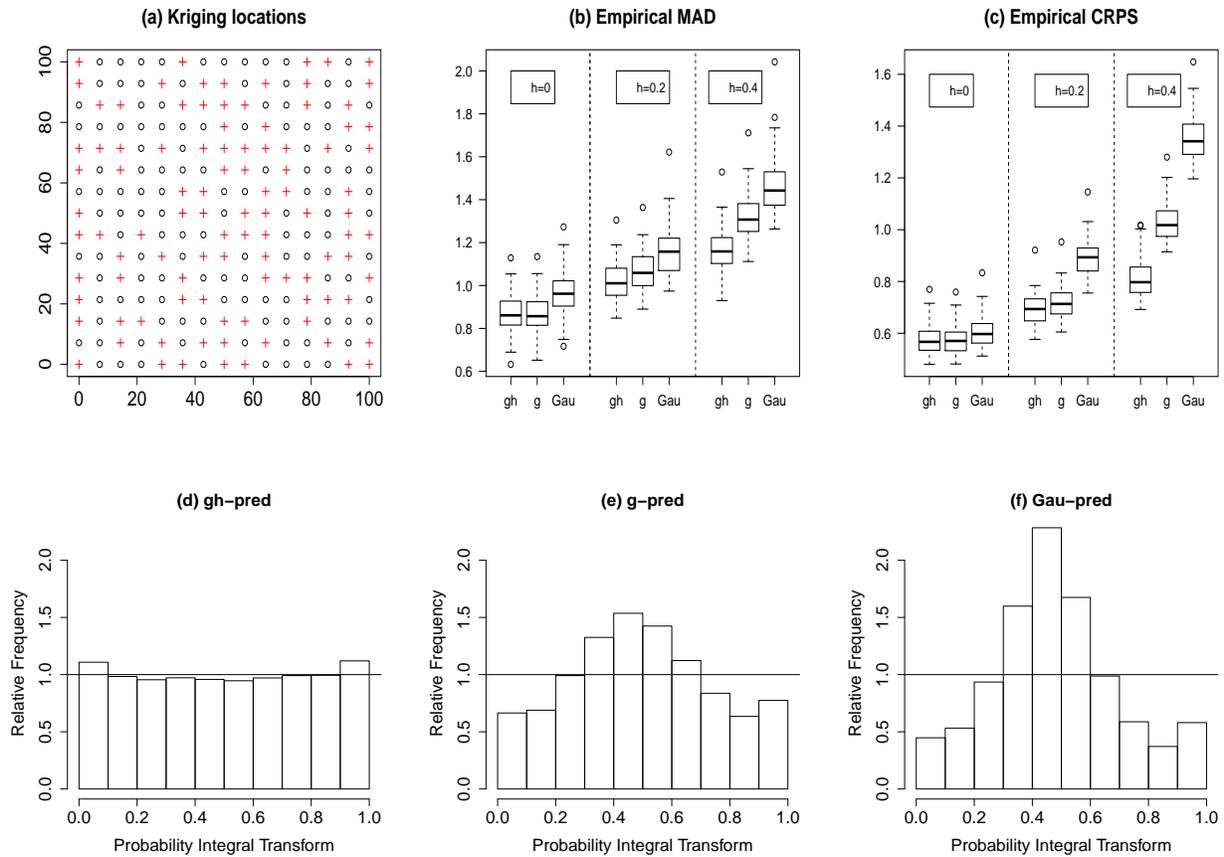
Figure 4: (a) Spatial locations: "∘" training sites and "+" prediction sites; (b)-(c) $MAD(\mathbf{s}_i^p)$'s and $mCRPS(\mathbf{s}_i^p)$'s of three kriging approaches for $g = 0.5$ and $h = 0, 0.2, 0.4$ at 100 prediction locations; (d)–(f) PIT histograms of three predictive distributions for $g = 0.5$ and $h = 0.2$.
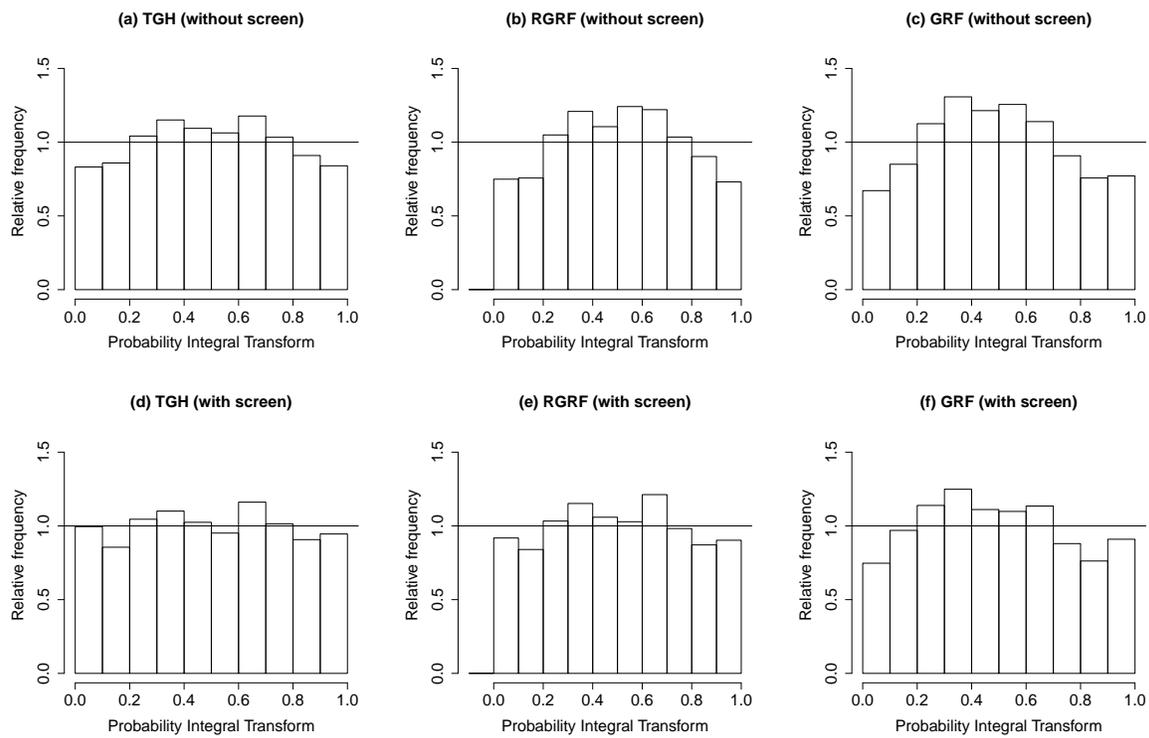
Figure 5: The PIT histogram of the probabilistic predictions made by three models: (a)-(c) without screen; (d)-(f) with screen.
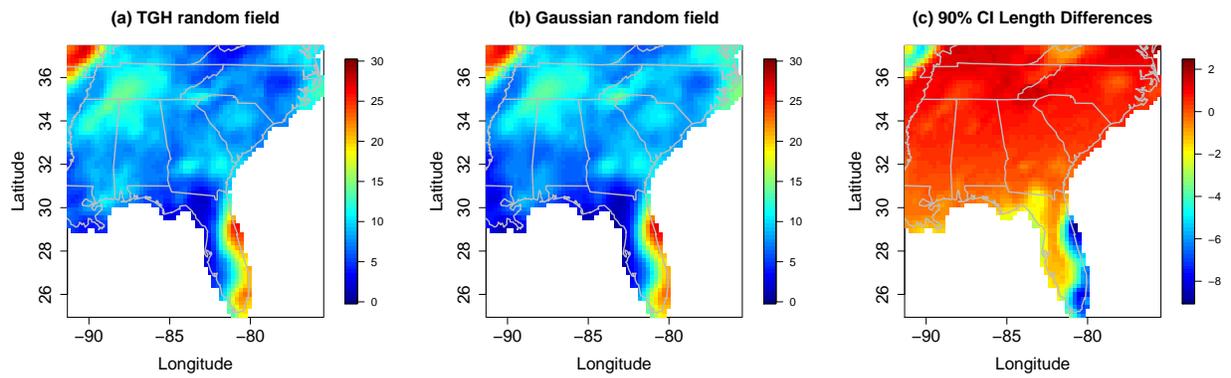
40

Figure 6: Predicted values using (a) TGH random field (with screen) and (b) Gaussian random field; (c) Length differences of 90% prediction intervals (Gaussian−TGH).