



Multivariate transformed Gaussian processes

Yuan Yan¹ · Jaehong Jeong² · Marc G. Genton³

Received: 24 July 2019 / Accepted: 26 November 2019 / Published online: 26 December 2019
© Japanese Federation of Statistical Science Associations 2019

Abstract

We set up a general framework for modeling non-Gaussian multivariate stochastic processes by transforming underlying multivariate Gaussian processes. This general framework includes multivariate spatial random fields, multivariate time series, and multivariate spatio-temporal processes, whereas the respective univariate processes can also be seen as special cases. We advocate joint modeling of the transformation and the cross-/auto-correlation structure of the latent multivariate Gaussian process, for better estimation and prediction performance. We provide two useful models, the Tukey g -and- h transformed vector autoregressive model and the sinh-arcsinh-transformed multivariate Matérn random field. We evaluate them with a simulation study. Finally, we apply the two models to a wind data set for modeling the two perpendicular components of wind speed vectors. Both the simulation study and data analysis show the advantages of the joint modeling approach.

Keywords Heavy tails · Multivariate random fields · Multivariate time series · Non-Gaussian · Skewness

1 Introduction

Along with the abundance of data collected in space and/or time, the need for analyzing multiple spatially and/or temporally correlated response variables at the same time is ever-increasing in various fields. To model a multivariate stochastic process

✉ Yuan Yan
yuan.yan@dal.ca

Jaehong Jeong
jaehong.jeong@maine.edu

Marc G. Genton
marc.genton@kaust.edu.sa

¹ Department of Mathematics and Statistics, Dalhousie University, Halifax, NS B3H 4R2, Canada

² Department of Mathematics and Statistics, University of Maine, Orono, ME 04469, USA

³ Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

jointly, it is essential to take the cross-correlation structure between variables into account in addition to the auto-correlation structure for each of the univariate processes. For example, in climate science, surface pressure and temperature are known to be negatively correlated (Gneiting et al. 2010); in environmental science, the concentrations of different pollutants are correlated (Genton and Kleiber 2015); in astronomy, to simulate multifrequency sky maps, backgrounds, sources, and noise are correlated (Vio et al. 2002).

The Gaussian process is the core in stochastic process modeling because its distribution is entirely determined by specifying the mean and covariance function, and properties of likelihood-based inference have been studied extensively. However, real data collected in different areas often show traits of skewness and a lighter or heavier tail that could not be accommodated by a Gaussian model. Therefore, flexible multivariate non-Gaussian spatio-temporal models that are also computationally feasible are needed to analyze big data with multiple variables collected in space and/or time.

As a general setting, we refer to a stochastic process $\mathbf{Y}(\mathbf{x}) = \{Y_1(\mathbf{x}), \dots, Y_k(\mathbf{x})\}^T$ as a collection of random variables $Y \in \mathbb{R}^k$ indexed by elements in a topological space $\mathbf{x} \in \mathbb{R}^q$, which include spatial random fields when the indices are spatial location, $\mathbf{x} = \mathbf{s} \in D_s \subset \mathbb{R}^2$, time series when $\mathbf{x} = t \in D_t = \{1, 2, 3, \dots\}$, and spatio-temporal processes when $\mathbf{x} = (\mathbf{s}, t)$. The stochastic process $\mathbf{Y}(\mathbf{x})$ could be a univariate process when $k = 1$ or a process with multiple variables when $k \geq 2$. In this paper, we provide a general framework for modeling a non-Gaussian multivariate stochastic process $\mathbf{Y}(\mathbf{x})$ by applying a component-wise transformation $\boldsymbol{\psi} : \mathbb{R}^k \rightarrow \mathbb{R}^k$ to an underlying multivariate Gaussian process $\mathbf{Z}(\mathbf{x}) = \{Z_1(\mathbf{x}), \dots, Z_k(\mathbf{x})\}^T$:

$$\mathbf{Y}(\mathbf{x}) = \boldsymbol{\psi}\{\mathbf{Z}(\mathbf{x})\}, \quad (1)$$

or more explicitly, $Y_j(\mathbf{x}) = \psi_j\{Z_j(\mathbf{x})\}$, $\psi_j : \mathbb{R} \rightarrow \mathbb{R}$, $j = 1, \dots, k$.

Each component of $\boldsymbol{\psi}$ could come from the same parametric family of monotonic transformations with different values of parameters for each variable $\psi_j = \psi_{\theta_1^j}$, to account for various shapes of the marginal distribution shown by different variables. We further suppose the dependence structure of the underlying Gaussian process $\mathbf{Z}(\mathbf{x})$ can be characterized by another set of parameters $\boldsymbol{\theta}_2$. Given n observations $\mathbf{Y} = \{\mathbf{Y}^T(\mathbf{x}_1), \dots, \mathbf{Y}^T(\mathbf{x}_n)\}^T$ from the non-Gaussian process $\mathbf{Y}(\mathbf{x})$, the joint log-likelihood can be derived:

$$\log L(\boldsymbol{\theta}; \mathbf{Y}) = f_\phi(\mathbf{Z}_{\boldsymbol{\theta}_1}, \boldsymbol{\theta}_2) - \sum_{j=1}^k \sum_{i=1}^n \log \left[\psi'_{\theta_1^j} \{Z_{\theta_1^j}(\mathbf{x}_i)\} \right], \quad (2)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)^T$ is the full set of parameters, f_ϕ denotes the log-likelihood of a zero mean multivariate Gaussian distribution with covariance matrix determined by $\boldsymbol{\theta}_2$, $\psi'_{\theta_1^j}(\cdot)$ is the derivative of the function $\psi_{\theta_1^j}(\cdot)$, $Z_{\theta_1^j}(\mathbf{x}_i) = \psi_{\theta_1^j}^{-1}\{Y_j(\mathbf{x}_i)\}$, $j = 1, \dots, k$ are from the latent Gaussian process obtained by applying the corresponding inverse transformation to the observations, and $\mathbf{Z}_{\boldsymbol{\theta}_1} = \{\mathbf{Z}_{\theta_1}^T(\mathbf{x}_1), \dots, \mathbf{Z}_{\theta_1}^T(\mathbf{x}_n)\}^T$ is the vector of

all the Gaussian-scale observations. Maximum likelihood estimation (MLE) of θ can be found by maximizing (2).

Prediction of the non-Gaussian process at an unobserved index \mathbf{x}_0 is based on the conditional distribution $Y(\mathbf{x}_0)|Y$. Since the transformation is monotonic, this conditional distribution problem in the transformed scale can be translated directly to the Gaussian scale. Assuming the true model is known, let Z and $Z(\mathbf{x}_0)$ be the latent Gaussian realizations corresponding to Y and $Y(\mathbf{x}_0)$. It follows that $Z(\mathbf{x}_0)|Z \sim \mathcal{N}_k(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$, where the conditional mean $\tilde{\boldsymbol{\mu}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}$ are determined by the dependence structure of the underlying Gaussian process. Then,

$$Y(\mathbf{x}_0)|Y \sim \boldsymbol{\psi}\{Z(\mathbf{x}_0)|Z\}, \quad (3)$$

the right hand side of which is a shorthand for the distribution of transforming a k -dimensional random variable that follows the conditional Gaussian distribution. Therefore, realizations of $Y(\mathbf{x}_0)|Y$ are the same as realizations of $\boldsymbol{\psi}\{Z(\mathbf{x}_0)|Z\}$. Thus, point and probabilistic prediction of $Y(\mathbf{x}_0)$ can be formulated. In particular, the optimal point prediction that minimizes the absolute loss is the conditional median $\hat{Y}_j^{\text{opt}}(\mathbf{x}_0) = \arg \min_c E\left[\left\{|Y_j(\mathbf{x}_0) - c|\right\}|Y\right] = \text{med}\{Y_j(\mathbf{x}_0)|Y\} = \psi_j(\tilde{\mu}_j)$.

This general trans-Gaussian approach is flexible, interpretable, and parsimonious by taking advantage of both a family of transformations and the appealing properties of the Gaussian process. In Sect. 2, we provide a literature review of existing approaches for modeling non-Gaussian processes, and compare them with, or put them within, our general framework in Eq. (1). In Sect. 3, we present two useful non-Gaussian models for multivariate times series and multivariate spatial random fields: the multivariate Tukey g -and- h (TGH) transformed vector autoregressive (VAR) model and the multivariate sinh–arcsinh (SAS) transformed Matérn random field. Section 4 illustrates estimation and prediction performance of the bivariate SAS-transformed parsimonious Matérn random fields with a simulation study. We then, in Sect. 5, apply the two models to a wind data set for modeling bivariate wind speeds that come from the two perpendicular directions. Both the simulation study and data analysis show the advantages of the joint modeling approach. Finally, in Sect. 6, we discuss a potential area for further research.

2 Literature review on non-Gaussian processes

For non-Gaussian data, the transformation approach is straightforward and widely used among practitioners to, first, Gaussianize the data and then examine the latent Gaussian process on the transformed scale. In light of the general framework given by Eq. (1), it is equivalent to Gaussianizing the data with $\boldsymbol{\psi}^{-1}$. For example, Shumway and Stoffer (2011) used the logarithm transformation on glacial varves thicknesses time series and quarterly U.S. GNP time series whereas the square-root transformation was adopted by Johns et al. (2003) to normalize spatial precipitation data. However, a fixed transformation is not flexible enough to adapt to various non-Gaussian features exhibited by the data. A parametric form of transformations that allows modelers to find the optimal one within a family of transformations by

estimating the related parameters from the data is thus preferable. For example, the Box–Cox power transformation (Box and Cox 1964) is a family of transformations with one parameter that includes the logarithm, square- and cubic-root, but it can only be applied to positive values. Instead of the two-stage procedure, to first estimate θ_1 without considering the dependent structure of the observations, and then model the Gaussianized process, modeling the trans-Gaussian process directly with joint likelihood (2) is beneficial for estimating the optimal transformation as well as making a prediction on the original scale. Snelson et al. (2004) outlined a general procedure and coined the term ‘warped Gaussian processes’ for univariate transformed Gaussian processes to include the transformation as an integral part of the model, rather than an ad-hoc step. In particular, Block et al. (1990) presented a non-Gaussian time series model with arbitrary continuous marginal distributions that belong to a parametric family by transforming a Gaussian autoregressive (AR) time series by the composition of the inverse marginal cumulative distribution function (CDF) and the Gaussian CDF. Cressie (1993) introduced log-normal random fields, developed log-normal kriging methods, and further generalized the methodology to trans-Gaussian kriging to derive unbiased predictors with unknown mean but known covariance structure of the latent Gaussian random fields. De Oliveira et al. (1997) considered Box–Cox transformed random fields for spatial data under a Bayesian framework and applied the model to predict spatial weekly rainfall amounts. Zammit-Mangion et al. (2016), for modelling mole-fraction and flux fields, relaxed the Gaussian assumption through Box–Cox transformations for the spatio-temporal bivariate inversion problem and performed Bayesian inference via approximation of the likelihood under the hierarchical framework. Recently, Xu and Genton (2017) and Yan and Genton (2019a) introduced the TGH random fields and the TGH-AR process, that can accommodate different levels of skewness and tail heaviness in marginal distributions. Zhang and Yeung (2010) attempted to extend the general warped Gaussian processes (Snelson et al. 2004) to a multivariate setting, with an application in personalized age estimation. However, their model essentially utilizes one latent univariate Gaussian process with white noise of different variance added to different variables. Vio et al. (2001, 2002) discussed the associations between the cross-correlation functions in Gaussian-scale \mathbf{Z} and the transformed-scale \mathbf{Y} under the general framework (2), for the purpose of simulating multivariate random fields with prescribed marginal distributions and cross-correlation structure.

Another popular approach of modeling non-Gaussian data is via the generalized linear model (GLM), which is especially useful for analyzing binary and count data. Under the spatial setting, the GLM approach was first proposed by Gotway and Stroup (1997) and then formalized by Diggle et al. (1998) with a latent Gaussian random field as a random effect. The hierarchical flavor of the GLM approach fits naturally within a Bayesian framework, in which the latent Gaussian random field can be regarded as a prior put on the space of functions, see Banerjee et al. (2014) and Cressie and Wikle (2011) for book-length overviews. Chagneau et al. (2011) extended the hierarchical Bayesian spatial GLM idea to build multivariate non-Gaussian random fields. In the spatial GLM models, a random field is added as a random effect in the linear predictor; therefore, the likelihood function involves high-dimensional integration, which makes computational evaluation challenging.

Instead of putting a Gaussian prior, Bradley et al. (2019) introduced the ‘conjugate multivariate distribution’ as prior for the latent process at the n observed locations to facilitate Bayesian inference of dependent data from the exponential families. In the time series context, Benjamin et al. (2003) introduced the generalized autoregressive moving average (ARMA) model. Unlike the spatial case, no underlying Gaussian time series is added as a random effect, thanks to the unidirectional property of time. However, the GLM approach is not flexible enough to model continuous data with arbitrary distribution shape. Cordeiro and de Andrade (2009) combined both the transformation and GLM approaches and introduced the transformed generalized linear models.

The third class of models for constructing a non-Gaussian process involves mixing a Gaussian process with a random variable (common mixing variable) or another process (individual mixing variable) or a mixture of several independent processes. Wong and Li (2000) introduced the mixture AR models with a mixture of m Gaussian AR components, which is a generalization of the Gaussian mixture transition distribution models (GMTD) (Le et al. 1996). Later, Wong et al. (2009) extended the Gaussian mixture AR models to the Student t -mixture AR model. Palacios and Steel (2006) proposed scale mixing Gaussian random fields by combining a Gaussian random field with spatially correlated scale parameters (a log-Gaussian random field). Later, Fonseca and Steel (2011) extended the idea to non-Gaussian processes in space and time with non-separable covariance functions. Ma (2009) introduced a general class of non-Gaussian random fields, the so-called elliptically contoured random fields, by multiplying a Gaussian random field with a random scale factor. However, Genton and Zhang (2012) pointed out identifiability problems related to the above construction. Ma (2010) built the χ^2 random process by summation of the squares of m independent Gaussian processes and extended it to the multivariate case (Ma 2011). A series of papers with the same author combined the above two ideas and constructed many special cases for non-Gaussian vector random fields, such as the hyperbolic vector random fields (Du et al. 2012) and the K-distributed process (Ma 2013). Bolin et al. (2019) presented multivariate latent Gaussian random fields mixture models that combined a discrete Markov random field with a latent multivariate Gaussian random field and used it for classification purpose. Yin and Craigmire (2018) introduced the heteroscedastic asymmetric spatial process (HASP) formed by transformation and mixing of two dependent univariate Gaussian random fields. All the above models, except those by mixing with a random variable, involve more underlying processes than the resulting one, which can be seen as (1) with the modification that $\boldsymbol{\psi}$ has a more general form of transformations $\boldsymbol{\psi} : \mathbb{R}^m \rightarrow \mathbb{R}^k, m \geq k$, instead of the element-wise transformation.

Besides the above three classes, other non-Gaussian spatial models include: the skew-Gaussian random fields (Zhang and El-Shaarawi 2010; Rimstad and Omre 2014) and log-skew-elliptical random fields (Marchenko and Genton 2010) as extensions from the skew-elliptical family (Azzalini 1985; Azzalini and Dalla Valle 1996; Genton 2004); spatial copula models (Gräler 2014; Krupskii et al. 2018); and non-Gaussian Matérn fields based on stochastic partial differential equations (Wallin and Bolin 2015); additional models for time series are mainly based on the ARMA model with non-Gaussian noise (Lawrance and Lewis 1980; Gaver and Lewis 1980;

Davies et al. 1980; Li and McLeod 1988; Tarami and Pourahmadi 2003). See also the review paper by Jones (2015) for constructing non-Gaussian distributions for a univariate random variable. We attempted to make the literature review as thorough as possible, however, it is still not exhaustive. Other approaches for non-Gaussian processes might exist in other fields, such as machine learning, signal processing and econometrics, that we are unaware of.

3 Two multivariate transformed Gaussian models

In this section, we provide two useful models for multivariate time series and multivariate random fields under our general framework (1). The time series model is an extension of the TGH-AR time series (Yan and Genton 2019a) to a multivariate setting, while the spatial model uses the SAS transformation (Jones and Pewsey 2009) in modeling multivariate trans-Gaussian random fields.

3.1 Multivariate TGH-VAR time series models

The Tukey's g -and- h transformation (Tukey 1977) is a function with two parameters g and h :

$$\tau_{g,h}(z) = \begin{cases} g^{-1} \{ \exp(gz) - 1 \} \exp(hz^2/2), & g \neq 0, \quad z \in \mathbb{R}, \\ z \exp(hz^2/2), & g = 0, \quad z \in \mathbb{R}, \end{cases} \quad (4)$$

which is a strictly monotonic function of z when $h \geq 0$.

Tukey (1977) introduced this transformation and used it on a standard Gaussian random variable $Z \sim \mathcal{N}(0, 1)$ to yield $Y = \tau_{g,h}(Z)$, in order to construct skewed distributions with heavy tails and model the quantiles of Y directly rather than the density function (Yan and Genton 2019b). The resulting random variable Y is said to follow the TGH distribution, with parameter g and h controlling the skewness and tail heaviness, respectively. Shapes of the TGH distribution with three sets of parameter values are illustrated in the top row of Fig. 1. Martinez and Iglewicz (1984) studied various properties of the TGH distributions and gave an explicit form of their finite moments. For parameter estimation, since the inverse transformation does not have an explicit form (except when either g or h is equal to 0), the likelihood inference is difficult. Earlier works focused on quantile-based method (Hoaglin 1985; Dutta and Babbal 2002) and later, Xu and Genton (2015) proposed an efficient parameter estimation algorithm for the independent TGH distribution using an approximated likelihood. The maximum approximated likelihood estimation (MALE) substantially improved the parameter estimation performance compared to the moment or quantile-based methods without compromising the computational speed. Field (2004) applied the univariate TGH distribution to model extreme wind speed and showed its adaptability by comparing it with the classical generalized extreme value (GEV) distribution and the generalized Pareto distribution (GPD). Field and Genton (2006) and He and Raghunathan (2012) used different approaches to extend the univariate TGH distribution to a multivariate setting. Recently, Xu and Genton (2017) further

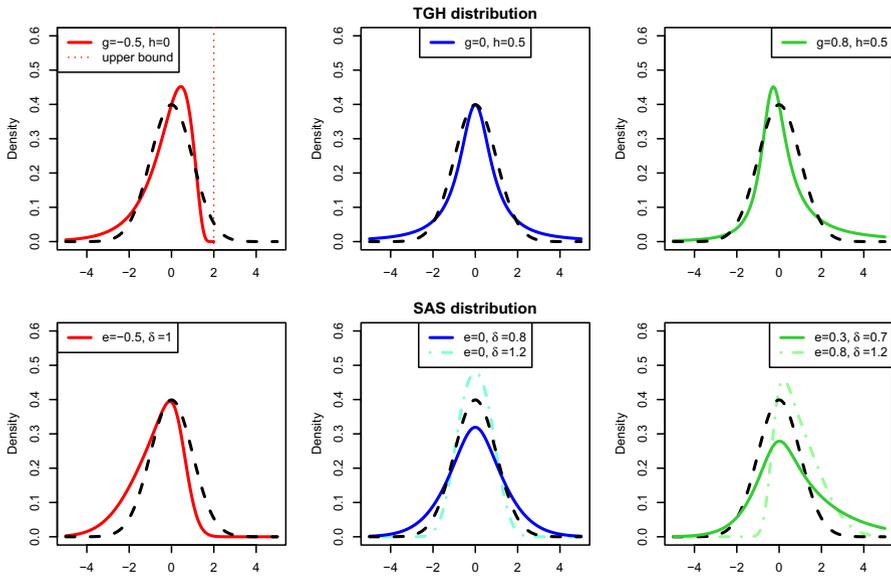


Fig. 1 Density shapes of the TGH distributions (top row) and SAS distributions (bottom row) with different parameters values to control skewness and tail-heaviness. A standard normal density (obtained when $g = h = 0$) is shown in each panel with dashed black line for reference

generalized the TGh distribution to a spatial configuration and constructed TGh random fields, while Yan and Genton (2019a) extended it in a time series setting and introduced the TGh-AR model. Jeong et al. (2019) used the TGh-AR process for building a stochastic generator for global monthly wind speed as a computationally efficient way for uncertainty quantification of wind energy potential.

For this multivariate time series model example, we consider transforming each component of a latent stable Gaussian VAR(1) process Z_t by the TGh transformation with extra location parameters:

$$Y_t = \xi + \tau_{g,h}(Z_t), \quad Z_t = \mathbf{A}Z_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{i.i.d.}{\sim} \mathcal{N}_k(\mathbf{0}, \Psi) \tag{5}$$

where the vector $\xi = (\xi_1, \dots, \xi_k)^T$ contains the location parameters, the vectors $g = (g_1, \dots, g_k)^T \in \mathbb{R}^k, h = (h_1, \dots, h_k)^T \in \mathbb{R}_+^k, \tau_{g,h} : \mathbb{R}^k \rightarrow \mathbb{R}^k$ denotes the element-wise application of the TGh transformation with the corresponding parameters, \mathbf{A} is a $k \times k$ transition matrix for the underlying Gaussian VAR(1) process, and Ψ is the $k \times k$ covariance matrix of the Gaussian white noise process. Notice that, unlike the previous TGh models, in our TGh-VAR(1) model (5), the TGh transformation is applied to an underlying Gaussian time series without the constraints for each component to have variance 1. In fact, the covariance function of the underlying stable Gaussian process Z_t is implied by the dynamic VAR(1) structure and depends on \mathbf{A} and Ψ . Instead of transforming a Gaussian process with unit variance and multiplying by a scale parameter outside the transformation like previous models, we transform the underlying Gaussian VAR process as a whole, which can have

arbitrary marginal variances. Since $\tau_{g,h}(\sigma Z) = \sigma \tau_{g\sigma,h\sigma^2}(Z)$, the scale parameter can be put either inside or outside the TGH transformation and results in the same TGH distribution family.

For model (5), each component of the multivariate time series Y_t follows a TGH distribution marginally with different g and h parameters. Mean and variance of the TGH-VAR process can be computed in similar ways as in Xu and Genton (2017) by the equivalence relationship $\tau_{g,h}(\sigma Z) = \sigma \tau_{g\sigma,h\sigma^2}(Z)$, i.e., with g and h in formulae for the standard TGH distribution substituted by $g\sigma$ and $h\sigma^2$. The mean exists for $h_j\sigma_j^2 < 1$, where σ_j^2 is the marginal variance of Z_{jt} :

$$\mu_{Y_j} = E(Y_{jt}) = \xi_j + \frac{1}{g_j\sqrt{1-h_j\sigma_j^2}} \left[\exp \left\{ \frac{g_j^2\sigma_j^2}{2(1-h_j\sigma_j^2)} \right\} - 1 \right],$$

and the variance exists for $h_j\sigma_j^2 < \frac{1}{2}$:

$$\begin{aligned} \sigma_{Y_j}^2 = \text{var}(Y_{jt}) &= \frac{1}{g_j^2\sqrt{1-2h_j\sigma_j^2}} \left[\exp \left(\frac{2g_j^2\sigma_j^2}{1-2h_j\sigma_j^2} \right) \right. \\ &\quad \left. - 2 \exp \left\{ \frac{g_j^2\sigma_j^2}{2(1-2h_j\sigma_j^2)} \right\} + 1 \right] - (\mu_{Y_j} - \xi_j)^2. \end{aligned}$$

For the cases $g_j = 0$, their values are defined as the limit as $g_j \rightarrow 0$. The cross-covariance can be derived similarly as the covariance matrix in He and Raghunathan (2012):

$$\begin{aligned} \text{cov}(Y_{j_1t_1}, Y_{j_2t_2}) &= \frac{1}{g_{j_1}g_{j_2}\sqrt{(1-h_{j_1}\sigma_{j_1}^2)(1-h_{j_2}\sigma_{j_2}^2) - \rho^2\sigma_{j_1}^2\sigma_{j_2}^2h_{j_1}h_{j_2}}} \\ &\times \left(\exp \left[\frac{\{1-h_{j_2}\sigma_{j_2}^2(1-\rho^2)\}g_{j_1}^2\sigma_{j_1}^2 + \{1-h_{j_1}\sigma_{j_1}^2(1-\rho^2)\}g_{j_2}^2\sigma_{j_2}^2 + 2\rho\sigma_{j_1}\sigma_{j_2}g_{j_1}g_{j_2}}{2\{(1-h_{j_1}\sigma_{j_1}^2)(1-h_{j_2}\sigma_{j_2}^2) - \rho^2\sigma_{j_1}^2\sigma_{j_2}^2h_{j_1}h_{j_2}\}} \right] \right. \\ &\quad - \exp \left[\frac{\sigma_{j_1}^2g_{j_1}^2}{2} \frac{1-h_{j_2}\sigma_{j_2}^2(1-\rho^2)}{(1-h_{j_1}\sigma_{j_1}^2)(1-h_{j_2}\sigma_{j_2}^2) - \rho^2\sigma_{j_1}^2\sigma_{j_2}^2h_{j_1}h_{j_2}} \right] \\ &\quad - \exp \left[\frac{\sigma_{j_2}^2g_{j_2}^2}{2} \frac{1-h_{j_1}\sigma_{j_1}^2(1-\rho^2)}{(1-h_{j_1}\sigma_{j_1}^2)(1-h_{j_2}\sigma_{j_2}^2) - \rho^2\sigma_{j_1}^2\sigma_{j_2}^2h_{j_1}h_{j_2}} \right] + 1 \Big) \\ &\quad - (\mu_{Y_{j_1}} - \xi_{j_1})(\mu_{Y_{j_2}} - \xi_{j_2}), \end{aligned}$$

where $\sigma_{j_1}^2$ and $\sigma_{j_2}^2$ are the marginal variances of Z_{j_1t} and Z_{j_2t} , $\rho = \rho_{j_1j_2}(t_1, t_2)$ is the cross-correlation between the two variables at temporal points t_1 and t_2 , and the above terms depend on \mathbf{A} and $\mathbf{\Psi}$. The pairwise covariance exists if $(1-h_{j_1}\sigma_{j_1}^2)(1-h_{j_2}\sigma_{j_2}^2) - \rho^2\sigma_{j_1}^2\sigma_{j_2}^2h_{j_1}h_{j_2} > 0$.

This model is flexible and can accommodate different tail and skewness behavior of each variable separately. There are $3k$ parameters for θ_1 related to the location TGH

transformation and $k^2 + k(k + 1)/2$ parameters for θ_2 that contains \mathbf{A} and Ψ . Parameter estimation for model (5) can be done by MALE (Xu and Genton 2015), obtained by maximizing the approximated likelihood with piece-wise linearization of the inverse transformation $\tau_{g,h}^{-1}$ involved in the log-likelihood (2). Forecasting of \mathbf{Y}_{t+1} can be based on (3) with the estimated θ and conventional forecasting of Gaussian VAR(1) processes, see standard multivariate time series textbooks (e.g., Lütkepohl 2007).

3.2 Multivariate SAS Matérn random fields

The sinh–arcsinh transformation (Jones and Pewsey 2009) is defined as:

$$S_{e,\delta}(z) = \sinh\{\delta \sinh^{-1}(z) - e\}, \quad e \in \mathbb{R}, \delta \in \mathbb{R}_+. \tag{6}$$

Jones and Pewsey (2009) associated the standard normal random variable Z and a random variable $Y_{e,\delta}$ that follows a SAS distribution by $Z = S_{e,\delta}(Y_{e,\delta})$ or $Y_{e,\delta} = S_{e,\delta}^{-1}(Z) = S_{-e/\delta,1/\delta}(Z)$. The parameter e determines the skewness, and δ controls the tail weight of the SAS distribution. $Y_{e,\delta}$ is positively (negatively) skewed if $e > 0$ ($e < 0$), and has lighter (heavier) tails than the normal if $\delta > 1$ ($\delta < 1$). Shapes of the SAS distribution with three sets of parameter values are shown in the bottom row of Fig. 1. The advantages of the SAS distribution over the TGH distribution are that both the SAS and its inverse transformation belong to the same family and it can model distributions with lighter tails. However, transforming a Gaussian random variable with variance not equal to 1 by the SAS transformation does not result in the canonical SAS distribution. Hence, in order to construct marginally SAS distributed multivariate random fields, we apply the SAS transformation to a latent k -dimensional Gaussian process $\mathbf{Z}(\mathbf{s})$ of mean 0 and variance 1 for each component and some cross-covariance function $\rho_{ij}(\mathbf{s}_k, \mathbf{s}_l)$ by the location-scale version of the SAS transformation:

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\xi} + \boldsymbol{\omega} \circ \mathbf{S}_{e,\delta}^{-1}\{\mathbf{Z}(\mathbf{s})\}, \tag{7}$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_k)^T$, $\boldsymbol{\omega} = (\omega_1, \dots, \omega_k)^T$, $\mathbf{e} = (e_1, \dots, e_k)^T$, $\boldsymbol{\delta} = (\delta_1, \dots, \delta_k)^T$, \circ is the element-wise (Hadamard) product, and $\mathbf{S}_{e,\delta}^{-1} : \mathbb{R}^k \rightarrow \mathbb{R}^k$ denotes the element-wise application of the inverse SAS transformation with the corresponding parameters. Moments of the univariate SAS distribution were given in Jones and Pewsey (2009), however, the derivation is tedious and explicit formulas for cross-moments of the SAS random fields seem difficult to obtain.

Valid models for cross-covariance functions $\rho_{ij}(\mathbf{s}_k, \mathbf{s}_l)$ can be found in the review by Genton and Kleiber (2015). For example, the multivariate Matérn covariance function (Gneiting et al. 2010; Apanasovich et al. 2012) has not only a marginal Matérn correlation but also flexible cross-correlation structures, and it has a form as follows:

$$C_{ii}(\mathbf{h}) = \sigma_i^2 \mathcal{M}(\mathbf{h}|v_i, \phi_i) \text{ and } C_{ij}(\mathbf{h}) = \rho_{ij} \sigma_i \sigma_j \mathcal{M}(\mathbf{h}|v_{ij}, \phi_{ij}), \quad \mathbf{h} \in \mathbb{R}^q.$$

Here, σ_i is the marginal standard deviation, ρ_{ij} measures the strength of the correlation between two random variables at the same location ($\mathbf{h} = \mathbf{0}$), and

$\mathcal{M}(\mathbf{h}|\nu, \phi) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\|\mathbf{h}\|/\phi)^\nu \mathcal{K}_\nu(\|\mathbf{h}\|/\phi)$ is the Matérn correlation function at distance $\|\mathbf{h}\|$ where \mathcal{K}_ν is a modified Bessel function of the second kind of order ν , $\phi > 0$ is a spatial range parameter, and $\nu > 0$ is the smoothness parameter. In particular, the parsimonious version assumes that $\nu_{ij} = 0.5(\nu_i + \nu_j)$ and $\phi_i = \phi_j = \phi_{ij} = \phi$ for all i ; for more details, see Gneiting et al. (2010). On the other hand, Royle and Berliner (1999) and Cressie and Zammit-Mangion (2016) suggested modeling the cross-variable dependence structure via conditional specification in multivariate spatial fields, when known causal relationships imply a natural conditional ordering of the variables.

The MLE for model (7) can be obtained from maximizing the log-likelihood (2). Spatial prediction of $Y(\mathbf{s}_0)$ given $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$ can be made based on (3) and co-kriging of the underlying multivariate Gaussian random field treating the MLE of θ as the true parameters. Co-kriging is a multivariate variant of the ordinary kriging, and it performs predictions for the primary variable (or a poorly sampled variable) with the help of a more readily observed variable. For discussion and technical details, see Myers (1982) and Wackernagel (2013).

Figure 2 shows one realization of the bivariate SAS random field on a unit square for each variable as well as their histograms. Here, variable 1 has a left-skewed lighter tailed distribution, and variable 2 is right-skewed with heavy tails, respectively. The underlying bivariate Gaussian random field has a parsimonious bivariate Matérn cross-covariance function. The true SAS density and the corresponding Gaussian density with the same mean and variance are also shown in the histograms, where we can see a noticeable departure from the Gaussian assumption. In the next section, we run a simulation study for the estimation and prediction performance of the bivariate SAS random field with parsimonious bivariate Matérn cross-covariance function.

4 Simulation study

We perform a Monte Carlo simulation study to assess the estimation and prediction performance of the bivariate SAS parsimonious Matérn random fields and compare the results with alternative methods. We generate observations from the bivariate SAS model with parsimonious Matérn on a perturbed 15×15 grid within a unit square, similar to the design in Yan and Genton (2018). We use 200 points for estimation and the remaining 25 locations for prediction. For the bivariate case with parsimonious Matérn cross-covariance function, there are 8 parameters related to the transformation θ_1 , and 4 parameters involved for the spatial cross-correlation structure θ_2 : the common range parameter ϕ , smoothness parameters ν_1, ν_2 , and ρ for the correlation between the two variables. The model parameters chosen in this simulation study are $\xi_1 = \xi_2 = 0, \sigma_1 = \sigma_2 = 1, e_1 = 0.3, e_2 = 0.1, \delta_1 = 1.2, \delta_2 = 0.7, \phi = 0.05, \nu_1 = 0.5, \nu_2 = 1, \rho = 0.7$, for which variable 1 is right-skewed with lighter tail and variable 2 is heavy tailed and slightly right-skewed. We use the following four methods for estimation and prediction:

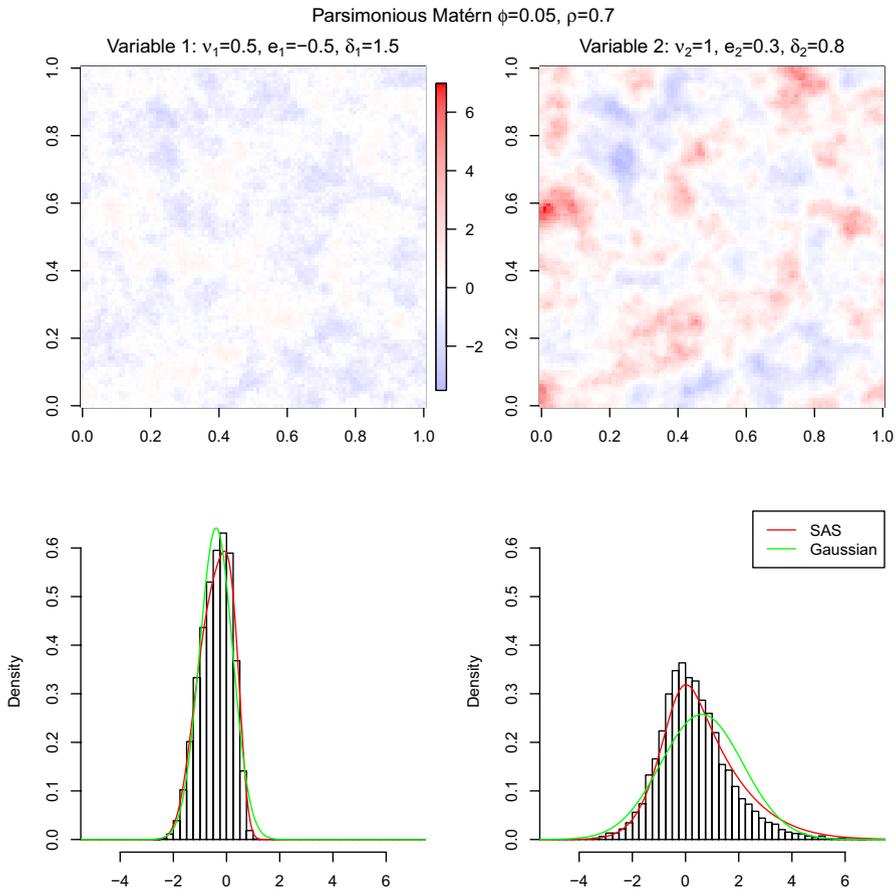


Fig. 2 One realization and histograms of the bivariate SAS random field for each variable (without location-scale parameters) with parsimonious bivariate Matérn cross-covariance function for $\phi = 0.05, \rho = 0.7, v_1 = 0.5, v_2 = 1, e_1 = -0.5, e_2 = 0.3, \delta_1 = 1.5, \delta_2 = 0.8$. Histograms are overlaid with the SAS density (red line) and corresponding Gaussian density with the same mean and variance (green line) (color figure online)

Method 1 (Gaussian): MLE and prediction based on bivariate Gaussian random fields with parsimonious Matérn and unknown constant mean;

Method 2 (Independent SAS): MLE and prediction for each variable separately based on two independent SAS random fields with full Matérn covariance function;

Method 3 (Sequential): two-stage estimation procedure for the bivariate SAS random fields to first estimate parameters related to the SAS transformation for each variable (same as method 2) and then an additional step to estimate parameters related to the bivariate parsimonious Matérn from the underlying Gaussian random fields obtained by transforming the observations with estimated parameters;

Method 4 (Joint): estimate all parameters for the bivariate SAS fields with parsimonious Matérn by maximizing the joint log-likelihood function (2).

All the optimizations for the four estimation methods were done by the R software (Development Core Team 2019) with the Template Model Builder (TMB) package (Kristensen et al. 2016), which utilizes automatic differentiation (Griewank and Walther 2008) to assist optimization.

For estimation, Fig. 3 shows boxplots of the 12 estimated parameters by Method 3 and Method 4 from 500 replicates that are without convergence problem. Parameters in Method 1 and Method 2 are not directly comparable with the 12 parameters in the bivariate parsimonious Matérn SAS fields, and thus boxplots are not shown alongside. The first step in the sequential procedure for estimating the transformation-related parameters θ_1 ignores the correlation of the two variables, which results in slightly larger estimation variability than the joint method and biased estimation for ρ , shown by the boxplots. We also notice that estimation is more unstable for the light-tailed variable 1 than the heavy-tailed variable 2, which can be explained by the fact that the distribution shape does not change much with the skewness parameter when the tail is light. Table 1 shows the number of parameters, mean log-likelihood, and Bayesian Information Criterion (BIC) for estimation by the four methods. The joint estimation has the maximum mean log-likelihood and minimum BIC among the four estimation methods.

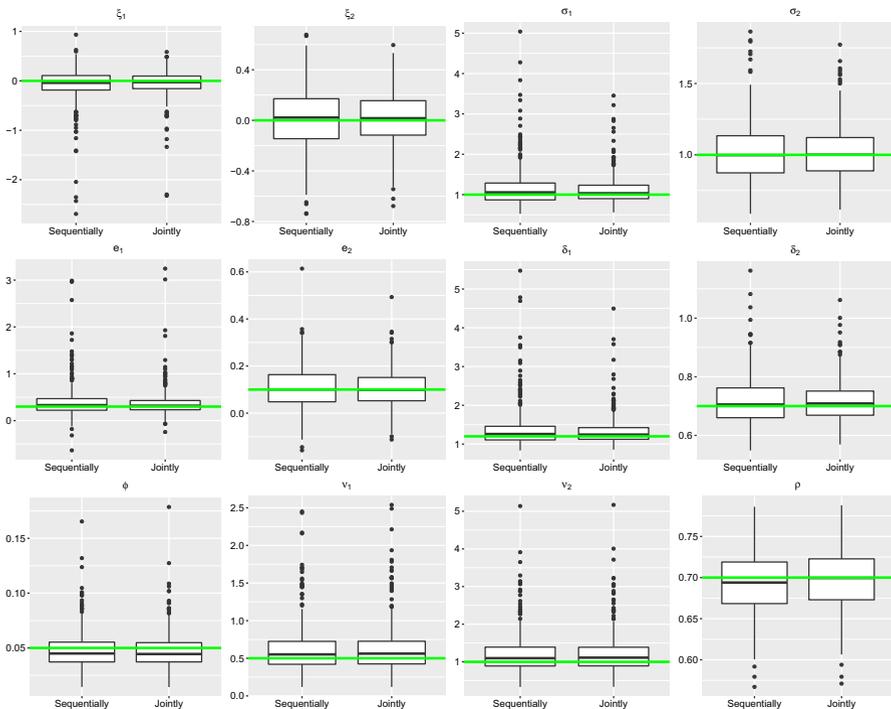


Fig. 3 Boxplots of estimated parameters for bivariate SAS parsimonious Matérn random fields in 500 simulations by Method 3 and Method 4 (green lines represent the true values $\xi_1 = \xi_2 = 0, \sigma_1 = \sigma_2 = 1, e_1 = 0.3, e_2 = 0.1, \delta_1 = 1.2, \delta_2 = 0.7, \phi = 0.05, \nu_1 = 0.5, \nu_2 = 1, \rho = 0.7$) (color figure online)

Table 1 Comparison of the four methods for bivariate SAS parsimonious Matérn random fields in terms of the number of parameters, mean log-likelihood and BIC; MAE, width of 95% CI and empirical coverage of the 95% CI for prediction of each variable (left/right column for variable 1/2)

	# pars	Log-likelihood	BIC	MAE		CI width		Coverage	
				Var 1	Var 2	Var 1	Var 2	Var 1 (%)	Var 2 (%)
Method 1	8	- 513.5	1069.3	14.019	23.481	2.68	4.72	95.2	93.9
Method 2	12	- 565.9	1195.4	14.102	23.481	2.63	4.73	94.4	94.1
Method 3	12	- 498.0	1059.6	13.993	23.314	2.61	4.70	94.5	94.1
Method 4	12	- 497.3	1058.1	13.985	23.309	2.61	4.69	94.6	94.2

The best method is in bold

For prediction, Method 1 ignores the non-Gaussian characteristics of the two variables while Method 2 omits the correlation between the two variables. Prediction by Method 3 and Method 4 differs only in the estimated parameters. Table 1 presents prediction performance by the four methods in terms of mean absolute error (MAE), the width of the 95% confidence interval (CI) and empirical coverage of the 95% CI, for each variable. The joint procedure has the minimum MAE and CI width while maintaining empirical coverage close to the claimed level. Note that for the best method with respect to CI width and coverage, we use combined criteria and for both variables, e.g., 94.6% and 94.2% with shorter CI widths by Method 4 are better than 95.2% and 93.9% by Method 1. The empirical coverage is always slightly less than the claimed 95% (except for the lighter tailed variable by Method 1) since we treat the estimated parameters as the true values for prediction without taking the uncertainty in estimation into consideration.

Additional simulations with different parameter settings and a simulation study for the trivariate SAS parsimonious Matérn random field show similar results (see supplementary material). In summary, for modeling bivariate or trivariate random fields, ignoring either the non-Gaussian features of the variables or the correlation between the variables is detrimental for the prediction purpose. For the bivariate or trivariate SAS random fields, it is beneficial to estimate all the parameters simultaneously with the joint likelihood rather than the sequential procedure. However, as the number of parameters to estimate grows as k grows, optimization becomes much more difficult for joint estimation. If optimization becomes problematic when estimating θ jointly for a high-dimensional vector process, we suggest estimation in an iterative manner of θ_1 and θ_2 until convergence as in Xu and Genton (2017) or sequentially (iterate once).

5 Wind data examples

5.1 Large ENSemble Project (LENS)

We consider wind data sets at daily and monthly levels from the LENS, which is an ensemble of runs of the Community Earth System Model (CESM) from the National

Center for Atmospheric Research. The LENS is a set of climate model simulations intended for understanding the internal variability, and it includes 40 ensemble members of simulations from 1910 to 2010. The spatial resolution of the ensemble members is $0.9375^\circ \times 1.25^\circ$ (latitude \times longitude) and each run was generated under the Representative Concentration Pathway (RCP) 8.5 (van Vuuren et al. 2011; Kay et al. 2015). Since future wind trends are often of interest, here, we work on daily and monthly near-future wind components, eastward U , and northward V components, above the ground level from 2006 to 2100. In the following two examples of bivariate time series and spatial data sets, we focus on demonstrating the benefits of using multivariate transformations in terms of either a model fit or a prediction.

5.2 Example 1: bivariate time series data (monthly LENS data)

At first, we examine the temporal properties of global monthly wind components with all 288 longitudes, but 134 latitudes between 62° N and 62° S. For a total of 95 years, we have 1,140 temporal points at each location and run. For each site, we calculate the empirical skewness and kurtosis of the wind components, U and V , after removing the climatology (i.e., the average across five selected runs for each space-time location). We consider the significance test of the third and fourth central moments, the skewness and kurtosis, of residuals over time (Bai and Ng 2005). We observe that both residuals, RU and RV , clearly show features of skewness and heavier tails compared to the Gaussian distribution in many spatial locations (only locations where p -values of the significance test are smaller than 0.05 are reported), as shown in Fig. 4. That is, the first two moments, such as the mean and variance, are not enough in characterizing the temporal patterns of both residuals. Thus, we conclude that the Gaussianity assumption is not suitable for modeling, and we fit univariate TGH-AR(1) and bivariate TGH-VAR(1) models marginally and jointly for the RU and RV residuals, respectively. Estimated values of TGH-AR(1) such as \hat{g}_i , \hat{h}_i , and $\hat{\phi}_i$ (autoregressive coefficients) for $i = 1, 2$ are shown in Fig. 5. In particular, \hat{g} and \hat{h} for both RU and RV were estimated with nonzero values over most locations, suggesting non-Gaussianity.

Now we focus on the region with longitudes between 26.25° E and 68.75° E and latitudes 8.953° N and 36.283° N near Saudi Arabia. Figure 6 shows boxplots of some selected parameters for TGH-AR(1) and TGH-VAR(1). The first two boxplots in Fig. 6a indicate the distribution of estimates \hat{g}_i for RU from the two models and the remaining two boxplots represent the distribution of \hat{g}_i for RV from two models. Similarly, Fig. 6b reports the distribution of \hat{h}_i from the two models. We observe that the ranges in the boxplots of \hat{g}_i and \hat{h}_i for TGH-VAR(1) are substantially longer compared to TGH-AR(1), i.e., the parameters for the skewness and tail behavior were estimated higher in TGH-VAR(1) compared to TGH-AR(1). The boxplots of estimates of the VAR coefficient matrix, \mathbf{A} , for the TGH-VAR(1) are reported in Fig. 6c.

Figure 7 presents the BIC values between the two models at each site from one ensemble member. Positive and negative values indicate a better and worse fit of the TGH-VAR(1) compared to the TGH-AR(1). In terms of the BIC values, TGH-VAR(1) outperforms TGH-AR(1) over many spatial locations. In particular,

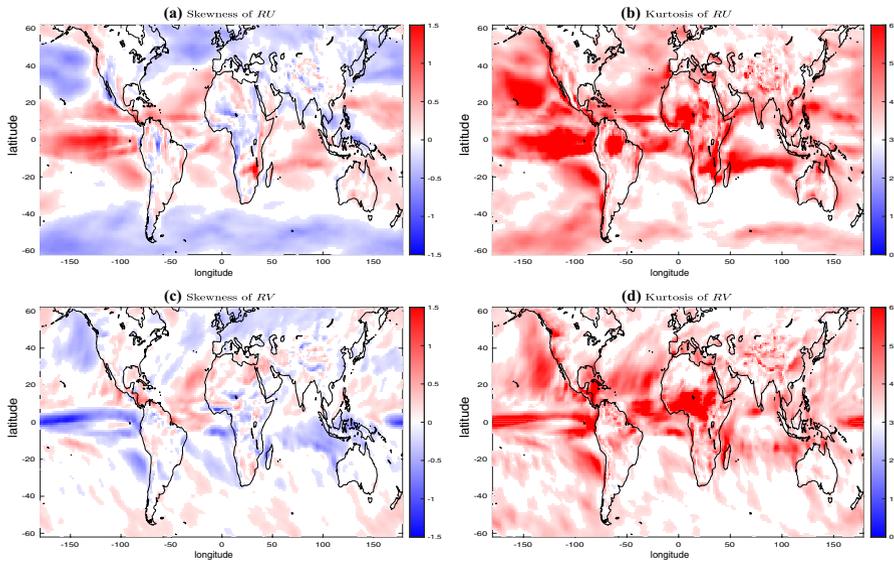


Fig. 4 The empirical skewness and kurtosis of (a, b) RU and (c, d) RV from one ensemble member after removing the trend (from five runs) are reported

locations near the Red Sea and the Persian Gulf have significantly higher values than other locations. Since the TGH-VAR(1) considers the relationship between the two variables and provides additional flexibility, it has a better model fit compared to the TGH-AR(1). Here, we show that the TGH-VAR(1) model can be a better fit for the wind component than the TGH-AR(1) model. However, there is room for improvement. For example, one may think that the bivariate time series model is not entirely appropriate for the horizontal and vertical components. A small change in both horizontal and vertical components can result in a profound difference in wind direction. Thus, a state-dependent model, e.g., threshold vector autoregressive time-series model (Lo and Zivot 2001; Wikle et al. 2019) for the \mathbf{A} and $\mathbf{\Psi}$ matrices might be more appropriate and preferred.

5.3 Example 2: bivariate spatial data (daily LENS data)

We examine the spatial behaviors of daily wind components. For the data sets, we select the residuals of two components, RU and RV , on September 23, 2020. Their histograms from one ensemble member are reported in Fig. 8. Based on the histograms and empirical kurtosis values, we observe that RU has heavier tails, whereas RV has slightly lighter tails compared to the Gaussian distribution.

Again, we pay attention to the region near Saudi Arabia from one ensemble member after removing the trend on September 23, 2020, resulting in $n = 1050$ spatial locations. Figure 8 displays the maps of RU and RV over the spatial domain we consider and their histograms. Since the two histograms of RU and

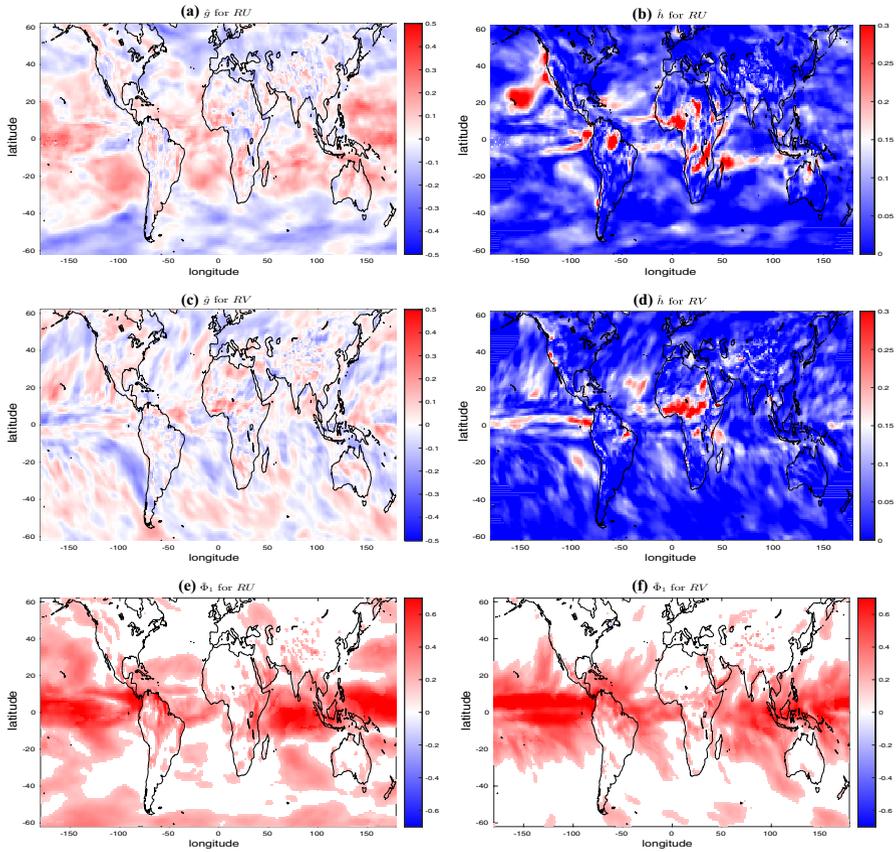


Fig. 5 The estimates of g and h for (a, b) RU and (c, d) RV from one ensemble member are reported. The estimates of $\hat{\Phi}_1$ for the TGH-AR(1) of (e) RU and (f) RV

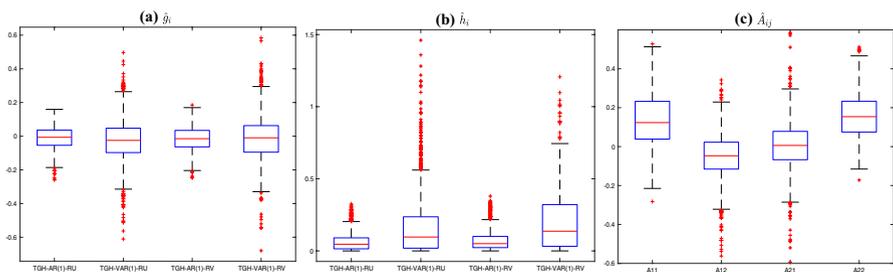


Fig. 6 a, b Boxplots of \hat{g}_i and \hat{h}_i 's for TGH-AR(1) and TGH-VAR(1) near Saudi Arabia. c Boxplots of the estimates for the \mathbf{A} matrix for the TGH-VAR(1)

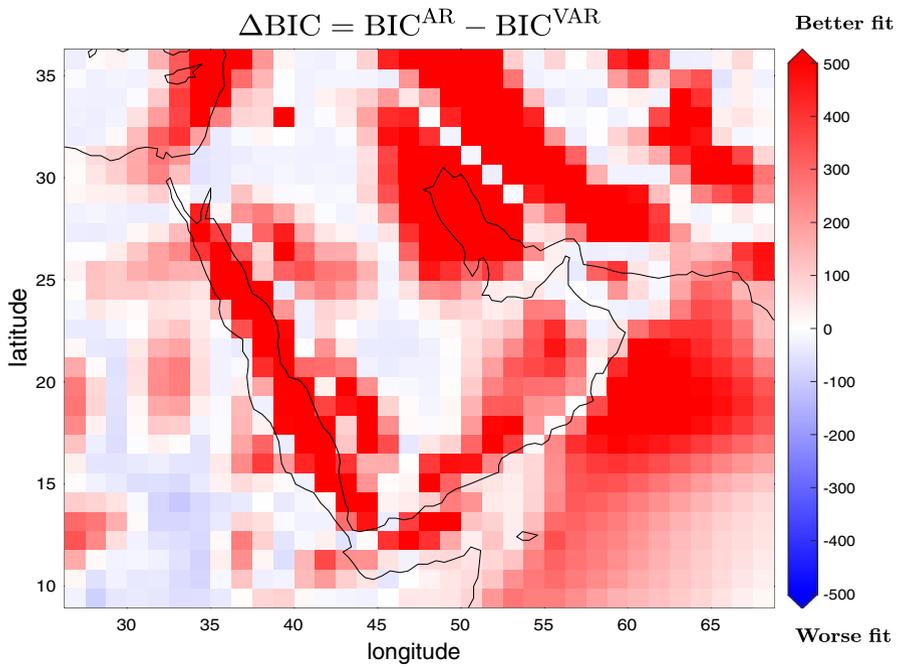


Fig. 7 Map of differences in the BIC values between the TGH-VAR(1) and TGH-AR(1)

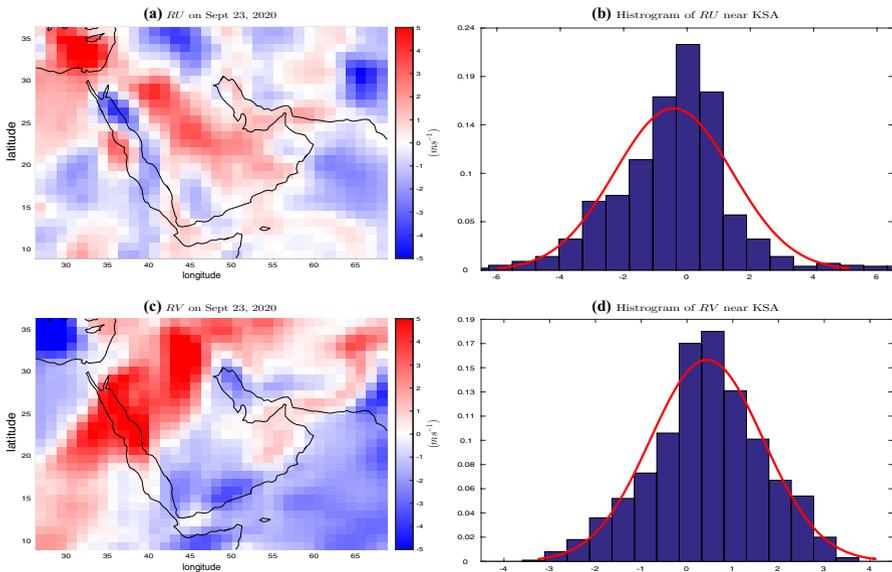


Fig. 8 The maps of **a** *RU* and **c** *RV* near Saudi Arabia from one ensemble member after removing the trend (from five runs) on September 23, 2020. The histograms of **b** *RU* and **d** *RV* on September 23, 2020. Red curves are corresponding normal fits. The empirical skewness and kurtosis are (0.26, 5.17) for *RU* and (− 0.21, 2.90) for *RV*

RV show lighter and heavier tail behaviors at the same time, we consider the SAS transformation instead of the TGH distribution for spatial data sets. Because the area of this region is not very broad, we consider the chordal distance instead of great-circle distance. Here, the maximum distance between two locations is approximately 5100 km. We consider the following list of models:

Model A: two independent SAS random fields with Matérn covariance function;

Model B: a bivariate Gaussian random field with a parsimonious Matérn covariance function;

Model C: a bivariate SAS random field with a parsimonious Matérn covariance function.

The maximum likelihood estimates for all these models are reported in Table 2. The estimated correlation coefficient, $\hat{\rho}_{12}$, between the two components for the Models B and C are 0.36 and 0.29, so the relationship between the two variables cannot be neglected. According to Table 2, the distribution of RU has a heavier tail compared to the Gaussian distribution ($\hat{\delta}_1 < 1$) and is not significantly skewed (\hat{e}_1 is close to 0), whereas that of RV has a lighter tail compared to the Gaussian distribution ($\hat{\delta}_2 > 1$) and is skewed ($\hat{e}_2 > 0$).

Table 2 presents the number of parameters, log-likelihood, and BIC values for the models. Model B shows the worst model fit regarding the log-likelihood and BIC values among models. It is evident that the Gaussian assumption is not suitable. Between the SAS random fields, Model C outperforms Model A in terms of model fit. This result is not surprising because the correlation between the two different components cannot be neglected (the two components over the spatial domain have a positive empirical correlation coefficient, $\hat{\rho}_0 = 0.31$). Here, Model A has the advantage of its flexibility for fitting marginally with additional parameters for the spatial ranges, whereas Model C highlights the flexibility for modeling jointly along with cross-correlation between the two components. If we have a more flexible multivariate Matérn covariance function, i.e., the spatial range and smoothness parameters for the marginal and cross-covariance structures can be determined without restrictions, then it might have a better performance than Model C.

Finally, we compare the prediction performance among models using the co-kriging predictor (Wackernagel 2013). Here, we use a ten-fold cross-validation strategy and quantify the discrepancy between the simulated and predicted values for RU and RV . That is, we split the data set into ten groups and take one group as a test set (10% of locations) to perform interpolations over the spatial domain and take the remaining groups as a training set (90% of locations), respectively. Then, we fit a model on the training set and evaluate it on the test set. We repeat this procedure ten times by changing the test set and evaluating it in the same manner. For the prediction scores, we consider the MAE. Since, for the transformed random field, the conditional median is the optimal predictor, it is fair to compare the MAE values. Table 3 shows that Model B has the highest MAE value overall, which is consistent with the log-likelihood and BIC values, whereas Model C

Table 2 The number of parameters, log-likelihood, BIC, and maximum likelihood estimates for Models A, B, and C

Model	# of pars	$\hat{\mu}_1$	$\hat{\sigma}_1$	\hat{e}_1	$\hat{\delta}_1$	$\hat{\phi}_1$	$\hat{\nu}_1$
A	12	0.195 (0.164)	1.057 (0.077)	0.044 (0.037)	0.780 (0.024)	106.829 (9.450)	3.223 (0.211)
Log-likelihood	BIC	$\hat{\mu}_2$	$\hat{\sigma}_2$	\hat{e}_2	$\hat{\delta}_2$	$\hat{\phi}_2$	$\hat{\nu}_2$
- 233.1	549.7	1.995 (1.006)	3.865 (1.063)	0.103 (0.176)	1.510 (0.247)	181.651 (17.805)	2.305 (0.113)
Model	# of pars	$\hat{\mu}_1$	$\hat{\sigma}_1$	-	-	$\hat{\phi}$	$\hat{\nu}_1$
B	8	0.631 (0.364)	2.053 (0.149)	-	-	131.502 (7.371)	2.884 (0.113)
Log-likelihood	BIC	$\hat{\mu}_2$	$\hat{\sigma}_2$	-	-	$\hat{\rho}_{12}$	$\hat{\nu}_2$
- 268.2	597.6	0.682 (0.293)	1.685 (0.095)	-	-	0.356 (0.014)	2.758 (0.112)
Model	# of pars	$\hat{\mu}_1$	$\hat{\sigma}_1$	\hat{e}_1	$\hat{\delta}_1$	$\hat{\phi}$	$\hat{\nu}_1$
C	12	0.050 (0.197)	1.274 (0.101)	0.000 (0.045)	0.765 (0.025)	140.731 (8.378)	2.705 (0.109)
Log-likelihood	BIC	$\hat{\mu}_2$	$\hat{\sigma}_2$	\hat{e}_2	$\hat{\delta}_2$	$\hat{\rho}_{12}$	$\hat{\nu}_2$
- 198.8	489.4	0.510 (0.383)	2.068 (0.305)	0.224 (0.034)	1.121 (0.078)	0.286 (0.015)	2.622 (0.106)

Standard errors are given in parentheses. The best model is in bold

outperforms both Models A and B in prediction. Furthermore, we use (i) other randomly selected test sets and (ii) selected regions of two adjacent latitudinal bands over the domain as the prediction regions, and their prediction results are similar. In particular cases (from different ensemble members), Model A had a

Table 3 (A ten-fold cross-validation) Comparison between Models A, B, and C in terms of the number of parameters, log-likelihood, BIC, and prediction scores (averages across 10 folds are presented)

	# of parameters	Log-likelihood	BIC	MAE	MAE ₁	MAE ₂
Model A	12	- 3125.2 (48.4)	6332.7 (96.8)	0.916 (0.109)	0.849 (0.132)	0.983 (0.121)
Model B	8	- 3148.5 (59.5)	6358.3 (119.0)	1.270 (0.122)	1.380 (0.173)	1.159 (0.110)
Model C	12	- 3116.4 (50.8)	6323.4 (101.7)	0.831 (0.132)	0.847 (0.170)	0.815 (0.111)

Here, MAE_i represents the MAE associated to variable *i*, for *i* = 1, 2, i.e., *RU* and *RV*, respectively. Standard deviations of values across 10 folds are given in parentheses. The best model is in bold

slightly better model fit than Model C when the two spatial range parameters were estimated to be very different, and a correlation coefficient between some selected sets of RU and RV was negligible. However, the overall prediction performance of Model C tended to outperform Model A. Indeed, for this specific data set, Model C shows a better performance than Model A in terms of predictions.

6 Discussion

In this paper, we have discussed a general framework of multivariate transformed Gaussian processes. The framework can be applied to multivariate time series, multivariate spatial random fields, and possibly multivariate spatio-temporal processes. The proposed multivariate transformed Gaussian framework coupled with TGH or SAS transformations has shown a better statistical performance than using either multivariate Gaussian processes or independent univariate transformed Gaussian processes, for time series or spatial processes, as a benchmark. It is evident that, in both estimation and prediction, the joint modeling approach of the transformed processes has some advantages, e.g., flexibility in capturing distributional features of skewness and tail behavior and the relationship between variables, compared to either the independent modeling approach with multivariate processes or the joint modeling with multivariate Gaussian processes.

We have not included regression and white noise terms in our framework since the problem is not trivial and beyond the scope of this paper. One direction for incorporating additional trend and white noise terms is to adopt the data/process/parameter hierarchical framework (Cressie and Wikle 2011) by treating the transformed Gaussian model as a noiseless process while regression and noise terms can be included in the data hierarchy. Another direction could be to add a non-Gaussian noise term modeled by the transformed Gaussian process to the regression term. Furthermore, the framework could be applied beyond the spatial and/or temporal setting by extending the index \mathbf{x} to the more general covariate/feature space. Then a regression problem translates directly into a trans-Gaussian process, which is especially relevant to methods in the machine learning field. At last, domain-specific knowledge can be crucial for choices to incorporate regression and white noise terms.

Although for both simulation and data example only bivariate or trivariate processes ($k = 2$ or 3) with limited time points or spatial locations over the spatial domain were examined, applications with larger sample size, and more than two variables can be handled similarly. One issue, in particular for spatial and spatio-temporal processes, is that our framework is depending on the Gaussian process, and thus computation can be easily intensive as the dimensionality of the data sets and number of parameters for more than two variables increase. The computational issue is a critical topic in the spatial statistics and environmental science communities. There have been great efforts to model large data sets via various approaches. For more details and references, we refer the reader to Sun and Genton (2012), Genton and Kleiber (2015), and Heaton et al. (2019).

There are some further directions the authors are currently pursuing. We need to investigate theoretical properties for our new framework, such as restrictions for the cross-correlation structure of \mathbf{Y} after transformation. In general, the transformation weakens the correlation and the correlation $\text{corr}\{Y_{j_1}(\mathbf{x}_{i_1}), Y_{j_2}(\mathbf{x}_{i_2})\}$ is always weaker than $\text{corr}\{Z_{j_1}(\mathbf{x}_{i_1}), Z_{j_2}(\mathbf{x}_{i_2})\}$. Therefore, \mathbf{Y} cannot accept any prescribed valid cross-correlation structure, and the restriction depends on the transformation used. Another technical issue is to find optimal prediction considering uncertainty in parameter estimation, such as in De Oliveira (2006) for the log-Gaussian random fields. Also, when we perform the predictions based on co-kriging, it is necessary to evaluate the probabilistic prediction performance other than by an optimal point predictor. So more reasonable prediction scores are required, in particular, multivariate versions of continuous rank probability scores (Gneiting and Raftery 2007). In application, working with anisotropic or nonstationary multivariate covariance models with appropriate transformation is another exciting direction to investigate.

Acknowledgements This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2018-CRG7-3742.

References

- Apanasovich, T. V., Genton, M. G., & Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497), 180–193.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178.
- Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, 83(4), 715–726.
- Bai, J., & Ng, S. (2005). Tests for skewness, kurtosis, and normality for time series data. *Journal of Business & Economic Statistics*, 23(1), 49–60.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Boca Raton, FL: Chapman and Hall/CRC.
- Benjamin, M. A., Rigby, R. A., & Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association*, 98(461), 214–223.
- Block, H. W., Langberg, N. A., & Stoffer, D. S. (1990). Time series models for non-Gaussian processes. *Lecture Notes-Monograph Series*, 16, 69–83.
- Bolin, D., Wallin, J., & Lindgren, F. (2019). Latent gaussian random field mixture models. *Computational Statistics & Data Analysis*, 130, 80–93.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological)*, 26(2), 211–252.
- Bradley, J. R., Holan, S. H., & Wikle, C. K. (2019). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *Journal of the American Statistical Association*. (To appear).
- Chagneau, P., Mortier, F., Picard, N., & Bacro, J. N. (2011). A hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields. *Biometrics*, 67(1), 97–105.
- Cordeiro, G. M., & de Andrade, M. G. (2009). Transformed generalized linear models. *Journal of Statistical Planning and Inference*, 139(9), 2970–2987.
- Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley.
- Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Hoboken, NJ: Wiley.
- Cressie, N., & Zammit-Mangion, A. (2016). Multivariate spatial covariance models: A conditional approach. *Biometrika*, 103(4), 915–935.

- Davies, N., Spedding, T., & Watson, W. (1980). Autoregressive moving average processes with non-normal residuals. *Journal of Time Series Analysis*, 1(2), 103–109.
- De Oliveira, V. (2006). On optimal point and block prediction in log-Gaussian random fields. *Scandinavian Journal of Statistics*, 33(3), 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(440), 1422–1433.
- Diggle, P. J., Tawn, J. A., & Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3), 299–350.
- Du, J., Leonenko, N., Ma, C., & Shu, H. (2012). Hyperbolic vector random fields with hyperbolic direct and cross covariance functions. *Stochastic Analysis and Applications*, 30(4), 662–674.
- Dutta, K., & Babel, D. (2002). On measuring skewness and kurtosis in short rate distributions: The case of the US dollar London inter bank offer rates. Technical report, The Wharton School, University of Pennsylvania.
- Field, C. (2004). Using the gh distribution to model extreme wind speeds. *Journal of Statistical Planning and Inference*, 122(1), 15–22.
- Field, C., & Genton, M. G. (2006). The multivariate g -and- h distribution. *Technometrics*, 48(1), 104–111.
- Fonseca, T. C. O., & Steel, M. F. J. (2011). Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika*, 98(4), 761–774.
- Gaver, D. P., & Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability*, 12(3), 727–745.
- Genton, M. G. (2004). *Skew-elliptical distributions and their applications: A journey beyond normality*. Boca Raton, FL: Chapman and Hall/CRC.
- Genton, M. G., & Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science*, 30(2), 147–163.
- Genton, M. G., & Zhang, H. (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics*, 3(2), 171–179.
- Gneiting, T., Kleiber, W., & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491), 1167–1177.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Gotway, C. A., & Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(2), 157–178.
- Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 10, 87–102.
- Griewank, A., & Walther, A. (2008). *Evaluating derivatives: Principles and techniques of algorithmic differentiation* (2nd ed.). Philadelphia, PA: Society for Industrial and Applied Mathematics.
- He, Y., & Raghunathan, T. E. (2012). Multiple imputation using multivariate gh transformations. *Journal of Applied Statistics*, 39(10), 2177–2198.
- Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3), 398–425.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The g -and- h distributions. In D. Hoaglin, F. Mosteller, & J. Tukey (Eds.), *Exploring data tables, trends, and shapes, chapter 11* (pp. 461–513). New York: Wiley.
- Jeong, J., Yan, Y., Castruccio, S., & Genton, M. G. (2019). A stochastic generator of global monthly wind energy with Tukey g -and- h autoregressive processes. *Statistica Sinica*, 19, 1105–1126.
- Johns, C. J., Nychka, D., Kittel, T. G. F., & Daly, C. (2003). Infilling sparse records of spatial fields. *Journal of the American Statistical Association*, 98(464), 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-arsinh distributions. *Biometrika*, 96(4), 761–780.
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8), 1333–1349.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, 70(5), 1–21.

- Krupskii, P., Huser, R., & Genton, M. G. (2018). Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, *113*(521), 467–479.
- Lawrance, A. J., & Lewis, P. A. W. (1980). The exponential autoregressive-moving average EARMA (p, q) process. *Journal of the Royal Statistical Society Series B (Methodological)*, *42*(2), 150–161.
- Le, N. D., Martin, R. D., & Raftery, A. E. (1996). Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association*, *91*(436), 1504–1515.
- Li, W. K., & McLeod, A. I. (1988). ARMA modelling with non-Gaussian innovations. *Journal of Time Series Analysis*, *9*(2), 155–168.
- Lo, M. C., & Zivot, E. (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Macroeconomic Dynamics*, *5*(4), 533–576.
- Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Berlin: Springer.
- Ma, C. (2009). Construction of non-Gaussian random fields with any given correlation structure. *Journal of Statistical Planning and Inference*, *139*, 780–787.
- Ma, C. (2010). χ^2 random fields in space and time. *IEEE Transactions on Communications*, *58*(1), 378–383.
- Ma, C. (2011). Covariance matrix functions of vector χ^2 random fields in space and time. *IEEE Transactions on Communications*, *59*(9), 2554–2561.
- Ma, C. (2013). K-distributed vector random fields in space and time. *Statistics & Probability Letters*, *83*(4), 1143–1150.
- Marchenko, Y. V., & Genton, M. G. (2010). Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics*, *21*(3–4), 318–340.
- Martinez, J., & Iglewicz, B. (1984). Some properties of the Tukey g and h family of distributions. *Communications in Statistics-Theory and Methods*, *13*(3), 353–369.
- Myers, D. E. (1982). Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*, *14*(3), 249–257.
- Palacios, M. B., & Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, *101*(474), 604–618.
- R Development Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rimstad, K., & Omre, H. (2014). Skew-Gaussian random fields. *Spatial Statistics*, *10*, 43–62.
- Royle, J. A., & Berliner, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, *4*(1), 29–56.
- Shumway, R. H., & Stoffer, D. S. (2011). *Time series analysis and its applications: With R Examples*. New York: Springer.
- Snelson, E., Ghahramani, Z., & Rasmussen, C. E. (2004). Warped Gaussian processes. In S. Thrun, L. K. Saul, & B. Schölkopf (Eds.), *Advances in neural information processing systems* (Vol. 16, pp. 337–344). Cambridge: MIT Press.
- Sun, Y., Li, B., & Genton, M. G. (2012). Geostatistics for large datasets. In E. Porcu, J. M. Montero, & M. Schlather (Eds.), *Advances and challenges in space-time modelling of natural events* (pp. 55–77). Berlin: Springer.
- Tarami, B., & Pourahmadi, M. (2003). Multi-variate t autoregressions: Innovations, prediction variances and exact likelihood equations. *Journal of Time Series Analysis*, *24*(6), 739–754.
- Tukey, J. (1977). *Modern techniques in data analysis*. North Dartmouth, MA: In NSF-sponsored regional research conference at Southeastern Massachusetts University.
- van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., et al. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*, 5–31.
- Vio, R., Andreani, P., Tenorio, L., & Wamsteker, W. (2002). Numerical simulation of non-Gaussian random fields with prescribed marginal distributions and cross-correlation structure. II. Multivariate random fields. *Publications of the Astronomical Society of the Pacific*, *114*(801), 1281–1289.
- Vio, R., Andreani, P., & Wamsteker, W. (2001). Numerical simulation of non-Gaussian random fields with prescribed correlation structure. *Publications of the Astronomical Society of the Pacific*, *113*, 1009–1020.
- Wackernagel, H. (2013). *Multivariate geostatistics: An introduction with applications*. Berlin: Springer.
- Wallin, J., & Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, *42*(3), 872–890.
- Wikle, C. K., Zammit-Mangion, A., & Cressie, N. (2019). *Spatio-temporal statistics with R*. Boca Raton: CRC Press.

- Wong, C. S., Chan, W. S., & Kam, P. L. (2009). A Student t -mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika*, *96*(3), 751–760.
- Wong, C. S., & Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(1), 95–115.
- Xu, G., & 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., & Genton, M. G. (2017). Tukey g -and- h random fields. *Journal of the American Statistical Association*, *112*, 1236–1249.
- Yan, Y., & Genton, M. G. (2018). Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function. *Environmetrics*, *29*, e2458.
- Yan, Y., & Genton, M. G. (2019a). Non-Gaussian autoregressive processes with Tukey g -and- h transformations. *Environmetrics*, *30*, e2503.
- Yan, Y., & Genton, M. G. (2019b). The Tukey g -and- h distribution. *Significance*, *16*(3), 12–13.
- Yin, J., & Craigmile, P. F. (2018). Heteroscedastic asymmetric spatial processes. *Stat*, *7*, e206.
- Zammit-Mangion, A., Cressie, N., & Ganesan, A. L. (2016). Non-gaussian bivariate modelling with application to atmospheric trace-gas inversion. *Spatial Statistics*, *18*, 194–220.
- Zhang, H., & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, *21*, 33–47.
- Zhang, Y., & Yeung, D. (2010). Multi-task warped Gaussian process for personalized age estimation. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 2622–2629).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.