Received: 31 October 2017 | Revised: 5 March 2018 | Accepted: 15 April 2018
DOI: 10.1002/env.2503

SPECIAL ISSUE PAPER

WILEY

Non-Gaussian autoregressive processes with Tukey *g*-and-*h* transformations

Yuan Yan | Marc G. Genton

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

Correspondence
Yuan Yan, Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
Email: yuan.yan@kaust.edu.sa

Funding information
King Abdullah University of Science and Technology (KAUST), Grant/Award Number: OSR-2015-CRG4-2640

Abstract

When performing a time series analysis of continuous data, for example, from climate or environmental problems, the assumption that the process is Gaussian is often violated. Therefore, we introduce two non-Gaussian autoregressive time series models that are able to fit skewed and heavy-tailed time series data. Our two models are based on the Tukey g-and-h transformation. We discuss parameter estimation, order selection, and forecasting procedures for our models and examine their performances in a simulation study. We demonstrate the usefulness of our models by applying them to two sets of wind speed data.

KEYWORDS

autoregressive, heavy tails, non-Gaussian, skewness, time series, Tukey g-and-h transformation

1 | INTRODUCTION

To study climate change, it is essential to understand the temporal properties of the data of interest. Climate and environmental data, such as precipitation, temperature, thickness of glacial varves, wind speed, concentration of air pollutants, are continuous in nature. In a time series analysis of continuous random variables, the autoregressive moving-average (ARMA) model, especially the Gaussian one, is popularly adopted because of its simplicity and interpretability. However, data from climate or environmental science are often asymmetric with heavy tails. Therefore, a non-Gaussian time series analysis is needed. Because most climate or environmental data we come across are continuous, in this paper, we do not consider time series analysis of categorical or binary data, which are intrinsically non-Gaussian.

In linear models, three approaches are mainly used to deal with non-Gaussian data: transformation of the dependent variable, regression with non-Gaussian error, and generalized linear models (GLM). Similarly, we can also classify many of the existing non-Gaussian time series methods and models into three categories: (a) transformations to "Gaussianize" the data, (b) ARMA models with non-Gaussian noise, and (c) extension of the GLM to the time series context.

The transformation approach is widely used among practitioners to, first, Gaussianize data and then examine the latent Gaussian process. Equivalently, it assumes that the observed process Y_t is obtained after a transformation η of the latent Gaussian process Z_t, as follows:

$$Y_t = \eta(Z_t). \tag{1}$$

For example, precipitation data can be approximated by the gamma distribution; then, a square or cube root transformation is usually applied to normalize the data. Similarly, wind speed data can be assumed to follow a Weibull distribution, and taking a logarithm alleviates the departure from normality (Haslett & Raftery, 1989). The logarithm, square root, and cube root transformations all belong to the Box–Cox power transformation (Box & Cox, 1964), which is a family of transformations with one parameter (λ) and can be applied only to positive values. A parametric form of transformations allows users to find the optimal transformation η from among the members of the Box–Cox transformation family by estimating λ from the data, instead of attempting many different transformations. A large literature exists on the Box–Cox

© 2018 John Wiley & Sons, Ltd.

power transformation, as well as its modifications and its application in linear models and time series analysis. Nelson and Granger (1979) extensively explored the use of Box–Cox transformations in 21 time series from economics. Nonparametric transformations are an appealing alternative to assuming a parametric form of transformations due to their flexibility. Johns, Nychka, Kittel, and Daly (2003) explored nonparametric transformations on both the dependent variable and the predictor variables in regression. Block, Langberg, and Stoffer (1990) used the empirical distribution and probability integral transform (PIT) to form ARMA models with arbitrary and continuous marginal distributions. However, there is a trade-off between flexibility and parsimony.

Instead of using transformations to Gaussianize the data, an ARMA model with non-Gaussian noise may be specified, whether it is the marginal distribution of the process or the distribution of the error term. Given the marginal distribution, the distribution of the error term can be found by linking their moment-generating function, namely, the Laplace transformation of the probability density function. Many models were previously proposed following this direction, for example, ARMA models with exponential marginals (Lawrance & Lewis, 1980) and autoregressive (AR) models with marginal gamma distribution (Gaver & Lewis, 1980). On the other hand, specifying the distribution of the non-Gaussian error term is easier for data simulation and maximum-likelihood inference because the conditional likelihood can then be written down in closed form. ARMA models that are driven by non-Gaussian innovations and their properties were discussed by Davies, Spedding, and Watson (1980) and Li and McLeod (1988).

For the third class of existing non-Gaussian time series methods and models, Benjamin, Rigby, and Stasinopoulos (2003) introduced the generalized ARMA model, extending the GLM to a time series setting. Earlier exploration of GLM in time series can be found in Benjamin et al. (2003) and references therein. Cordeiro and de Andrade (2009) proposed a model combining the GLM idea with transformations and considered its use in time series.

Apart from the three categories summarized above, there exist many other specialized models for non-Gaussian, nonstationary time series. The most famous ones are the autoregressive conditional heteroskedasticity (ARCH) (Engle, 1982) and the generalized autoregressive conditional heteroskedasticity (GARCH) (Bollerslev, 1986) models. These models have been used canonically for analyzing financial time series that exhibit changes in volatility. Other models include the zeroth-order self-exciting threshold autoregressive (SETAR) model (Tong, 1990), the multipredictor autoregressive time series (MATS) models (Martin, 1992), the Gaussian mixture transition distribution models (GMTD) (Le, Martin, & Raftery 1996), the mixture autoregressive models (Wong & Li, 2000), and their extension to the Student t-mixture autoregressive model (Wong, Chan, & Kam, 2009).

Considering the trade-offs between model flexibility and parsimony, in this paper, we use an alternative parametric family of transformations, the Tukey g-and-h (TGH) transformation, which is more flexible and interpretable than the Box–Cox family and at the same time keeps the model fairly simple compared with the nonparametric approaches. We build two autoregressive models for non-Gaussian time series via the TGH transformation in both the data transformation and non-Gaussian error approaches. Our first model uses the TGH transformation for η in (1). Our second model assumes the non-Gaussian error term in the AR process to be a TGH transformation of Gaussian white noise.

The remainder of this paper is organized as follows. In Section 2, we review the TGH transformation, discuss properties of the univariate TGH distribution and its extensions, and introduce our two AR models based on the TGH transformation. In Section 3, we describe the parameter estimation and forecasting algorithm for our models. In Section 4, we demonstrate the estimation and prediction performances of our models through a simulation study. In Section 5, we illustrate the usefulness of our models by applying them to two wind speed data sets. We summarize our findings and prospects for future research in Section 6.

2 | TWO TGH AUTOREGRESSIVE MODELS

2.1 | TGH transformations and properties

The TGH transformation (Tukey, 1977) is defined as a strictly monotonic increasing function of $z \in \mathbb{R}$ for $h \geq 0$ and $g \in \mathbb{R}$, as follows:

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

When applying the TGH transformation to a univariate standard normal random variable $Z \sim \mathcal{N}(0,1)$, the transformed random variable $T = \tau_{g,h}(Z)$ is said to follow the TGH distribution. The support of T is the real line when $h \neq 0$ or $h = g = 0$; for $h = 0$ and $g \neq 0$, it has a lower (upper) bound of $-1/g$ when $g > 0$ ($g < 0$). The univariate TGH distribution

is a flexible family of distributional models for non-Gaussian data. It has two interpretable parameters, g and h, which respectively control the skewness and tail heaviness of Y. The tails become heavier as h increases, and the qth moment of Y exists only for $h < 1/q$. The level of skewness increases as $|g|$ increases (becoming right skewed when $g > 0$). Martinez and Iglewicz (1984) included the case of $h < 0$, for which the TGH transformation is no longer monotonic and the resulting TGH distribution has lighter tails than the Gaussian one. However, their method is unconventional, so we only consider the case of $h \geq 0$. Also, from a practical point of view, real data are more often heavy tailed rather than light tailed. Special cases of the TGH distribution include the shifted log-normal random variable for $h = 0$ and $g > 0$, and a Pareto-like distribution for $g = 0$ and $h > 0$, which was discussed in detail by Goerg (2015). The qth moments of the TGH distribution were derived by Martinez and Iglewicz (1984), for $h < 1/q$, as follows:

$$\mathbb{E}(T^q) = \begin{cases} \frac{1}{g^q\sqrt{1-qh}} \sum_{i=0}^{q} (-1)^i \binom{q}{i} \exp\left\{\frac{(q-i)^2 g^2}{2(1-qh)}\right\}, & g \neq 0, \\ \frac{q!}{2^{q/2}(q/2)!}(1-qh)^{-\frac{q+1}{2}}, & g = 0, q \text{ even}, \\ 0, & g = 0, q \text{ odd}. \end{cases} \tag{3}$$

We illustrate various aspects of the TGH distributions by a visuanimation (Genton et al., 2015) in the Supporting Information.

Extensions of the univariate TGH distribution to the multivariate case can be found in Field and Genton (2006) and He and Raghunathan (2012). Xu and Genton (2016) used the TGH transformation to build a new max-stable process for the modeling of spatial extremes. Recently, Xu and Genton (2017) further extended the TGH distribution to the spatial case and defined TGH random fields. Those models were found to be useful in practice and have been applied to air pollution, wind speed, rainfall, and economic data. In this paper, we apply the TGH transformation in the time series context and build two models that take advantage of the AR structure.

2.2 | TGH transformation of a latent AR process

Our first model transforms a latent Gaussian process similar to that of Xu and Genton (2017). Let $Z_t \sim \mathrm{AR}(p), t = 1, 2, \ldots,$ be a stationary Gaussian AR process of order p with zero mean and unit variance, and let $T_t = \tau_{g,h}(Z_t)$ by applying the standard TGH transformation to Z_t. Our model 1, the TGH transformed autoregressive process of order p (TGH-AR(p)-t) Y_t, has the same form as (1), with η being a generalized version of the TGH transformation, as follows:

$$Y_t = \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta} + \xi + \omega\tau_{g,h}(Z_t), \tag{4}$$

where $\boldsymbol{X}_t$ and $\boldsymbol{\beta}$ are the covariates and the corresponding coefficients, respectively, and ξ and ω are the location and scale parameters, respectively. In a time series analysis, the covariates are usually functions of t to model the trend or the seasonality, which may also be incorporated through a seasonal and integrated AR model. In our first model, unlike the commonly adopted Box–Cox approach, the deterministic part that consists of the location parameter, covariates, and their coefficients is outside the transformation. In this way, the trend (also seasonality or periodicity) is attributed directly to the process Y_t rather than the underlying Gaussian process Z_t. We control the skewness and tail behavior of the stochastic part of the process Y_t by applying the TGH transformation to Z_t with different values of g and h. The moments of Y_t can be computed as the moments for T_t in (3) with modification for the additional location and scale parameters. The autocovariance function $C_T(\tau)$ of T_t can be found in terms of the autocorrelation function (ACF) $\rho_Z(\tau)$ of the underlying Gaussian process Z_t, which was derived from Xu and Genton (2017), as follows:

$$C_T(\tau) = \frac{\exp\left[\frac{1+\rho_Z(\tau)}{1-h\{1+\rho_Z(\tau)\}}g^2\right] - 2\exp\left[\frac{1-h\{1-\rho_Z^2(\tau)\}}{(1-h)^2-h^2\rho_Z^2(\tau)}\frac{g^2}{2}\right] + 1}{g^2\sqrt{(1-h)^2 - \rho_Z^2(\tau)h^2}} - \{\mathbb{E}(T_t)\}^2.$$

The autocorrelation of Y_t is always weaker than the autocorrelation of Z_t, and in the Supporting Information, we illustrate this fact via Visuanimation S1.

When $p = 1$, the TGH-AR(1)-t process is of the form (4), with $Z_t = \phi Z_{t-1} + \epsilon_t$, and where ϵ_t is a Gaussian white-noise process with zero mean and constant variance $\sigma_\epsilon^2 = 1 - \phi^2$. There is a constraint that σ_ϵ^2 is a function of ϕ so that Z_t has a unit variance. This dependency may seem strange at first, but it is effectively equivalent to standardizing Z_t to a process with a unit variance by multiplication by a scaling constant. For example, the usual AR(1) process with a Gaussian error

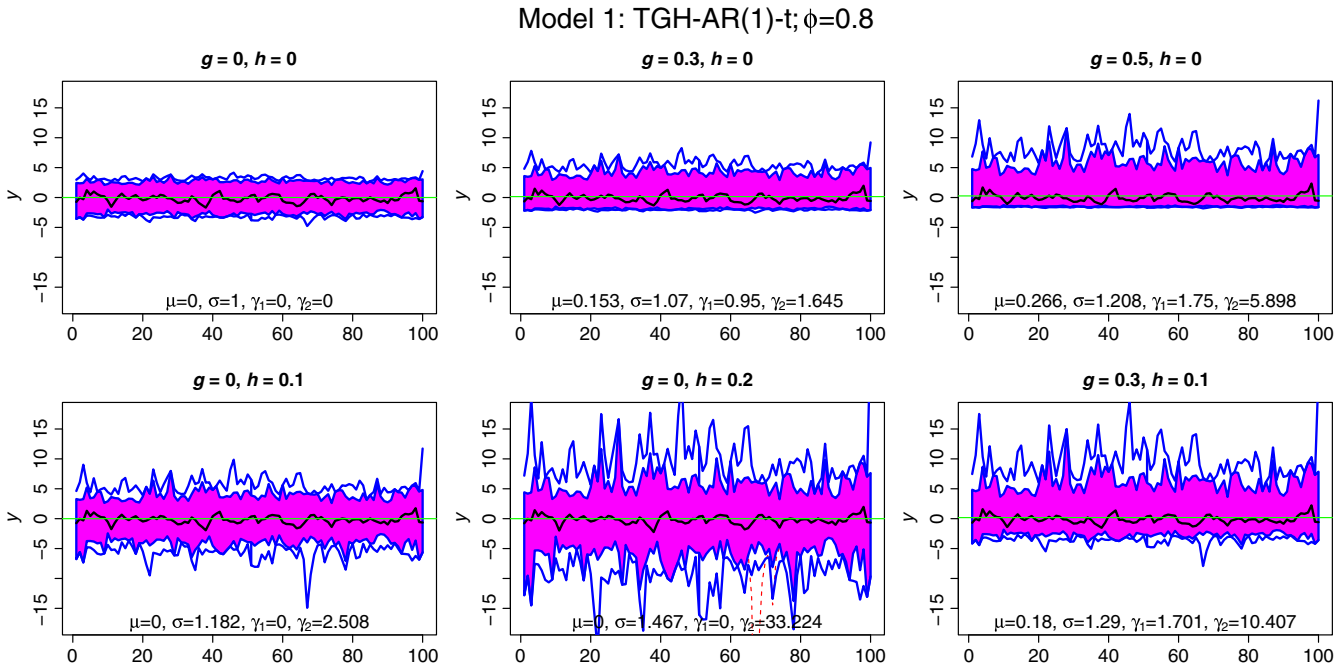

FIGURE 1 Functional boxplots of 1,000 realizations of sample size $n = 100$ from model 1, that is, TGH-AR(1)-t, without covariates, with ξ=0, ω=1, an AR coefficient of ϕ=0.8, and six pairs of values for g and h, labeled with values of the mean (μ, green line), standard deviation (σ), skewness (γ_1), and excess kurtosis (γ_2) of the process

term ϵ_t of variance σ_ω^2 will result in a process Z_t' of variance $\frac{\sigma_\omega^2}{1-\phi^2}$, and Z_t' can be standardized to have a unit variance by multiplying the rescaling factor by $\frac{\sqrt{1-\phi^2}}{\sigma_\omega}$. This is equivalent to an AR(1) process with an error term of variance $1-\phi^2$.

To simulate samples from the TGH-AR(p)-t model, we first simulate data from the underlying Gaussian AR(p) process Z_t for $t = 1, 2, \ldots, n$. Then, we transform the process by applying the generalized version of the TGH transformation to Z_t at each time point. Figure 1 shows the functional boxplots (Sun & Genton, 2011) of 1,000 realizations from the TGH-AR(1)-t model of sample size $n = 100$ without covariates, with $\xi = 0$, $\omega = 1$, $\phi = 0.8$, $\sigma_\epsilon^2 = 0.36$, and six pairs of values for g and h. The six pairs of values for g and h include the Gaussian AR case when $g = h = 0$. The functional boxplots are labeled with values of the marginal mean (μ, green line), standard deviation (σ), skewness (γ_1) and excess kurtosis (γ_2) for the transformed process Y_t. From the functional boxplots of the sample paths of Y_t, we see clearly when $h = 0$, as g increases from 0.3 to 0.5, the process is more right skewed; when $g = 0$, as h increases from 0.1 to 0.2, extreme values occur more in the time series. With $g = 0.3, h = 0.1$, we see the effect of both g and h on the process.

2.3 | AR process with TGH error

Instead of applying the transformation to the time series itself, in our second model, we transform the independent and identically distributed (i.i.d) error term $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ in an AR process to a non-Gaussian error term $T_t = \tau_{g,h}(\epsilon_t)$ that follows a standard TGH distribution. Thus, model 2, the AR(p) process with TGH error (the TGH-AR(p)-e model) is defined as follows:

$$Y_t = \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta} + \xi + \phi_1\tilde{Y}_{t-1} + \cdots + \phi_p\tilde{Y}_{t-p} + \omega\tau_{g,h}(\epsilon_t), \tag{5}$$

where $\tilde{Y}_t = Y_t - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta} - \xi$ is a process of median 0 (however, not a zero-mean process when $g \neq 0$); $\boldsymbol{X}_t$ and $\boldsymbol{\beta}$ are the covariates and corresponding coefficients, respectively; ξ and ω are the location and scale parameters, respectively; and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^{\mathrm{T}}$ are the AR coefficients under constraints such that the AR process is weakly stationary. We express (5) in a moving average form: $\tilde{Y}_t = \psi(B)(\omega T_t)$, where $\psi(B) = \sum_{j=0}^{\infty}\psi_j B^j$ of the backshift operator B. Although the marginal distribution of Y_t in the TGH-AR(p)-e model cannot be written down in closed form, the moments of $\tilde{Y}_t$ can be derived with respect to moments of the standardized error term T_t: $\mathbb{E}(\tilde{Y}_t) = \omega\sum_{j=0}^{\infty}\psi_j\mathbb{E}(T_t) = \frac{\omega}{1-\sum_{j=1}^{p}\phi_j}\mathbb{E}(T_t)$, $\mathbb{V}(\tilde{Y}_t) = \omega^2\sum_{j=0}^{\infty}\psi_j^2\mathbb{V}(T_t)$.

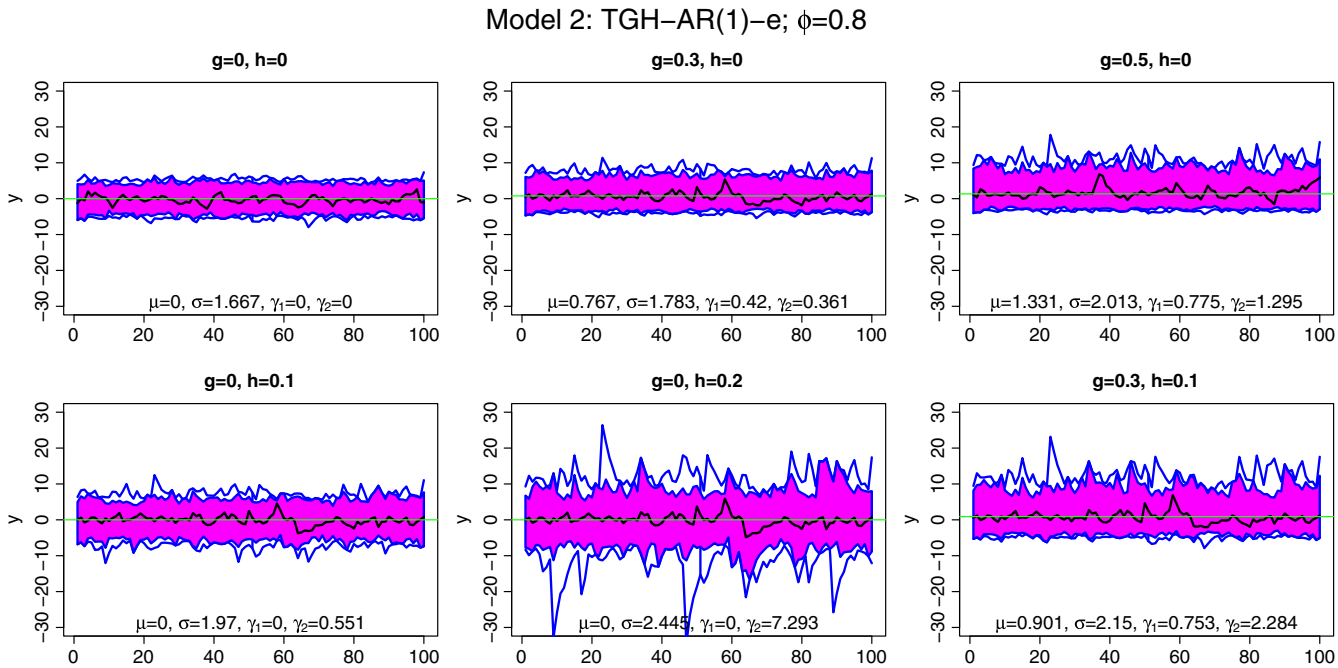

FIGURE 2 Functional boxplots of 1,000 realizations of sample size $n = 100$ from model 2, that is, TGH-AR(1)-e, without covariates, with ξ=0, ω=1, an AR coefficient of ϕ=0.8, and six pairs of values for g and h, labeled with values of the mean (μ, green line), standard deviation (σ), skewness (γ_1), and excess kurtosis (γ_2) of the process

Furthermore, $\gamma_1(\tilde{Y}_t) = \frac{\sum_{j=0}^{\infty}\psi_j^3}{(\sum_{j=0}^{\infty}\psi_j^2)^{3/2}}\gamma_1(T_t)$, $\gamma_2(\tilde{Y}_t) = \frac{\sum_{j=0}^{\infty}\psi_j^4}{(\sum_{j=0}^{\infty}\psi_j^2)^{2}}\gamma_2(T_t)$, as long as the summation converges. These relationships conform with the algebraic relations between the skewness and kurtosis of the process and error term in ARMA processes derived by Davies et al. (1980). Figure 2 shows the functional boxplots of 1,000 realizations from the TGH-AR(1)-e model with a sample size of 100, and the same parameters as in Figure 1. Again, we see that skewness and tail behavior of the process can be controlled by g and h.

2.4 | Comparison of the two models

Comparing Figures 1 and 2, we notice that, given the same values for g and h, the skewness and kurtosis of the TGH-AR(1)-t process are larger than those of the TGH-AR(1)-e process, while the variance is smaller. This agrees with the findings of Davies et al. (1980), where the ARMA process has skewness and kurtosis less than the error process.

We interpret the difference between the two models from the viewpoint of the data-generating mechanism. In model 1, TGH-AR-t, we assume that there is an underlying latent Gaussian process Z_t and that the observed process Y_t is a realization of a nonlinear transformation of Z_t. Goerg (2011) justified this idea using financial data. Overreactions to changes in the stock market skew the underlying symmetric process of financial news and result in skewed log-returns of the stock market; see also De Luca, Genton, and Loperfido (2006). For model 2, the TGH-AR-e model, we do not assume such a latent process, but Y_t itself is an AR model with non-Gaussian noise that follows the TGH distribution. The difference between the two models can be seen most obviously by plotting the relationship between $Y_t - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta}$ and $Y_{t-1} - \boldsymbol{X}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}$ (Figure 3). The relationship is linear for model 2 and nonlinear for model 1. The exploratory plots shown in Figure 3 for a sample size of $n = 500$ and different values of g, h, and ϕ should be consulted as a reference for deciding whether model 1 or 2 is more suitable for the time series data to be analyzed. In practice, one should take into account both a reasonable data-generating mechanism for the process and the empirical behavior of $Y_t - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta}$ versus $Y_{t-1} - \boldsymbol{X}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}$ to decide which model to use for a specific data set.

We want to emphasize that our time series models are not simply variations of the spatial case (Xu & Genton, 2017). We utilize the unique properties of the AR structure for discrete time processes, which the random field approach does not possess. One main difference between the AR time series and the random fields is that the correlation structure of the random process is induced by the AR coefficients, whereas in the random fields, it is modeled directly. Moreover, the

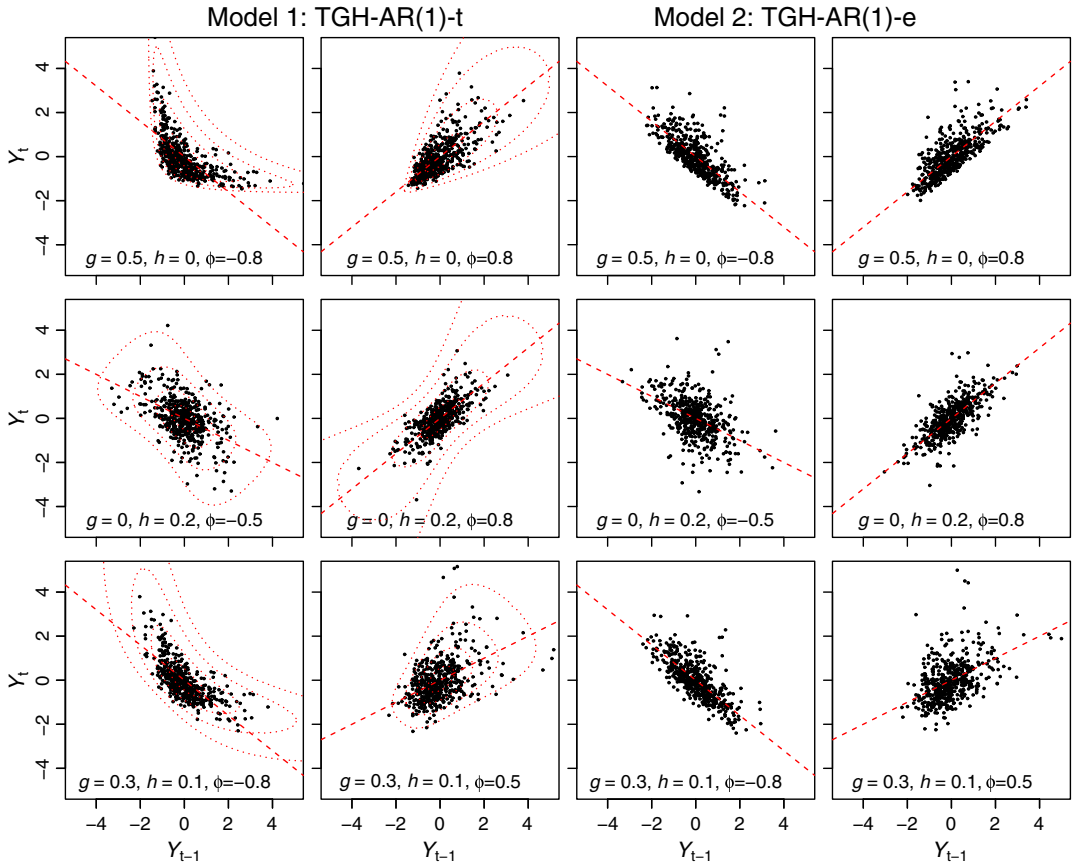

FIGURE 3 Illustration of the nonlinear relationship between Y_t and Y_{t-1} for the TGH-AR(1)-t model, overlapped with theoretical contours of the joint density (two left columns); illustration of the linear relationship between Y_t and Y_{t-1} for the TGH-AR(1)-e model with skewed and/or heavy-tailed noise (two right columns). Each panel is plotted with one realization of sample size $n = 500$ from our models without covariates, with ξ=0, ω=1, and different values for g, h, and ϕ

property of the conditional independence of an AR process allows us to write down the likelihood explicitly instead of a matrix form as for the random fields. The AR structure also makes it possible for us to build the TGH-AR-e model, which is not possible in the random fields setting.

3 | INFERENCE

3.1 | Parameter estimation

A drawback of the TGH transformation is that its inverse transformation does not have an explicit form (except when either g or h is equal to 0), and thus, the likelihood inference is difficult. Earlier estimation procedures rely on matching sample quantiles or sample moments with their population counterparts to avoid this difficulty. For example, Hoaglin (1985) developed an estimation procedure by matching a sequence of sample quantiles and theoretical quantiles, which we refer to as the letter-value-based method. Recently, Xu and Genton (2015) proposed an efficient parameter estimation algorithm for the independent TGH distribution by using an approximated likelihood. Their algorithm greatly improved the parameter estimation performance compared with the moment- or quantile-based methods without compromising the computational speed.

The parameters involved in our two TGH-AR models are $\xi, \omega, g, h, \boldsymbol{\beta}$, and $\boldsymbol{\phi}$. Because the maximum-likelihood estimator (MLE) is well known to possess many good asymptotic properties, we prefer the likelihood-based estimator over the moment- or quantile-based estimators. The vector of the parameters related to the TGH transformation is denoted by $\boldsymbol{\theta}_1 = (\xi, \omega, g, h, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$, the autoregressive parameters by $\boldsymbol{\phi}$, and set $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\phi}^{\mathrm{T}})^{\mathrm{T}}$.

For model 1, the log-likelihood function given n observations $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$ can be written as follows:

$$\begin{aligned} l_n(\boldsymbol{\theta}) =& f_{\boldsymbol{\phi}}(z_{1,\theta_1}) + f_{\boldsymbol{\phi}}\left(z_{2,\theta_1}|z_{1,\theta_1}\right) + \cdots + f_{\boldsymbol{\phi}}\left(z_{n,\theta_1}|z_{n-1,\theta_1}, \ldots, z_{n-p,\theta_1}\right) - n\log\omega \\ & - \frac{h}{2}\sum_{t=1}^{n} z_{t,\theta_1}^2 - \sum_{t=1}^{n}\log\left[\exp(gz_{t,\theta_1}) + \frac{h}{g}\left\{\exp(gz_{t,\theta_1}) - 1\right\} z_{t,\theta_1}\right], \end{aligned} \tag{6}$$

where $z_{t,\theta_1} = \tau_{g,h}^{-1}\{\frac{y_t - \xi - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\beta}}{\omega}\}$, $f_{\boldsymbol{\phi}}(\cdot|\cdot)$ is the conditional log-likelihood for the underlying Gaussian AR(p) process Z_t, which is also Gaussian.

Because an explicit form of the likelihood of Y_t is intractable for model 2, we consider the conditional likelihood given the first $k \geq p$ observations. The conditional log-likelihood of model 2 can be written as follows:

$$\begin{aligned} l_n(\boldsymbol{\theta}|\, y_1, \ldots, y_k) =& -\frac{h+1}{2}\sum_{t=k+1}^{n}\epsilon_{t,\theta}^2 - \sum_{t=k+1}^{n}\log\left[\exp(g\epsilon_{t,\theta}) + \frac{h}{g}\left\{\exp(g\epsilon_{t,\theta}) - 1\right\}\epsilon_{t,\theta}\right] \\ & - (n-k)\log\omega - \frac{n-k}{2}\log(2\pi), \end{aligned} \tag{7}$$

where $\epsilon_{t,\theta} = \tau_{g,h}^{-1}\{\frac{\tilde{y}_t - \phi_1\tilde{y}_{t-1} - \cdots - \phi_p\tilde{y}_{t-p}}{\omega}\}$ and $\tilde{y}_t = y_t - \xi - \boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\beta}$.

Although we can find the MLE for $\boldsymbol{\theta}$ in our two models by maximizing Equation (6) or (7) with respect to $\boldsymbol{\theta}$ in principle, it is computationally expensive to find the inverse $\tau_{g,h}^{-1}(\cdot)$ numerically for each data point and iteration of the optimization because $\tau_{g,h}^{-1}(\cdot)$ does not have a closed form. To bypass this computational challenge, we borrow the idea of approximated likelihood from Xu and Genton (2015) for the univariate TGH distribution. Xu and Genton (2017) extended this idea to the random field case and proposed an algorithm for iteratively estimating both the parameters related to the TGH transformation and the parameters in the model of the covariance function. In essence, they linearized the inverse transformation function piecewisely and maximized the approximated likelihood function with the piecewisely linearized inverse function instead of the exact one.

Equipped with this linearization procedure, we now numerically maximize the approximated log-likelihood by using the approximated inverse transformation function $\tilde{\tau}_{g,h}^{-1}$ instead of $\tau_{g,h}^{-1}$ in either Equation (6) or (7). Thus, we obtain the maximum approximated likelihood estimator (MALE) of the parameters in our models much faster. We also found that iteratively optimizing $\boldsymbol{\theta}_1$ and $\boldsymbol{\phi}$ is better than optimizing $\boldsymbol{\theta}$ directly; however, it is about 20 times slower on a Lenovo computer with 16 2.50GHZ Intel Xeon(R) CPUs. Hence, we optimize all of the parameters for estimation directly in our algorithm.

3.2 | Order selection

The order selection for an AR model is usually performed via the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). In a simulation study (all simulations are done on the same computer mentioned above), we found that the algorithm based on BIC performs much better than the one based on Akaike information criterion for our models. Therefore, we use the order selection procedure based on BIC with our estimation algorithm. For model 1, BIC$=-2l_n(\boldsymbol{\theta}) + (p+4)\log n$; for model 2, BIC$=-2l_n(\boldsymbol{\theta}|\, y_1, \ldots, y_k) + (p+4)\log(n-k)$, and again, we use the approximated likelihood instead of the exact one.

We evaluate the empirical correct order selection rate by BIC with the approximated likelihood for our two models through a simulation study. We generate data from both models with $\xi = 0$, $\omega = 1$, $g = 0.3, h = 0.1$ without covariates, with sample sizes $n = 100$ and $n = 500$, and a true AR order of $p = 0, 1, 2$. For $p = 1$ and $p = 2$, we use multiple values for the AR coefficient(s), which exhibit different behaviors of the ACF and different intensities for the spectral density. We get correct order selection rate from 1,000 simulations. Table 1 summarizes the correct order selection rate by BIC for our two models with different sample sizes n and orders p for different values of $\phi(\boldsymbol{\phi})$.

The correct order selection rate improved and reached over 98% by increasing the sample size from 100 to 500 in all cases. Order selection performance is similar with the data generated by all the six pairs of values for g and h as used in the functional boxplots. We conclude that the overall correct order selection rate is satisfactory for our order selection procedure based on BIC with the approximated likelihood.

TABLE 1 Correct order selection rate for both the TGH-AR(p)-t and TGH-AR(p)-e models without covariates and with $\xi = 0$, $\omega = 1$, $g = 0.3, h = 0.1$, and a true AR order of 0, 1, and 2 for $n = 100$ and $n = 500$

		TGH-AR(p)-t		TGH-AR(p)-e	
		$n = 100$	$n = 500$	$n = 100$	$n = 500$
$p = 0$		0.959	0.988	0.938	0.990
$p = 1$	$\phi = -0.5$	0.951	0.987	0.934	0.987
	$\phi = 0.5$	0.950	0.988	0.932	0.981
	$\phi = -0.8$	0.952	0.982	0.940	0.985
	$\phi = 0.8$	0.961	0.987	0.942	0.992
$p = 2$	$\boldsymbol{\phi} = (-0.5, -0.3)^{\mathrm{T}}$	0.825	0.991	0.858	0.987
	$\boldsymbol{\phi} = (0.2, 0.4)^{\mathrm{T}}$	0.532	0.986	0.592	0.986
	$\boldsymbol{\phi} = (0.8, -0.25)^{\mathrm{T}}$	0.706	0.987	0.777	0.983
	$\boldsymbol{\phi} = (1.5, -0.75)^{\mathrm{T}}$	0.948	0.982	0.937	0.983

3.3 | Forecasting

In this section, we derive one-step-ahead point and probabilistic forecasts for our two models.

For model 1, $Y_t = \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta} + \xi + \omega T_t$, given p observations $y_t, \ldots, y_{t-p+1}$, we derive the following conditional distribution:

$$T_{t+1}|y_t, \ldots, y_{t-p+1} \sim \tau_{g,h}(\tilde{\mu} + \tilde{\sigma}Z), Z \sim \mathcal{N}(0,1),$$

where $\tilde{\mu}$ and $\tilde{\sigma}^2$ are the conditional mean and variance, respectively, for $Z_{t+1}|Z_t, \ldots, Z_{t-p+1}$ of the underlying Gaussian AR(p) process Z_t. Here, $\tilde{\mu}$ and $\tilde{\sigma}^2$ are determined by $\boldsymbol{\phi}$ and can be found efficiently by the Durbin–Levinson algorithm. With this conditional distribution, the point predictors for minimizing the absolute prediction error (conditional median) and the squared prediction error (conditional mean) are $\hat{Y}_{t+1} = \xi + \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \omega\tau_{g,h}(\tilde{\mu})$ and $\hat{Y}_{t+1} = \xi + \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \frac{\omega}{g\sqrt{1-h\tilde{\sigma}^2}}\exp\{\frac{h\tilde{\mu}^2}{2(1-h\tilde{\sigma}^2)}\}[\exp\{\frac{g^2\tilde{\sigma}^2+2g\tilde{\mu}}{2(1-h\tilde{\sigma}^2)}\} - 1]$, respectively.

For model 2, the conditional distribution is as follows:

$$Y_{t+1}|y_t, \ldots, y_{t-p+1} \sim \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \xi + \phi_1\tilde{y}_t + \cdots + \phi_p\tilde{y}_{t-p+1} + \omega\tau_{g,h}(Z), Z \sim \mathcal{N}(0,1).$$

The point predictors for minimizing the absolute loss (conditional median) and squared loss (conditional mean) are $\hat{Y}_{t+1} = \xi + \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \phi_1\tilde{y}_t + \cdots + \phi_p\tilde{y}_{t-p+1}$ and $\hat{Y}_{t+1} = \xi + \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \phi_1\tilde{y}_t + \cdots + \phi_p\tilde{y}_{t-p+1} + \frac{\omega}{g\sqrt{1-h}}[\exp\{\frac{g^2}{2(1-h)}\} - 1]$, respectively. We note that the conditional median has the same form as that of an AR model with Gaussian error. In practice, we need to estimate the parameters first to make forecasts for future values of the time series based on those estimated parameters. Thus, the difference between the point predictions based on the conditional medians of our model and the Gaussian AR model comes from the difference in their respective estimations of $\boldsymbol{\beta}, \xi$, and $\boldsymbol{\phi}$.

Prediction confidence intervals (CIs) can be found easily from the conditional distributions of our two models. There are different ways to form the two-sided CI for a given distribution. One popular choice for the $(1-\alpha)100\%$ CI is to exclude $\alpha/2$ from both tails and then use the central $1-\alpha$ probability interval of the distribution, which we refer to as the symmetric weight CI. Another choice is to find the shortest $1-\alpha$ probability interval, which we refer to as the minimum-length CI. The minimum-length CI coincides with the symmetric weight CI for symmetric distributions. For model 1, the lower and upper bounds of the 95% prediction CI can be found by transforming the lower and upper bounds of a 95% probability interval of the underlying normal distribution of mean $\tilde{\mu}$ and variance $\tilde{\sigma}^2$. The symmetric weight prediction interval is $[\boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \xi + \omega\tau_{g,h}(\tilde{\mu} - z_{1-\alpha/2}\tilde{\sigma}), \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \xi + \omega\tau_{g,h}(\tilde{\mu} + z_{1-\alpha/2}\tilde{\sigma})]$, and the minimum-length prediction interval is $[\boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \xi + \omega\tau_{g,h}(\tilde{\mu} - z_{1-\gamma^{\mathrm{opt}}}\tilde{\sigma}), \boldsymbol{X}_{t+1}^{\mathrm{T}}\boldsymbol{\beta} + \xi + \omega\tau_{g,h}(\tilde{\mu} + z_{1-\alpha+\gamma^{\mathrm{opt}}}\tilde{\sigma})]$, where γ^{opt} needs to be optimized numerically for different values of g and h for given $\tilde{\mu}$ and $\tilde{\sigma}$. For model 2, the symmetric weight prediction interval is $[\hat{Y}_{t+1} + \omega\tau_{g,h}(-z_{1-\alpha/2}), \hat{Y}_{t+1} + \omega\tau_{g,h}(z_{1-\alpha/2})]$, and the minimum-length prediction interval is $[\hat{Y}_{t+1} + \omega\tau_{g,h}(-z_{1-\gamma^{\mathrm{opt}}}), \hat{Y}_{t+1} + \omega\tau_{g,h}(z_{1-\alpha+\gamma^{\mathrm{opt}}})]$, where $\hat{Y}_{t+1}$ is the conditional median of $Y_{t+1}|y_t, \ldots, y_{t-p+1}$.

The prediction intervals given above are derived with respect to the true parameters of the two models. However, in practice, those parameters need to be estimated, and the prediction interval with the estimated parameters should be inflated, taking the uncertainty in parameter estimation into account. Research on mean squared prediction error (MSPE) with estimated parameters can be referred to a vast literature in the context of linear models, time series models, and spatial models (Zimmerman & Cressie, 1992). In particular, for a Gaussian AR(p) model, the MSPE of a one-step-ahead forecast with estimated autoregressive coefficients is inflated by a factor of $1 + p/n$, where n is the sample size (Yamamoto, 1976).

Many authors (e.g., Toyooka, 1982) have concluded that the bias of the estimated MSPE is asymptotically negligible, and we see later in the simulation study that the empirical coverage of the prediction CIs by ignoring the uncertainty in parameter estimation is close to the nominal level.

4 | SIMULATION STUDY

We perform a Monte Carlo simulation study to assess the performance of our models from different aspects. We consider six pairs of values for g and h (same values as in the functional boxplots) and sample sizes $n = 100, 250, 500$. For each of our TGH-AR(1)-t and TGH-AR(1)-e models, each sample size n, and each pair of values for g and h, we first generate one realization from the model with $\xi = -3$, $\omega = 1.5$, $\boldsymbol{\beta} = (3, -2)^{\mathrm{T}}$, $\boldsymbol{X}_t = \{\cos(2\pi t/24), \sin(2\pi t/24)\}^{\mathrm{T}}$, and $\phi = 0.8$. Next, we carry out parameter estimation by three methods: the MALE for both of our models without order selection, and the MLE for a Gaussian AR model without order selection. Finally, we use again the three methods based on the TGH-AR-t, TGH-AR-e, and Gaussian AR models, to produce point and probabilistic forecasts with the estimated parameters. For each scenario, we run the above procedure 1,000 times.

In Section 4.1, we compare the MALE for the TGH-AR(1)-t model with two other estimators for the independent TGH distribution when data are generated from the TGH-AR(1)-t model. In Section 4.2 we evaluate the finite sample behavior of the MALE for both models, and in Section 4.3, we compare the forecasting performances of our models with the performance of a Gaussian AR model.

4.1 | Comparison of estimation methods

Using data generated from the TGH-AR-t model, we first check the improvement of estimation performance by using the MALE to simultaneously estimate $\boldsymbol{\theta}_1$ and $\boldsymbol{\phi}$ rather than sequentially estimating $\boldsymbol{\theta}_1$ from Y_t, treating them as independent realizations, and then estimate $\boldsymbol{\phi}$ from the transformed process $\tilde{Z}_t = \tau^{-1}_{\hat{g},\hat{h}}(\frac{Y_t - \hat{\xi} - \boldsymbol{X}_t^{\mathrm{T}}\hat{\boldsymbol{\beta}}}{\hat{\omega}})$. The latter approach is often adopted by practitioners to first find the optimal transformation and then estimate the temporal structure from the "Gaussianized" process. For this comparison, we use two additional methods for estimation in the simulation procedure as described before: the letter-value-based method (Hoaglin, 1985) and the MALE for independent TGH distributions (Xu & Genton, 2015) to estimate $\boldsymbol{\theta}_1$ first and then ϕ by the transformed process $\tilde{Z}_t$.

Figure 4 presents boxplots of the parameters estimated by the three different estimators for 1,000 replications with $g = 0.3, h = 0.1$ for three sample sizes. We see that the bias and variance of the estimators improve dramatically as the sample size grows from 100 to 500 for all three methods. Our joint estimation procedure outperforms the other two sequential methods, especially for h and ϕ, which indicate that estimating the optimal transformation by ignoring the dependency structure of the underlying process deteriorates the estimation. Hence, the comparison result justifies the need to develop such a tailored estimation procedure for the TGH distribution in the time series context.

4.2 | Estimation performance of MALE

Next we check the finite sample behavior of the MALE for both models. Figure 5 shows the estimation results from our two TGH-AR(1) models with $\phi = 0.8$, $n = 100$ for six pairs of different values of g and h. For $g = h = 0$, which corresponds to a Gaussian AR(1) process, we also include boxplots of the MLE based on the Gaussian AR model for $\xi, \omega, \beta_1, \beta_2$, and ϕ. Visuanimations of boxplots in the same manner as sample size grows ($n = 100, 250, 500$) for both models can be found in the Supporting Information (Visuanimations S2 and S3). Results of estimation performance for the two TGH-AR(2) models with $\phi_1 = 0.8, \phi_1 = -0.25$ are also included in the Supporting Information (Visuanimations S4 and S5).

From Figure 5, we can see that with sample size $n = 100$, the MALE for ω, h, and ϕ is a bit biased. Nevertheless, as shown by Visuanimations S2 and S3 in the Supporting Information, the estimation improves greatly as the sample size increases to $n = 250$. In addition, under $g = h = 0$ when the true distribution is Gaussian, with two more parameters, g and h, to estimate, the estimators based on the approximated TGH-AR likelihood do not deteriorate much from the Gaussian MLE. The visuanimations also give us an empirical guideline about how the sample size affects the estimation performance. To get an overall satisfying estimation performance by the MALE, we suggest a sample size no less than $n = 250$.

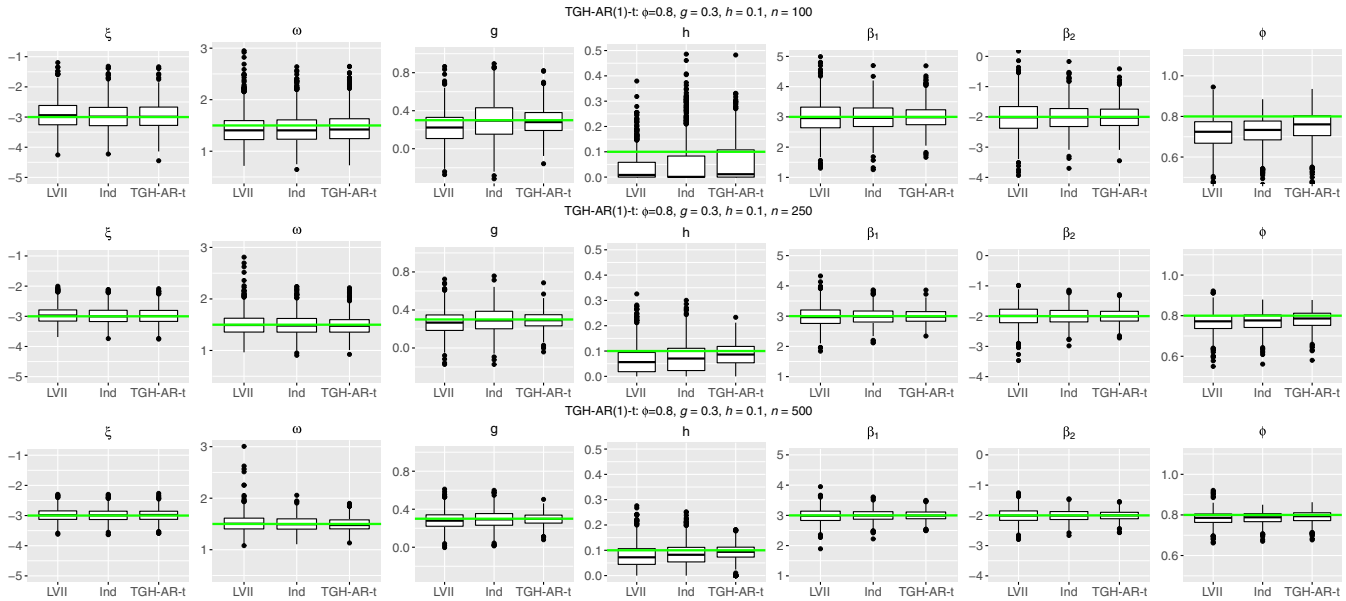

FIGURE 4 Comparison of three estimators: letter-value-based method (LVII), maximum approximated likelihood estimator based on independent Tukey g-and-h (TGH) distribution (Ind), and maximum approximated likelihood estimator for the TGH-AR-t model (TGH-AR-t), with boxplots of each estimator for each parameter from 1,000 replications with data generated from the TGH-AR(1)-t with true parameters indicated by green line

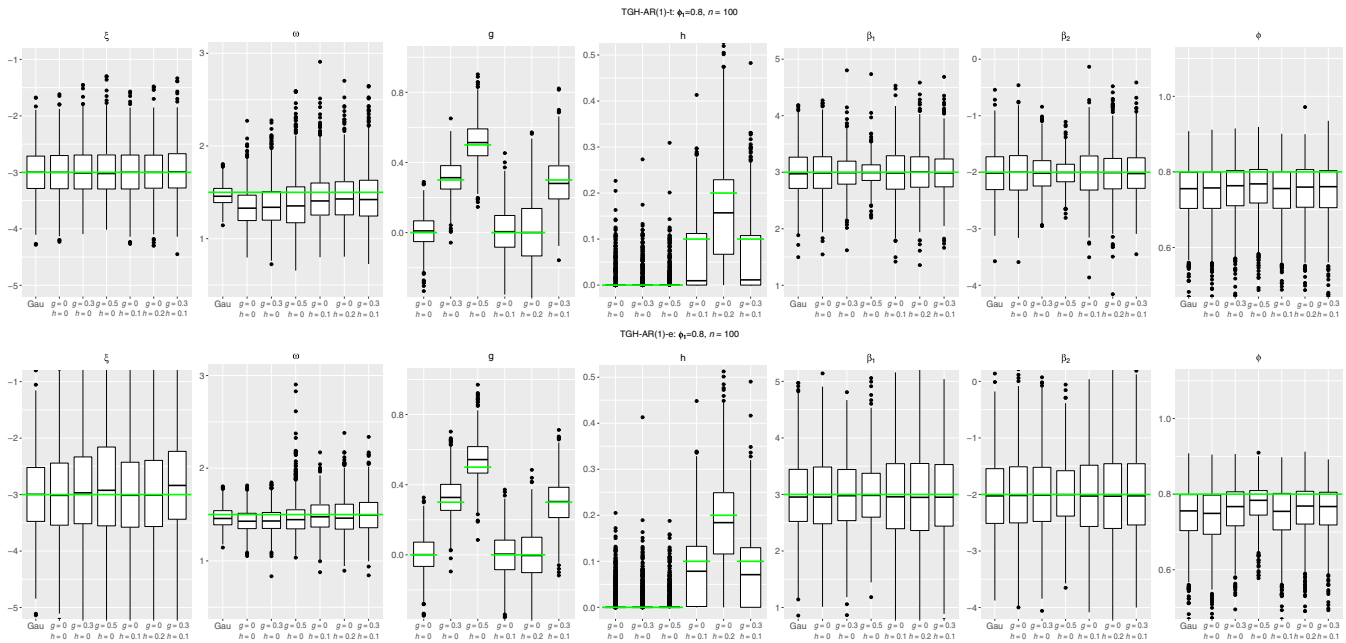

FIGURE 5 Boxplots of the maximum approximated likelihood estimator in 1,000 simulations for the two TGH-AR(1) models with $n = 100$ for six pairs of different values of g and h for each parameter with true value indicated by green line. For $g = h = 0$, boxplots of the maximum-likelihood estimator based on the Gaussian AR(1) model (indicated by "Gau") for $\xi, \omega, \beta_1, \beta_2$, and ϕ are also included

4.3 | Forecasting performance

Finally, we evaluate the forecasting performance for our two models with parameters estimated by the corresponding MALE without order selection. Table 2 summarizes the performance of the point forecasts from each of the three methods based on the TGH-AR(1)-t, TGH-AR(1)-e, and Gaussian AR(1) models, and for data generated from each of our two TGH-AR(1) models with $\xi = -3, \omega = 1.5, \boldsymbol{\beta} = (3, -2)^{\mathrm{T}}, \boldsymbol{X}_t = \{\cos(2\pi t/24), \sin(2\pi t/24)\}^{\mathrm{T}}$ and $\phi = 0.8$ for $g = 0.3, h = 0.1$,

TABLE 2 Summary of point forecast performance of the three methods based on the TGH-AR(1)-t, TGH-AR(1)-e, and Gaussian AR(1) models, and for data generated from each of our two TGH-AR(1) models with $\xi = -3$, $\omega = 1.5, \boldsymbol{\beta} = (3, -2)^{\mathrm{T}}, \boldsymbol{X}_t = \{\cos(2\pi t/24), \sin(2\pi t/24)\}^{\mathrm{T}}$, and $\phi = 0.8$ for $g = 0.3, h = 0.1$, and $n = 500$

Data generated from	Model 1: TGH-AR(1)-t			Model 2: TGH-AR(1)-e		
Modeled by	TGH-AR(1)-t	TGH-AR(1)-e	Gau-AR(1)	TGH-AR(1)-t	TGH-AR(1)-e	Gau-AR(1)
MAE	**0.860**	0.869	0.870	1.374	**1.365**	1.379
RMSE	**1.124**	1.142	1.139	1.798	**1.795**	1.801
95% CI coverage	**95.6%**	95.2%	94.6%	95.0%	**95.6%**	95.6%
95% CI width	**4.30**	4.60	4.68	7.40	**7.24**	7.55

Note. MAE = mean absolute error; RMSE = root mean square error; CI = confidence interval.

and $n = 500$. The mean absolute errors (MAEs) are calculated from the conditional median, and the root mean square errors (RMSEs) from the conditional mean as a point predictor. It also shows the empirical coverage and average length of the 95% minimum-length CI from each method.

The probabilistic forecasts of the three methods can be evaluated through a histogram of the PIT values, which is to apply the conditional cumulative distribution function to the true value of the process at the time point for forecasting. If the conditional distribution assumed by a certain model conforms with the true conditional distribution, the histogram should be flat (uniform). Another metric for evaluating probabilistic forecast is the continuous ranked probability score (CRPS); more research on probabilistic forecasts and CRPS for a Gaussian distribution can be found in Gneiting and Katzfuss (2014) and Gneiting, Larson, Westrick, Genton, and Aldrich (2006). Xu and Genton (2017) derived the CRPS for a TGH distribution. The lower the CRPS is, the better the probabilistic forecast is. A plot of the empirical versus nominal

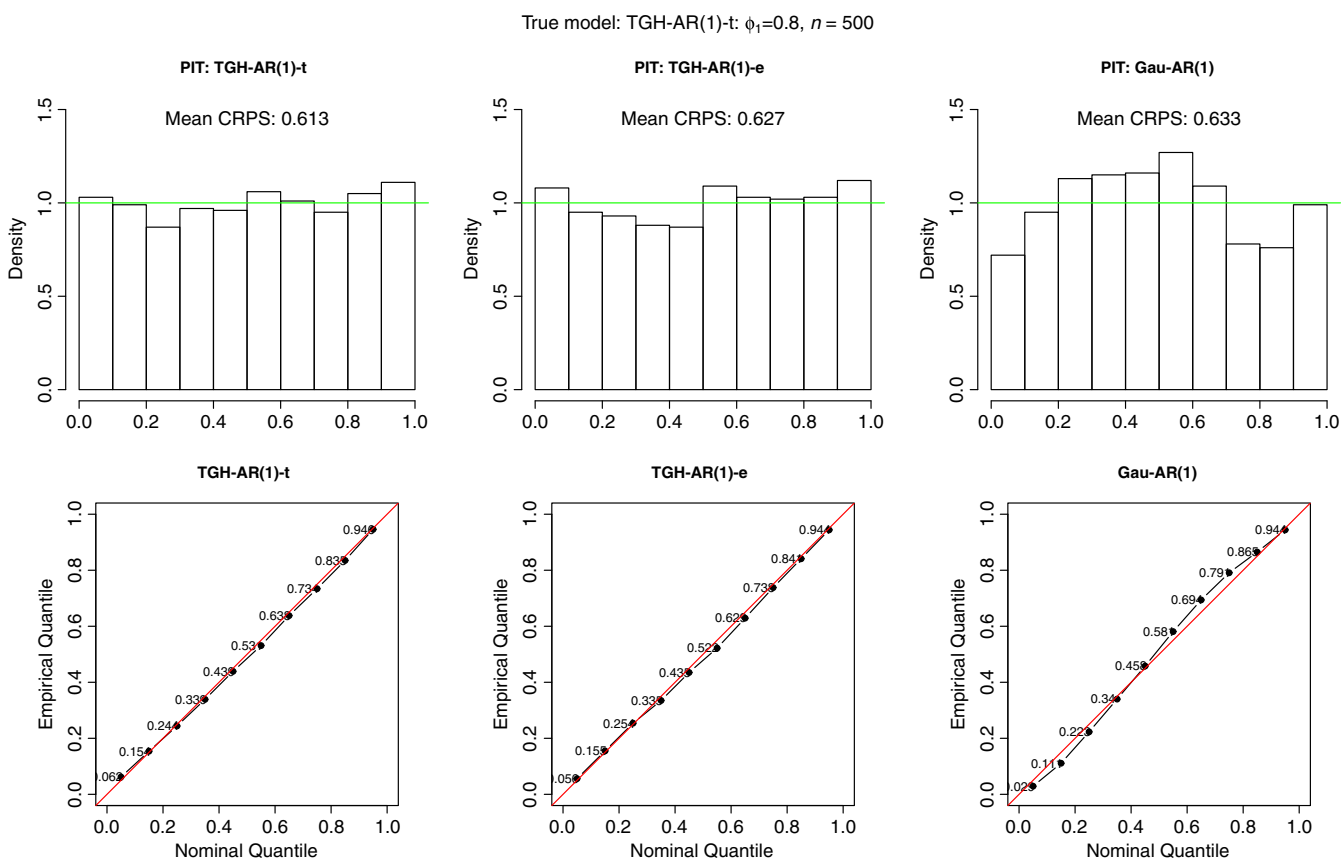

FIGURE 6 Comparison of probabilistic forecast performances via probability integral transform (PIT) and reliability plots when data are generated from the TGH-AR(1)-t model with $\xi = -3, \omega = 1.5, \boldsymbol{\beta} = (3, -2)^{\mathrm{T}}, \boldsymbol{X}_t = \{\cos(2\pi t/24), \sin(2\pi t/24)\}^{\mathrm{T}}$, and $\phi = 0.8$ for $g = 0.3, h = 0.1$, and $n = 500$ and fitted for the three methods: TGH-AR(1)-t, TGH-AR(1)-e, and Gaussian AR(1). Mean continuous ranked probability score (CRPS) value is also labeled

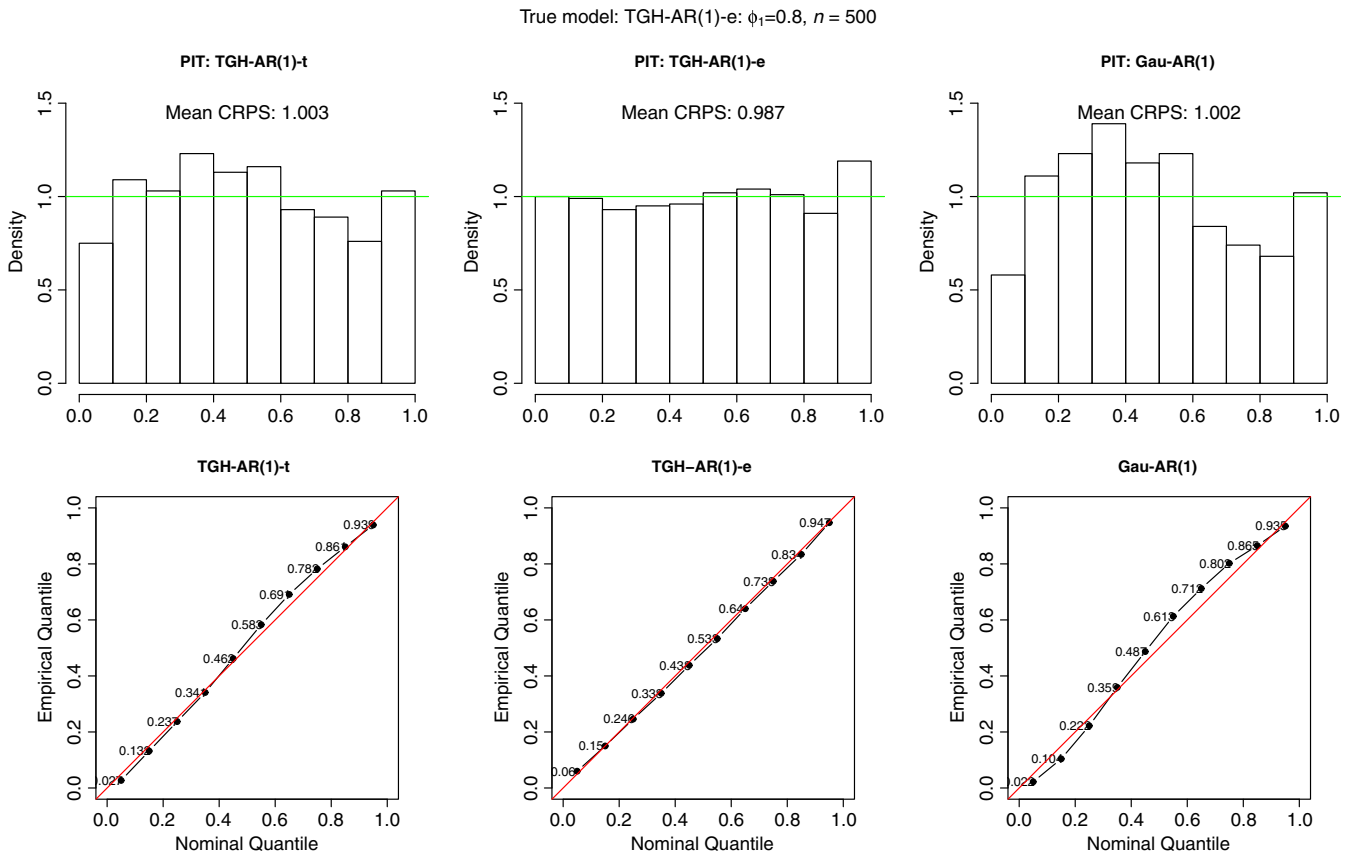

FIGURE 7 Comparison of probabilistic forecast performances via probability integral transform (PIT) and reliability plots when data are generated from TGH-AR(1)-e model with $\xi = -3$, $\omega = 1.5$, $\boldsymbol{\beta} = (3, -2)^{\mathrm{T}}$, $\boldsymbol{X}_t = \{\cos(2\pi t/24), \sin(2\pi t/24)\}^{\mathrm{T}}$, and $\phi = 0.8$ for $g = 0.3$, $h = 0.1$, and $n = 500$ and fitted for the three methods: TGH-AR(1)-t, TGH-AR(1)-e, and Gaussian AR(1). Mean continuous ranked probability score (CRPS) value is also labeled

quantile, that is, a reliability plot, can be used to check the quality of the quantile prediction. Figures 6 and 7 show the histograms of the PIT values (labeled with the mean CRPS value) and reliability plots for all three forecasting methods applied to the data generated from each of our TGH-AR(1) model. Figures S1 and S2 in the Supporting Information show probabilistic forecast results for the two TGH-AR(2) models with $\phi_1 = 0.8, \phi_1 = -0.25$.

Even though in Table 2, the MAEs and RMSEs do not differ much between the models, superiority of a model can be seen evidently by comparison of the PIT histograms. Not surprisingly, the best forecast is achieved by using the same model that the data are generated from: The forecast has the smallest MAE and RMSE, a flatter PIT histogram, and a smaller mean CRPS; the reliability plot is close to a 45° straight line; the empirical coverage of the 95% CI is close to the nominal level while the width is shorter than that of the Gaussian AR predictor. Note that, although we used the prediction CIs derived in Section 3.3 by treating the estimated parameters as the true parameters, Table 2 shows that the empirical coverage of the 95% CIs is close to the nominal level. We also notice that the TGH-AR-t and TGH-AR-e models produce quite different forecasts via their distinctive histograms of the PIT values, which means that the two models are not interchangeable.

5 | APPLICATION TO WIND SPEED DATA

The analysis of wind data is a crucial step in simulations for climate science. Accurate forecasting of wind speeds and quantifying the forecast uncertainty are also important for exploiting wind as clean energy. In this section, we illustrate the usefulness of our two non-Gaussian time series models under both wind speed simulation and wind speed forecasting scenarios. In Section 5.1, we fit daily wind speed data with the TGH-AR-t model to get parameter estimation for the purpose of fast wind speed simulation. In Section 5.2, we apply the TGH-AR-e model in order to make better wind speed

forecasts. For the two data sets in the two subsections, the reason to use one specific model is based on plots of $Y_t - \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta}$ versus $Y_{t-1} - \boldsymbol{X}_{t-1}^{\mathrm{T}}\boldsymbol{\beta}$ and with reference to Figure 3 to see whether the relationship is linear or not.

5.1 | Wind speed simulation

Climate models can produce multiple outputs of spatiotemporal wind speed data over the globe. Statistical models have been developed to reproduce the output from climate models for the sake of fast simulation instead of using computationally expensive physical models. In order to fit a space–time statistical model for wind field over the entire world, a four-step multiresolution method has been proposed by Castruccio and Genton (2016), in which the first step is to model the time series of wind speed at each location individually; see Castruccio and Genton (2018) for the general principles for analyzing big spatiotemporal data from climate models. For the four-step multiresolution method, our TGH-AR time series models can be used as a modification of the first step when time series data show non-Gaussian features.

To illustrate the usefulness of our TGH-AR-t model, we consider a publicly available Large Ensemble Project data set that consists of 30 ensembles of daily wind speed over the globe with a spatial resolution of around 1° longitude and 1° latitude from the year 1920 to 2100 (Kay et al., 2015). We select one ensemble from the complete data set and use the historical 86 years (1920–2005; $n = 31,390$) at 22×22 gridded locations over Saudi Arabia (bounded roughly by $13 - 34$°N and $32 - 59$°E). For each day in a year for each location, we estimate the seasonality by taking the average of the wind speed across 86 years at that location. For each location, we remove the seasonal effect from each time point and analyze the residual wind speed data, which show a clear AR pattern by the ACF and the partial ACF (PACF). We use the TGH-AR-t model because the residual wind speed time series clearly show features of skewness and heavy tails, and a plot of Y_t versus Y_{t-1} for the residual wind speed shows a nonlinear relationship. We get parameter estimation with order selection by fitting the TGH-AR-t model at each location. With the TGH-AR-t model and the estimated parameters at each location, wind speed data can be generated rapidly without using the physical model. The estimated parameters also give insights into the pattern of the distributional properties for the wind speed residuals.

Figure 8 shows maps of the estimated values of the four parameters related to the TGH transformation and two autoregressive coefficients. Because the autoregressive order selected is 2 for the majority of the locations, maps of estimations

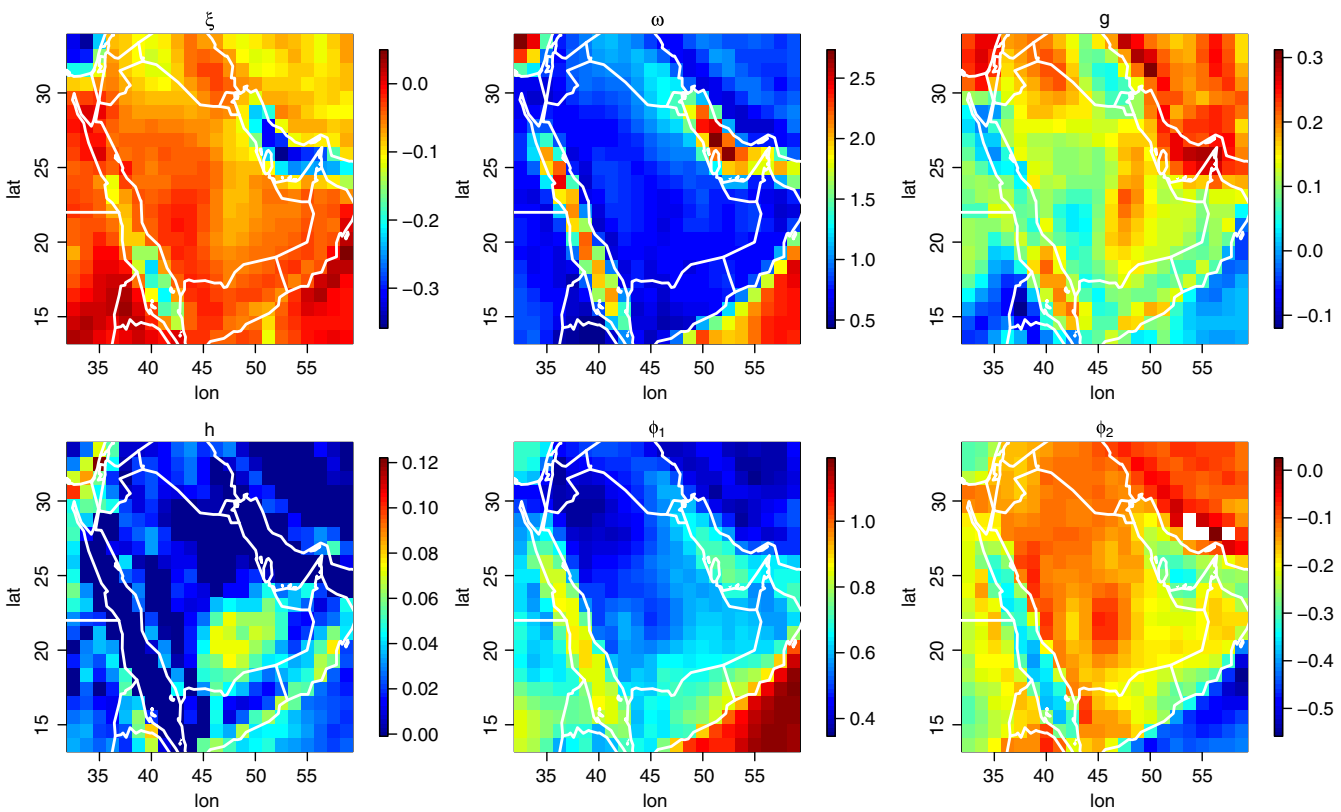

FIGURE 8 Maps of the estimated parameters from fitting the TGH-AR-t model with order selection to daily wind speed residuals

of higher order autoregressive coefficients are omitted. We may notice from these plots that $\hat{\xi}$, $\hat{\omega}$, $\hat{\phi}_1$, and $\hat{\phi}_2$ (white area indicates that the order selected is only 1) are distinctively different between land and ocean. We also observe that the $\hat{g}$ and $\hat{h}$ estimates show interesting patterns that are closely related to elevation and geographical features of a location. For the full three-step multiresolution analysis (without the altitudinal dependence) of monthly wind speed over the globe, where the TGH-AR-t model is used in the first step, see Jeong, Yan, Castruccio, and Genton (2017).

5.2 | Hourly wind speed data at a meteorological station

We consider hourly wind speed data observed at a meteorological tower in Sunnyside, Oregon, from December 1, 2013 to December 31, 2014 (Kazor & Hering, 2015). First, we use the hourly wind speed observed in June 2014 ($n = 720$) of this data set to demonstrate the suitability of using the TGH-AR-e model. We notice that there exists a diurnal pattern in the hourly wind speed, so we include harmonics with periods of 12 and 24 h as the covariates in the model. We find the MALE with order selection by fitting the wind speed time series from that month with the TGH-AR-e model. The estimated parameters are $\hat{\xi} = 7.84$, $\hat{\omega} = 1.65$, $\hat{g} = 0.11$, $\hat{h} = 0.06$, $\hat{\phi}_1 = 1.01$, $\hat{\phi}_2 = -0.15$, and $\hat{\boldsymbol{\beta}} = (-0.48, 2.31, -0.14, 0.19)^{\mathrm{T}}$ with $\boldsymbol{X}_t = \{\cos(2\pi t/12), \sin(2\pi t/12), \cos(2\pi t/24), \sin(2\pi t/24)\}^{\mathrm{T}}$.

Figure 9 shows the original hourly wind speed time series overlapped with a diurnal pattern (green line) estimated using the TGH-AR-e model. The histogram and normal Q–Q plot of the residual wind speeds after removing periodicity obviously deviate from Gaussianity. The skewness and kurtosis values presented with the histogram further support using our TGH-AR-e model to adapt to right-skewed and heavy-tailed data. The residual wind speed at time t is plotted against time $t-1$ in the first panel of the middle row in Figure 9, showing a strong linear relationship, which is the reason why we choose to transform the error term using the TGH-AR-e model rather than the process itself. The ACF and the PACF indicate an absence of seasonality after removing the diurnal pattern with the harmonics by the TGH-AR-e model. The ACF and the PACF also validate the appropriateness of using an AR(2) model, selected by BIC, for the wind speed

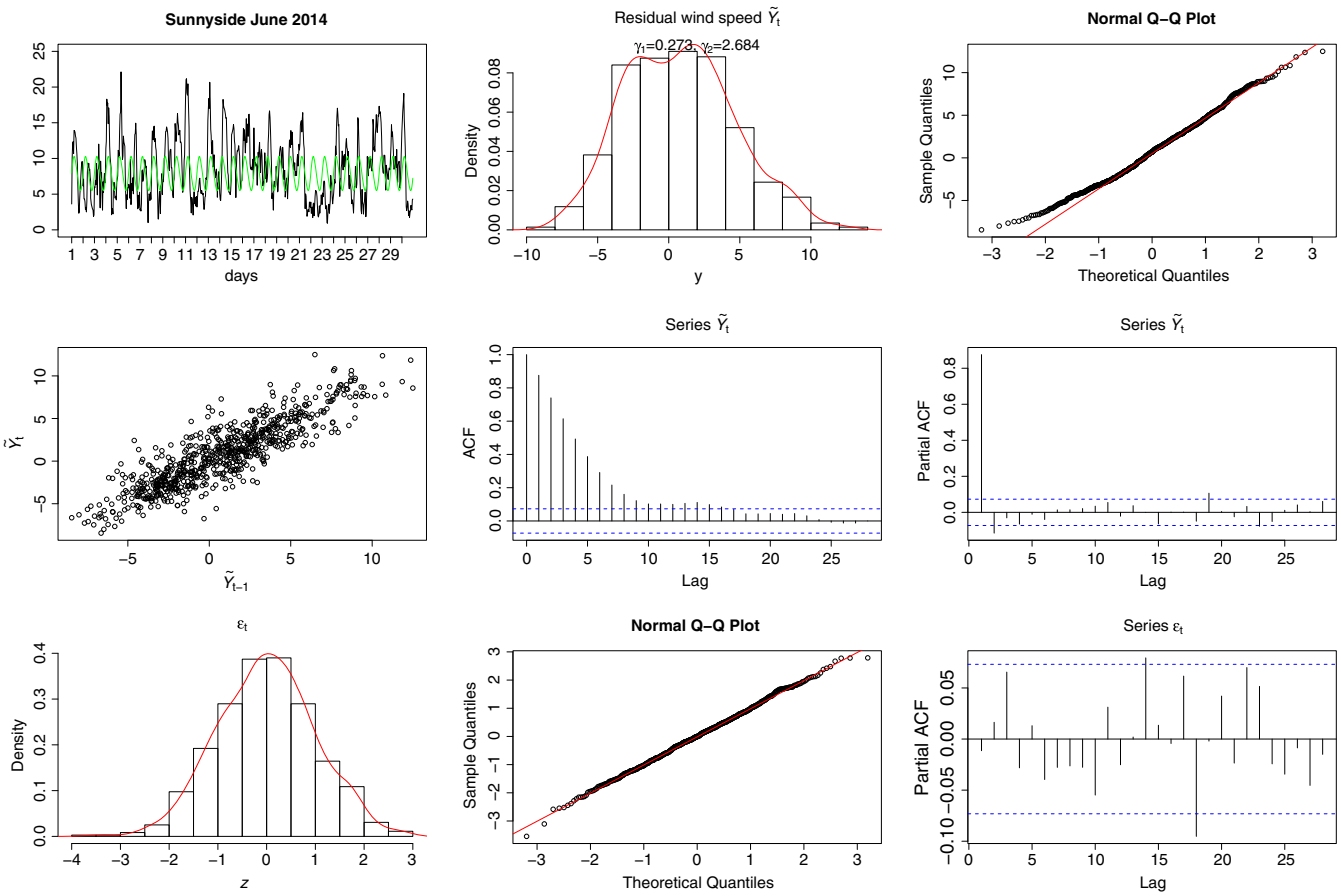

FIGURE 9 Various diagnostic plots of applying the TGH-AR-e model to the hourly wind speed data in June 2014. ACF = autocorrelation function; PACF = partial ACF

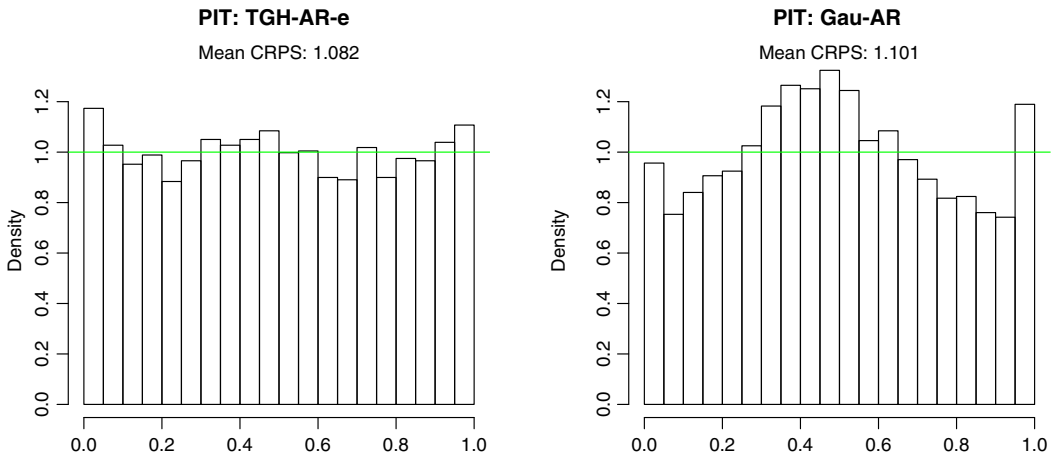

FIGURE 10 Histograms of the probability integral transform (PIT) values of probabilistic forecasts by the TGH-AR-e and Gaussian AR model for 1-h-ahead forecasts over the whole year of 2014. CRPS = continuous ranked probability score

residuals process. The bottom row of Figure 9 shows a histogram, a normal Q–Q plot, and the PACF for the estimated back-transformed Gaussian error term $\tilde{\epsilon}_{t,\boldsymbol{\theta}} = \tau^{-1}_{\hat{g},\hat{h}}\{\frac{\hat{\tilde{y}}_t - \hat{\phi}_1 \hat{\tilde{y}}_{t-1} - \hat{\phi}_2 \hat{\tilde{y}}_{t-2}}{\hat{\omega}}\}$, where $\hat{\tilde{y}}_t = y_t - \hat{\xi} - \boldsymbol{x}_t^{\mathrm{T}}\hat{\boldsymbol{\beta}}$. These plots confirm the validity of using a TGH-AR(2)-e model in which $\epsilon_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.

Next, we compare the forecast performance from our TGH-AR-e model with those of the Gaussian AR model for this data set. We make one-step-ahead forecasts for the whole year of 2014, with a rolling window length of 30 days ($n = 720$) for parameter estimation to account for the nonstationarity caused by seasonal effect. To estimate the parameters for the Gaussian AR model, we first remove the diurnal pattern using linear regression with the same harmonics as in the TGH-AR-e model. Then, we find the MLE of the autoregressive parameters from the residuals.

The MAE of the conditional median from the TGH-AR-e model is 4.05; the Gaussian AR forecast has an MAE of 4.28. The RMSEs for the conditional means of the TGH-AR-e model and the Gaussian AR model are 1.47 and 1.50, respectively. The empirical coverage of the 95% minimum-length CI is 94.2% for the TGH-AR-e model and 93.3% for the Gaussian model. We conclude that the point forecasts and CIs based on the TGH-AR-e model are better than those based on the Gaussian AR model for these hourly wind speed data. However, it is difficult to see the differences between the forecast results from only these numbers. Figure 10 shows histograms of the PIT values from forecasts based on the TGH-AR-e and Gaussian AR models. By comparing these histograms, we see evidently the superiority of fitting the wind speed using a non-Gaussian TGH error in the AR model rather than a Gaussian error.

6 | DISCUSSION

In this paper, we applied the TGH transformation in a time series context and built two flexible non-Gaussian autoregressive models. The TGH-AR-t model assumes a latent Gaussian process, which is transformed as whole, whereas the TGH-AR-e model transforms the white-noise error term in an autoregressive regression model. The intrinsic difference between the two models can be explained by different data-generating mechanisms. For a given time series, a plot such as that in Figure 3 can best help in deciding which model is more suitable. We described an efficient parameter estimation procedure for our two models that approximates the MLE and an order selection procedure based on BIC with the approximated likelihood. We derived formulas for point and probabilistic forecasts using the two models. We illustrated the empirical performances of estimation, order selection, and forecasting of our models through a simulation study. We found that by estimating all parameters at once by our estimation procedure, the performance drastically improved compared with sequential estimation that ignores the temporal dependence. Another finding was that the two models yielded different forecasts when calibrated on the same sample, hence proving that the two models are not interchangeable. Simulations also suggested a sample size no less than 250 for a satisfying performance of our models. Finally, we demonstrated the usefulness of our models by applying them to two wind speed data sets at different temporal resolutions. Our

TGH-AR models provided a fast simulation method that could emulate wind speed outcome of a gridded climate model and produced competitive forecast results for an observational hourly wind speed data set from a meteorological station.

The AR models considered in this paper cannot incorporate measurement errors. Extensions of the TGH-AR models to ARMA or state-space models need further research. Also, we feel that exhaustive comparisons of many existing transformations, including the Box–Cox, TGH, and sinh-arcsinh transformations (Jones & Pewsey, 2009), would be welcome. In a future study, the parameters g and h should be allowed to change smoothly across time, either by imposing a parametric function of t for g and h or by a penalty of smoothness, instead of the moving window scheme we used here. Extension of the TGH framework to a spatiotemporal setting is also promising.

ACKNOWLEDGEMENT

This publication is based upon the work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Grant OSR-2015-CRG4-2640.

REFERENCES

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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

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.

Castruccio, S., & Genton, M. G. (2016). Compressing an ensemble with statistical models: An algorithm for global 3D spatio-temporal temperature. *Technometrics*, *58*(3), 319–328.

Castruccio, S., & Genton, M. G. (2018). Principles for statistical inference on big spatio-temporal data from climate models. *Statistics & Probability Letters*, *136C*,92–96.

Cordeiro, G. M., & de Andrade, M. G. (2009). Transformed generalized linear models. *Journal of Statistical Planning and Inference*, *139*(9), 2970–2987.

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 Luca, G., Genton, M. G., & Loperfido, N. (2006). A multivariate skew-GARCH model. In *Econometric analysis of financial and economic time series (Advances in econometrics, Volume 20 Part 1)* (pp. 33–57). Bingley, UK: Emerald Publishing Limited.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Field, C., & Genton, M. G. (2006). The multivariate g-and-h distribution. *Technometrics*, *48*(1), 104–111.

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., Castruccio, S., Crippa, P., Dutta, S., Huser, R., Sun, Y., & Vettori, S. (2015). Visuanimation in statistics. *Stat*, *4*(1), 81–96.

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

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., & Aldrich, E. (2006). Calibrated probabilistic forecasting at the stateline wind energy center. *Journal of the American Statistical Association*, *101*(475), 968–979.

Goerg, G. M. (2011). Lambert, W random variables - a new family of generalized skewed distributions with applications to risk estimation. *The Annals of Applied Statistics*, *5*(3), 2197–2230.

Goerg, G. M. (2015). The Lambert way to Gaussianize heavy-tailed data with the inverse of Tukey's h transformation as a special case. *The Scientific World Journal*, *2015*, 1–16. https://doi:10.1155/2015/909231

Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *38*(1), 1–50.

He, Y., & Raghunathan, T. E. (2012). Multiple imputation using multivariate gh transformations. *Journal of Applied Statistics*, *39*(10), 2177–2198.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In *Data analysis for tables, trends, and shapes*. Hoboken, NJ: John Wiley & Sons.

Jeong, J., Yan, Y., Castruccio, S., & Genton, M. G. (2017). A stochastic generator of global monthly wind energy with Tukey g-and-h autoregressive processes. Retrieved from http://arxiv.org/abs/1711.03930.

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., & Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, *96*(4), 761–780.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., … Vertenstein, M. (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.

Kazor, K., & Hering, A. S. (2015). Assessing the performance of model-based clustering methods in multivariate time series with application to identifying regional wind regimes. *Journal of Agricultural, Biological, and Environmental Statistics*, *20*(2), 192–217.

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.

Martin, V. L. (1992). Threshold time series models as multimodal distribution jump processes. *Journal of Time Series Analysis*, *13*(1), 79–94.

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.

Nelson, H. L., Jr., & Granger, C. W. J. (1979). Experience with using the Box-Cox transformation when forecasting economic time series. *Journal of Econometrics*, *10*(1), 57–69.

Sun, Y., & Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, *20*(2), 316–334.

Tong, H. (1990). *Non-linear time series*. Oxford, UK: Oxford University Press.

Toyooka, Y. (1982). Prediction error in a linear model with estimated parameters. *Biometrika*, *69*(2), 453–459.

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

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. (2016). Tukey max-stable processes for spatial extremes. *Spatial Statistics*, *18*(Pt. B), 431–443.

Xu, G., & Genton, M. G. (2017). Tukey g-and-h random fields. *Journal of the American Statistical Association*, *112*(519), 1236–1249.

Yamamoto, T. (1976). Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *25*(2), 123–127.

Zimmerman, D. L., & Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics*, *44*(1), 27–43.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Yan Y, Genton MG. Non-Gaussian autoregressive processes with Tukey g-and-h transformations. *Environmetrics*. 2019;30:e2503. https://doi.org/10.1002/env.2503