

This is an Accepted Manuscript of an article published by Taylor & Francis in *Communications in Statistics: Case Studies, Data Analysis and Applications* on 06 Jan 2021, available online:

<https://doi.org/10.1080/23737484.2020.1869629>

© 2021 Taylor & Francis Group, LLC. This manuscript version is made available under the CC BY-NC 4.0 license <http://creativecommons.org/licenses/by-nc/4.0/>

Functional time series forecasting of extreme values

Han Lin Shang*

Department of Actuarial Studies and Business Analytics
Macquarie University
Research School of Finance, Actuarial Studies and Statistics
Australian National University

Ruofan Xu

Department of Econometrics and Business Statistics
Monash University
Research School of Finance, Actuarial Studies and Statistics
Australian National University

December 20, 2020

Abstract

We consider forecasting functional time series of extreme values within a generalised extreme value distribution (GEV). The GEV distribution can be characterised using the three parameters (location, scale and shape). As a result, the forecasts of the GEV density can be accomplished by forecasting these three latent parameters. Depending on the underlying data structure, some of the three parameters can either be modelled as scalars or functions. We provide two forecasting algorithms to model and forecast these parameters. To assess the forecast uncertainty, we apply a sieve bootstrap method to construct pointwise and simultaneous prediction intervals of the forecasted extreme values. Illustrated by a daily maximum temperature dataset, we demonstrate the advantages of modelling these parameters as functions. Further, the finite-sample performance of our methods is quantified using several Monte-Carlo simulated data under a range of scenarios.

Keywords: Generalised extreme value distribution; Generalised additive extreme value model; Dimension reduction; Maximum daily temperature.

*Postal address: Department of Actuarial Studies and Business Analytics, Level 7, 4 Eastern Road, Macquarie University, Sydney, NSW 2109, Australia; Telephone: +61(2) 9850 4689; Email: hanlin.shang@mq.edu.au; ORCID: <https://orcid.org/0000-0003-1769-6430>

1 Introduction

For planning government infrastructure, forecasting extreme values is extremely useful in many scientific disciplines, including but not limited to, Economics (e.g., [Calabrese & Osmetti 2013](#)), Hydrology (e.g., [Tyrallis et al. 2019](#)) and Meteorology (e.g., [de Oliveira et al. 2011](#)). Motivated by the need to forecast the extreme temperatures, we propose a functional time series method to forecast the GEV density with an application to Sydney daily maximum temperatures from 1900 to 2019.

Monitoring the temperature extremes is of critical importance in the study of climate impact, as the extreme weather events generally have large negative impacts on both social and environmental systems ([Field et al. 2012](#)). Temperature extremes are changing over time. Over the last two decades, numerous studies have identified changes in temperature extremes both in observations (e.g., [Naveau et al. 2014](#)) and in general circulation model simulations of future climate (e.g., [Kharin et al. 2013](#)). Therefore, it is crucial not only to identify but also forecast the temperature extremes.

The definition of “extreme” may vary in the literature across multiple disciplines. However, from a statistical point of view, one common definition is based on occurrences in the far tail of the distribution of the quantity of interest. In statistical literature, extreme value theory provides a mathematical framework for studying these far tails ([Fisher & Tippett 1928](#)). One common approach making use of the extreme value theory is based on the “block extremes”, the maxima (minima) of some climate variable over given blocks of time. Under certain regularity conditions, the magnitudes of extremes over sufficiently long blocks approximately follow a GEV distribution.

The GEV distribution is a well-developed distribution in statistical literature (e.g., see [De Haan & Ferreira 2007](#), Chapter 1 for details). In general, the GEV distribution can be fully characterised using only three parameters: the location, scale and shape parameters. As a result, forecasting GEV can be reduced to the forecasting of these three latent parameters. However, the classical forecasting methods for multivariate time series are not directly applicable, since the scale parameter is constrained to be non-negative. In contrast, the location and shape parameters can take arbitrary real values. A natural way to deal with such a constraint is to perform an invertible transformation, for example, the Box-Cox transformation ([Box & Cox 1964](#)) and the Yeo-Johnson transformation ([Yeo & Johnson 2000](#)), that maps the shape parameter onto the real line first and then make the prediction before inverting back.

In the field of meteorology, much attention has been focused on the study of annual extremes under a GEV distribution with three scalar parameters, since blocks of one year are common and sensible (see, e.g., [Huang et al. 2016](#), [Cooley 2009](#)), and annual extremes are generally

weakly dependent and stationary. In contrast, little attention has been paid to the study of daily extremes, which is of great value as it gives a detailed guideline of how extremes change among four seasons in a year and across years. Because of the inherent seasonality, it is not reasonable to model the daily maximum temperature within a year using a GEV distribution with *constant* scalar parameters. To this end, a generalised additive extreme value (GAEV) model, which is a GEV model with parameters modelled with generalised additive models, is often considered (see, e.g., [Yee & Stephenson 2007](#), [Chavez-Demoulin & Davison 2005](#)). To capture the daily variation in the annual extreme temperature, time can be treated as a covariate. In the statistical literature, GEV with covariate effects has been studied using different approaches. For example, [Eastoe & Tawn \(2009\)](#) transformed the sample to remove the effects of covariates prior to extreme value analysis; [Jonathan et al. \(2013\)](#) considered a conditional extremes model to incorporate covariate effects in all of the threshold selection, marginal and dependence modelling. In this paper, we follow [Davison & Smith \(1990\)](#) and parameterise extreme value model parameters in terms of the covariate time, so that the temporal dependency within the year can be preserved. It is worth noting that more sophisticated techniques have been developed for extreme value analysis of time series. [Chavez-Demoulin & Davison \(2012\)](#) provided a detailed review of the recent development in the time series's extremes. One of the main stream is to "decluster" the time series into approximately independent blocks given covariates. A basis reference for this approach is [Ferro & Segers \(2003\)](#), which proposed an automatic declustering scheme. Another approach is to use the physics of the problem to create a hierarchical model involving extreme value analysis of the "peaks" of whole heatwaves only, and then different models to characterise the evolution of a heatwave relative to its peak. This Bayesian modelling approach is often used in an oceanographic setting to model the evolution of severe ocean storms (see, e.g., [Randell et al. 2016](#), [Tendijck et al. 2019](#)).

We focus on the one-year-ahead forecast of daily maximum temperatures. We aim to model and forecast the cumulative distribution function (CDF) of extreme values and produce point and interval forecasts of extreme values associated with certain quantiles of the CDF. One natural way is to construct a sliced functional time series by breaking the long daily temperature data set into annual temperature curves with daily temperatures being the finite realisations on each curve. That is, the observed data are of the form $\{\mathcal{X}_t(\tau_j), t = 1, \dots, T, j = 1, \dots, 366\}$, where $\mathcal{X}_t(\tau_j)$ is the maximum temperature on the j^{th} day in the t^{th} year. Since the annual data contains 365 observations in common years and 366 observations in leap years, we treat the observation on the missing day, February 29, in common years as a missing value and use the average of the adjacent observations as its estimated value. Then, the yearly objects are observed on a common grid of 366 days. By treating the univariate time series as a functional

time series, the one-year-ahead daily maximum temperatures can be predicted as a whole. For each year, a GAEV model can be fitted to the daily maximum temperature curve. With the estimated parameters in the GAEV model, a time series forecasting algorithm is applied to forecast the future parameters in the GAEV model.

The rest of the paper is organised as follows. In Section 2, We first introduce the background of the GEV distribution and then provide a forecasting algorithm for the GEV distribution with scalar parameters. This scalar GEV method will be considered as one of the benchmarks and later compared with the GAEV model using a simulation study and an empirical data analysis. In Section 3, we introduce our forecasting method for the GAEV model and compare the forecasting method with a number of existing methods using a daily maximum temperature data set in Section 4. To measure forecast uncertainty, we apply a sieve bootstrap method of Paparoditis & Shang (2020) to construct pointwise and simultaneous prediction intervals of the forecasted extreme values. Further, we assess forecast accuracy using the simulated data sets in Section 5. Finally, we summarise the research findings, along with some ideas on how the methodology can be further extended in Section 6.

2 Forecasting extremes using GEV with scalar parameters

We consider a functional time series $\{\mathcal{X}_t(\tau), t = 1, 2, \dots, T, \tau \in \mathcal{I}\}$, where T denotes the sample size, $\mathcal{I} \subset \mathbb{R}$ is a compact interval on the real line, and the observed data is $\mathcal{X}_t(\tau_j)$ for $j = 1, 2, \dots, J$, where J denotes the number of discrete data points in a curve. In the maximum temperature data, $\mathcal{I} = [1, 366]$ and $J = 366$.

In this section, we assume that $\mathcal{X}_t(\cdot)$ follows a GEV distribution with time-specific scalar parameters, i.e. $\mathcal{X}_t(\cdot) \sim \text{GEV}(\mu_t, \sigma_t, \xi_t)$. We consider the simplest case that (μ_t, σ_t, ξ_t) are scalar parameters with respect to τ . In Section 3, these latent parameters are considered to be generalised additive models of a functional variable τ . When the GEV parameters are function-valued, each GEV parameter function can be well approximated by cubic regression splines. These spline basis functions are able to model temporal dependence exhibited in each GEV parameter.

As the GEV densities are not directly observable, we first estimate the time-varying parameters in the GEV densities and then model these estimated GEV parameters $\{(\mu_t, \sigma_t, \xi_t), t = 1, \dots, T\}$ to forecast the future GEV density. In Section 2.1, we present a brief overview of the GEV distribution and the corresponding parameter estimation method. Then, we introduce our time series forecasting method for the time-varying GEV scalar parameters in Section 2.2. In Section 2.3, a complete forecasting algorithm is summarised.

2.1 The GEV distribution and its parameter estimation

The GEV distribution arises in many cases of natural data. When considering the daily maximum temperature extremes within a year, let $(Y_{j,1}, \dots, Y_{j,K})$ be the temperatures recorded densely within day j , so that $M_j = \max(Y_{j,1}, \dots, Y_{j,K})$ is the daily maximum temperature. The extreme value theorem states that if random variables $(Y_{j,1}, \dots, Y_{j,K})$ are independent and identically distributed (i.i.d.), with “block length” K sufficiently large, the maxima M_j converge to a GEV distribution as $j \rightarrow \infty$ (Fisher & Tippett 1928, Coles et al. 2001). Moreover, there are theoretical justification that the i.i.d condition on $(Y_{j,k})_{k=1}^K$ can be relaxed to weakly dependent stationary time series (see, e.g., Hsing 1991, Einmahl et al. 2016), which is also the case we consider here. Note that M_j corresponds to one daily observation $\mathcal{X}_t(\tau_j)$ on the sliced functional time series $\mathcal{X}_t(\cdot)$ in day j at year t .

According to Coles et al. (2001), the probability density function of the GEV distribution is defined as

$$f(x|\mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} (1 + \xi \frac{x-\mu}{\sigma})^{-1/\xi-1} \exp \left[-(1 + \xi \frac{x-\mu}{\sigma})^{-1/\xi} \right], & \xi \neq 0, \\ \frac{1}{\sigma} \exp(-\frac{x-\mu}{\sigma}) \exp \left[-\exp(-\frac{x-\mu}{\sigma}) \right], & \xi = 0, \end{cases} \quad (1)$$

where $u_+ = \max(0, u)$, $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. When $\xi > 0$, the GEV distribution has a finite upper tail for the shape parameter. In contrast, there is no upper bound when $\xi \leq 0$.

Because of its relation with quantile function, we choose to work with the CDF. The CDF for a random variable $X \sim \text{GEV}(\mu, \sigma, \xi)$ is given by

$$F(x|\mu, \sigma, \xi) = \begin{cases} \exp \left[-(1 + \xi \frac{x-\mu}{\sigma})_+^{-1/\xi} \right], & \xi \neq 0, \\ \exp \left[-\exp \left(-\frac{x-\mu}{\sigma} \right) \right], & \xi = 0. \end{cases}$$

Since the CDF is invertible, the quantile function for the GEV distribution has an explicit expression, namely, for any probability $p \in [0, 1]$, the quantile is given by

$$Q(p|\mu, \sigma, \xi) = \begin{cases} \mu + \frac{\sigma [(-\ln(p))^{-\xi} - 1]}{\xi}, & \xi > 0, p \in [0, 1]; \xi < 0, p \in (0, 1], \\ \mu - \sigma \ln[-\ln(p)], & \xi = 0, p \in (0, 1). \end{cases}$$

In the statistical literature, several methods have been used to estimate the parameters of the GEV distribution, for example, the method of moments (e.g., Christopheit 1994), the Bayesian method (e.g., Coles & Tawn 2005) and the maximum likelihood method (e.g., Smith & Naylor 1987). Among them, the maximum likelihood method is the most popular one, because it allows

additional components to the fitting of covariates, such as trends and cycles (Katz et al. 2002). Therefore, we use the maximum likelihood method to estimate the parameters of the GEV distribution as follows.

If $M_1, \dots, M_J \stackrel{\text{i.i.d}}{\sim} \text{GEV}(\mu, \sigma, \xi)$ with $\xi \neq 0$, the likelihood function \mathcal{L} is given by

$$\begin{aligned} \mathcal{L}(\mu, \sigma, \xi; M) &= \prod_{i=1}^J \frac{1}{\sigma} \left[1 + \xi \frac{m_i - \mu}{\sigma} \right]^{-\frac{1}{\xi} - 1} \exp \left\{ - \left[1 + \xi \frac{m_i - \mu}{\sigma} \right]^{-\frac{1}{\xi}} \right\} \\ &= \sigma^{-j} \exp \left\{ \sum_{i=1}^J - \left[1 + \xi \frac{m_i - \mu}{\sigma} \right]^{-\frac{1}{\xi}} \right\} \prod_{i=1}^J \left[1 + \xi \frac{m_i - \mu}{\sigma} \right]^{-\frac{1}{\xi} - 1}, \end{aligned}$$

By taking the natural-log transformation, the log-likelihood is

$$l(\mu, \sigma, \xi; M) = -J \ln \sigma - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^J \ln \left(1 + \xi \frac{m_i - \mu}{\sigma} \right) - \sum_{i=1}^J \left[1 + \xi \frac{m_i - \mu}{\sigma} \right]^{-\frac{1}{\xi}}. \quad (2)$$

The maximum likelihood estimation (MLE) is then obtained by maximising (2) under the constraints that $1 + \xi \frac{x - \mu}{\sigma} > 0$ and $\sigma > 0$. There is no analytic solution to this optimisation problem no matter whether the constraints are met. However, the numerical solution $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ can be obtained using the `fgev` function of the `evd` package (Stephenson 2002).

2.2 Forecasting approach for the time-varying GEV parameters

Since the daily maximum follows a GEV distribution within each year, the GEV parameters (mean: μ_t , scale: σ_t , shape: ξ_t) are estimated for each year t independently, and subscript t aims to distinguish the year. Furthermore, $\{(\mu_t, \sigma_t, \xi_t), t = 1, \dots, T\}$ forms a vector-valued time series with the constraint that σ_t is positive for all $t = 1, \dots, T$. Classic time series models, which have no constraint on the domain, can not be directly applied to this vector-valued time series. To this end, we first map σ_t onto the real line with an invertible transformation, such as the Box-Cox transformation. As the transformed time series may not be stationary, additional transformation, such as log-transformation, de-trending or differencing, may also be necessary to obtain an approximately stationary vector-valued time series θ_t . A stationary time series model can be fitted and then the h -step-ahead forecasts, $\hat{\theta}_{T+h}$, can be obtained accordingly. One simplest model in this case would be the vector autoregressive model, whose order can be selected via the corrected Akaike information criterion (AICC) (see, e.g., Hurvich & Tsai 1993). Finally, using the components of $\hat{\theta}_{T+h}$, we can compute the predicted parameters $(\hat{\mu}_{T+h}, \hat{\sigma}_{T+h}, \hat{\xi}_{T+h})$ through an inverse transformation of $\hat{\theta}_{T+h}$.

2.3 An algorithm for the GEV density prediction

The time series forecasting method for the GEV density can be summarised as follows.

- 1) For each $t = 1, 2, \dots, T$, maximise (2) under the constraints that $1 + \xi \frac{x-\mu}{\sigma} > 0$ and $\sigma > 0$ to obtain the MLEs $\hat{\mu}_t$, $\hat{\sigma}_t$ and $\hat{\xi}_t$.
- 2) Transform the estimated parameters $\{(\mu_t, \sigma_t, \xi_t)\}$ to obtain a stationary unconstrained vector time series $\{\boldsymbol{\theta}_t\}$.
- 3) Fit an appropriate time series model to $\{\boldsymbol{\theta}_t\}$.
- 4) Compute the h -step-ahead forecast $\hat{\boldsymbol{\theta}}_{T+h}$ for $h \geq 1$.
- 5) Using $\hat{\boldsymbol{\theta}}_{T+h}$ to compute the predicted parameters $(\hat{\mu}_{T+h}, \hat{\sigma}_{T+h}, \hat{\xi}_{T+h})$ via back-transformation.
- 6) Obtain the predicted GEV densities $\hat{f}_{T+h}(\tau) = f(u|\hat{\mu}_{T+h}, \hat{\sigma}_{T+h}, \hat{\xi}_{T+h})$ according to (1).

3 Forecasting extremes with a GAEV model

When the GEV parameters are modelled as scalars, it has great limitation and is often unrealistic, as observations on a curve are rarely i.i.d. Also, it is not reasonable to assume that all realisations on each curve follow the same GEV distribution. To capture the temporal dynamics in the GEV distribution along a curve, a GAEV model is considered. The GAEV is a GEV model with parameters characterised by generalised additive models (see, e.g., [Gilli et al. 2006](#), [Chavez-Demoulin & Davison 2005](#)). To be more specific,

$$\mathcal{X}_t(\tau) \sim \text{GEV}[\mu_t(\tau), \sigma_t(\tau), \xi_t(\tau)], \quad \forall t \in 1, \dots, T, \quad \tau \in \mathcal{I}, \quad (3)$$

where $\sigma_t(\tau)$ is a non-negative function since the scale parameter is required to be non-negative.

Let η_t^* denote any of the three functional parameters, namely $\eta_t^\mu(\tau) := \mu_t(\tau)$, $\eta_t^\sigma(\tau) := \ln[\sigma_t(\tau)]$ and $\eta_t^\xi(\tau) := \xi_t(\tau)$, the generalised additive model for η_t^* can be represented using a basis expansion

$$\eta_t^*(\tau) = \beta_{t,0}^* + \sum_{i=1}^{d^*} \beta_{t,i}^* b_i^*(\tau), \quad (4)$$

where d^* is a pre-determined positive integer, $\{b_i^*(\tau), i = 1, \dots, d^*\}$ are the pre-determined basis functions and $\{\beta_{t,i}^*, i = 0, \dots, d^*\}$ are the time-varying coefficients. In this paper, we use the cubic regression spline basis as the basis functions.

For the GAEV model in (3), the GEV parameters of the generalised additive model (GAM) form are estimated by maximising a penalised likelihood. Such estimation has been implemented in the R package `evgam` (Youngman 2020).

As the GEV functional parameters can be fully characterised by the basis coefficients $\{\beta_{t,i}^*, i = 0, \dots, d^*\}$, the forecast for the GAEV model can be performed through the prediction of those basis coefficients, which forms a vector time series $\beta_t = [\beta_{t,i}^\mu, \beta_{t,j}^\sigma, \beta_{t,l}^\xi, i = 0, \dots, d^\mu, j = 1, \dots, d^\sigma, k = 1, \dots, d^\xi]$. We fit a vector autoregressive (VAR) model to $\{\beta_t, t = 1, \dots, T\}$ and obtain the h -step-ahead time series prediction $\{\hat{\beta}_{T+h,i}^*, i = 0, \dots, d^*\}$, then the forecast $\hat{\eta}_{T+h}^*(\tau)$ can be computed via (4) with $\beta_{t,i}^*$ replaced by $\hat{\beta}_{T+h,i}^*$ and estimated basis functions. Through an inverse log-transformation, we obtain $\hat{\eta}_{T+h}^\sigma(\cdot)$. The h -step-ahead GAEV model prediction is obtained as $\hat{\mathcal{X}}_{T+h}(\tau) \sim \text{GEV}\{\hat{\eta}_{T+h}^\mu(\tau), \exp[\hat{\eta}_{T+h}^\sigma(\tau)], \hat{\eta}_{T+h}^\xi(\tau)\}$.

As the dimension d^* are unknown, in this paper, we apply the leave-one-out cross-validation technique to determine the dimension d^μ , d^σ and d^ξ jointly. We set the upper bound of d^* to be 10 (and the lower bound is 3 for cubic regression spline), then for all possible combinations of (d^μ, d^σ, d^ξ) , we fit the GAEV model using the first $T - 1$ data points, make one-step-ahead forecast $\hat{\mathcal{X}}_T(\tau)$. Since we only observe one single sample for each τ instead of a complete distribution, we evaluate $\hat{\mathcal{X}}_T(\tau)$ at 50% quantile for each τ , $\hat{\mathcal{X}}_T^{0.5}(\tau)$, and compute the Jensen-Shannon divergence (JSD) between the observed $\mathcal{X}_T(\tau)$ and $\hat{\mathcal{X}}_T^{0.5}(\tau)$ (see Eq. (5)). We choose the set of (d^μ, d^σ, d^ξ) resulting in the minimum JSD to be the appropriate dimension parameters to fit the complete data.

4 Application to daily maximum temperatures in Sydney

We consider the daily maximum temperature data recorded from a weather station in Sydney from 1900 to 2019. We assume that the daily maximum temperatures in a given year follow a GEV distribution, which is common when studying the temperature data (see also Stein 2017, Huang et al. 2016). We evaluate the one-year-ahead forecast of the daily maximum temperatures under the GEV model with scalar parameters in Section 2 and compare the forecast with the one obtained from the GAEV model in Section 3. As a benchmark, we also consider a GAEV model applied to the univariate time series record. This is a traditional method for forecasting extreme values in the univariate time series literature (see, e.g., Coles et al. 2001, McNeil & Frey 2000). We first introduce the motivating data set in Section 4.1, then the forecasting approaches in Section 4.2, followed by discussions in Section 4.3. In Section 4.4, we provide a sieve bootstrap 95% pointwise and simultaneous prediction intervals for one-year-ahead daily maximum temperature extremes at the 99.9% quantile.

4.1 Data set

The data set was obtained from the [Australian Bureau of Meteorology \(2020\)](#). The raw data contains the daily maximum temperature from January 1, 1900, to December 31, 2019, collected from the weather station (station number 66062) in the south of Sydney. This data set can be obtained upon request from the corresponding author. Since the temperatures show apparent annual cycle at the location, the univariate daily temperature records are split into 120 yearly records. That is, when considering the maximum temperature extremes, the random variables $\{\mathcal{X}_t(\tau_j), t = 1, \dots, T, j = 1, \dots, J\}$ are the j^{th} daily maximum temperature in year t with $T = 120$ and $J = 366$.

Figure 1 presents a univariate time-series plot for the raw daily temperature from 1900 to 2019 and the rainbow plot of the sliced daily maximum temperature curves. The long time series record on the left panel suggests that the time series is roughly stationary. On the right panel, the seasonal difference in the maximum temperature can be easily spotted along each curve, which implies that a GAVEM may be more appropriate compared to the GEV with scalar parameters.

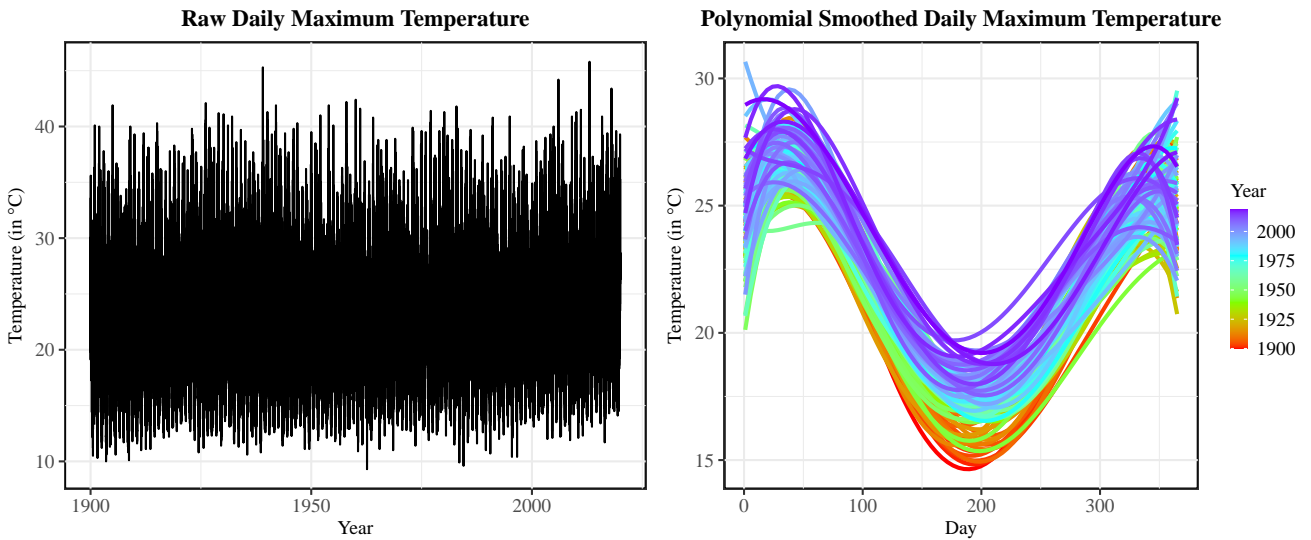


Figure 1 The raw and smoothed daily maximum temperatures from 1900 to 2019

4.2 Forecasting methods

In the empirical data analysis, comparisons are made based on the following three GEV models.

4.2.1 A GEV model with scalar parameters on functional time series (fGEV)

We first consider a GEV model with scalar parameters, which ignores the seasonal effect in the daily temperature. The forecasting algorithm is presented in Section 2.3. We make a few

specifications regarding steps 2 to 4 in the algorithm: log-transformation will be performed on $\{\widehat{\sigma}_t, t = 1, \dots, T\}$ to remove the positive constraint on the scalar parameter. Then one-step-ahead forecasts for the transformed parameters are performed using the VAR model.

4.2.2 A GAEV model on univariate scalar time series (tsGAEVM)

To capture the mean shift and seasonality in the GEV distribution, we also consider a naïve GAEV model performed on the original daily maximum temperature from January 1, 1990, to December 31, 2019. With such a long univariate time series, the upward trend may dominant the model, consequently resulting in a poor forecast result. Hence, in Section 4.3, we only use the latest one-year data, that is the daily temperature data in the year 2019, to fit the GAEV model. That is $X_t \sim \text{GEV}(\mu(t), \sigma(t), \xi)$ for $t = 1, \dots, 366$. One-to 366-step-ahead forecasts are performed to obtain the daily maximum temperature forecast in the year 2020. The model fitting and prediction are implemented using the `evgam` package (Youngman 2020) in R (R Core Team 2020).

4.2.3 A GAEV model on functional time series (fGAEVM)

We consider our recommended methodology, GAEV model on functional time series. We assume that $\mathcal{X}_t(\tau) \sim \text{GEV}[\mu_t(\tau), \sigma_t(\tau), \xi_t]$. Similar to the tsGAEVM mentioned above, we assume that the location and scale parameter follows a GAM, while the shape parameter is a scalar. The GAM for the location and scale parameter aim to capture the mean shift and scale change of the GEV distribution resulting from the seasonal effect. In contrast, the shape parameter is only fitted with intercept for the following two reasons: Firstly, data generally provide little information on the shape parameter. Secondly, estimation is numerically fraught when parameters are allowed to be too flexible (Yee & Stephenson 2007).

Having estimated $[\widehat{\mu}_t(\tau), \widehat{\sigma}_t(\tau), \widehat{\xi}_t]$, the one-step-ahead forecasts for the three parameters are obtained jointly by fitting a VAR model to $[\widehat{\mu}_t(\tau), \ln(\widehat{\sigma}_t(\tau)), \widehat{\xi}_t]$ for $t = 1, \dots, T$; the order of the VAR model is determined by AICC.

4.3 Results and Discussion

As the extreme temperatures are of particular interest in meteorology, Figure 2 presents the forecast maximum daily temperature curve at a 99.9% quantile of the GEV distribution. Plots from left to right correspond to the three forecasting models mentioned in section 4.2. The forecasts using the GEV with scalar parameters is a horizontal line, as all temperatures follow the same GEV, which contradicts to the real case. On the other hand, the forecasts using the

tsGAEVM and fGAEVM exhibit ‘U’ shape, which reflects the seasonal effect, and hence is more realistic compared to the GEV model. Compared with the tsGAEVM, the smooth forecast using our proposed fGAEVM is more convincing. Although we can not measure the goodness of fit, that wired rough shape in the tsGAEVM is hard to interpret.

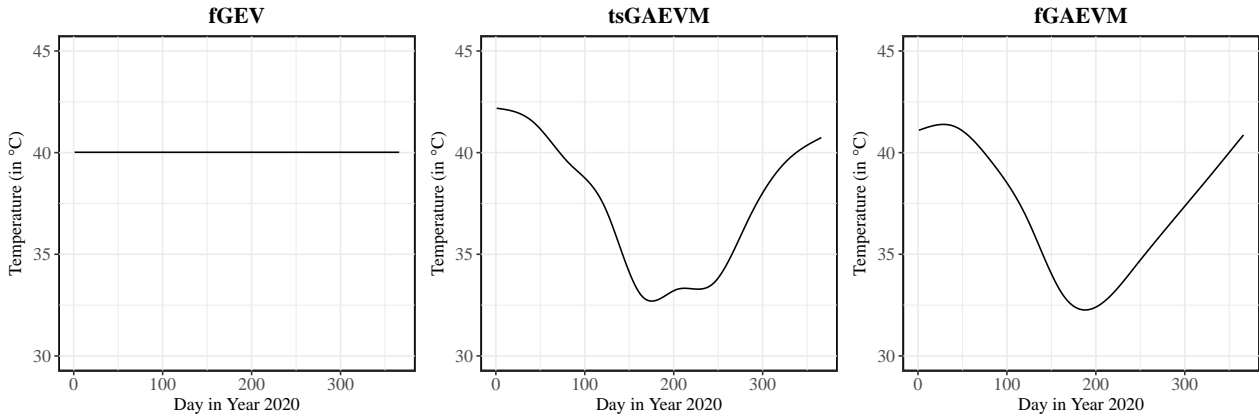


Figure 2 The forecast result for the extreme daily maximum temperatures in the Year 2020 at a 99.9% quantile

4.4 Bootstrap prediction interval

To quantify the forecast uncertainty associated with the point forecast, we construct 95% pointwise and simultaneous prediction intervals using a sieve bootstrap procedure of [Paparoditis & Shang \(2020\)](#). The advantage of the sieve bootstrap is its ability to take into account the misspecification error of a forecasting model. Through the sieve bootstrap, we construct $B = 1,000$ one-step-ahead bootstrap forecasts, from which we construct the pointwise and simultaneous prediction intervals for the 99.9% quantile. For the 99.9% quantile, the 95% pointwise prediction interval and 95% simultaneous prediction band are displayed in [Figure 3](#).

5 Simulation study

The forecast accuracy can not be measured in the real data analysis, since the actual GEV distribution is unknown. To investigate the finite-sample performance, in this section, we utilise a Monte-Carlo simulation study to compare the forecast accuracy for the three GEV models used in the empirical data analysis (see [Section 4.2](#)).

5.1 Data generating processes

We generate the functional time series $\{\mathcal{X}_t(\tau) \sim \text{GEV}(\mu_t(\tau), \sigma_t(\tau), \xi_t(\tau)), t = 1, \dots, T, \tau \in [0, 1]\}$ that follow a GAEV model in [\(3\)](#) and [\(4\)](#) and consider three settings regarding the GEV

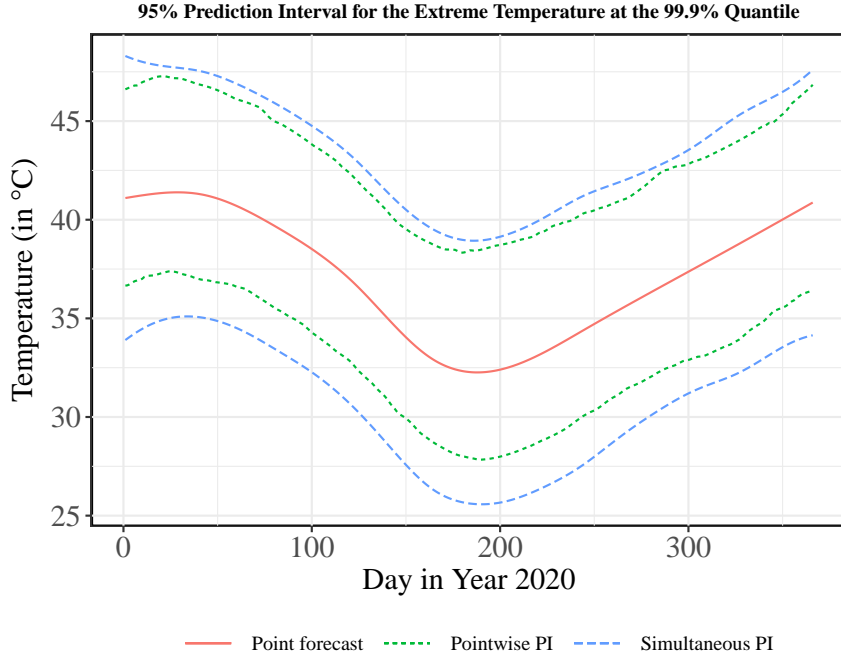


Figure 3 The 95% pointwise prediction interval and 95% simultaneous prediction band for the extreme daily maximum temperatures in 2020 at the 99.9% quantile

parameters as follows.

Setting 1: All parameters are scalars

We consider the first setting that all parameters are scalars, that is,

$$\mu_t(\tau) = \mu_t, \sigma_t(\tau) = \sigma_t, \text{ and } \xi_t(\tau) = \xi_t,$$

where all parameters μ_t , $\ln(\sigma_t)$ and ξ_t follow an autoregressive (AR) model of order 1 (AR(1)) independently. The AR coefficients are set randomly between -1 and 1, such that all GEV parameters are stationary with respect to t .

Setting 2: Location parameter and scale parameter are functional

In the second setup, we consider the location and shape parameters being function-valued, while the shape parameter is a scalar, i.e.,

$$\begin{aligned} \mu_t(\tau) &= \beta_{t,0} + \sum_{i=1}^{d_1} \beta_{t,i}^{\mu} b_i(\tau), \\ \ln(\sigma_t(\tau)) &= \sigma_{t,0} + \sum_{i=1}^{d_2} \beta_{t,i}^{\sigma} b_i(\tau), \\ \xi_t(\tau) &= \xi_t, \end{aligned}$$

where $b_i(\cdot)$ is a cubic regression spline basis.

For the location and scale parameters, all the coefficients $\beta_{t,i}^\mu$ and $\beta_{t,i}^\sigma$ are generated from AR(1) models with respect to time t independently, and we consider $d_1 = d_2 = 5$. All the AR coefficients are set randomly between -1 and 1, such that the basis coefficients are stationary with respect to t . For the scale parameter, we generate ξ_t from a stationary AR(1).

Setting 3: All parameters are functional

We also consider the most general setting that all GEV parameters are function-valued, that is

$$\begin{aligned}\mu_t(\tau) &= \beta_{t,0}^\mu + \sum_{i=1}^{d_1} \beta_{t,i}^\mu b_i(\tau), \\ \ln(\sigma_t(\tau)) &= \beta_{t,0}^\sigma + \sum_{i=1}^{d_2} \beta_{t,i}^\sigma b_i(\tau), \\ \xi_t(\tau) &= \beta_{t,0}^\xi + \sum_{i=1}^{d_3} \beta_{t,i}^\xi b_i(\tau),\end{aligned}$$

where $b_i(\cdot)$ is a cubic regression spline basis. Similar to setting 2, all the basis coefficients are generated from stationary AR(1) models independently, and $d_1 = d_2 = d_3 = 5$.

Once we obtained all the sampled GEV coefficients $(\mu_t(\tau_j), \sigma_t(\tau_j), \xi_t(\tau_j))$ for each grid point τ_j on the curve, where $\tau_j \in [0, 1], j = 1, \dots, 30$, we draw a random sample $\mathcal{X}_t(\tau_j)$ from $\text{GEV}(\mu_t(\tau_j), \sigma_t(\tau_j), \xi_t(\tau_j))$. In this simulation study, we consider $T = 50$.

5.2 Density error criteria

We measure the discrepancy between the forecast GEV density and the actual future GEV density by considering the discrete version of the JSD (Shannon 1948) and Kullback–Leibler divergence (KLD) (Kullback & Leibler 1951).

The JSD measures the loss of information when we choose an approximation. For the actual and predicted probability density functions, denoted by $f_j(\cdot)$ and $\hat{f}_j(\cdot)$, the discrete version of the JSD is given by

$$\begin{aligned}\text{JSD}_j &= \frac{1}{2} D_{\text{KL}}(f_j | \delta_j) + \frac{1}{2} D_{\text{KL}}(\hat{f}_j | \delta_j) \\ &= \frac{1}{2} \sum_{i=1}^K f_j(v_i) [\ln(f_j(v_i)) - \ln(\delta_j(v_i))] + \frac{1}{2} \sum_{i=1}^K \hat{f}_j(v_i) [\ln(\hat{f}_j(v_i)) - \ln(\delta_j(v_i))],\end{aligned}\tag{5}$$

where $\{v_i, i = 1, \dots, K\}$ are the finite realisations on the density function and $\delta(\cdot)$ measures a common quantity between $f_j(\cdot)$ and $\hat{f}_j(\cdot)$. We consider the simple mean given by $\delta_j(v) = f_j(v) + \hat{f}_j(v)$. The JSD is locally proportional to the Fisher information metric, and is similar

to the Hellinger metric, in the sense that it induces the same affine connection on a statistical manifold, and is equal to half the so-called Jeffreys divergence (Fuglede & Topsoe 2004).

Alternatively, the discrepancy can be measured by the KLD, give by

$$\text{KLD}_j = D_{\text{KL}}(f_j|\hat{f}_j) + D_{\text{KL}}(\hat{f}_j|f_j), \quad (6)$$

which is symmetric and non-negative.

Under the GAEV model, for each year i in the forecasting period, denoted by $\hat{\mathcal{X}}_i(\cdot)$ there is a unique GEV distribution for each τ_j . Hence, we measure the $\text{JSD}_j^{(i)}$ and $\text{KLD}_j^{(i)}$ for each realisation $j = 1, \dots, J$ on the curve and then take the average as the divergence measure for a testing sample i in the forecasting period, that is

$$\begin{aligned} \text{JSD}^{(i)} &= \frac{1}{J} \sum_{j=1}^J \text{JSD}_j^{(i)}, \\ \text{KLD}^{(i)} &= \frac{1}{J} \sum_{j=1}^J \text{KLD}_j^{(i)}. \end{aligned}$$

5.3 Results and Discussion

For each setting in the data generating process, we draw 250 Monte-Carlo samples to compare the forecasting performance. Within each iteration, we use an expanding window approach to measure the forecast accuracy. We use the last 20% functional data as the testing sample and the rest as the training sample. We then produce iterative one-step-ahead prediction based on the three methods (see Section 4.2 for details), where the training sample increases one data at a time. The error is measured as the mean divergence across all the test samples using the JSD and KLD. To be more specific, let N be the total number of the testing data within one iteration, and $\text{JSD}^{(i)}$ and $\text{KLD}^{(i)}$ be the JSD and KLD measure for the i^{th} testing data respectively. Then, the mean error for one iteration is

$$\begin{aligned} \text{JSD} &= \frac{1}{N} \sum_{i=1}^N \text{JSD}^{(i)}, \\ \text{KLD} &= \frac{1}{N} \sum_{i=1}^N \text{KLD}^{(i)}. \end{aligned}$$

The mean and standard deviation (in parentheses) of the averaged JSD and KLD across 250 repetitions are reported in Table 1. The corresponding boxplots are also presented in Figure 4.

Our method fGAEVM surpasses the other two methods when the data are generated from a true GAEV model (setting 2 and 3), that is a model with at least one function-valued parameter. Compared to the fGAEVM method, the fGEV fails to capture the true underlying dynamic

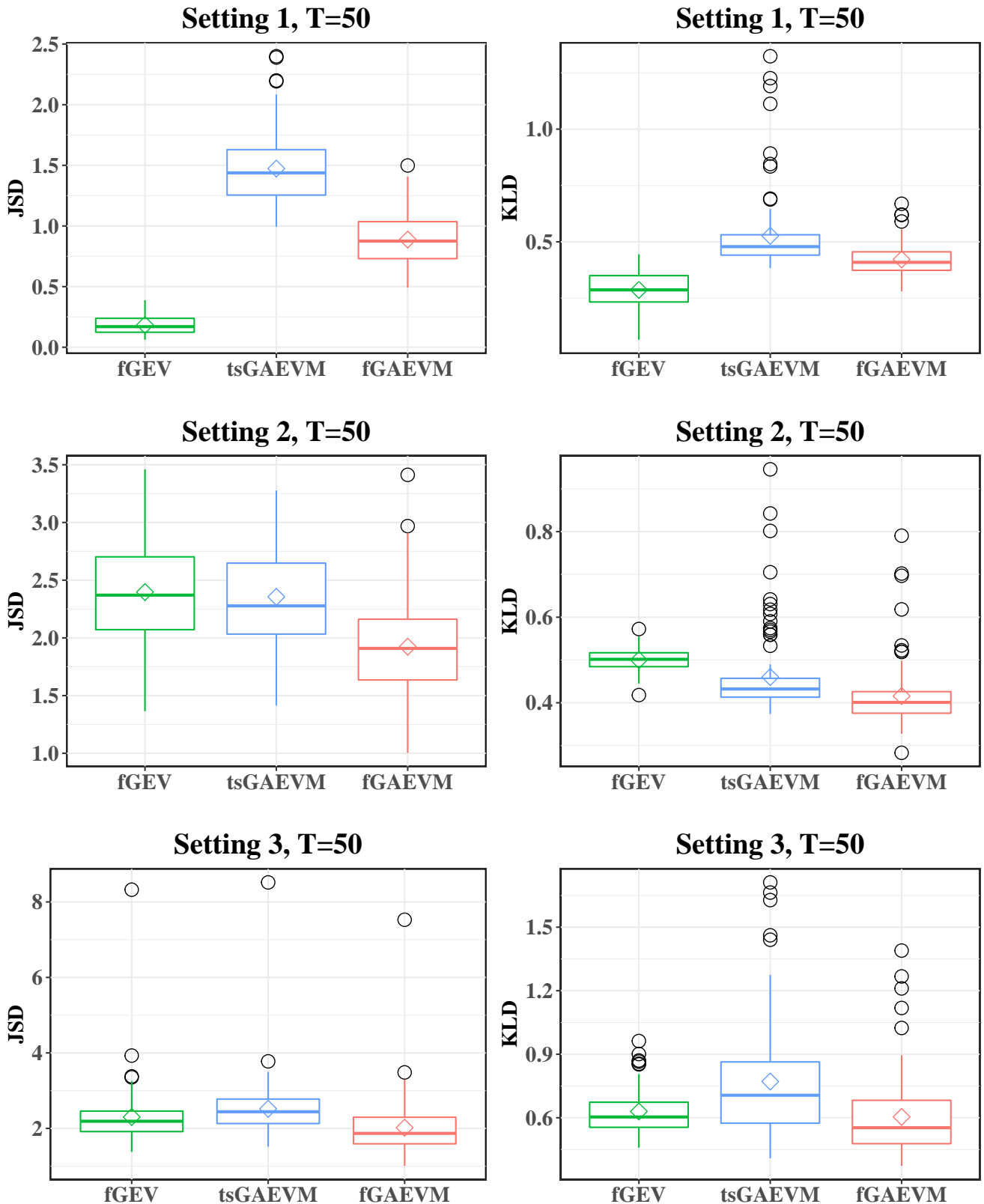


Figure 4 Boxplots for the point forecast accuracy under the fGEV, tsGAEVM and fGAEVM with the simulated 250 Monte-Carlo repetitions under Setting 1, 2, and 3 for $T = 50$ or 100

Table 1 Summary statistics of the averaged KLD and JSD based on the 250 simulated data

	fGEV		tsGAEVM		fGAEVM	
	JSD	KLD	JSD	KLD	JSD	KLD
Setting 1	0.18(0.08)	0.28(0.08)	1.47 (0.16)	0.82 (0.29)	0.88(0.21)	0.42(0.07)
Setting 2	2.39(0.47)	0.49(0.02)	2.35(0.41)	0.46(0.09)	1.92(0.41)	0.41(0.07)
Setting 3	2.30(0.77)	0.63(0.10)	2.51(0.76)	0.77(0.27)	2.02(0.78)	0.60(0.18)

structure in the parameters, while tsGAEVM treats the functional realisations as a univariate time series. Therefore, to obtain the one-year-ahead forecast for the functional object $\{\hat{\mathcal{X}}_{t+1}(\tau_j), j = 1, \dots, J\}$, using tsGAEVM requires the forecasts from one until J -step, the result of which generally becomes poorer as the forecast horizon increases.

However, when data follows a GEV distribution with scalar parameters (setting 1), the fGEV model gives the best forecasting results. The fGAEVM and tsGAEVM perform on par with the fGEV, subject to all the GAM basis coefficients for the parameters are set to zero, or the number of basis coefficients equals zero. In this case, d_1 and d_2 can not be appropriately determined using the leave-one-out cross-validation with the lower bound of the parameter setting to 3 (the least degree of freedom required for a cubic regression spline). Hence, the tsGAEVM and fGAEVM model both suffer from the overfitting problem when the data follows a GEV model.

From the forecasting results in all three scenarios, we may conclude that fGAEVM gives satisfactory results in most of the scenarios, even if the true underlying nonstationary data structure is different from the model.

Since the tsGAEVM always results in worse forecasts than the fGAEVM regardless of the underlying structure, modelling the data as a functional object is advantageous than a long univariate time series.

6 Conclusion

We proposed two general forecasting frameworks for functional time series of extreme values. One is a GEV model with scalar parameters (fGEV; see Section 2), and another one is a GAEV model (fGAEVM, see Section 3), which is a GEV model with parameters follow a GAM. As the tsGAEVM always results in worse forecasts than the fGAEVM regardless of the underlying structures (Section 5.3), the importance of treating the data as a sliced functional time series rather than a long univariate time series has been revealed. Illustrated by the maximum temperature data in Sydney in Section 4.3 and simulated data in Section 5.3, when the underlying data follows a GAEV model, the fGEV gives a senseless result of a horizontal line along with

the functional object. At the same time, the fGAEVM can capture the dynamics along with the functional object. On the other hand, if the data indeed follows a static GEV distribution, the large flexibility of the fGAEVM may reduce the forecast accuracy compared to the fGEV model. Since the static GEV distribution and the GAEV distribution can be distinguished based on prior knowledge of the data set or using data visualisations, we can choose between the fGEV and fGAEVM adaptively.

There are a few ways in which this paper could be extended, and we briefly discuss three. Firstly, We can model the temporal dependency along day-in-year in a more precise way. In this study, we consider daily maxima. Instead of block maxima, we may also consider identifying “independent” clusters and then fitting a GEV model to cluster maxima or fit the GEV under a Bayesian framework. Secondly, we use the cubic regression spline basis to represent the generalised additive model for the GEV parameters. In future research, other bases may be utilised, and comparison can be made. Finally, in the empirical analysis, the forecasting uncertainty is assessed using a sieve bootstrapped prediction interval method proposed for functional time series ([Paparoditis & Shang 2020](#)). Bootstrap method tailored for the fGAEVM density may be a promising future research topic.

References

- Australian Bureau of Meteorology (2020), *Climate data online*. Data retrieved from Australian Data Archive for Meteorology on May 4, 2020.
URL: <http://www.bom.gov.au/climate/data/>
- Box, G. E. & Cox, D. R. (1964), 'An analysis of transformations', *Journal of the Royal Statistical Society: Series B (Methodological)* **26**(2), 211–243.
- Calabrese, R. & Osmetti, S. A. (2013), 'Modelling small and medium enterprise loan defaults as rare events: the generalized extreme value regression model', *Journal of Applied Statistics* **40**(6), 1172–1188.
- Chavez-Demoulin, V. & Davison, A. (2012), 'Modelling time series extremes', *REVSTAT-Statistical Journal* **10**(ARTICLE), 109–133.
- Chavez-Demoulin, V. & Davison, A. C. (2005), 'Generalized additive modelling of sample extremes', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(1), 207–222.
- Christopeit, N. (1994), 'Estimating parameters of an extreme value distribution by the method of moments', *Journal of Statistical Planning and Inference* **41**(2), 173–186.
- Coles, S., Bawa, J., Trenner, L. & Dorazio, P. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, New York.
- Coles, S. & Tawn, J. (2005), 'Bayesian modelling of extreme surges on the UK east coast', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **363**(1831), 1387–1406.
- Cooley, D. (2009), 'Extreme value analysis and the study of climate change', *Climatic Change* **97**(1-2), 77.
- Davison, A. C. & Smith, R. L. (1990), 'Models for exceedances over high thresholds', *Journal of the Royal Statistical Society: Series B (Methodological)* **52**(3), 393–425.
- De Haan, L. & Ferreira, A. (2007), *Extreme Value Theory: An Introduction*, Springer Science & Business Media, New York.
- de Oliveira, M. M. F., Ebecken, N. F. F., de Oliveira, J. L. F. & Gilleland, E. (2011), 'Generalized extreme wind speed distributions in South America over the Atlantic Ocean region', *Theoretical and Applied Climatology* **104**(3-4), 377–385.

- Eastoe, E. F. & Tawn, J. A. (2009), 'Modelling non-stationary extremes with application to surface level ozone', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **58**(1), 25–45.
- Einmahl, J. H., de Haan, L. & Zhou, C. (2016), 'Statistics of heteroscedastic extremes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**(1), 31–51.
- Ferro, C. A. & Segers, J. (2003), 'Inference for clusters of extreme values', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**(2), 545–556.
- Field, C. B., Barros, V., Stocker, T. F. & Dahe, Q. (2012), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, New York.
- Fisher, R. A. & Tippett, L. H. C. (1928), 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Mathematical Proceedings of the Cambridge Philosophical Society* **24**(2), 180–190.
- Fuglede, B. & Topsoe, F. (2004), Jensen-Shannon divergence and Hilbert space embedding, in 'International Symposium on Information Theory, ISIT 2004. Proceedings.', IEEE, p. 31.
- Gilli, M. et al. (2006), 'An application of extreme value theory for measuring financial risk', *Computational Economics* **27**(2-3), 207–228.
- Hsing, T. (1991), 'On tail index estimation using dependent data', *The Annals of Statistics* **19**(3), 1547–1569.
- Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S. & Moyer, E. J. (2016), 'Estimating changes in temperature extremes from millennial scale climate simulations using generalized extreme value (GEV) distributions', *Advances in Statistical Climatology, Meteorology and Oceanography* **2**, 79–103.
- Hurvich, C. M. & Tsai, C.-L. (1993), 'A corrected Akaike information criterion for vector autoregressive model selection', *Journal of time series analysis* **14**(3), 271–279.
- Jonathan, P., Ewans, K. & Randell, D. (2013), 'Joint modelling of extreme ocean environments incorporating covariate effects', *Coastal Engineering* **79**, 22–31.
- Katz, R. W., Parlange, M. B. & Naveau, P. (2002), 'Statistics of extremes in hydrology', *Advances in Water Resources* **25**(8-12), 1287–1304.
- Kharin, V. V., Zwiers, F., Zhang, X. & Wehner, M. (2013), 'Changes in temperature and precipitation extremes in the CMIP5 ensemble', *Climatic Change* **119**(2), 345–357.

- Kullback, S. & Leibler, R. A. (1951), 'On information and sufficiency', *The Annals of Mathematical Statistics* **22**(1), 79–86.
- McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**(3-4), 271–300.
- Naveau, P., Guillou, A. & Rietsch, T. (2014), 'A non-parametric entropy-based approach to detect changes in climate extremes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**(5), 861–884.
- Paparoditis, E. & Shang, H. L. (2020), Bootstrap Prediction Bands for Functional Time Series, Working paper, University of Cyprus.
URL: <https://arxiv.org/abs/2004.03971>
- R Core Team (2020), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: <https://www.R-project.org/>
- Randell, D., Turnbull, K., Ewans, K. & Jonathan, P. (2016), 'Bayesian inference for nonstationary marginal extremes', *Environmetrics* **27**(7), 439–450.
- Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell System Technical Journal* **27**(3), 379–423.
- Smith, R. L. & Naylor, J. C. (1987), 'A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **36**(3), 358–369.
- Stein, M. (2017), 'Should annual maximum temperatures follow a generalized extreme value distribution?', *Biometrika* **104**(1), 1–16.
- Stephenson, A. G. (2002), 'evd: Extreme value distributions', *R News* **2**(2), 31.
- Tendijck, S., Ross, E., Randell, D. & Jonathan, P. (2019), 'A model for the directional evolution of severe ocean storms', *Environmetrics* **30**(1), e2541.
- Tyralis, H., Papacharalampous, G. & Tantane, S. (2019), 'How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset', *Journal of Hydrology* **574**, 628–645.
- Yee, T. W. & Stephenson, A. G. (2007), 'Vector generalized linear and additive extreme value models', *Extremes* **10**(1-2), 1–19.

Yeo, I.-K. & Johnson, R. A. (2000), 'A new family of power transformations to improve normality or symmetry', *Biometrika* **87**(4), 954–959.

Youngman, B. D. (2020), *evgam: An R package for generalized additive extreme value models*, Working paper, University of Exeter.

URL: <https://arxiv.org/abs/2003.04067>