

## Forecasting functional time series using weighted likelihood methodology

Ufuk Beyaztas & Han Lin Shang

To cite this article: Ufuk Beyaztas & Han Lin Shang (2019): Forecasting functional time series using weighted likelihood methodology, Journal of Statistical Computation and Simulation, DOI: [10.1080/00949655.2019.1650935](https://doi.org/10.1080/00949655.2019.1650935)

To link to this article: <https://doi.org/10.1080/00949655.2019.1650935>



Published online: 01 Aug 2019.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)



# Forecasting functional time series using weighted likelihood methodology

Ufuk Beyaztas <sup>a</sup> and Han Lin Shang <sup>b</sup>

<sup>a</sup>Department of Statistics, Bartin University, Bartin, Turkey; <sup>b</sup>Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, Australia

## ABSTRACT

Functional time series whose sample elements are recorded sequentially over time are frequently encountered with increasing technology. Recent studies have shown that analyzing and forecasting of functional time series can be performed easily using functional principal component analysis and existing univariate/multivariate time series models. However, the forecasting performance of such functional time series models may be affected by the presence of outlying observations which are very common in many scientific fields. Outliers may distort the functional time series model structure, and thus, the underlying model may produce high forecast errors. We introduce a robust forecasting technique based on weighted likelihood methodology to obtain point and interval forecasts in functional time series in the presence of outliers. The finite sample performance of the proposed method is illustrated by Monte Carlo simulations and four real-data examples. Numerical results reveal that the proposed method exhibits superior performance compared with the existing method(s).

## ARTICLE HISTORY

Received 5 March 2019  
Accepted 29 July 2019

## KEYWORDS

Bootstrap; functional principal components; functional time series; weighted likelihood

## 2010 MATHEMATICS SUBJECT CLASSIFICATIONS

62G35; 62M10; 62P12

## 1. Introduction

A functional time series arises when functional objects are collected sequentially over time. In other words, let  $\{y_t\}$  ( $t = 1, \dots, N$ ) denote a sequence of observed functional data; then, it is termed an functional time series if each functional element  $y_t(\tau)$  is defined on a bounded interval with a continuous variable  $\tau$ ;  $\tau \in [a, b]$ . Denote by  $\{y_t(x_j)\}$ , for  $t = 1, \dots, N$  and  $j = 1, \dots, J$ , the noisy functional time series data observed at time points  $\{x_1, \dots, x_J\}$ . It is assumed that the functional time series is characterized by a continuous smooth function  $f$  and an error process  $\varepsilon$  such that

$$y_t(x_j) = f_t(x_j) + \sigma_t(x_j)\varepsilon_{t,j} \quad (1)$$

where  $\varepsilon_{t,j}$  is an independently and identically distributed (i.i.d.) Gaussian random variable with zero mean and unit variance and  $\sigma_t(x_j)$  allowing for heteroskedasticity.

Forecasting unobservable future realizations of functional time series is of great interest. In practice, forecasts can be obtained in the form of point and/or interval forecasts. A point

forecast corresponds to an estimate (conditionally on the available data) of the unknown future realization of the underlying process. However, this approach may not produce reliable inferences for future observations since it does not provide any information about the degree of uncertainty associated with the point forecasts. By contrast, interval forecasts, such as the prediction interval, provide better inferences taking into account the uncertainty associated with point forecasts; see for example [1–3]. In the context of functional time series, Hyndman and Ullah [4] propose a functional data approach to obtain point and interval forecasts for age-specific mortality and fertility rates observed over time. While doing so, they consider that the functional data  $\{y_t(x_j)\}$  consist of random functions separated by consecutive and continuous time intervals. Their approach is as follows: (i) Approximate the smooth functions  $\{f_t(x)\}$  separately using a nonparametric smoothing technique on each function  $y_t(x)$  ( $t = 1, \dots, N$ ). (ii) Decompose the smoothed functional time series data into  $K$  orthonormal principal components  $\phi_k$  ( $k = 1, \dots, K$ ) and associated uncorrelated scores  $\beta_{k,t}$  ( $t = 1, \dots, N$ ) using a functional principal component model. (iii) Apply a univariate time series model to each principal component score to obtain their future values. (iv) Calculate the future realization of the functional time series by multiplying principal components with the forecasted part of the principal component scores. (v) Obtain the prediction intervals under the assumption of Gaussian distributed error terms. This approach (or a modified version) has received extensive attention in the literature and has successfully been used in many areas; see for example, [5–11] and references therein.

The aforementioned literature uses non-robust time series estimation methodologies when forecasting future values of the principal component scores. However, this approach is affected by the outliers which are common in real data sets; see, for example, [12]. An outlier is an observation that has been generated by a stochastic process with a distribution different from that of the vast majority of the remaining observations [13]. In the context of functional time series, three types of outliers are observed: (1) *magnitude outlier*, which is a point far from the bulk of the data; (2) *shape outlier*, which falls within the range of the data but differs in shape from the bulk of the data; and (3) the combination of both outlier types. For more information about the outliers in functional time series, see [13–16].

In the case of outlying observation(s), non-robust techniques produce biased estimates, and high forecasting errors correspond to outlying observations. In such cases, the high forecasting errors may severely affect point forecasts as well as prediction intervals and could lead to unreliable inferences. We propose a robust functional time series forecasting method based on the minimum density power divergence estimator of Basu et al. [17]. The proposed method uses a modified version of the usual maximum likelihood score equations, called weighted score equations, to estimate the model parameters. The weighted score equations are defined as a function of the Pearson residuals for which large values are obtained when the observations diverge from the underlying model. Hence, this approach makes it possible to check whether the maximum likelihood estimators are affected by a set of observations that are inconsistent with the model, and provides robust estimates by downweighting such observations. It also provides weighted residuals, which are used to obtain point and/or interval forecasts. We use several Monte Carlo experiments and real-data examples to compare the finite sample performance of the proposed and existing methods. Our numerical records, which are discussed in Sections 3 and 4, reveal that the proposed method provides finite sample performance competitive with that of the existing methods when no outlier is present in the observed data. In addition, when outliers are

present in the data, its performance is shown to be superior compared with that of available techniques.

The remaining of this paper is organized as follows. Section 2 provides an overview of the functional time series methods considered and the weighted likelihood estimation methodology. Several Monte Carlo experiments under different scenarios are conducted to evaluate the finite sample performance of the proposed method, and the results are presented in Section 3. Section 4 reports the findings obtained by applying the proposed method to some real-data examples. Section 5 concludes the paper.

## 2. Methodology

Let us consider a sequence of stationary functional time series  $\{y_t(x) : t \in \mathcal{Z}, x \in J\}$  where  $J$  is a bounded interval. It is assumed that the functions  $y_t$  are elements of the metric, semi-metric, Hilbert or the Banach space, in general. We assume that the functions are elements of a square-integrable function  $y \in L^2[0, 1]$  residing in Hilbert space  $\mathcal{H}$  satisfying  $\int y^2(x) dx < \infty$  with an inner product  $\langle y, z \rangle = \int y(x)z(x) dx, \forall y, z \in L^2[0, 1]$ . Denote by  $(\Omega, \Sigma, P)$  the probability space where  $\Omega, \Sigma$  and  $P$  represent the sample space,  $\sigma$ -algebra on  $\Omega$  and the probability measure on  $\Sigma$ , respectively. Then the random functional variable  $y$  is defined as  $y : (\Omega, \Sigma, P) \rightarrow \mathcal{H}$  so that  $y$  is assumed to be an element of  $L^2$  and  $y^{-1}(\mathcal{B}) \in \Sigma$  where  $\mathcal{B}$  is a Borel set of the Borel  $\sigma$ -algebra generated by  $L^2$ . We further assume that the random variable  $y \in L^2(\Omega)$  with finite second-order moment is a second-order stochastic process so that  $E[|y|^2] = \int_{\Omega} |y|^2 dP < \infty$ . The mean and covariance functions of the random variable  $y$  are defined as in (2) and (3), respectively; see [18].

$$\mu(x) = E[y(x)] = \int_{\Omega} y(x) dP \quad (2)$$

$$\begin{aligned} C(x, s) &= E[(y(x) - \mu(x))(y(s) - \mu(s))] \\ &= \int_{\Omega} [(y(x) - \mu(x))(y(s) - \mu(s))] dP \end{aligned} \quad (3)$$

Let  $\{y_t(x) : t = 1, \dots, N, x \in J\}$  be an observed functional time series of size  $N$  with the same distribution as  $y$ . Then, the sample mean and sample covariance functions are given by (4) and (5), respectively.

$$\bar{\mu}(x) = N^{-1} \sum_{t=1}^N y_t(x) \quad (4)$$

$$\hat{C}(x, s) = (N-1)^{-1} \sum_{t=1}^N (y_t(x) - \bar{y}(x))(y_t(s) - \bar{y}(s)) \quad (5)$$

Functional principal component analysis is frequently used to analyse functional time series. Briefly, it represents the data by a linear combination of orthonormal principal components  $\phi_k$  and their associated scores  $\beta_{k,t}$  ( $k = 1, 2, \dots$ ). In doing so, it decomposes the covariance operator given in (3) into orthogonal bases of eigenfunctions. Let  $\psi_k$  and  $\lambda_k$ , respectively, denote the  $k$ th eigenfunction and eigenvalue. Then, the covariance operator

is decomposed as follows:

$$C(x, s) = \sum_{k=1}^{\infty} \lambda_k \psi_k(x) \psi_k(s).$$

The  $k$ th principal component score is then defined as  $\beta_{k,t} = \int y_t(x) \psi_k(x) dx$ . In what follows, the random functions are expressed using Karhunen-Loève expansion as:

$$y_t(x) = \sum_{k=1}^{\infty} \beta_{k,t} \psi_k(x)$$

See [19–21] for more details about functional principal component analysis and its practical demonstration.

Let  $\{y_t(x_j)\}$ , for  $t = 1, \dots, N$  and  $j = 1, \dots, J$  be the observed functional time series (with noise) characterized by a continuous smooth function  $f$  and an error process  $\varepsilon$ :

$$y_t(x_j) = f_t(x_j) + \sigma_t(x_j) \varepsilon_{t,j}$$

Each function can be decomposed using a basis function expansion, as follows:

$$f_t(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\phi}_k(x) \hat{\beta}_{k,t} + \epsilon_t(x) \quad (6)$$

where  $\hat{\mu}(x) = N^{-1} \sum_{t=1}^N f_t(x)$  is the estimated mean of functions,  $\hat{\phi}_k(x)$  and  $\hat{\beta}_{k,t}$  denote the estimated functional principal components and their associated principal component scores, respectively,  $\epsilon_t(x)$  is the error function with mean zero and  $K$  is the number of basis functions with  $K < N$ . The choice of the basis function in (6) is arbitrary under orthogonality restriction. As Hyndman and Ullah [4] highlight, each principal component score can be forecasted independently using univariate time series methods since they are uncorrelated to each other. They also note that there may be cross-correlations at non-zero lags but these can be negligible. However, multivariate time series methods, such as vector autoregression, can be used to take into account the cross-correlations; see [22]. Let  $h$  represent the forecast horizon. Denote by  $\hat{\beta}_{k,N+h|N}$  the  $h$ -step-ahead forecast of  $\beta_{k,N+h}$  conditionally on the available data up to time  $N$ . Then, the point predictor of  $y_{N+h}(x)$  is obtained by multiplying the forecasted principal component scores with the estimated functional principal components, as follows:

$$\hat{y}_{N+h|N} = \hat{\mu}(x) + \sum_{k=1}^K \hat{\phi}_k(x) \hat{\beta}_{k,N+h|N}$$

As noted in Section 1, point forecasts do not provide any information about the uncertainty of future realization of the functional time series. Conversely, the prediction interval is capable of producing valid inferences taking into account the uncertainty of each forecast. Hyndman and Ullah [4] propose a  $(1 - \alpha)100\%$  prediction interval for  $y_{N+h}$  under the assumption of normality. However, this approach may seriously be affected by any departure from the normality assumption, which is not known in practice. In such cases,

the bootstrap method is commonly used method to overcome this issue since it does not require full knowledge of the underlying distributional assumption. Therefore, we consider only the bootstrap prediction interval. Hyndman and Shang [7] introduce a bootstrap approach to construct prediction intervals for the future values of the functional time series. Based on the functional time series model discussed above, we incorporate three error sources to construct the bootstrap prediction interval: (i) smoothing error  $\varepsilon_{t,j}$  given in (1), (ii) the error caused by the functional principal component decomposition  $\epsilon_t(x)$  in (6), and (iii) the error occurring owing to forecasting principal component scores  $\beta_{k,t}$ . Let  $\xi_{k,h,t}$  denote the  $h$ -step-ahead forecast error  $\hat{\xi}_{k,h,t} = \beta_{k,t} - \hat{\beta}_{k,t|t-h}$  ( $t = h + 1, \dots, N$ ). Then, the algorithm of the bootstrap procedure proposed by Hyndman and Shang [7] is as follows.

*Step 1.* Smooth the entire functional time series  $y_t(x) = f_t(x) + \sigma_t(x)\varepsilon_t$  for  $t = 1, \dots, N$  to obtain smooth functions  $\hat{f}_t(x)$ , smoothing error  $\hat{\varepsilon}_t$  and variance component  $\hat{\sigma}_t(x)$ .

*Step 2.* Decompose the smooth functions into  $k = 1, \dots, K$  orthonormal functional principal components and associated principal component scores and obtain the fitted functions  $\hat{f}_t(x)$ . Then, calculate the error functions  $\hat{\epsilon}_t(x) = f_t(x) - \hat{f}_t(x)$ .

*Step 3.* Obtain  $h$ -step-ahead forecasts of the principal component scores  $\beta_{k,t}$ ,  $\hat{\beta}_{k,N+h|N}$  for  $k = 1, \dots, K$  and  $t = N + 1, \dots, N + h$  using a univariate time series method. In addition, obtain the  $h$ -step-ahead forecast errors  $\hat{\xi}_{k,h,t}$ .

*Step 4.* Calculate the future bootstrap values as follows:

$$\hat{y}_{N+h|N}^*(x) = \hat{\mu}(x) + \sum_{k=1}^K \hat{\beta}_{k,N+h|N}^* + \hat{\epsilon}_{N+h|N}^*(x) + \hat{\sigma}_{N+h}^*(x)\hat{\varepsilon}_{N+h|N,j}^*$$

where  $\hat{\beta}_{k,N+h|N}^* = \hat{\beta}_{k,N+h|N} + \xi_{k,h,*}$ ,  $\xi_{k,h,*}$  is an i.i.d. random sample from  $\{\xi_{k,h,t}\}$ ,  $\hat{\epsilon}_{N+h|N}^*(x)$ ,  $\hat{\varepsilon}_{N+h|N,j}^*$  and  $\hat{\sigma}_{N+h}^*(x)$  are random samples with replacement from  $\{\hat{\epsilon}_t(x)\}$ ,  $\{\hat{\varepsilon}_t\}$  and  $\{\hat{\sigma}_t(x)\}$ , respectively.

*Step 5.* Repeat Step 4.  $B$  times by drawing random samples of the error terms to obtain  $B$  sets of bootstrap replicates of  $\{\hat{y}_{N+h|N}^*(x), \{\hat{y}_{N+h|N}^{*,1}(x), \dots, \hat{y}_{N+h|N}^{*,B}(x)\}\}$  for each  $h$ , where  $B$  denotes the number of bootstrap simulations.

The  $(1 - \alpha)100\%$  bootstrap prediction intervals for  $y_{N+h}$  are then obtained by the  $\alpha/2$ th and  $(1 - \alpha/2)$ th quantiles of the bootstrap replicates  $\{\hat{y}_{N+h|N}^{*,1}(x), \dots, \hat{y}_{N+h|N}^{*,B}(x)\}$ . This approach works well when the functional time series is relatively smooth, and no outlier is present in the data. It has successfully been used in a wide range of applications; see, for example, [23–30]. By contrast, the traditional univariate time series methods used for modelling the functional principal component scores may produce biased estimates as well as high forecasting errors, leading to unreliable results when outliers are present in the data. To overcome this issue, we propose a robust forecasting functional time series method. It is based on replacing the traditional estimators and residuals by the minimum power divergence estimator of Basu et al. [17] and weighted residuals, respectively.

## 2.1. The weighted likelihood methodology

To employ the weighted likelihood methodology in the forecasting of functional time series, we consider the stationary autoregressive model of order  $p$  (AR( $p$ )). Let us suppose

that the  $k$ th principal component score  $\beta_{k,t}$  ( $t = 1, \dots, N$ ) is characterized by a zero mean AR( $p$ ) process, as follows:

$$\beta_{k,t} = \phi_1 \beta_{t-1,k} + \dots + \phi_p \beta_{t-p,k} + \zeta_t, \quad t = 1, \dots, N \quad (7)$$

where  $\Phi = (\phi_1, \dots, \phi_p)$  denotes the parameter vector and  $\zeta_t$  is an i.i.d. white noise sequence with mean zero and variance  $\sigma_\zeta^2$ . Under the assumption of normality, the probability density function for the model (7) is given by

$$P(\zeta_N | \Phi; \sigma_\zeta^2) = \left( \frac{1}{2\pi\sigma_\zeta^2} \right)^{-(N-p)/2} \exp \left\{ -\frac{1}{2\sigma_\zeta^2} \sum_{t=p+1}^N \zeta_t^2(\Phi) \right\} \quad (8)$$

where  $\zeta_t(\Phi) = \beta_{k,t} - \sum_{i=1}^p \phi_i \beta_{k,t-i}$  and  $\zeta_N = \langle \zeta_{p+1}, \dots, \zeta_N \rangle$ . Let  $\ell(\zeta_t(\Phi); \sigma_\zeta^2) = \log \mathcal{L}(\zeta_t(\Phi); \sigma_\zeta^2)$  denote the conditional log-likelihood function. Then, the maximum likelihood estimators of  $\Phi$  and  $\sigma_\zeta^2$  conditionally on the first  $p$  observations, respectively, are obtained by the solution of the score functions  $u_\Phi$  and  $u_{\sigma_\zeta}$ , as follows:

$$\begin{aligned} u_\Phi(\zeta_t(\Phi); \sigma_\zeta^2) &= \frac{\partial}{\partial \Phi} \ell(\zeta_t(\Phi); \sigma_\zeta^2) \\ u_{\sigma_\zeta}(\zeta_t(\Phi); \sigma_\zeta^2) &= \frac{\partial}{\partial \sigma_\zeta} \ell(\zeta_t(\Phi); \sigma_\zeta^2) \end{aligned}$$

The weighted likelihood methodology is proposed by Markatou [31] and Basu et al. [17] to construct efficient and robust estimators by replacing the usual score functions with weighted score equations that measure the discrepancy between the estimated and hypothesized model densities. It has been extended to a wide variety of statistical inference problems, see, for example, [32–36]. Let us consider the probability density function given by (8) and let  $f^*(\zeta_t(\Phi), \hat{F}_N(\Phi))$  and  $m^*(\zeta_t(\Phi), \sigma_\zeta^2)$  denote a kernel density estimator with bandwidth  $h$  based on the empirical distribution function  $\hat{F}_N(\Phi)$  and the kernel smoothed model density, as follows:

$$\begin{aligned} f^*(\zeta_t(\Phi), \hat{F}_N(\Phi)) &= \int k(\zeta_t(\Phi); r, g) d\hat{F}_N(r; \Phi) \\ m^*(\zeta_t(\Phi), \sigma_\zeta^2) &= \int k(\zeta_t(\Phi); r, g) dM(r; \sigma_\zeta^2) \end{aligned}$$

where  $M(\sigma_\zeta^2)$  and  $k(\zeta_t(\Phi); r, g)$  represent the normal distribution function with mean zero and variance  $\sigma_\zeta^2$  and a kernel density with bandwidth  $g$ , respectively. Denote by  $\delta_t(\zeta_t(\Phi); M(\sigma_\zeta^2), \hat{F}_N(\Phi)) = f^*(\zeta_t(\Phi), \hat{F}_N(\Phi)) / m^*(\zeta_t(\Phi), \sigma_\zeta^2)$  and  $\omega(\delta_t) = \min\{1, [A(\delta_t) + 1]^+ / (\delta_t + 1)\}$  the Pearson residual and the weight function, respectively, where  $[\cdot]^+$  denotes the positive part and  $A(\cdot)$  is the residual adjustment function. Note that we consider the Hellinger residual adjustment function of Lindsay [37],  $A(\delta) = 2[(\delta + 1)^{0.5} - 1]$ . Based on these definitions, the conditional weighted likelihood of the parameters  $\Phi$  and  $\sigma_\zeta$  are obtained by solving the following estimating equations:

$$(N-p)^{-1} \sum_{t=p+1}^N \omega(\delta_t) u_\Phi(\zeta_t(\Phi); \sigma_\zeta^2)$$

$$(N-p)^{-1} \sum_{t=p+1}^N \omega(\delta_t) u_{\sigma_\zeta} \left( \zeta_t(\Phi); \sigma_\zeta^2 \right)$$

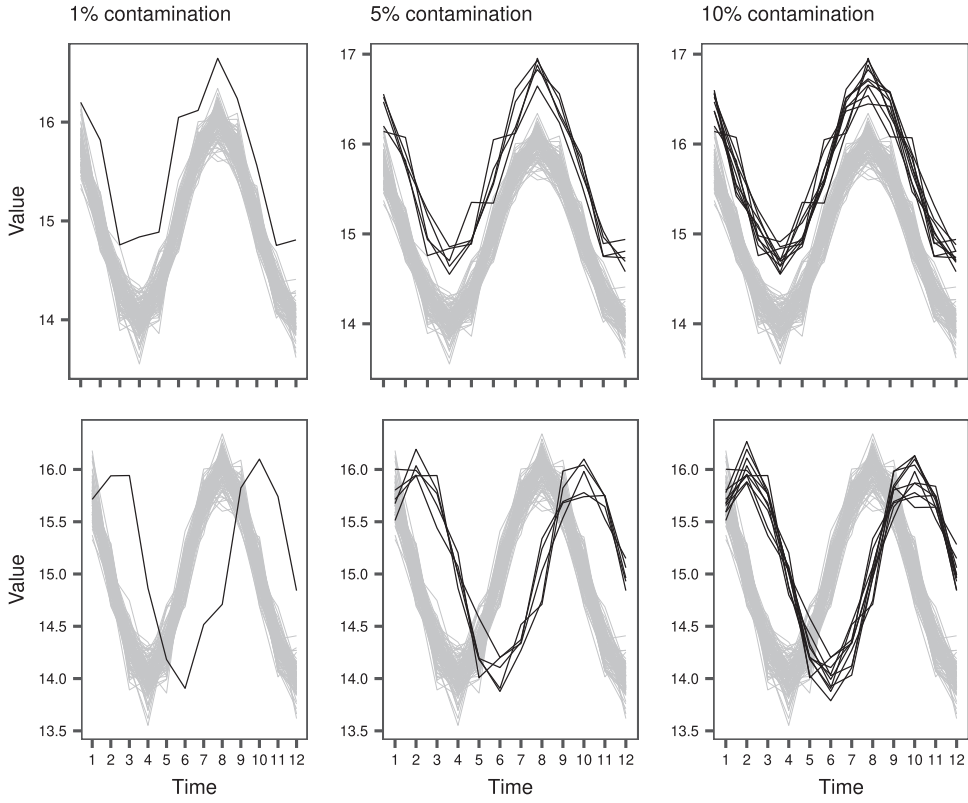
Now let  $\hat{\Phi}^\omega = (\hat{\phi}_1^\omega, \dots, \hat{\phi}_p^\omega)$  denote the estimated parameter vector of AR( $p$ ) process using the weighted likelihood methodology. Denote by  $\hat{\beta}_{k,N+h|N}^\omega = \sum_{i=1}^p \hat{\phi}_i^\omega \hat{\beta}_{k,N+h-i|N}^\omega$  the weighted likelihood version of  $h$ -step ahead forecast of  $\beta_{k,N+h}$ . Then, the weighted likelihood-based  $h$ -step-ahead point forecasts and bootstrap prediction intervals for  $y_{N+h|N}$  are obtained similarly as in the algorithm given in the previous subsection. In the proposed forecasting strategy, the following holds for the final weights obtained from the full model,  $\hat{\omega}_t$ , when the model is accurately specified and no outlying observation is present in the data;  $\sup_t |\hat{\omega}_t - 1| \xrightarrow{P} 0$ , see [38]. Following by [33], it can be shown that  $|\sum_{t=1}^N \hat{\omega}_t \zeta_t(\hat{\Phi}^\omega)^2 - \sum_{t=1}^N \zeta_t(\hat{\Phi})^2| = o_p(N)$ . This result indicates that (i) the weighted likelihood method tends to perform similar to the maximum likelihood method when no outlier is present in the data, and (ii) the weighted likelihood is expected to have better performance than the maximum likelihood method when the data are contaminated by the outlier(s) (since it downweights the ‘bad’ forecast errors caused by the contamination).

### 3. Numerical results

This section reports the finite sample performance of the proposed functional time series forecasting method via several Monte Carlo experiments. Throughout the experiments, two different simulation scenarios are considered: (1) The data are generated from a relatively smooth process with no outlying observations, and (2)  $\gamma = [1\%, 5\%, 10\%]$  of the generated data is contaminated by the deliberately inserted magnitude and shape outlier(s). The following process is used to generate the data;  $y_t(x_j) = 15 + \cos(\pi j/4) + N(0, 0.15^2)$  ( $t = 1, \dots, N = 100$  and  $j = 1, \dots, J = 12$ ). The magnitude outlier(s) are generated by contaminating  $N\gamma$  randomly selected function(s) by a random function  $y_t^c(x_j) = |N(0.75, 0.15^2)|$ , that is  $y_t(x_j) = y_t(x_j) + y_t^c(x_j)$ . Conversely, the shape outliers are generated from the following process:  $15 + \sin(\pi j/4) + N(0, 0.15^2)$ . Examples of the simulated data (with magnitude and shape outliers) are presented in Figure 1.

We divide the generated data into two parts so that we can construct the model using the first  $N-h$  functions to obtain point forecasts and the last  $h$  functions for the bootstrap prediction intervals, where  $h = [1, 5, 10]$ . To construct an functional time series model, first, the noisy functional time series is converted to a smooth function by the smoothing spline method. Then, the hybrid principal component model of Hyndman and Ullah [4] is used to decompose the smooth functions. The nominal level  $\alpha$  is set to 0.05 to calculate 95% bootstrap prediction intervals. For each scenario,  $MC = 1000$  Monte Carlo simulations with  $B = 999$  bootstrap resamples are performed, and only the first  $K = 3$  principal components and their scores are used to obtain forecasts. The performance of the proposed method is compared with that of the autoregressive integrated moving average (ARIMA) model as used by Hyndman and Ullah [4] and most of the references cited. To evaluate the forecasting performance of the proposed method, we calculate several performance metrics, including the average mean squared error (AMSE), coverage probability (Cp) and the





**Figure 1.** Example of the simulated functional time series (grey lines) with outliers (black lines): magnitude outlier (first row) and shape outlier (second row).

average interval scores (Score), as follows:

$$\begin{aligned}
 \text{AMSE} &= \frac{1}{h \times 12} \sum_{i=1}^h \sum_{j=1}^{12} \{y_{N+h}(x_j) - \hat{y}_{N+h|N}(x_j)\}^2 \\
 \text{Cp} &= \frac{1}{h \times 12} \sum_{i=1}^h \sum_{j=1}^{12} 1 \left\{ \hat{y}_{N+h|N}^{*,0.025}(x_j) \leq y_{N+h}(x_j) \leq \hat{y}_{N+h|N}^{*,0.975}(x_j) \right\} \\
 \text{Score} &= \frac{1}{h \times 12} \sum_{i=1}^h \sum_{j=1}^{12} \left\{ \left( \hat{y}_{N+h|N}^{*,0.975}(x_j) - \hat{y}_{N+h|N}^{*,0.025}(x_j) \right) \right. \\
 &\quad + \frac{2}{\alpha} \left( \hat{y}_{N+h|N}^{*,0.025}(x_j) - y_{N+h}(x_j) \right) \mathbb{1} \left\{ y_{N+h}(x_j) < \hat{y}_{N+h|N}^{*,0.025}(x_j) \right\} \\
 &\quad \left. + \frac{2}{\alpha} \left( y_{N+h}(x_j) - \hat{y}_{N+h|N}^{*,0.975}(x_j) \right) \mathbb{1} \left\{ y_{N+h}(x_j) > \hat{y}_{N+h|N}^{*,0.975}(x_j) \right\} \right\}
 \end{aligned}$$

where  $\mathbb{1}\{\cdot\}$  denotes the indicator function and  $\hat{y}_{N+h|N}^{*,\alpha}(x_j)$  is the  $\alpha$ th quantile of the bootstrap replicates.

**Table 1.** Simulation results: Estimated AMSE, Cp and Score values for the proposed (weighted likelihood estimator (WLE)) and ARIMA models when the data have magnitude (MO) and shape (SO) outliers.

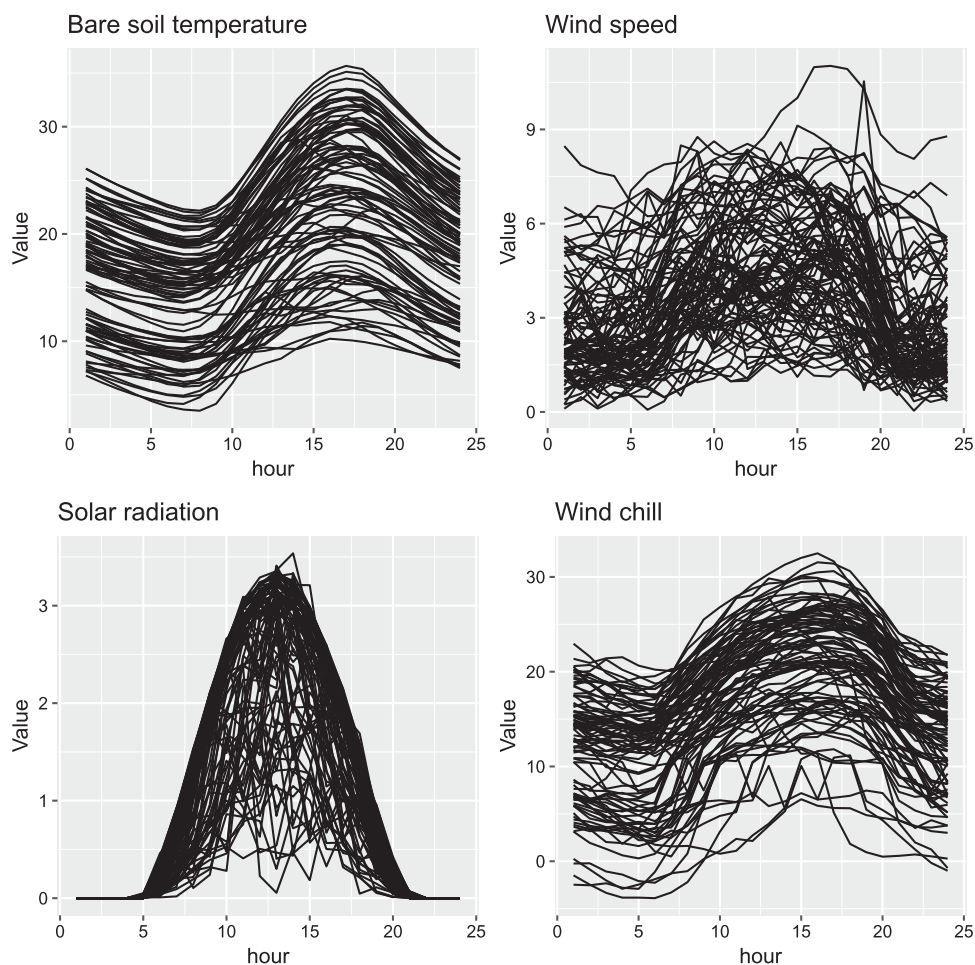
Contamination	$h$	Outlier	Method	AMSE	Cp	Score	
0%	$h = 1$		WLE	0.0228	0.9365	0.7244	
			ARIMA	0.0229	0.9119	0.7419	
	$h = 5$		WLE	0.0228	0.9360	0.7258	
			ARIMA	0.0228	0.9122	0.7413	
	$h = 10$		WLE	0.0229	0.9350	0.7312	
			ARIMA	0.0228	0.9106	0.7492	
1%	$h = 1$	MO	WLE	0.0229	0.9375	0.7299	
			ARIMA	0.0229	0.9245	0.7507	
		SO	WLE	0.0238	0.9153	0.7525	
			ARIMA	0.0242	0.9086	0.7749	
	$h = 5$	MO	WLE	0.0231	0.9438	0.7303	
			ARIMA	0.0246	0.9553	1.1374	
		SO	WLE	0.0225	0.9366	0.7315	
			ARIMA	0.0258	0.9413	1.3039	
	$h = 10$	MO	WLE	0.0242	0.9467	0.7404	
			ARIMA	0.0296	0.9594	1.2357	
	5%		SO	WLE	0.0228	0.9373	0.7398
				ARIMA	0.0343	0.935	1.4226
$h = 1$		MO	WLE	0.0229	0.9378	0.7277	
			ARIMA	0.0230	0.9246	0.7415	
		SO	WLE	0.0233	0.9167	0.7537	
			ARIMA	0.0242	0.9089	0.7847	
$h = 5$		MO	WLE	0.0233	0.9420	0.7329	
			ARIMA	0.0249	0.9557	1.1344	
		SO	WLE	0.0297	0.9306	0.7375	
			ARIMA	0.0263	0.9342	1.3158	
$h = 10$		MO	WLE	0.0233	0.9507	0.7298	
			ARIMA	0.0295	0.9592	1.2390	
10%		SO	WLE	0.0247	0.9346	0.7583	
			ARIMA	0.0361	0.9322	1.4349	
	$h = 1$	MO	WLE	0.0227	0.9377	0.7265	
			ARIMA	0.0228	0.9244	0.7476	
		SO	WLE	0.0228	0.9197	0.7601	
			ARIMA	0.0238	0.9071	0.8843	
	$h = 5$	MO	WLE	0.0232	0.9437	0.7315	
			ARIMA	0.0249	0.9546	1.1469	
		SO	WLE	0.0231	0.9225	0.7652	
			ARIMA	0.0276	0.9217	1.2838	
	$h = 10$	MO	WLE	0.0236	0.9491	0.7324	
			ARIMA	0.0303	0.9582	1.2457	
	SO	WLE	0.0232	0.9286	0.7769		
		ARIMA	0.0374	0.9338	1.4522		

The results obtained from the Monte Carlo experiments are reported in Table 1. Our findings show that, regardless of the forecast horizon, both weighted likelihood and ARIMA-based forecasting models tend to have similar AMSE values for the point forecasts when no outliers are present in the data. However, the proposed method produces better coverage probabilities and interval scores than those obtained from ARIMA. These results indicate that compared with ARIMA, we can obtain more reliable and more precise prediction intervals using the weighted likelihood-based bootstrap method even if the outliers do not contaminate the data. When the data has outlier(s) (magnitude and/or shape), both methods still have similar AMSE values for the point forecasts when  $h = 1$  but the proposed method has better coverage probabilities and interval scores compared

with those of ARIMA. By contrast, for long-term forecast horizons ( $h=5$  and  $h=10$ ), the proposed method has smaller AMSE values than ARIMA, and the difference becomes more prominent as the forecast horizon increases. For the prediction intervals, while it seems that both methods have similar coverage performance, the weighted likelihood-based bootstrap method produces significantly narrower prediction intervals than those of ARIMA. This is because the bootstrap method based on ARIMA is considerably affected

**Table 2.** Simulation results: Estimated AMSE, Cp and Score values for the proposed (weighted likelihood estimator (WLE)) and ARIMA models when the data have large magnitude outliers and forecast horizon  $h = 1$ .

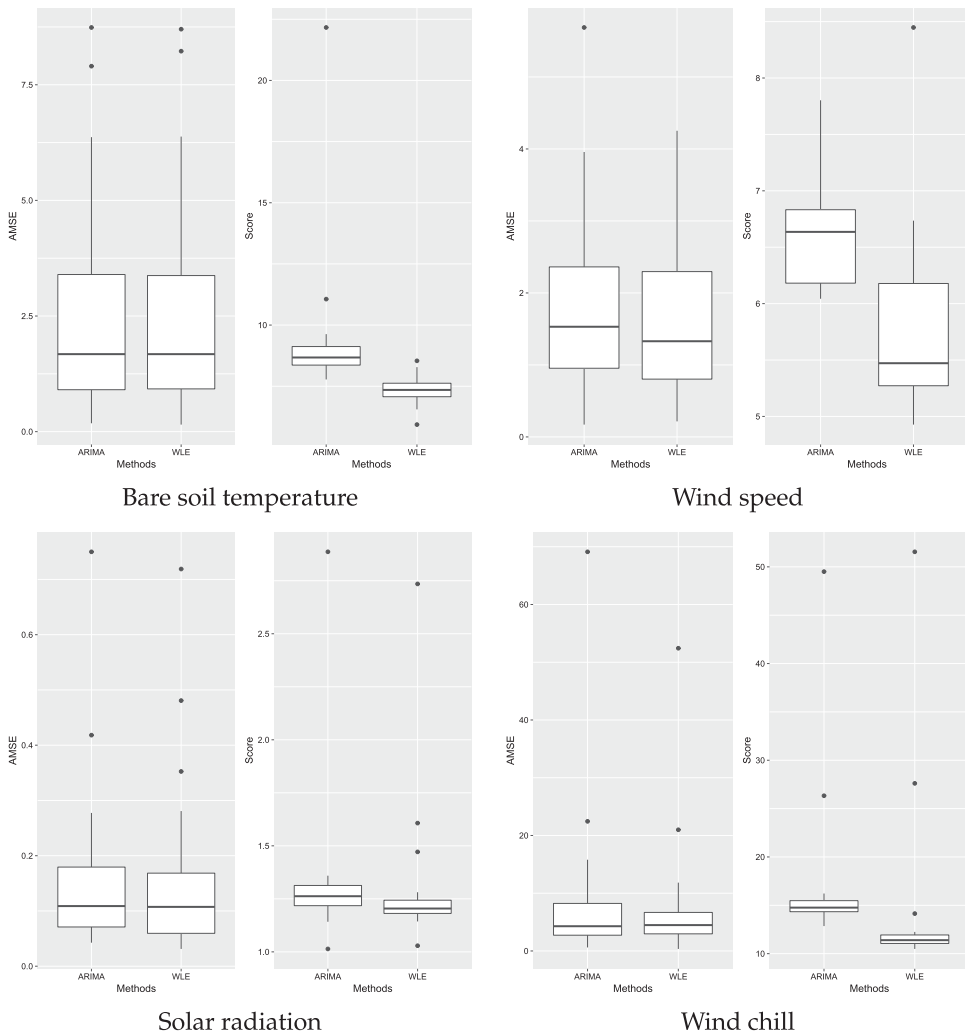
Contamination	Method	AMSE	Cp	Score
1%	WLE	0.0300	0.9352	0.7575
	ARIMA	0.0321	0.9100	0.8003
5%	WLE	0.0229	0.9533	0.7564
	ARIMA	0.0636	0.9710	4.1816
10%	WLE	0.0305	0.9612	0.9618
	ARIMA	0.1805	0.9780	4.4548



**Figure 2.** Functional time series representations of the environmental variables.

by the large forecast errors produced by the outliers. However, the proposed method down-weights the effects of outliers, and the structure of the bootstrap prediction intervals are not distorted.

The finite-sample performance of a maximum likelihood based forecasting method, such as ARIMA, depends on the magnitude of outliers. Via several Monte-Carlo experiments, we examine the effects of outlier size on forecasting accuracy of the ARIMA and weighted likelihood based functional time series methods. Our results indicate that the performance of the ARIMA becomes worse as the magnitude of the outliers increases while the proposed weighted likelihood based functional time series method produces consistent results. For instance, when the magnitude outliers are generated from  $y_t^c(x_j) = |N(3.75, 0.15^2)|$ , the one-step-ahead forecasting performances of the ARIMA and proposed weighted likelihood based functional time series methods are presented in Table 2.

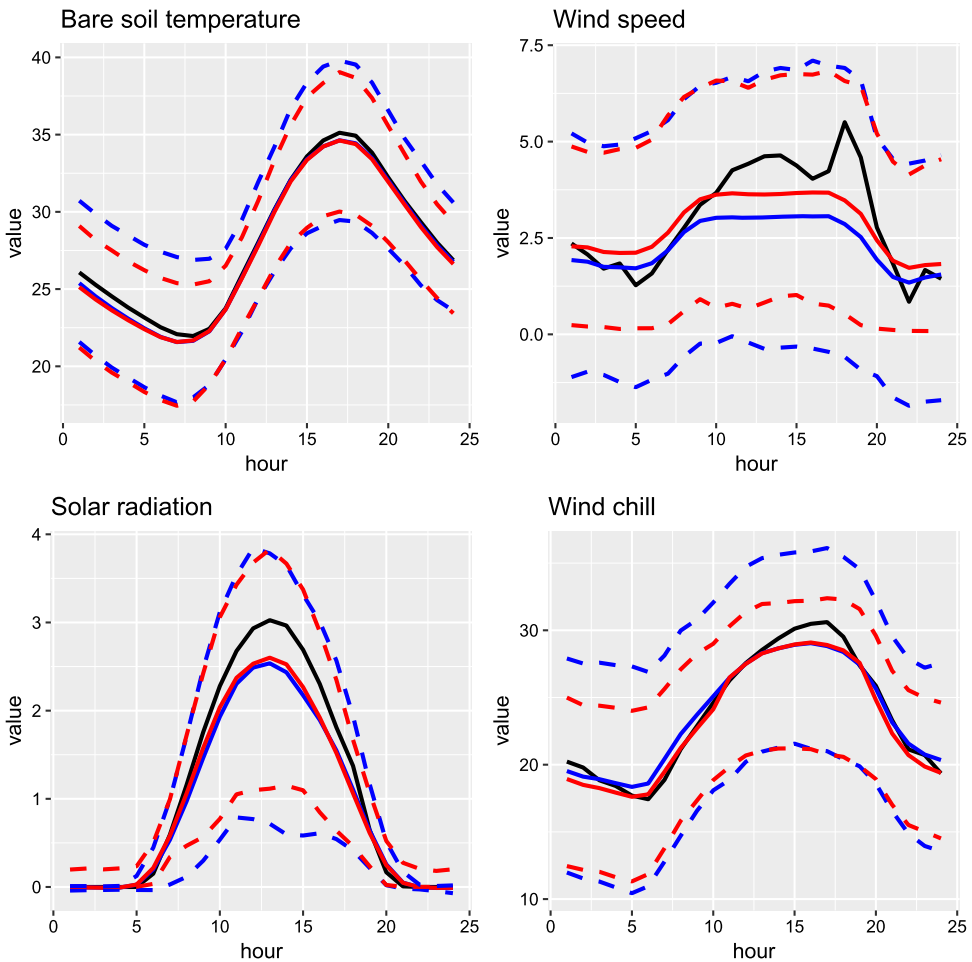


**Figure 3.** Obtained AMSE and Score values of the ARIMA and proposed (WLE) models for the real-data examples.

Compared to the results reported in Table 1, Table 2 shows that when [1%, 5%, 10%] of the data are contaminated with larger magnitude outliers, the ARIMA method produced about [1.40, 2.76, 7.91] times larger AMSE than the ones obtained when the data are contaminated with small magnitude outliers and [1.06, 5.64, 5.96] times larger score values. On the other hand, the proposed method produced almost the similar AMSE and score values in both cases. Similar results can be obtained from the corresponding author upon request for the case when the data have large shape outliers.

#### 4. Real-data examples

This section evaluates the finite sample performance of the proposed method using four environmental datasets – hourly bare soil temperature, wind speed, solar radiation, and



**Figure 4.** Obtained one-step-ahead point forecasts: ARIMA (blue solid lines) and weighted likelihood (red solid lines), and bootstrap 95% prediction intervals: ARIMA (blue dashed lines) and weighted likelihood (red dashed lines), together with the observed function (black solid lines), please see the online version of the article to see coloured plots.

wind chill from 1 May 2017 to 31 July 2017 (92 days in total) – which are collected from the Michigan weather station (the data are obtained from North Dakota Agricultural Weather Network Center <https://ndawn.ndsu.nodak.edu/>). The functional time series representation of the datasets is presented in Figure 2. It is clear from this figure that the bare soil temperature functions are reasonably smooth and have no clear outlier. Conversely, the functions of the functional time series of wind speed, solar radiation, and wind chill are noisy, and all three datasets have several outlying functions. For all four datasets, we obtain only one-step-ahead forecasts. The forecasting performance of the ARIMA and of the proposed method are compared using the rolling holdout testing samples, in line with Shang [12] and Hyndman and Ullah [4]; thus, 80% of the datasets are used as training samples and the remaining 20% are used for validation. In the modelling step, the number of principal components is also determined based on rolling holdout validation samples. The optimal value of  $K$  is found as 4 for the bare soil temperature dataset and 6 for the other three datasets.

The obtained AMSE and Score values are presented in Figure 3. The results indicate that compared with ARIMA, the proposed method produces slightly better AMSE values for the one-step-ahead point forecast. The results also show that the weighted likelihood-based bootstrap method produces better prediction intervals than those obtained from the ARIMA-based bootstrap procedure especially for the datasets contaminated by outliers. Figure 4 presents a graphical representation of the one-step-ahead forecasts for which the functional time series models are constructed based on the first 91 functions. This figure supports the results presented in Figure 3.

## 5. Conclusion

Functional time series are frequently observed in many scientific fields owing to novel data collection tools. Consequently, several functional time series models have been developed to analyse such datasets and to obtain forecasts of their unobserved realizations. Recent studies have shown that the univariate time series models together with functional principal component regression can be used to obtain valid point forecasts and bootstrap prediction intervals. However, the traditional univariate time series models such as ARIMA, which are commonly used in the analyses, may severely be affected in the presence of outliers, leading to poor forecasting results. Therefore, we propose a robust functional time series forecasting approach based on the weighted likelihood methodology. We evaluate the finite sample performance of the proposed method using several Monte Carlo experiments and four environmental datasets. The numerical results produced showed that the proposed method is a good competitor for ARIMA and would be widely adopted because of its narrower prediction intervals.

## Acknowledgments

We thank the anonymous referee and AE for their careful reading of our manuscript and valuable suggestions and comments, which have helped us produce a significantly better paper.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The second author also acknowledges the financial support from a research grant at the College of Business and Economics at the Australian National University.

## ORCID

Ufuk Beyaztas  <http://orcid.org/0000-0002-5208-4950>

Han Lin Shang  <http://orcid.org/0000-0003-1769-6430>

## References

- [1] Chatfield C. Calculating interval forecasts. *J Bus Econ Stat.* **1993**;11(2):121–135.
- [2] Jore AS, Mitchell J, Vahey SP. Combining forecast densities from vars with uncertain instabilities. *J Appl Econom.* **2010**;25(4):621–634.
- [3] Kim JH. Bootstrap after bootstrap prediction intervals for autoregressive models. *J Bus Econ Stat.* **2001**;19(1):117–128.
- [4] Hyndman RJ, Ullah MS. Robust forecasting of mortality and fertility rates: a functional data approach. *Comput Stat Data Anal.* **2007**;51(10):4942–4956.
- [5] Curceac S, Ternynck C, Ouarda TBMJ, et al. Short-term air temperature forecasting using nonparametric functional data analysis and SARMA models. *Environ Model Softw.* **2019**;111:394–408.
- [6] Gao Y, Shang HL. Multivariate functional time series forecasting: application to age-specific mortality rates. *Risks.* **2017**;5(21):1–18.
- [7] Hyndman RJ, Shang HL. Forecasting functional time series. *J Korean Stat Soc.* **2009**;38(3):199–211.
- [8] Kosiorowski D. Functional regression in short-term prediction of economic time series. *Stat Transit.* **2014**;15(4):611–626.
- [9] Shang HL. Functional time series approach for forecasting very short-term electricity demand. *J Appl Stat.* **2013**;40(1):152–168.
- [10] Shang HL, Hyndman RJ. Nonparametric time series forecasting with dynamic updating. *Math Comput Simul.* **2011**;81(7):1310–1324.
- [11] Wagner-Muns IM, Guardiola IG, Samaranayake VA, et al. A functional data analysis approach to traffic volume forecasting. *IEEE Trans Intell Transp Syst.* **2019**;19(3):878–888.
- [12] Shang HL. A robust functional time series forecasting method. *J Stat Comput Simul.* **2019**;89(5):795–814.
- [13] Raña P, Aneiros G, Vilar JM. Detection of outliers in functional time series. *Environmetrics.* **2015**;26(3):178–191.
- [14] Febrero M, Galeano P, Gonzalez-Manteiga W. Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics.* **2007**;19(4):331–345.
- [15] Hyndman RJ, Shang HL. Rainbow plots, bagplots, and boxplots for functional data. *J Comput Graph Stat.* **2010**;19(1):29–45.
- [16] Sun Y, Genton MG. Functional boxplots. *J Comput Graph Stat.* **2011**;20(2):316–334.
- [17] Basu A, Harris IR, Hjort NL, et al. Robust and efficient estimation by minimising a density power divergence. *Biometrika.* **1998**;85(3):549–559.
- [18] Ramsay JO, Dalzell C. Some tools for functional data analysis. *J R Stat Soc Ser B.* **1991**;53(3):539–572.
- [19] Ramsay JO, Hooker G, Graves S. *Functional data analysis with R and MATLAB.* New York: Springer; **2009**.
- [20] Ramsay JO, Silverman BW. *Applied functional data analysis: methods and case studies.* New York: Springer; **2002**.
- [21] Shang HL. A survey of functional principal component analysis. *AStA Adv Stat Anal.* **2014**;98(2):121–142.

- [22] Aue A, Norinho DD, Hormann S. On the prediction of stationary functional time series. *J Am Stat Assoc.* [2015](#);110:378–392.
- [23] Aneiros G, Vilar J, Raña P. Short-term forecast of daily curves of electricity demand and price. *Int J Elec Power Energy Syst.* [2016](#);80:96–108.
- [24] Aneiros G, Vilar JM, Cao R, et al. Functional prediction for the residual demand in electricity spot markets. *IEEE Trans Power Syst.* [2013](#);28(4):4201–4208.
- [25] Canale A, Vantini S. Constrained functional time series: applications to the Italian gas market. *Int J Forecast.* [2016](#);32(4):1340–1351.
- [26] Husin WZW, Ramli NM, Zainol MS, et al. Forecasting Malaysian mortality rates: a functional data analysis approach. *Adv Sci Lett.* [2016](#);2(12):4259–4262.
- [27] Hyndman RJ, Booth H, Yasmeen F. Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography.* [2013](#);50(1):261–283.
- [28] Kearney F, Cummins M, Murphy F. Forecasting implied volatility in foreign exchange markets: a functional time series approach. *Eur J Finance.* [2018](#);24:1–18.
- [29] Shang HL. Forecasting intraday S&P500 index returns: a functional time series approach. *Forecasting < /DIFdel > J Forecast.* [2017](#);36(7):741–755.
- [30] Shang HL, Yang Y, Kearney F. Intraday forecasts of a volatility index: functional time series method with dynamic updating. *Ann Oper Res.* [2018](#). DOI:10.1007/s10479-018-3108-4
- [31] Markatou M. Robust statistical inference: weighted likelihoods or usual mestimation? *Comm Statist Theory Methods.* [1996](#);25(11):2597–2613.
- [32] Agostinelli C. Robust model selection in regression via weighted likelihood methodology. *Stat Probab Lett.* [2002](#);56(3):289–300.
- [33] Agostinelli C. Robust stepwise regression. *J Appl Stat.* [2002](#);29(6):825–840.
- [34] Agostinelli C. Robust time series estimation via weighted likelihood. In: Dutter R, Filzmoser P, Gather U, Rousseeuw PJ, editors. *Developments in robust statistics.* Heidelberg: Springer; 2003.
- [35] Agostinelli C, Bisaglia L. ARFIMA processes and outliers: a weighted likelihood approach. *J Appl Stat.* [2010](#);37(9):1569–1584.
- [36] Agostinelli C, Markatou M. Test of hypotheses based on the weighted likelihood methodology. *Stat Sin.* [2001](#);11(2):499–514.
- [37] Lindsay BG. Efficiency versus robustness: the case for minimum hellinger distance and related methods. *Ann Stat.* [1994](#);22(2):1018–1114.
- [38] Agostinelli C. *Inferenza statistica robusta basata sulla funzione di verosimiglianza pesata: alcuni sviluppi* [Phd thesis]. Dipartimento di Scienze Statistiche, Università di Padova; 1998.