

Multi-step-ahead prediction interval for locally stationary time series with application to air pollutant concentration data

Jie Li¹  | Qirui Hu¹  | Fengying Zhang²

¹Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China

²China National Environmental Monitoring Center, Beijing, 10012, China

Correspondence

Qirui Hu, Center for Statistical Science, Tsinghua University, Huaye Building 3702, Beijing 100084, China.
Email: hqr20@mails.tsinghua.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 11771240, 12026242; China National Environmental Monitoring Center

Locally stationary time series frequently appears in both finance and environmental sciences (e.g., daily air pollutant concentration or financial returns). However, constructing the multi-step-ahead prediction interval for such time series remains an open question. Hence, we extend the nonparametric regression model with autoregressive errors for equally spaced designs to the time series setup. We propose a B-spline estimator for the trend function and a kernel estimator for the variance function to implement the model. The prediction interval of multi-step-ahead future observations is also constructed after fitting the autoregressive model of errors and obtaining the quantile of prediction residuals. The proposed method is illustrated by various simulation studies and an example of air pollutant data, containing 8 years of daily air pollutant concentrations in Xi'an. Our results demonstrate that our method outperforms others owing to its higher prediction accuracy and versatility.

KEYWORDS

air pollutants concentration, B-spline, kernel, locally stationary time series, prediction interval

1 | INTRODUCTION

A fundamental task of time series analysis is forecasting for its wide application in environment, economics and other disciplines. Prediction interval (PI), the estimate of an interval wherein a future observation will fall with a certain probability, is indispensable for predictive inference (Brockwell & Davis, 1991; Fan & Yao, 2008).

Many recent papers investigated PI based on various time series models. Particularly, Lori and William (1990) applied nonparametric bootstrap to predict autoregression. Wang et al. (2014) and Kong et al. (2018) proposed a kernel estimator for the distribution function of unobserved errors and multi-step-ahead prediction errors in autoregressive time series, respectively, based on residuals computed by estimating autoregressive coefficients using the Yule–Walker method. All of them successfully constructed multi-step-ahead PIs; however, they only focused on the simple AR(p) model, which is inadequate for fitting most real data. Aneiros-Pérez et al. (2011) adopted a nonparametric view for time series prediction using functional data techniques based on homogeneous errors and bootstrap residuals. De Livera et al. (2011) incorporated Box–Cox transformations, Fourier representations with time-varying coefficients to complex seasonal time series and derived analytical expressions for point forecasts and interval predictions under the assumption of Gaussian errors. Although inference results of the above models have been well developed, predicting performance varies in different scenarios. As these models are insufficient for explaining the complex structure of some data, they may obtain a non-smooth estimation owing to model misspecification. Moreover, they usually assume that residuals are normal or asymptotically normal for short-term predictions, which is quite restrictive. This sometimes leads to a much wider PI than usual in practice (Section 5).

To address the above issue, assuming a slowly changing stochastic structure is more realistic: that is, a locally stationary model (Dahlhaus, 2012). Dette and Wu (2022) developed an estimator for the high-dimensional covariance matrix of a locally stationary process with a smoothly varying trend. They then used this statistic to derive consistent predictors in nonstationary time series. The proposed predictor did not rely on fitting an autoregressive model or required a vanishing trend; however, they prioritized the point predictor rather

than the PI. Little literature is available on PIs of locally stationary time series. We believe that the only relevant paper is Das and Politis (2021), which is the first to combine one-step-ahead point predictors and PIs for model-free or model-based scenarios in the context of locally stationary time series. The bootstrap method was used in constructing PIs; thus, they failed to establish a multi-step-ahead PI owing to the lack of error distribution estimation. In our paper, we introduce a novel and applicable method for forecasting multi-step-ahead future observations and establishing its corresponding PI under the locally stationary time series setting. Our proposed method is applied to construct future air pollutants concentration PIs based on a large dataset comprising 8 years of daily air pollutant concentration data in Xi'an. Our final results show that proposed PI achieves superior performance in comparison with its counterpart derived by the seasonal ARIMA method.

Some commonly used methods for assessing model forecasting performance, such as out-of-sample and cross-validation techniques, do not work in time series settings as the testing and training sets cannot be split randomly in the presence of dependence among temporally ordered variables. PI performance is a good alternative approach for model assessment. Kong et al. (2018) highlighted that an "ideal" PI must be accurate and effective. First, PI should be accurate as the probability of the unknown quantity being contained in the PI should be close to a predetermined nominal level, $1 - \alpha$. Second, PI must be effective, namely, the interval should be sufficiently narrow and, hence, useful in locating the unknown quantity. Thus, we can compare the empirical coverage frequency of true values in the testing set and investigate the length and boundary of the PI to assess PI efficiency.

The rest of the paper is organized as follows. Section 2 introduces the nonparametric regression model for a locally stationary time series and provides an estimate of each component to construct multi-step-ahead PIs. Implementation details and simulation findings are reported in Sections 3 and 4, respectively. In Section 5, we illustrate the proposed method by applying it to air pollutant concentration data and comparing its performance with the seasonal ARIMA method.

2 | METHODOLOGY

Consider a real-valued time series dataset $\{Y_t\}_{t=1}^T$ spanning a long time interval, which usually contains complex nonstationary structures. To consider nonstationarity, we extended the nonparametric regression model for equally spaced designs to the time series setup. Specifically, the observed time series $\{Y_t\}_{t=1}^T$ is a realization from the following model:

$$Y_t = m(t/T) + \sigma(t/T)Z_t, \quad t = 1, \dots, T, \quad (1)$$

where $m(\cdot)$ represents a slowly varying function known as a trend component possessing some degree of smoothness and $\sigma^2(\cdot)$ is the variance function, allowing heteroscedasticity at different time points. Errors $\{Z_t\}_{t=1}^T$, assumed as stationary and weakly independent time series, satisfy $\mathbb{E}Z_t = 0$ and $\mathbb{E}Z_t^2 = 1$ for model identification and are fitted an autoregressive model with order p as follows:

$$Z_t = \sum_{k=1}^p \phi_k Z_{t-k} + \varepsilon_t, \quad t = p+1, \dots, T, \quad (2)$$

where the white noise $\{\varepsilon_t\}_{t=p+1}^T$ is independent and identically distributed (IID) with a mean of 0. Using this formulation, we can first separate the smooth trend from noisy stochastic errors via smoothing techniques.

2.1 | Estimating the trend function $m(\cdot)$

For model (1), B-spline is applied to approximate the trend function $m(\cdot)$. We provide a brief introduction to the B-spline smoothing method in the following.

B-spline is widely used in nonparametrics for its computational simplicity and derivable asymptotic theory (Cai et al., 2019; Cao et al., 2012, 2016; Gu et al., 2014; Gu & Yang, 2015; Liu & Yang, 2010, 2016; Song & Yang, 2009; Wang & Yang, 2007, 2009; Wang et al., 2020; Xue & Yang, 2006) for its applications in different scenarios. To describe the spline functions, we denote by $\{t_\ell\}_{\ell=1}^{J_s}$ a sequence of equally spaced points, $t_\ell = \ell / (J_s + 1)$, $1 \leq \ell \leq J_s$, $0 < t_1 < \dots < t_{J_s} < 1$, called interior knots, which divides the interval $[0, 1]$ into $(J_s + 1)$ equal subintervals $l_0 = [0, t_1]$, $l_\ell = [t_\ell, t_{\ell+1}]$, $\ell = 1, \dots, J_s - 1$, $l_{J_s} = [t_{J_s}, 1]$. For any positive integer p , let $t_{1-p} = \dots = t_0 = 0$ and $1 = t_{J_s+1} = \dots = t_{J_s+p}$ be auxiliary knots. Let $\mathcal{H}^{(p-2)} = \mathcal{H}^{(p-2)}[0, 1]$ be the polynomial spline space of order p on l_ℓ , $\ell = 0, \dots, J_s$, which consists of all $(p-2)$ times continuously differentiable functions on $[0, 1]$, which are polynomials of degree $(p-1)$ on subintervals l_ℓ , $\ell = 0, \dots, J_s$. We then denote by $\{B_{\ell,p}(x), 1 \leq \ell \leq J_s + p\}$ the p -th order B-spline basis functions of $\mathcal{H}^{(p-2)}$, hence $\mathcal{H}^{(p-2)} = \left\{ \sum_{\ell=1}^{J_s+p} \lambda_{\ell,p} B_{\ell,p}(x) \mid \lambda_{\ell,p} \in \mathbb{R}, x \in [0, 1] \right\}$.

We propose to estimate the trend function $m(\cdot)$ using the following formula:

$$\hat{m}(\cdot) = \arg \min_{g(\cdot) \in \mathcal{H}^{(p-2)}} \sum_{t=1}^T \{Y_t - g(t/T)\}^2.$$

The definition of $\hat{m}(\cdot)$ in (2.1) leads to

$$\hat{m}(\cdot) \equiv \sum_{\ell=1}^{J_s+p} \hat{\beta}_{\ell,p} \mathbf{B}_{\ell,p}(\cdot), \quad (3)$$

with coefficients $\{\hat{\beta}_{1,p}, \dots, \hat{\beta}_{J_s+p,p}\}^\top$, solving the following least-squares problem:

$$\{\hat{\beta}_{1,p}, \dots, \hat{\beta}_{J_s+p,p}\}^\top = \arg \min_{\{\beta_{1,p}, \dots, \beta_{J_s+p,p}\} \in \mathbb{R}^{J_s+p}} \sum_{t=1}^T \left\{ Y_t - \sum_{\ell=1}^{J_s+p} \beta_{\ell,p} \mathbf{B}_{\ell,p}(t/T) \right\}^2.$$

Applying elementary algebra, one obtains

$$\hat{m}(\cdot) = \mathbf{B}(\cdot)^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}, \quad (4)$$

where $\mathbf{Y} = (Y_1, \dots, Y_T)^\top$ and the $T \times (J_s + p)$ design matrix \mathbf{X} for spline regression is

$$\mathbf{X} = \{\mathbf{B}(1/T), \dots, \mathbf{B}(T/T)\}^\top,$$

with $\mathbf{B}(\cdot) = \{\mathbf{B}_{1,p}(\cdot), \dots, \mathbf{B}_{J_s+p,p}(\cdot)\}^\top$.

2.2 | Estimating the variance function $\sigma^2(\cdot)$

Variance function $\sigma^2(\cdot)$ in model (1) measures the heteroscedastic variation in the errors e_t : where $e_t = Y_t - m(t/T)$, and $t = 1, \dots, T$. Errors $\{e_t\}_{t=1}^T$ are unobservable as $m(\cdot)$ is unknown. Following Cai et al. (2019), we replace $m(\cdot)$ with $\hat{m}(\cdot)$ in (4) and propose a kernel estimator of $\sigma^2(\cdot)$ as follows:

$$\hat{\sigma}^2(x) = \frac{\sum_{t=1}^T K_h(t/T - x) \hat{e}_t^2}{\sum_{t=1}^T K_h(t/T - x)}, \quad (5)$$

where $\hat{e}_t = Y_t - \hat{m}(t/T)$, $h = h_T > 0$ is the bandwidth, and K is a kernel function with $K_h(u) = K(u/h)/h$.

2.3 | Autoregressive coefficients estimation

Parameters of interest in model (2) are autoregressive coefficients $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$. According to Eq. (8.1.1) of Brockwell and Davis (1991), they satisfy

$$\boldsymbol{\phi} = \boldsymbol{\Gamma}_p^{-1} \boldsymbol{\gamma}_p, \quad \boldsymbol{\Gamma}_p = \{\gamma(i-j)\}_{i,j=1}^p, \quad \boldsymbol{\gamma}_p = (\gamma(1), \dots, \gamma(p)),$$

wherein $\gamma(l) = \mathbb{E}(Z_t Z_{t+l})$, $l = 0, \pm 1, \pm 2, \dots$, representing the autocovariance function of $\{Z_t\}_{t=1}^T$. We denote $\hat{Z}_t = \hat{e}_t / \hat{\sigma}(t/T)$ and the sample autocovariance function as

$$\hat{\gamma}(l) = T^{-1} \sum_{t=1}^{T-l} \hat{Z}_t \hat{Z}_{t+l}, \quad 0 \leq l \leq T-1.$$

The classic Yule–Walker estimator of $\boldsymbol{\phi}$, a method of the moment estimator based on residuals $\{\hat{Z}_t\}_{t=1}^T$, is defined by

$$\hat{\phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p, \hat{\Gamma}_p = \{\hat{\gamma}(i-j)\}_{i,j=1}^p, \hat{\gamma}_p = (\hat{\gamma}(1), \dots, \hat{\gamma}(p)). \quad (6)$$

2.4 | Constructing PI for Y_{T+k}

Following Kong et al. (2018), the k -step-ahead linear predictor $\hat{Z}_{T+k}^{[k]}$ for Z_{T+k} , $k \geq 1$ based on $\{Z_t\}_{t=1}^T$ is defined recursively by

$$\hat{Z}_{T+k}^{[k]} = \phi_1 \hat{Z}_{T+k-1}^{[k-1]} + \dots + \phi_p \hat{Z}_{T+k-p}^{[k-p]}, \quad (7)$$

and satisfies

$$\hat{Z}_{T+k}^{[k]} = \phi_1^{[k]} Z_T + \dots + \phi_p^{[k]} Z_{T-p+1},$$

where the coefficient vector $\phi^{[k]} = (\phi_1^{[k]}, \dots, \phi_p^{[k]})^\top$ is a polynomial function g_k of $\phi = (\phi_1, \dots, \phi_p)^\top$, $\phi^{[k]} = g_k(\phi)$, where g_k is defined recursively by repeated applications of (7).

Using the Yule-Walker estimator $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)^\top$ of ϕ in (6), we obtain a plug-in estimate $\hat{\phi}^{[k]} = (\hat{\phi}_1^{[k]}, \dots, \hat{\phi}_p^{[k]})^\top = g_k(\hat{\phi})$ of $\phi^{[k]} = g_k(\phi)$. We denote $\hat{Z}_{T+k}^{[k]}$ as the data version of the linear predictor $\hat{Z}_{T+k}^{[k]}$:

$$\hat{Z}_{T+k}^{[k]} = \hat{\phi}_1^{[k]} \hat{Z}_T + \dots + \hat{\phi}_p^{[k]} \hat{Z}_{T-p+1}, \quad (8)$$

and $\hat{\varepsilon}_{T+k}^{[k]} = Z_{T+k} - \hat{Z}_{T+k}^{[k]}$ as the k -step-ahead prediction residuals. Let $F^{[k]}(x)$ denote the k -step-ahead prediction residual distribution and its α -th quantile as $q_\alpha^{[k]}$.

Furthermore, we propose an estimator $\hat{q}_{n,\alpha}^{[k]} = (\hat{F}^{[k]})^{-1}(\alpha) = \inf\{x : \hat{F}^{[k]}(x) \geq \alpha\}$ of $q_\alpha^{[k]}$ based on a two-step plug-in kernel distribution estimator (KDE) $\hat{F}^{[k]}(x)$ of $F^{[k]}(x)$:

$$\hat{F}^{[k]}(x) = \int_{-\infty}^x T^{-1} \sum_{t=k}^T \tilde{K}_{\tilde{h}}(u - \hat{\varepsilon}_t^{[k]}) du, x \in \mathbb{R}, \quad (9)$$

where $\tilde{h} = \tilde{h}_T > 0$ is the bandwidth, and \tilde{K} is a kernel function with $\tilde{K}_{\tilde{h}}(u) = \tilde{K}(u/\tilde{h})/\tilde{h}$, and $\hat{\varepsilon}_t^{[k]} = \hat{Z}_t - \hat{Z}_{t-k}^{[k]} = \hat{Z}_t - \hat{\phi}_1^{[k]} \hat{Z}_{t-k} - \dots - \hat{\phi}_p^{[k]} \hat{Z}_{t-p+1}$, $k \leq t \leq T$ are prediction residuals.

As both $m(\cdot)$ and $\sigma(\cdot)$ are slowly-varying functions, approximating the value of the trend function and the variance function at time point $T+k$ with their values at time point T , namely, $m(1)$ and $\sigma(1)$, respectively, when k was small, is reasonable. Combining the k -step-ahead predictor $\hat{Z}_{T+k}^{[k]}$ and its corresponding quantile estimator, the $(1-\alpha)$ -th PI for the k -step-ahead observation Y_{T+k} is constructed as

$$\left[\hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,\alpha/2}^{[k]} \right), \hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,1-\alpha/2}^{[k]} \right) \right]. \quad (10)$$

3 | IMPLEMENTATION

To realize the proposed method, we need to obtain estimates of unknown parts $m(\cdot)$, $\sigma(\cdot)$ and quantile $q_{n,\alpha/2}^{[k]}$ in (10).

Estimating the trend function $m(\cdot)$ mainly involves choosing the number of interior knots J_s and the spline order p . The number of interior knots J_s , often considered as an unknown tuning parameter, is crucial for spline smoothing as the spline fitting can be sensitive to knots selection. The spline estimator $\hat{m}(\cdot)$ is obtained from (3), with the number of interior knots considered as $J_s = \lceil cT^{1/4} \log \log T \rceil$, where c is a tuning constant and $\lceil a \rceil$ denotes the integer part of a . The default value of the order is $p = 4$, namely, the cubic spline.

To derive $\hat{\sigma}^2(x)$ in (5), the quartic kernel $K(u) = 15(1-u^2)^2 I\{|u| \leq 1\}/16$ is chosen with the bandwidth $h = ch_{rot} \times \log^{-1/2} T$, where c is a tuning constant and the rule-of-thumb bandwidth is

$$h_{rot} = \left[\frac{35 \sum_{t=1}^T \left\{ \hat{\epsilon}_t^2 - \sum_{k=0}^4 \hat{a}_k(t/T)^k \right\}^2}{n \sum_{t=1}^T \left\{ 2\hat{a}_2 + 6\hat{a}_3(t/T) + 12\hat{a}_4(t/T)^2 \right\}^2} \right]^{1/5}, \quad (11)$$

in which $(\hat{a}_k)_{k=0}^4 = \operatorname{argmin}_{(a_k)_{k=0}^4 \in \mathbb{R}^5} \sum_{t=1}^T \left\{ \hat{\epsilon}_t^2 - \sum_{k=0}^4 a_k(t/n)^k \right\}^2$. We have found in extensive simulations in which $J_s = \lceil 6T^{1/4} \log \log T \rceil$ and $h = 0.2h_{rot} \times \log^{-1/2} T$ work quite well and are what we recommended.

Estimating the AR time series (2), which includes lag selection and parameter estimation, is conducted using \hat{Z}_t instead of Z_t . The order p is determined by the Akaike information criterion (AIC). After acquiring the Yule–Walker estimator of $\hat{\phi} = (\hat{\phi}_1, \dots, \hat{\phi}_p)^\top$, we can obtain $\hat{\phi}^{[k]}$ by the following recursive formula:

$$\hat{\phi}_m^{[k]} = \hat{\phi}_1^{[k-1]} \hat{\phi}_m + \hat{\phi}_{m+1}^{[k-1]}, \quad 1 \leq m \leq p-1, \quad (12)$$

$$\hat{\phi}_p^{[k]} = \hat{\phi}_1^{[k-1]} \hat{\phi}_p, \quad (13)$$

with $\hat{\phi}_m^{[0]} = \hat{\phi}_m$ for $m = 1, \dots, p$.

To estimate quantile $a_\alpha^{[k]}$, the same kernel $\tilde{K}(u) = 15(1-u^2)^2 I\{|u| \leq 1\}/16$ is used in (9). The bandwidth \tilde{h} is denoted as $\tilde{h} = (4/3T)^{1/5} \hat{s}$, where \hat{s}^2 is the sample variance. Finally, the $(1-\alpha)$ -th PI for future observations Y_{T+k} is constructed as (10). Details of the computational algorithm are provided in the following table.

TABLE 1 The 95% and 90% PIs' average length (inside the parentheses) and coverage frequencies of future points over 1000 replications with normal distribution errors $\{\epsilon_t\}_{t=2}^{T+5}$

Point	T	95% proposed PI	95% normal PI	90% proposed PI	90% normal PI
Y_{T+1}	1000	0.841(1.715)	0.837(1.705)	0.775(1.433)	0.776(1.431)
	2000	0.880(1.831)	0.876(1.819)	0.824(1.529)	0.834(1.527)
	4000	0.897(1.886)	0.896(1.876)	0.836(1.577)	0.835(1.574)
	8000	0.911(1.915)	0.909(1.906)	0.850(1.602)	0.852(1.600)
	16,000	0.944(1.974)	0.944(1.969)	0.904(1.654)	0.904(1.652)
	32,000	0.949(2.001)	0.945(2.002)	0.906(1.678)	0.902(1.676)
	64,000	0.945(2.012)	0.924(2.009)	0.904(1.686)	0.905(1.685)
	Y_{T+2}	1000	0.800(2.036)	0.803(2.071)	0.735(1.736)
2000		0.854(2.248)	0.858(2.262)	0.787(1.902)	0.776(1.898)
4000		0.891(2.360)	0.890(2.362)	0.844(1.986)	0.840(1.982)
8000		0.902(2.421)	0.902(2.419)	0.851(2.034)	0.846(2.030)
16,000		0.945(2.512)	0.940(2.509)	0.912(2.108)	0.910(2.105)
32,000		0.941(2.553)	0.928(2.557)	0.895(2.144)	0.891(2.141)
64,000		0.934(2.573)	0.932(2.571)	0.894(2.157)	0.895(2.155)
Y_{T+3}		1000	0.751(2.153)	0.763(2.226)	0.672(1.858)
	2000	0.846(2.429)	0.852(2.469)	0.750(2.074)	0.758(2.072)
	4000	0.899(2.587)	0.902(2.604)	0.842(2.190)	0.838(2.185)
	8000	0.900(2.678)	0.902(2.684)	0.843(2.257)	0.838(2.252)
	16,000	0.946(2.793)	0.947(2.793)	0.884(2.348)	0.884(2.344)
	32,000	0.931(2.846)	0.924(2.852)	0.877(2.400)	2.858(2.388)
	64,000	0.936(2.870)	0.932(2.868)	0.901(2.408)	0.902(2.406)
	Y_{T+5}	1000	0.682(2.209)	0.701(2.329)	0.612(1.930)
2000		0.775(2.554)	0.778(2.629)	0.728(2.202)	0.726(2.206)
4000		0.884(2.769)	0.892(2.807)	0.821(2.361)	0.817(2.356)
8000		0.905(2.903)	0.910(2.919)	0.834(2.456)	0.834(2.450)
16,000		0.926(3.046)	0.929(3.053)	0.882(2.568)	0.886(2.562)
32,000		0.930(3.116)	0.925(3.126)	0.863(2.631)	0.855(2.618)
64,000		0.934(3.149)	0.935(3.148)	0.880(2.644)	0.880(2.641)

Algorithm: Construction of the $(1 - \alpha)$ -th PI for future observation Y_{T+k}

Input: Data $\{Y_t\}_{t=1}^T$

Step 1: Estimate the trend function $\hat{m}(\cdot)$ based on (4) with $J_s = \lceil 6T^{1/4} \log \log T \rceil + 1$ and $p = 4$.

Step 2: Denote $\hat{\varepsilon}_t = Y_t - \hat{m}(t/T)$ and compute the bandwidth $h = 0.2h_{rot} \times \log^{-1/2} T$ with h_{rot} in (11).

Step 3: Estimate the variance function based on (5) with the quartic kernel $K(u)$.

Step 4: Denote $\hat{Z}_t = \hat{\varepsilon}_t / \hat{\sigma}(t/T)$ and obtain the auto-regressive coefficients of ϕ in (2) based on residuals \hat{Z}_t through (6).

Step 5: Compute $\hat{\phi}^{[k]}$ by the recursive formula (12) and obtain the data version predictor $\hat{Z}_{T+k}^{[k]}$ by (8) and k -step-ahead prediction residuals $\hat{\varepsilon}_t^{[k]}$, $k \leq t \leq T$.

Step 6: Calculate the cumulative distribution function of k -step-ahead prediction residuals $\hat{\varepsilon}_t^{[k]}$ from (9) and obtain its α -th quantile $q_\alpha^{[k]}$.

Step 7: Compute the $(1 - \alpha)$ -th PI for k -step-ahead observation Y_{T+k} :

$$\left[\hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,\alpha/2}^{[k]} \right), \hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,1-\alpha/2}^{[k]} \right) \right].$$

Output: The $(1 - \alpha)$ -th PI for k -step-ahead observation Y_{T+k} .

TABLE 2 The 95% and 90% PIs' average length (inside the parentheses) and coverage frequencies of future points over 1000 replications with mixture normal distribution errors $\{\varepsilon_t\}_{t=2}^{T+5}$

Point	T	95% proposed PI	95% normal PI	90% proposed PI	90% normal PI
Y_{T+1}	1000	0.801(2.357)	0.820(2.419)	0.735(2.002)	0.740(2.030)
	2000	0.866(2.451)	0.877(2.526)	0.822(2.090)	0.828(2.120)
	4000	0.902(2.566)	0.911(2.655)	0.847(2.195)	0.850(2.228)
	8000	0.915(2.619)	0.934(2.723)	0.866(2.248)	0.870(2.285)
	16,000	0.954(2.681)	0.959(2.794)	0.908(2.306)	0.912(2.345)
	32,000	0.941(2.220)	0.942(2.226)	0.896(1.879)	0.896(1.878)
	64,000	0.950(2.224)	0.951(2.242)	0.889(1.889)	0.888(1.888)
	Y_{T+2}	1000	0.799(2.855)	0.811(2.938)	0.744(2.452)
2000		0.822(3.076)	0.828(3.136)	0.742(2.629)	0.750(2.632)
4000		0.876(3.292)	0.881(3.344)	0.822(2.807)	0.821(2.806)
8000		0.915(3.406)	0.918(3.455)	0.844(2.902)	0.844(2.900)
16,000		0.920(3.509)	0.924(3.560)	0.881(2.990)	0.880(2.987)
32,000		0.923(2.841)	0.922(2.841)	0.894(2.402)	0.891(2.399)
64,000		0.946(2.873)	0.945(2.876)	0.882(2.412)	0.882(2.415)
Y_{T+3}		1000	0.734(3.031)	0.773(3.156)	0.657(2.627)
	2000	0.790(3.338)	0.804(3.422)	0.738(2.867)	0.736(2.871)
	4000	0.867(3.628)	0.872(3.687)	0.823(3.098)	0.824(3.094)
	8000	0.928(3.790)	0.928(3.834)	0.854(3.224)	0.850(3.218)
	16,000	0.937(3.923)	0.939(3.963)	0.886(3.332)	0.886(3.326)
	32,000	0.923(3.168)	0.922(3.171)	0.897(2.680)	0.898(2.676)
	64,000	0.944(3.208)	0.946(3.212)	0.892(2.699)	0.89(2.696)
	Y_{T+5}	1000	0.698(3.116)	0.718(3.300)	0.641(2.733)
2000		0.754(3.516)	0.763(3.642)	0.706(3.045)	0.702(3.056)
4000		0.847(3.895)	0.854(3.976)	0.784(3.341)	0.785(3.337)
8000		0.900(4.118)	0.909(4.171)	0.842(3.509)	0.840(3.500)
16,000		0.938(4.293)	0.938(4.332)	0.855(3.644)	0.848(3.635)
32,000		0.931(3.470)	0.930(3.476)	0.871(2.938)	0.875(2.933)
64,000		0.943(3.522)	0.944(3.525)	0.908(2.963)	0.906(2.960)

4 | SIMULATION STUDIES

In this section, we describe Monte Carlo simulations to examine the finite sample performance of the proposed method. Our data are generated from the following model:

$$Y_t = m\left(\frac{t}{T+5}\right) + \sigma\left(\frac{t}{T+5}\right)Z_t, t = 1, \dots, T+5, \quad (14)$$

$$Z_t = 0.8Z_{t-1} + \varepsilon_t, t = 2, \dots, T+5. \quad (15)$$

We set $m(x) = 5 + 4\cos(2.5\pi x)$ and $\sigma(x) = (5 - \exp(-x))/(5 + \exp(-x))$ in (14). The ID errors $\{\varepsilon_t\}_{t=2}^N$ follow three different distributions: the normal distribution $N(0, 0.6^2)$, the mixture normal distribution $0.5N(-0.5, 0.6^2) + 0.5N(0.5, 0.6^2)$ and the Laplace distribution $Laplace(0, 0.6/\sqrt{2})$, which can ensure $\mathbb{E}Z_t^2 = 1$.

The sample size $(T+5)$ is denoted as 1005, 2005, 4005, 8005 and 16,005, and realizations $\{Z_t\}_{t=-999}^{T+5}$ of size $T+1005$ are generated from (15), with the first 1000 values deleted to guarantee the strict stationarity of $\{Z_t\}_{t=1}^{T+5}$.

The dataset $\{Y_t\}_{t=1}^{T+5}$ is divided into a testing set $\{Y_t\}_{t=T+1}^{T+5}$ and a training set $\{Y_t\}_{t=1}^T$. In the following, we construct 90% and 95% the proposed and normal PIs and compare their performance over 1000 replications.

TABLE 3 The 95% and 90% PIs' average length (inside the parentheses) and coverage frequencies of future points over 1000 replications with Laplace distribution errors $\{\varepsilon_t\}_{t=2}^{T+5}$

Point	T	95% proposed PI	95% normal PI	90% proposed PI	90% normal PI
Y_{T+1}	1000	0.887(1.801)	0.869(1.654)	0.814(1.422)	0.801(1.388)
	2000	0.901(1.934)	0.882(1.772)	0.834(1.506)	0.832(1.487)
	4000	0.926(2.016)	0.911(1.849)	0.876(1.558)	0.874(1.552)
	8000	0.953(2.100)	0.934(1.933)	0.883(1.619)	0.885(1.622)
	16,000	0.950(2.140)	0.931(1.973)	0.900(1.647)	0.902(1.655)
	32,000	0.947(2.163)	0.934(1.997)	0.896(1.654)	0.896(1.666)
	64,000	0.946(2.156)	0.934(1.992)	0.892(1.664)	0.888(1.678)
Y_{T+2}	1000	0.823(2.070)	0.812(2.017)	0.765(1.722)	0.758(1.693)
	2000	0.887(2.307)	0.872(2.207)	0.826(1.881)	0.816(1.852)
	4000	0.912(2.455)	0.897(2.331)	0.854(1.975)	0.854(1.956)
	8000	0.932(2.584)	0.918(2.453)	0.878(2.066)	0.877(2.059)
	16,000	0.947(2.647)	0.932(2.514)	0.899(2.109)	0.898(2.110)
	32,000	0.925(2.682)	0.910(2.550)	0.892(2.122)	0.894(2.128)
	64,000	0.948(2.679)	0.942(2.548)	0.890(2.147)	0.888(2.137)
Y_{T+3}	1000	0.767(2.155)	0.768(2.172)	0.694(1.835)	0.690(1.823)
	2000	0.858(2.462)	0.848(2.413)	0.797(2.048)	0.790(2.025)
	4000	0.902(2.660)	0.895(2.573)	0.842(2.178)	0.844(2.159)
	8000	0.910(2.826)	0.898(2.723)	0.847(2.294)	0.851(2.285)
	16,000	0.951(2.906)	0.94(2.799)	0.894(2.350)	0.882(2.349)
	32,000	0.920(2.954)	0.916(2.845)	0.877(2.368)	0.880(2.373)
	64,000	0.947(2.953)	0.942(2.845)	0.901(2.388)	0.902(2.397)
Y_{T+5}	1000	0.704(2.195)	0.710(2.279)	0.617(1.898)	0.618(1.912)
	2000	0.807(2.563)	0.806(2.575)	0.740(2.170)	0.738(2.161)
	4000	0.860(2.819)	0.857(2.777)	0.808(2.345)	0.806(2.331)
	8000	0.904(3.030)	0.892(2.962)	0.845(2.495)	0.844(2.486)
	16,000	0.915(3.139)	0.901(3.060)	0.852(2.569)	0.852(2.568)
	32,000	0.947(3.201)	0.938(3.119)	0.864(2.596)	0.865(2.601)
	64,000	0.934(3.207)	0.929(3.123)	0.888(2.621)	0.889(2.631)

90% the proposed PI: $\left[\hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,0.05}^{[k]} \right), \hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,0.95}^{[k]} \right) \right]$ and 95% of the proposed PI: $\left[\hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,0.025}^{[k]} \right), \hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + \hat{q}_{n,0.975}^{[k]} \right) \right]$ given by (10).
 90% the normal PI: $\left[\hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} - 1.64\hat{s}(k) \right), \hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + 1.64\hat{s}(k) \right) \right]$ and 95% the normal PI: $\left[\hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} - 1.96\hat{s}(k) \right), \hat{m}(1) + \hat{\sigma}(1) \left(\hat{Z}_{T+k}^{[k]} + 1.96\hat{s}(k) \right) \right]$, where $\hat{s}(k)$ is standard deviation of $\hat{\varepsilon}_t^{[k]}$ based on the naive assumption that $F^{[k]}(x)$ is normal.

Following one reviewer's suggestion, we also compare the proposed PI with the seasonal ARIMA PI, which is obtained by fitting the classical ARIMA model (16), with the seasonal lag and orders selected by the Bayesian information criterion (BIC) automatically.

Tables 1–3 show the above 95% and 90% PIs' average length and coverage frequencies of Y_{T+k} for $k = 1, 2, 3, 5$, namely, the percentage out of 1000 replications of the true value of Y_{T+k} being contained in the PI. Under the normal distribution errors $\{\varepsilon_t\}_{t=2}^{T+5}$, no significant difference can be seen in coverage frequencies between the proposed and normal PIs; both intervals approach the nominal confidence level as the sample size T increases (see Table 1). This is reasonable as the distribution of the errors $\{\varepsilon_t\}_{t=2}^{T+5}$ is inherently normal. The lengths of two PIs are also very similar. One can find that the coverage frequency of Y_{T+1} and Y_{T+2} performs better than that of Y_{T+3} and Y_{T+5} . It is reasonable as forecasts with fewer steps always possess higher accuracy.

From Tables 2 and 3, we note that the coverage frequency of normal PI is always greater than that of the proposed PI in the case of mixture normal distribution errors $\{\varepsilon_t\}_{t=2}^N$. Conversely, it is lower in the case of Laplace distribution errors. In both scenarios, coverage frequency of the proposed PI is closer to the predetermined level as T increases. This indicates that the length of the normal PI may be too wide in the mixture normal case and too narrow in the Laplace case. This partly explains why coverage frequencies of normal PI are lower or higher than the nominal

TABLE 4 The 95% and 90% ARIMA PIs' average length (inside the parentheses) and coverage frequencies of future points over 1000 replications with three different distribution errors $\{\varepsilon_t\}_{t=2}^{T+5}$: Normal, mixture normal and Laplace, respectively

Point	T	95%, normal	95%, mixture normal	95%, Laplace	90%, normal	90%, mixture normal	90%, Laplace
Y_{T+1}	1000	0.925(1.876)	0.932(2.638)	0.924(1.871)	0.866(1.574)	0.851(2.214)	0.876(1.570)
	2000	0.920(1.860)	0.926(2.622)	0.935(1.861)	0.846(1.561)	0.864(2.201)	0.907(1.562)
	4000	0.930(1.850)	0.927(2.612)	0.915(1.851)	0.865(1.553)	0.855(2.192)	0.866(1.553)
	8000	0.916(1.844)	0.933(2.605)	0.911(1.845)	0.865(1.548)	0.863(2.186)	0.868(1.548)
	16,000	0.915(1.840)	0.935(2.672)	0.910(1.840)	0.849(1.544)	0.860(2.242)	0.874(1.544)
	32,000	0.922(1.876)	0.926(2.096)	0.920(1.876)	0.864(1.573)	0.866(1.759)	0.874(1.574)
	64,000	0.918(1.873)	0.897(2.094)	0.920(1.873)	0.867(1.572)	0.846(1.757)	0.868(1.572)
Y_{T+2}	1000	0.942(2.460)	0.924(3.447)	0.930(2.454)	0.872(2.064)	0.868(2.892)	0.883(2.059)
	2000	0.927(2.419)	0.921(3.398)	0.928(2.420)	0.876(2.030)	0.847(2.852)	0.879(2.031)
	4000	0.939(2.392)	0.932(3.372)	0.927(2.393)	0.886(2.008)	0.871(2.830)	0.878(2.009)
	8000	0.933(2.376)	0.923(3.353)	0.936(2.377)	0.872(1.995)	0.863(2.814)	0.894(1.995)
	16,000	0.897(2.365)	0.919(3.501)	0.925(2.366)	0.833(1.985)	0.845(2.938)	0.866(1.985)
	32,000	0.932(2.436)	0.908(2.721)	0.904(2.437)	0.862(2.042)	0.876(2.284)	0.886(2.044)
	64,000	0.914(2.428)	0.901(2.715)	0.936(2.428)	0.856(2.037)	0.842(2.562)	0.884(2.038)
Y_{T+3}	1000	0.928(2.810)	0.929(3.920)	0.931(2.803)	0.879(2.358)	0.861(3.290)	0.883(2.352)
	2000	0.935(2.742)	0.927(3.839)	0.923(2.742)	0.877(2.301)	0.858(3.222)	0.880(2.302)
	4000	0.920(2.697)	0.917(3.796)	0.924(2.699)	0.851(2.263)	0.858(3.186)	0.876(2.265)
	8000	0.926(2.671)	0.917(3.765)	0.934(2.672)	0.866(2.242)	0.852(3.159)	0.886(2.242)
	16,000	0.905(2.653)	0.924(3.999)	0.920(2.653)	0.850(2.226)	0.848(3.356)	0.860(2.227)
	32,000	0.904(2.746)	0.900(3.065)	0.906(2.749)	0.846(2.300)	0.862(2.574)	0.860(2.305)
	64,000	0.911(2.731)	0.908(3.054)	0.920(2.732)	0.849(2.291)	0.866(2.562)	0.874(2.293)
Y_{T+5}	1000	0.940(3.222)	0.920(4.452)	0.948(3.214)	0.889(2.704)	0.872(3.373)	0.897(2.697)
	2000	0.943(3.093)	0.935(4.305)	0.918(3.096)	0.870(2.595)	0.856(3.613)	0.875(2.598)
	4000	0.929(3.013)	0.930(4.228)	0.924(3.016)	0.874(2.529)	0.861(3.549)	0.878(2.531)
	8000	0.931(2.967)	0.924(4.174)	0.925(2.968)	0.874(2.490)	0.860(3.503)	0.880(2.491)
	16,000	0.923(2.936)	0.923(4.637)	0.914(2.936)	0.851(2.464)	0.858(3.891)	0.863(2.464)
	32,000	0.910(3.037)	0.910(3.382)	0.928(3.045)	0.836(2.537)	0.830(2.842)	0.846(2.549)
	64,000	0.915(2.993)	0.908(3.348)	0.918(2.996)	0.856(2.512)	0.864(2.807)	0.860(2.515)

level. Such a difference in length provides strong evidence for using the proposed PI instead of the naive normal PI, in addition to coverage frequencies discussed above.

From Table 4, the coverage frequency of the seasonal ARIMA PIs cannot approach the nominal level even when the sample size T is quite large. Moreover, the average length of the seasonal ARIMA PI is much wider than that of the proposed PI in small sample size scenarios (see Tables 1–3). Thus, our proposed method exhibits superior performance compared to the seasonal ARIMA method.

5 | REAL DATA ANALYSIS

The dataset used in the paper comprises daily air pollutant concentrations in Xi'an between January 1, 2013, and July 31, 2020, for 8 years and 30 seasons. Six major air pollutants concentrations measured in tons per square kilometre (t/km^2) are CO, NO₂, O₃, PM₁₀, PM_{2.5} and SO₂. The Xi'an Environmental Monitoring Center and China National Environmental Monitoring Center provided the dataset. There are 2769 observations for each pollutant, and we remove invalid observations that only account for a small proportion, using the remainder as our dataset. Records of each pollutant concentration were split into a testing set (the last five observations) and a training set (other observations).

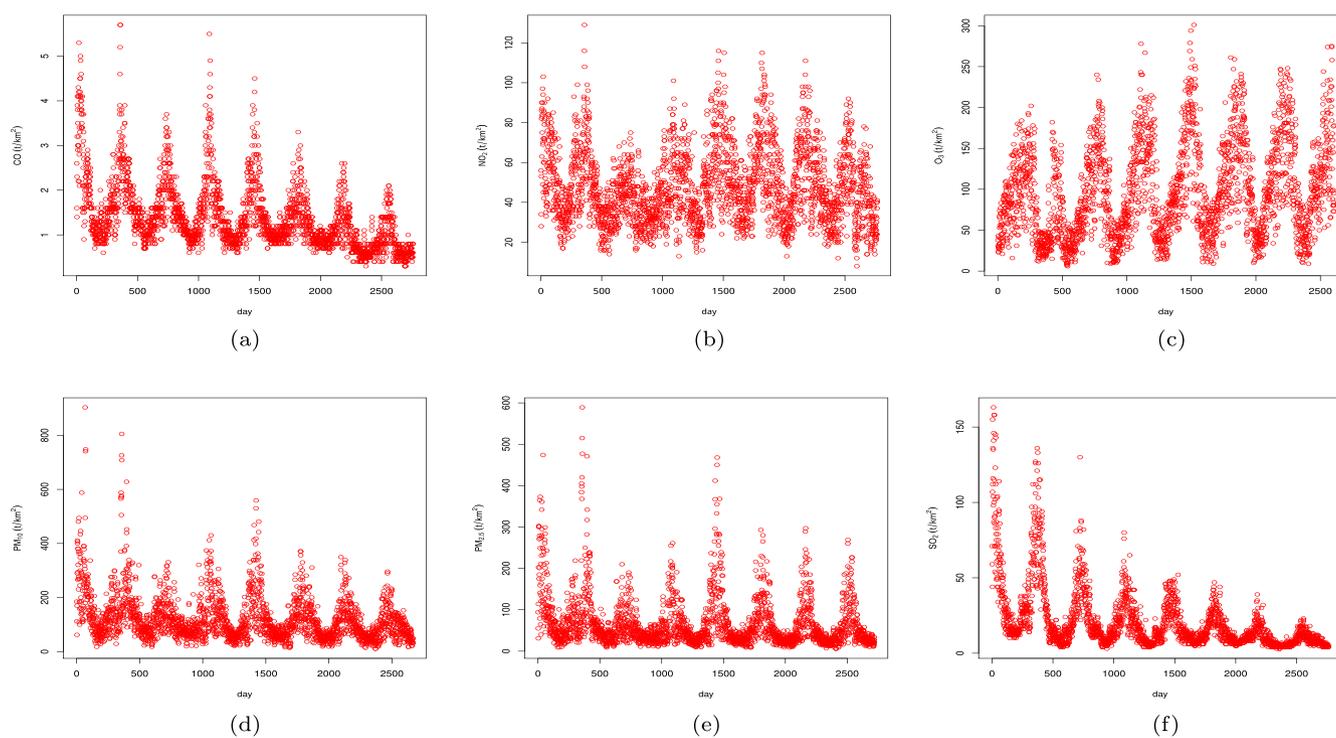


FIGURE 1 Scatterplot of each air pollutant concentration

TABLE 5 Descriptive statistics of each air pollutant concentration

Pollutant	CO	NO ₂	O ₃	PM ₁₀	PM _{2.5}	SO ₂
Minimum	0.3	8	6	11	6	3
Maximum	5.7	129	301	903	589	163
Mean	1.38	47.40	96.31	120.77	65.65	20.53
Std. dev.	0.73	18.37	56.39	85.39	60.59	20.56
Q1	0.9	33	50	65	29	8
Median	1.2	44	86	97	45	13
Q3	1.6	59	136	148.25	77	25
Skewness	1.71	0.73	0.65	2.49	2.78	2.79
Kurtosis	7.12	3.24	2.75	14.0	14.1	13.0

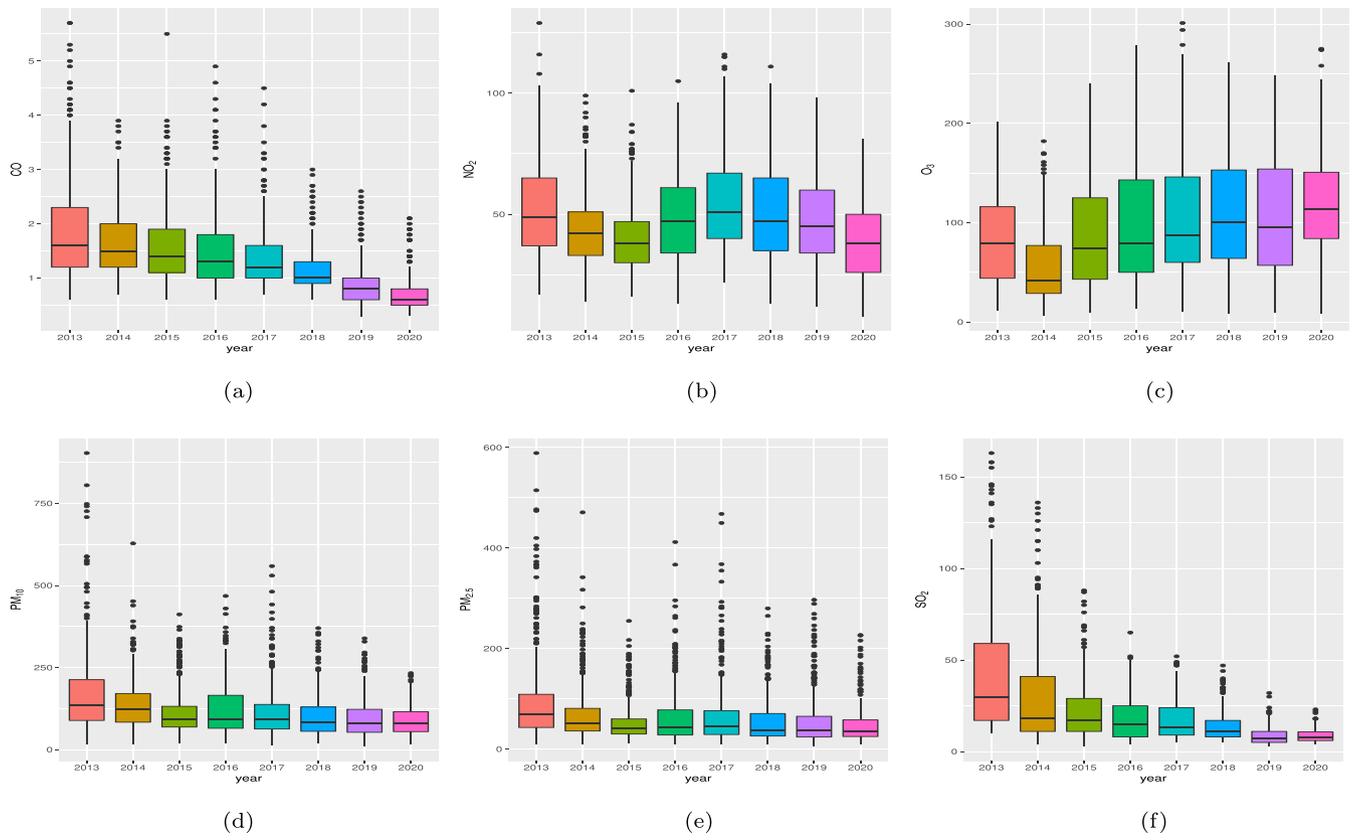


FIGURE 2 Box plot of each pollutant daily concentration. Black bars inside the boxes represent medians

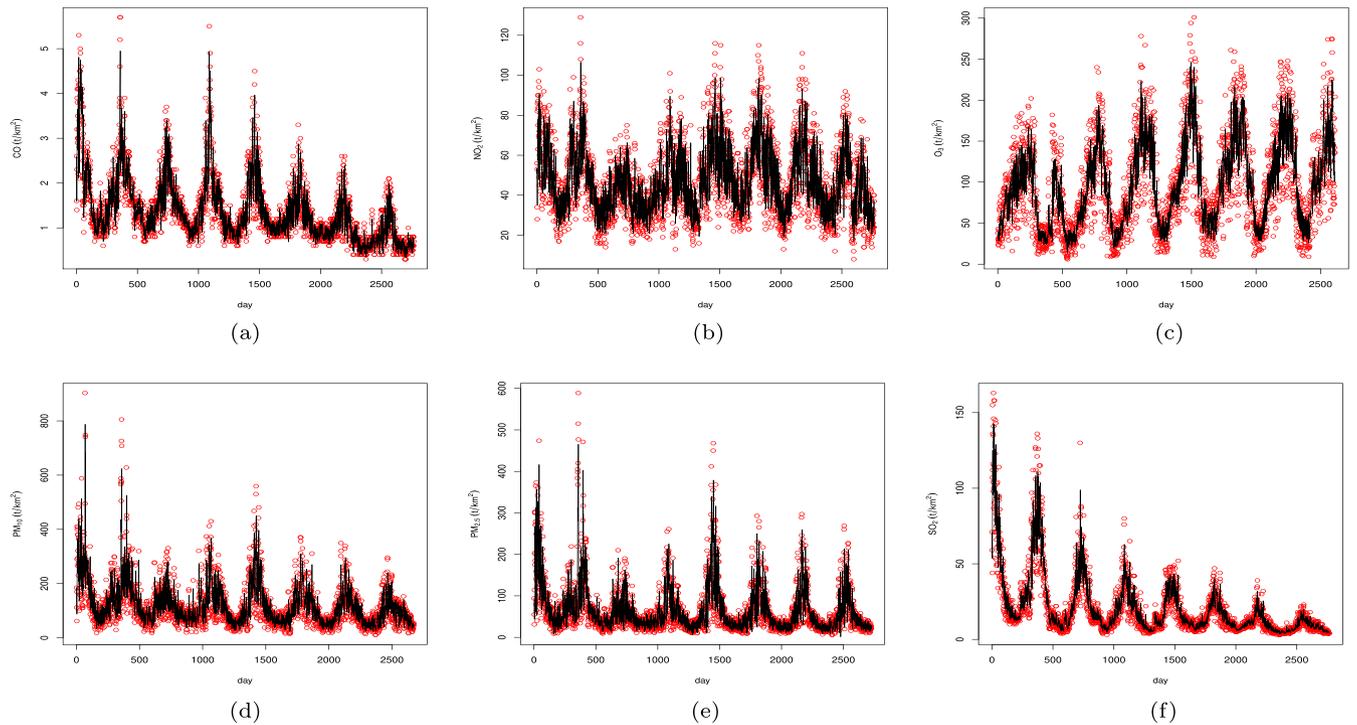


FIGURE 3 Fitted values of each air pollutant concentration with the seasonal ARIMA model

5.1 | Pre-analysis exploration

Figure 1 shows the scatterplot of six pollutant concentration records. For each air pollutant concentration, time trend and cyclic features are exhibited; hence, these are not stationary. Further augmented Dickey–Fuller (ADF) test for stationarity supports our findings. Table 5 gives brief descriptive statistics of each pollutant concentration: minimum, maximum, mean, standard deviation, quantile, skewness and kurtosis. Table 5 shows that higher values of skewness are those of PM_{10} , $PM_{2.5}$ and SO_2 , corresponding to rapid increases in the data, as presented in Figure 1. The kurtosis of the three pollutants is also large, relating to existing discontinuities in the data. Box plots are shown in Figure 2. O_3 levels in the warm season are higher than those in the cold season, whereas concentrations of the other five pollutants are most severe in winter. Therefore, all pollutant concentrations, except O_3 , show a decreasing trend.

5.2 | Forecasts of air pollutants concentration from seasonal ARIMA model

The general seasonal ARIMA is denoted as $ARIMA(p, d, q)(P, D, Q)_s$, where p is the number of parameters describing the autoregressive process, d is the order of differencing, q is the number of parameters for the moving average process, P is the number of seasonal autoregressive terms, D is the order of seasonal differencing, Q is the number of seasonal moving average terms and s is the number of periods in each season. Given air pollutant concentration observations $\{Y_t\}_{t=1}^T$, the seasonal ARIMA model can be written as

TABLE 6 95% the ARIMA PI and the proposed PI for future observations of each air pollutant

Pollutant	Future observation	ARIMA PI	Proposed PI
CO	Y_{T+1}	[0.0276, 1.211]	[0.555, 0.812]
	Y_{T+2}	[-0.057, 1.476]	[0.627, 0.888]
	Y_{T+3}	[-0.084, 1.533]	[0.632, 0.917]
	Y_{T+4}	[-0.107, 1.586]	[0.638, 0.950]
	Y_{T+5}	[-0.127, 1.635]	[0.606, 0.932]
NO_2	Y_{T+1}	[5.415, 48.356]	[16.881, 28.469]
	Y_{T+2}	[1.250, 55.604]	[16.312, 27.937]
	Y_{T+3}	[0.372, 57.816]	[15.745, 28.326]
	Y_{T+4}	[0.213, 58.584]	[15.286, 29.604]
	Y_{T+5}	[0.192, 59.012]	[16.085, 31.090]
O_3	Y_{T+1}	[70.028, 201.505]	[85.308, 190.703]
	Y_{T+2}	[47.000, 196.998]	[67.538, 174.426]
	Y_{T+3}	[36.341, 194.824]	[71.650, 188.302]
	Y_{T+4}	[32.698, 196.262]	[82.701, 205.975]
	Y_{T+5}	[38.401, 206.896]	[67.020, 195.309]
PM_{10}	Y_{T+1}	[-42.096, 152.525]	[36.663, 59.536]
	Y_{T+2}	[-69.120, 179.151]	[35.720, 58.666]
	Y_{T+3}	[-74.391, 184.265]	[33.157, 58.737]
	Y_{T+4}	[-74.739, 186.722]	[32.198, 60.000]
	Y_{T+5}	[-75.045, 189.103]	[32.410, 62.125]
$PM_{2.5}$	Y_{T+1}	[-43.433, 85.431]	[20.166, 33.182]
	Y_{T+2}	[-61.659, 110.086]	[24.184, 37.168]
	Y_{T+3}	[-64.626, 119.236]	[24.533, 38.780]
	Y_{T+4}	[-64.801, 121.879]	[23.427, 39.109]
	Y_{T+5}	[-65.069, 122.987]	[23.756, 40.127]
SO_2	Y_{T+1}	[-10.659, 22.008]	[4.049, 5.419]
	Y_{T+2}	[-13.846, 25.801]	[3.697, 5.068]
	Y_{T+3}	[-14.944, 27.342]	[3.446, 4.923]
	Y_{T+4}	[-15.506, 28.592]	[3.411, 5.052]
	Y_{T+5}	[-16.301, 30.210]	[3.472, 5.194]

$$\phi_p(L)\Phi_P(L^S)\nabla^d\nabla_s^D Y_t = \theta_q(L)\Theta_Q(L^S)\varepsilon_t, \tag{16}$$

where L is the lag distance operator satisfying $L^d Y_t = Y_{t-d}$, and ∇ is the differencing operator satisfying $\nabla = 1 - L$ and $\nabla_s = 1 - L^S$, ε_t is the white noise satisfying $\varepsilon_t \sim WN(0, \sigma^2)$, ϕ and Φ include non-seasonal and seasonal autoregressive parameters, θ and Θ include the non-seasonal and seasonal moving average parameters with $\phi(L) = 1 - \sum_{k=1}^P \phi_k L^k$, $\Phi(L^S) = 1 - \sum_{k=1}^P \phi_k (L^S)^k$, $\theta(L) = 1 - \sum_{k=1}^Q \theta_k L^k$ and $\Theta(L^S) = 1 - \sum_{k=1}^Q \theta_k (L^S)^k$.

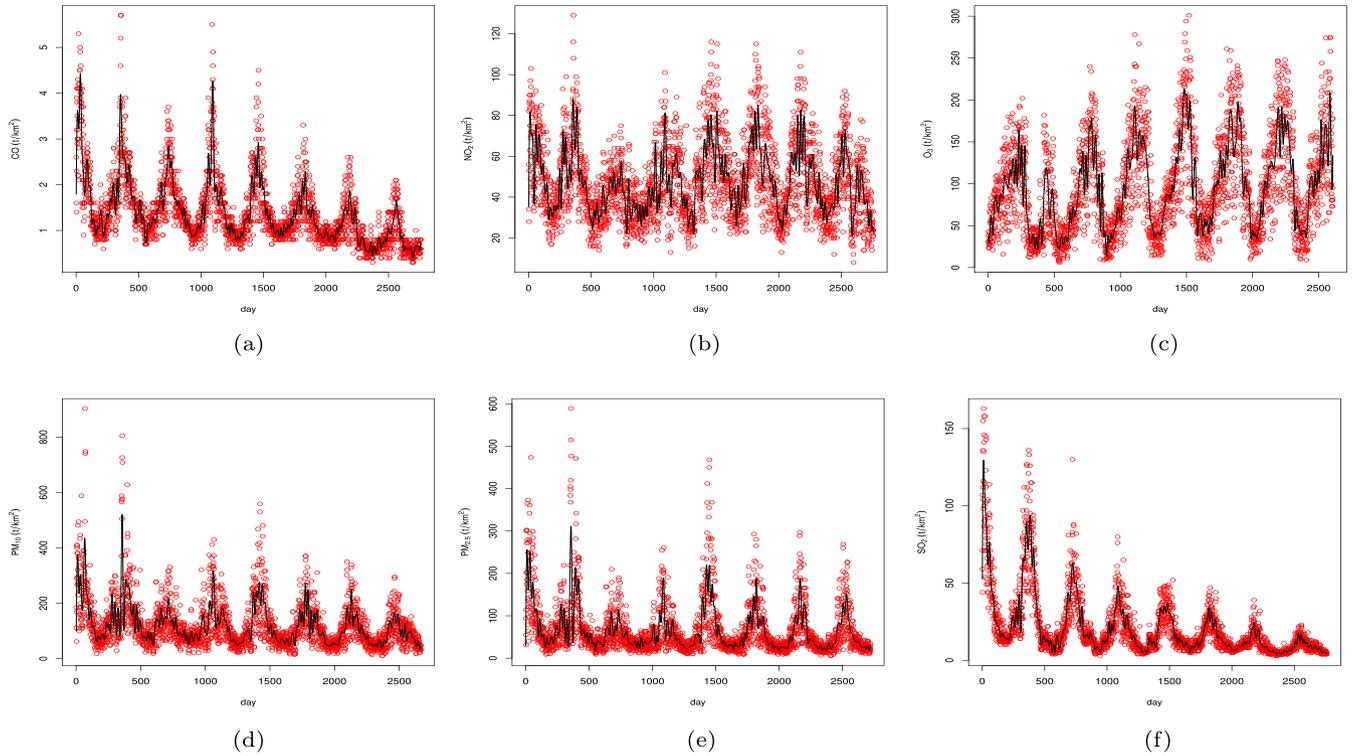


FIGURE 4 Each air pollutant concentration with its trend function estimator $\hat{m}(\cdot)$ (solid line)

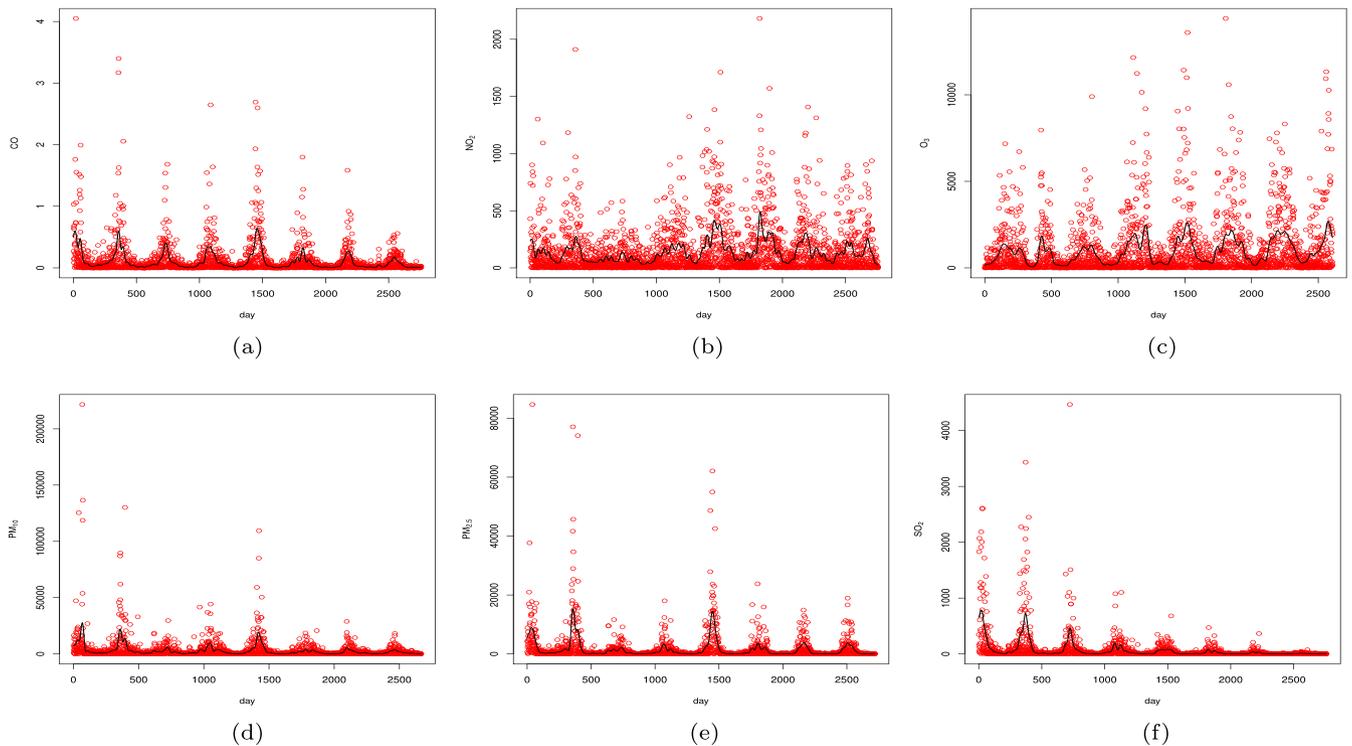


FIGURE 5 The scatterplot of $\hat{\varepsilon}_t^2$ for each pollutant, its variance function estimator $\hat{\sigma}^2(\cdot)$ (solid line)

Considering cyclic recurrences in the data, a seasonal ARIMA model with a lag $s = 365$ is applied for each air pollutant concentration, respectively. We specify the orders p , d , q , P , D and Q using the BIC and then estimate coefficients by the least-squares method. Figure 3 shows the estimation results with the original points in red and fitted values in black. Corresponding 95% PIs up to five steps ahead are displayed in the third column of Table 6.

The seasonal ARIMA model may perform well in fitting, but usually has a large predicting variance and much smaller lower prediction bound, sometimes even being a negative number. This is clearly unrealistic as the concentration value of air pollutants is always positive. Additionally, some PIs in Table 6 are so wide that they almost cover the entire range of data. Hence, all data points fall into the PI, making it lose effect in statistical inference.

5.3 | Forecasts of air pollutants concentration by the proposed method

We first separate the trend function and estimate the variance function in (1) via nonparametric regression. Figure 4 depicts the smooth estimate $\hat{m}(\cdot)$ for each pollutant concentrations in the training dataset. The estimator perfectly captures the pattern and reflects overall trend of air pollutant concentration data. Compared with the estimator in the seasonal ARIMA model, it is not that noisy and fluctuant. The variance estimator of each air pollutant concentration $\hat{\sigma}^2(\cdot)$ is displayed in Figure 5.

After fitting the autoregressive model of errors and obtaining the quantile of fitted residuals, we established PIs of up to five steps ahead for each air pollutant. The fourth column of Table 6 exhibits PIs for future concentrations Y_{T+1}, \dots, Y_{T+5} . Clearly, these PIs are much narrower than the derived counterparts by the seasonal ARIMA method, demonstrating high accuracy and practicality of our method. To further visualize the pointwise multi-step-ahead PI, Figure 6 shows plots of ARIMA estimation, the proposed estimation, 95% pointwise the ARIMA PI and the proposed PI of the last five observations (testing set) for each air pollutant concentration.

In all panels, the true air pollutant concentration values are entirely covered by ARIMA PIs, whereas some true values fall outside of the proposed PIs. ARIMA PIs seem to perform better at capturing future true values but at the price of precision. Table 6 and Figure 6 show that the ARIMA PI is much wider than the proposed PI on average. Therefore, they are substantially less useful in locating the whereabouts of the future value. The proposed PI achieves better balance between coverage probability and accuracy. Moreover, from Figure 6, one can easily find that the

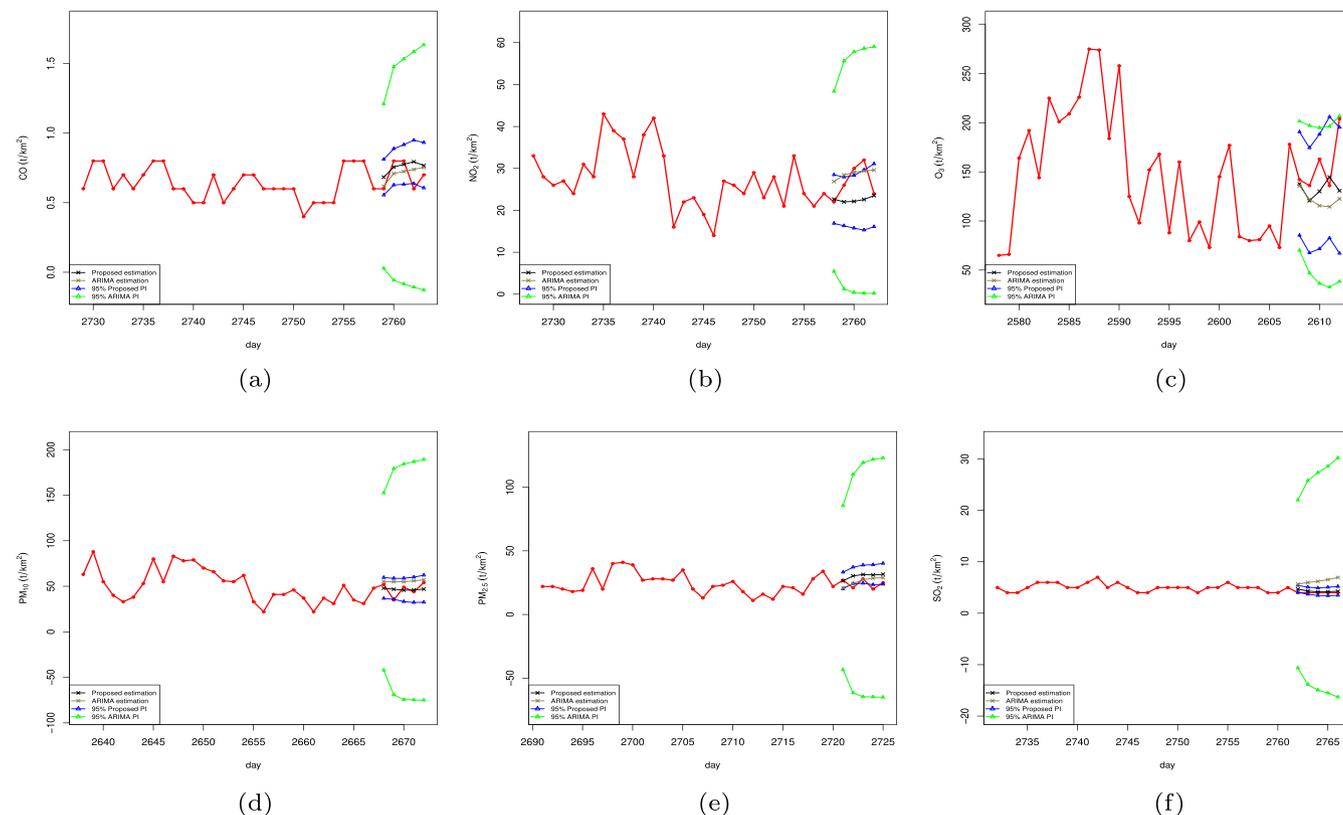


FIGURE 6 Plots of ARIMA estimation, proposed estimation, 95% pointwise the ARIMA PI and the proposed PI of the last five observations for each air pollutant concentration

proposed estimation has a smaller bias, which is closer to the future true value than the ARIMA estimation, which further reveals a positive confirmation of our proposed method.

6 | CONCLUDING REMARKS

Our study aims to construct a multi-step-ahead PI for a locally stationary time series. Hence, we propose a B-spline estimator for the trend function and a kernel estimator for the variance function. The quantile estimator is obtained after fitting the autoregressive model of errors and PIs for multi-step-ahead future observations are constructed using the estimated quantiles. We believe this is the first study of constructing reliable multi-step-ahead PIs by nonparametric regression in the local stationary time series setting. Our proposed method is applied for interpreting the underlying dynamics of air pollutant concentration data and forecasting future concentrations, which is helpful for pollutant management and early prevention.

ACKNOWLEDGEMENTS

The authors thank the China National Environmental Monitoring Center for providing data for the application. This research was supported by the National Natural Science Foundation of China awarded 11771240 and 12026242. The authors are truly grateful to the editor, associate editor, two reviewers, Prof. Lijian Yang and Mr. Yongzhen Feng from the Center for Statistical Science at Tsinghua University for their constructive comments and suggestions that led to significant improvements in the paper.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jie Li  <https://orcid.org/0000-0002-6553-4770>

Qirui Hu  <https://orcid.org/0000-0002-4846-3886>

REFERENCES

- Aneiros-Pérez, G., Cao, R., & Vilar-Fernández, J. M. (2011). Functional methods for time series prediction: A nonparametric approach. *Journal of Forecasting*, 30(4), 377–392.
- Brockwell, P., & Davis, R. (1991). *Time series: Theory and methods*. (2nd ed.). New York: Springer.
- Cai, L., Liu, R., Wang, S., & Yang, L. (2019). Simultaneous confidence bands for mean and variance functions based on deterministic design. *Statistica Sinica*, 29(1), 505–525.
- Cao, G., Wang, L., Li, Y., & Yang, L. (2016). Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Statistica Sinica*, 26(1), 359–383.
- Cao, G., Yang, L., & Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2), 359–377.
- Dahlhaus, R. (2012). Locally stationary processes, *In Handbook of statistics*, Vol. 30: Elsevier, pp. 351–413.
- Das, S., & Politis, N. (2021). Predictive inference for locally stationary time series with an application to climate data. *Journal of the American Statistical Association*, 116(534), 919–934.
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496), 1513–1527.
- Dette, H., & Wu, W. (2022). Prediction in locally stationary time series. *Journal of Business & Economic Statistics*. 40(1), 370–381. <https://doi.org/10.1080/07350015.2020.1819296>
- Fan, J., & Yao, Q. (2008). *Nonlinear time series: Nonparametric and parametric methods*: Springer Science and Business Media.
- Gu, L., Wang, L., Hrdle, W., & Yang, L. (2014). Simultaneous confidence corridor for varying coefficient regression with sparse functional data. *TEST*, 23(4), 806–843.
- Gu, L., & Yang, L. (2015). Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electronic Journal of Statistics*, 9(1), 1540–1561.
- Kong, J., Gu, L., & Yang, L. (2018). Prediction interval for autoregressive time series via oracally efficient estimation of multi-step-ahead innovation distribution function. *Journal of Time Series Analysis*, 39(5), 690–708.
- Liu, R., & Yang, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theory*, 26(1), 29–59.
- Liu, R., & Yang, L. (2016). Spline estimation of a semiparametric GARCH model. *Econometric Theory*, 32(4), 1023–1054.
- Lori, A., & William, R. (1990). Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85(410), 486–492.
- Song, Q., & Yang, L. (2009). Spline confidence bands for variance function. *Journal of Nonparametric Statistics*, 21(5), 589–609.
- Wang, J., Cao, G., Wang, L., & Yang, L. (2020). Simultaneous confidence band for the stationary covariance function of dense functional data. *Journal of Multivariate Analysis*, 176(104), 584.
- Wang, J., Liu, R., Cheng, F., & Yang, L. (2014). Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *Annals of Statistics*, 42, 654–668.

- Wang, J., & Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statistica Sinica*, 19(1), 325–342.
- Wang, L., & Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics*, 35(6), 2474–2503.
- Xue, L., & Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, 16(4), 1423–1446.

How to cite this article: Li, J., Hu, Q., & Zhang, F. (2022). Multi-step-ahead prediction interval for locally stationary time series with application to air pollutant concentration data. *Stat*, 11(1), e411. <https://doi.org/10.1002/sta4.411>