Check for updates

Modeling and forecasting of the high-dimensional time series data with functional data analysis and machine learning approaches

Yousuf Alkhezi[1], Hajar M. Alkhezi[2] and Ahmad Shafee[3]*

[1]Mathematics Department, College of Basic Education, Public Authority for Applied Education and Training (PAAET), Kuwait City, Kuwait, [2]Department of Statistics and Operations Research, Faculty of Science, Kuwait University, Safat, Kuwait, [3]Laboratory Technology Department, College of Technological Studies, PAAET, Kuwait City, Kuwait

Temperature impacts every part of the world. Meteorological analysis and weather forecasting play a crucial role in sustainable development by helping reduce the damage caused by extreme weather events. A key indicator of climate change is the change in surface temperature. This research focuses on developing and testing a lightweight and innovative weather prediction system that uses local weather stations and advanced functional time series (FTS) techniques to forecast air temperature (AT). The system is built on the latest functional autoregressive model of order one [FAR(1)]. Our results show that the proposed model provides more accurate forecasts than machine learning techniques. Additionally, we demonstrate that our model outperforms several benchmark methods in predicting AT.

1 Introduction

Meteorology attempts to achieve scientific knowledge by utilizing scientific methodology describing transitory climate for a specific region and time. Meteorological phases which, whilst ever changing, are such concepts as simply temperature, humidity, wind speed, precipitation, and atmospheric pressure data taken over a certain time, can yield quantitative comprehension of the atmosphere's present state. That helps forecasting the further state of the atmosphere concerned, with a grip on the science behind the processes involved. Weather forecasts allow activities to be planned according to future weather conditions. They help people and governments can thus take action to adopt measures to minimize harmful impacts in the case of snow, rain, heat waves, or floods, out of which they could warn a coming. These profiled interests represented in the weather sector last year. Farmers may rely on meteorological forecasts to plan for harvest and other activities in agriculture. Utility companies use forecasts to procure natural gas and electricity, whereas retailers use the forecasts to manage inventory based on expected demand. They thus help the public plan their outdoor activities better by issuing timely weather warnings that save lives and property.

Extreme natural phenomena such as heat waves, extreme cold, heavy snowfall, torrential rain, and droughts continue to increase worldwide with changing climate. These phenomena destroy large tracts, damage the ecosystem, and injure or kill people. The changes in AT are a significant measure of air quality by sudden changes [1]. The sudden changes in AT would affect crop growth and yield during different agricultural operations [2]. AT plays a vital role in modeling crop productivity to assure food safety under changed climate scenarios [3]. AT variations are also directly associated with energy consumption, water use, animal farming, agricultural activities, and meteorological events [4]. The prediction of AT can be highly beneficial in deciding the timing of agricultural activities carried out for a particular crop, such as planting and harvesting dates, irrigation, pruning, and frost protection.

Weather forecasts play an essential role in human life and have attracted the attention of scientists [5]. Therefore, much research has been done toward advancing the forecasting technique [6–8]. In the case of short-term predictions, such as hourly and daily weather forecasts, statistical analytical techniques and artificial intelligence give precise results. For longer-term predictions, observed temperatures are increasingly deviating from the predicted values [6]. In their work, Li et al. [9] introduced different statistical and machine learning techniques to improve climate predictions. They assessed the value of these techniques in increasing the accuracy of long-term prediction of average temperature. Several forecasting models have been proposed for improving the accuracy of average temperature as well as land surface temperature prediction. You can put these models into six groups: Spatio-Temporal Correlation models, deep learning models, machine learning models, combined models, statistical models, and physical models [10].

When data items characterized by functions are repeatedly observed over time, they are known as FTS. Research on FTS has mostly concentrated on applying low-dimensional multivariate time series (ts) or univariate ts techniques to the functional domain. Recent advancements in FTS analysis have addressed various challenges such as modeling long-range dependence [11], developing predictive methods for high-dimensional data [12], and incorporating functional principal components in time series forecasting [13, 14]. Due to the quick advancement of data collection technology, high-dimensional FTS datasets are becoming more and more prevalent. PM10 trajectories [15] gathered at various measuring stations, daily electricity load curves [16] for numerous households, cumulative intraday return trajectories [17] for a group of stocks, and functional volatility processes [18] for a collection of stocks are a few examples. Recent advancements in FTS include the following: zamani et al. [19], Meintanis et al. [20], Chang et al. [21], and Baek et al. [22]. Further research is necessary because the functional data analysis (FDA) application for AT and precipitation is uncommon. We used data from 2015 to 2021 to train a first-order FAR model with AT as the input. To check how well the model worked, the two accuracy metrics, rooted mean square error (RMSE) and mean absolute percentage error (MAPE), are used to check the model's performance. The paper is structured as follows: The rest of the paper is structured as follows: Section 2 provides an overview of the preliminaries of the functional model, focusing

on the application of the FAR(1), ANN, NNAR, and SVM models for FTS forecasting. Section 3 describes the dataset and the results, including a discussion on model performance and forecast accuracy. The paper concludes with Section 4, which summarizes the significant findings, addresses limitations, and suggests future research areas.

2 Functional data analysis

The name FDA was initially used by Ramsay [23]. The FDA framework has expanded and modified several traditional statistical techniques [24]. The FDA provides a way to analyze data that is not merely discrete observations but represented as curves, forms, or patterns. This method works exceptionally well when dealing with data that varies gradually over time or space [25]. Functional data are typically observed at discrete time points, often with high frequency, which facilitates their representation as functional objects through appropriate smoothing or basis function expansion.

$$Y(i) = \sum_{j=1}^{J} \alpha_j \psi_j(i), \quad , \quad i \in I \tag{1}$$

Where $Y(i)$ denotes the functional observation at point i, $\{\psi_j(i)\}_{j=1}^{J}$ are predefined basis functions (e.g., B-splines, Fourier, or wavelets), and $\alpha_j \in \mathbb{R}$ are the associated basis coefficients. We assume that each ψ_j belongs to a real separable Hilbert space $\mathcal{H} = \mathcal{L}^2(I)$, the space of square-integrable functions over a compact domain $I \subset \mathbb{R}$, endowed with the inner product:

$$\langle f, g \rangle = \int_I f(i)g(i)\, di.$$

In this framework, $Y(i) \in \mathcal{H}$, and the basis expansion allows for a finite-dimensional approximation of functional data. The number of basis functions will primarily care for the FDA concern. Generally, to optimally choose the number of basis functions, a penalized residual sum of squares criterion is employed. This criterion smoothens the curve while preventing it from yielding an insufficient fit to the data. Argument values refer to the discretized locations in the I domain where the function is evaluated [25]. Researchers often use popular mathematical functions, such as exponential, B-spline, Fourier, and polynomial models to handle various data types. The choice depends on the data's characteristics; for instance, B-splines work well for non-periodic data, while Fourier functions are better suited for periodic patterns. Given the periodic nature of the data in our study, we decided to employ the Fourier basis approach.

2.1 Functional autoregressive model

In the context of FTS, curve evolution over time is analyzed statistically using the FAR model. The traditional AR concept is expanded upon inside a useful framework. The FAR model states that the function's previous state determines its current state. This study uses a first-order FAR model within the setting of a Hilbert

space $\mathcal{H}$. While the method can be extended to other $\mathcal{L}^2$-spaces, the Hilbert space $\mathcal{L}^2[0, 1]$ is specifically taken into consideration. This procedure adheres to the AR equation and is completely stationary

$$Y_t(i) - \mu(i) = \beta(Y_{t-1}(i) - \mu(i)) + \epsilon_t(i) \qquad (2)$$

A bounded linear operator is represented by β, the mean curve is represented by $\mu(i)$, and the residual term is shown by $\epsilon(i)$. From $\mathcal{L}^2[0, 1]$ to itself, the AR operator β is seen as a positive, bounded, symmetric, compact Hilbert–Schmidt linear operator mapping. The compact support (i) will be left out of the functional framework for simplicity's sake.

2.1.1 Operators in the Hilbert space $\mathcal{L}^2[0, 1]$

The inner product of the space $\mathcal{H}$ yields the bounded linear operators in the Hilbert space $\mathcal{L}^2[0, 1]$, which are specified by a norm. The norm supremum of the operator over all unit vectors Y in $\mathcal{H}$ equals this norm. In operator theory (on Hilbert spaces), the operator norm should be written as $\|\beta\|$, and its expression is

$$\|\beta\| = \sup_{\|Y\| \leq 1} \|\beta(Y)\|$$

An operator β is considered compact if it can be expressed as follows concerning orthonormal bases v_i and f_i in $\mathcal{H}$ and a sequence γ_i that approaches zero:

$$\beta(Y) = \sum_{i=1}^{\infty} \gamma_i \langle Y, v_i \rangle f_i$$

If the sum of the squares of an operator's sequence γ_i is finite, then the operator is Hilbert–Schmidt.

$$\sum_{i=1}^{\infty} \gamma_i^2 < \infty$$

The separable space of Hilbert–Schmidt operators S has an inner product and corresponding norm:

$$\langle \beta_1, \beta_2 \rangle_S = \sum_{1 <= k, i <= \infty} \langle \beta_1(c_k), h_i \rangle \langle \beta_2(c_k), h_i \rangle$$

$$\|\beta\|_S = \left(\sum_i \gamma_i^2 \right)^{1/2}$$

If $\langle \beta(x), Y \rangle = \langle Y, \beta() \rangle$, then the operator is symmetric; if $\langle \beta(Y), x \rangle \geq 0$, then it is positive. A positive symmetry in the decomposition is accepted by the Hilbert–Schmidt operator.

$$\beta(Y) = \sum_{i=1}^{\infty} \gamma_i \langle Y, v_i \rangle v_i$$

If the sequence $\{\gamma_i\}$ contains only finitely many nonzero elements, then the operator β becomes finite-rank, which is a special case of compact operators. However, if the decay or behavior

of γ_i is not specified (e.g., not tending to zero), the compactness of β cannot be guaranteed.

$$\sum_i |\gamma_i| < \infty$$

$\| \cdot \|_N \geq \| \cdot \|_S \geq \| \cdot \|_\mathcal{L}$ is the relationship that explains these functions rules. β must be an integral function in $\mathcal{L}^2$, as defined by $\beta(Y)(i) = \int \beta(i, s) Y(s) ds$, where $\beta(., .)$ is a real kernel, to be a Hilbert–Schmidt function. This is true if and only if $\int \int \beta^2(t, s) dt ds < \infty$. The model is non-parametric due to the infinite-dimensional parameter β.

2.1.2 Estimation of the operator β

To estimate the AR function β in the method proposed, there are necessary assumptions to identify a stationary solution. These include two possible scenarios for such a stationary solution. The first assumption translates into having some integer $i_0 \geq 1$ such that $\|\beta^{i_0}\|_\mathcal{L} < 1$. The second condition needs the existence of constants $a > 0$ and $0 < b < 1$, such that for any integers $i_0 \geq 0$, one has the following: $\|\beta_0^i\|_\mathcal{L} \leq ab^i$. According to Bosq [14], such assumptions can lead to the uniqueness of a strictly stationary solution under certain conditions.

In estimate β, it is not possible with likelihood methods since Lebesgue measure does not exist within non-locally compact spaces, and applicable density for functional data sets does not exist as well. Rather, one has to revert to the moment's methods for estimating ($\beta = \hat{C}\Gamma^{-1}$ with "times" denoting the Kronecker product. It should be noted that C and Γ are covariance and cross-covariance operators of the process defined by $C = E(Y_l \otimes Y_{l+1}), \Gamma = E(Y_l \otimes Y_l)$. The sample versions then of these operators are denoted by $\hat{C}_l$ and $\hat{\Gamma}_l$.

Supposing the mean of the process, $E(Y_l) = 0$, is known for notational convenience; as indicated by the symbols, the following sample versions represent the covariance and cross-covariance operators: $\hat{\Gamma}_l = \frac{1}{N} \sum_{l=0}^{N-1} Y_l \otimes Y_l$ and $\hat{C}_l = \frac{1}{N} \sum_{l=0}^{N-1} Y_l \otimes N_{l+1}$. Let Γ be a positive definite, symmetric, compact covariance operator. The decomposition of such an operator is in terms of eigenvalues and eigenfunctions represented henceforth by the symbols (γ_l) and (v_l). This, however, does not hold for the operator (Γ^{-1}). To solve that issue, the unknown population principal components (PCs) are approximated with the first w empirical functional principal components (FPCs) in an approximation following relation:

$$\Gamma_l^{-1}(Y) = \sum_{i=1}^{w} \gamma_i \langle Y, \hat{v}_i \rangle \hat{v}_i,$$

The eigenvalues are represented by γ_i, while the estimated eigenfunctions are indicated by $\hat{v}_i$. The relationship is obtained by multiplying the ARH(1) equation by Y_N

$$Y_l \otimes Y_{l+1} = Y_l \otimes (\beta Y_l) + Y_l \otimes \epsilon_{l+1} = \beta Y_l \otimes Y_l + Y_l \otimes \epsilon_{l+1}$$

The connections $C = \beta \Gamma$ and $\beta = C\Gamma^{-1}$ are obtained by using the definitions of the covariance and cross-covariance operators

for ARH(1) and assuming that $(E(\epsilon_{l+1}) = 0$. Next, we obtain the estimate of (β) as follows:

$$\hat{\beta}_N(Y) = \frac{1}{N-1} \sum_{k=1}^{N-1} \sum_{i=1}^{p} \sum_{h=1}^{p} \hat{\gamma}_i^{-1} \langle Y, \hat{v}_i \rangle \langle y_h, \hat{v}_i \rangle \langle y_{k+1}, \hat{v}_h \rangle \hat{v}_h$$

To produce the final term, y_{k+1} and $\hat{v}_i$ are subjected to a smoothing process. The asymptotic convergence of the estimated eigenfunctions and the population eigenfunctions is determined. The performance of the estimator $\hat{\beta}$ must be assessed in order to ascertain how well it approximates the genuine population parameter (β). With regard to the FAR parameter (β), it has been demonstrated that this estimator is best in terms of mean square error (MSE) and mean absolute error (MAE) [26].

2.2 Artificial neural networks

An artificial neural network (ANNs) is one of the essential components of machine learning. Such networks are made up of layers of interconnected nodes: input, hidden, and output. This basic construction mimics the composition and functioning of the human brain. These nodes operate similarly to biological neurons by processing information over weighted connections. The processing of data occurs by activation above certain thresholds, where information is passed between layers. One wonderful application of ANNs is Google search, which shows their power in fast classification and clustering of data. Moreover, ANNs are excellent for imparting other functions such as image and speech recognition.

The ANNs consists of three main components: the input, hidden, and output layers. The input layer accepts various data types since the hidden layers, often called the "distillation layers" by some researchers, sift relevant information from redundancy toward the ends of improving network performance. The output layer accepts this refined information and generates a resultant response with the inputs routed through the hidden ones. The most popular type of ANN used in predictive applications is the multi-layer perceptron (MLP), which uses the three-layer structure connected by directed (acyclic) connections. This enables several hidden layers to be inserted, in which each node acts as a processing unit in its own right.

2.2.1 Neural network autoregressive

A valuable algorithm for deep learning that has been used in several domains, such as speech recognition, FTS forecasting, and linguistic processing, is the neural network autoregressive (NNAR). The entire process of this recursive model employs historical data to predict future values, wherein prediction output at each time step serves as input for the next step. NNAR can model forecasting systems using feed forward or recurrent neural networks, whereby any complex data set can robustly represent nonlinear relations.

In the NNAR(p, k) architecture, the hidden layer contains "k" nodes and "p" lagged inputs for forecasting. Weighted node connections, nonlinear activation functions, AR dependencies,

external influences, and error terms are all included in the NNAR model equation, which may be expressed as

$$X_{it} = f\left(\alpha_1 \sum_{j=1}^{N} w_{ij} X_{i(t-1)} + \lambda Z_{j(t-1)}^t\right) + \epsilon_t \qquad (3)$$

Where

- w_{ij} shows the strength of the connection between nodes i and j for $i, j = 1, \ldots, N$,
- λ is a scaling parameter for the exogenous input,
- $Z_{i(t-1)}^t$ captures the exogenous influence,
- $f()$ is an unknown smoothed link function,
- X_{it} represents the univariate response at node i and time j.

Local linear approximation and profile least squares estimation are two techniques for the NNAR that are used to optimize parameter estimates and represent the unknown link function. This model demonstrates flexibility and resilience in managing complex data patterns by iterating through past inputs to produce multi-step forecasts in time series analysis. We utilize NNAR(1,1), a straightforward autoregressive of order one with a single node in the hidden layer.

2.3 Support vector machine

Powerful supervised learning models called support vector machines (SVM) are frequently used for data analysis in tasks including regression, classification, and identification of outliers. Their proficiency in managing linear and nonlinear data splits makes them adaptable to various applications, including time series prediction, picture recognition, gene expression analysis, and text categorization. Notably, these machines were painstakingly created to effectively simplify training set outcomes in addition to grading them. Because of this unique feature, SVM approaches are widely used, particularly in time series forecasting.

Much of the work on this deep learning model is done in different fields. In support vector machines (SVM), the data input is a set of labeled feature vectors belonging to one of two classes. SVMs work for binary classification and can be extended to solve multi-class classification problems. Different kernel functions, such as linear, polynomial, and radial basis functions, are used to transform the feature vectors into a higher dimensional space. This transformation is required to assess the similarity between different feature vectors, thus enabling the SVM to tackle the challenge of classification of nonlinearly separable data.

2.3.1 Key elements and optimization in the SVM
- Class labels and training data: as an anchor for the model's learning process, the training data includes assigned class labels (Y_i) for binary classification, denoted within $\{-1, 1\}$.
- Decision boundary and margin: the SVM seeks to create a hyperplane that maximizes the class margin, which may be stated mathematically as

$$w^T Y + b = 0,$$

where w represents the weight vector and b is the bias.

- Support vectors and decision boundary: the data points that fall inside or on the margin serve as support vectors, which are crucial in forming the decision border. They are crucial in determining the decision limit.

- Optimization and decision function: to create the decision function, the optimization process adjusts essential parameters such as the weight vector, bias, and regularization factor.

- Optimization (mathematical model): an optimization problem is involved in determining the hyperplane: Under the constraints $Y_i(w^T X_i + b) \geq 1 - \xi_i$ with $\xi_i \geq 0$, where C indicates the regularization parameter and ξ_i are slack variables allowing misclassifications inside margin bounds,

$$\text{Minimizing}(\frac{1}{2}\|w\|^2 + C \sum \xi_i)$$

.

- Prediction and decision function: using the altered feature vector produced by the kernel function, the decision function predicts the class label for a new data point x. $f(x) = \text{sign}(w^T \phi(x) + b)$ is the representation of this function, where $\phi(x)$ indicates the converted input feature vector.

The Support Vector Machine (SVM) model used in this study is based on epsilon-regression and employs a radial basis function (RBF) kernel. It is parameterized with a cost value (C) of 1.0, which controls the trade-off between model complexity and training error. The γ parameter is set to 0.5 and defines the influence of a single training example–higher values imply a closer reach, while lower values mean a broader influence. The ϵ is set to 0.1 and specifies the margin of tolerance within which no penalty is given in the training loss function; that is, predictions within ± 0.1 of the true value are considered acceptable. The model takes two lagged values of ozone concentration as input features. To guarantee precise predictions, the model is supported by 1,191 support vectors.

3 Analysis and results

3.1 Out-of-sample air temperature forecasting

The study used an hourly AT data set collected in Berlin, Germany. The data set is gathered from sensors across Berlin, and an average value is provided for every hour [27]. Because hourly measures capture the dynamic character of AT fluctuations, they offer a more comprehensive picture than daily or monthly averages. The 6-year time frame covered by the data set is January 1, 2015, through December 31, 2021. For modeling and forecasting, the data collection is divided into two sections:

- Training data: From January, 2015 to December, 2019 (43,824 observations, covering i = 1,826 days).
- Testing data: From January, 2020 till December, 2020 (8,784 observations, covering i = 366 days).

In order to determine the annual one-day out-of-sample forecast, the models are estimated using the expanding window
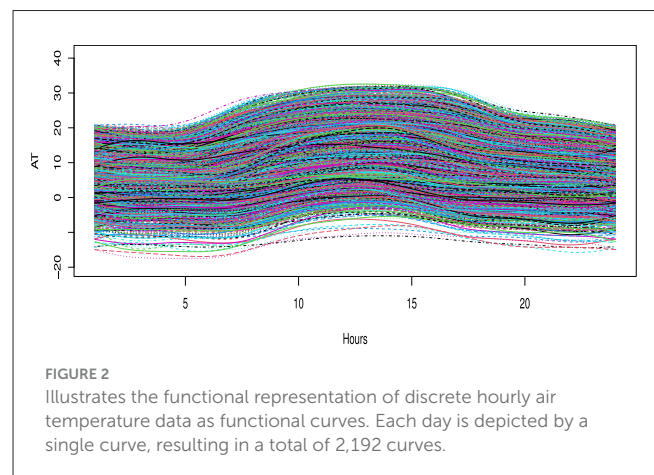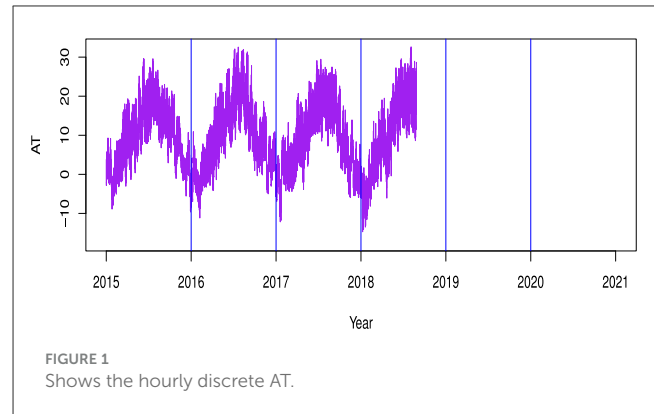
approach. The discrete hourly AT from 1*st* January 2015 to 31*st* December 2020, as shown on the right side of the Figure 1, amply demonstrates the periodic nature. A single day is represented by each curve in the Figure 2, which illustrates the conversion of discrete data into functional data. The 1-week functional representation of AT data is shown in the Figure 3.

3.2 Forecasting accuracy metrics

Two common error metrics, RMSE and MAPE, are used to assess the accuracy of the forecasting models. These error measures, which measure the degree of agreement between the actual values and the projections, are frequently employed in descriptive statistics. They are characterized in mathematics as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Y_{i,d} - \hat{Y}_{i,d} \right)^2}$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left(\left| \frac{Y_{i,d} - \hat{Y}_{i,d}}{Y_{i,d}} \right| \right) \times 100$$

The number of observations during the testing forecast period is denoted by n. The actual and expected AT values for the i-th day
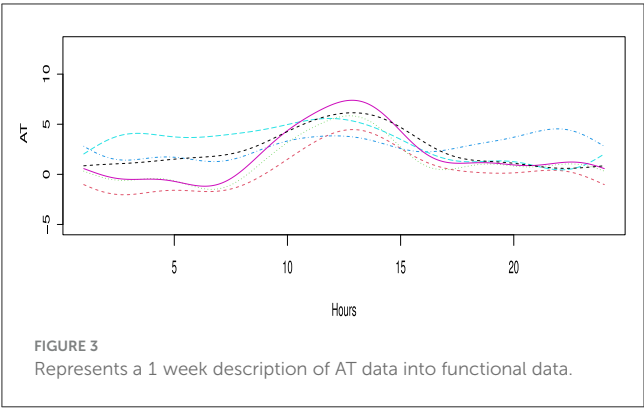
FIGURE 3
Represents a 1 week description of AT data into functional data.

TABLE 1 Overall forecast results for AT.

Model	RMSE	MAPE
ANN	112.66	263.47
NNAR	134.60	222.87
SVM	94.88	156.98
FAR(1)	56.98	77.39

$(i = 1, 2, \ldots, 366)$ and the d-th hour $(d = 1, 2, \ldots, 24)$ are shown by - $Y_{i,d}$ and $\hat{Y}_{i,d}$.

3.3 Results

Table 1 uses RMSE and MAPE as evaluation metrics to compare the forecasting accuracy and performance of various models. The results showed that the FAR(1) model outperformed the others due to its lowest error values. The SVM model came in second, the ANN model in third, and the NNAR model did the worst out of the four models that were assessed.

The RMSE and MAPE measures are used to compare the hourly forecasting errors of AT from four distinct models, as indicated in Table 2. Across all hours, the FAR(1) model exhibits the highest accuracy. The SVM model outperforms the ANN model, but the NNAR model does the poorest.

Table 3 gives a thorough example of the RMSE and MAPE values for each month for the models that were employed to predict AT in this study. The monthly errors during the testing year are averaged to determine these errors. The findings show that forecasting mistakes differ from month to month. Notably, compared to the other months of the year, May, September, and November exhibit comparatively higher rates. Again, in nearly every month, the FAR(1) model is the best, followed by the SVM model. Again, the NNAR model is the worst across all months, and the ANN model comes in third, suggesting that it is not a good option for AT predicting.

The weekly forecasting errors of the AT from four different models, measured using RMSE and MAPE as accuracy indicators, are presented in Table 4. It is observed that the proposed functional model outperforms the others, exhibiting the lowest forecasting errors across all days of the week. The SVM model ranks as the

TABLE 2 Hour-specific forecasting errors AT.

Hour	Errors	ANN	NNAR		FAR(1)
1	RMSE	38.49	44.30	28.03	19.88
	MAPE	42.23	48.03	32.98	24.57
2	RMSE	39.57	42.15	33.01	22.34
	MAPE	43.49	48.81	36.99	27.05
3	RMSE	33.69	27.19	22.88	19.44
	MAPE	34.98	29.77	25.66	22.07
4	RMSE	28.25	32.50	25.33	21.66
	MAPE	33.52	35.04	27.03	24.57
5	RMSE	39.65	51.02	32.05	25.33
	MAPE	43.24	55.75	35.98	29.66
6	RMSE	43.47	47.98	32.24	26.06
	MAPE	45.98	52.22	35.39	29.03
7	RMSE	69.39	93.86	46.55	26.75
	MAPE	90.84	114.55	71.05	55.88
8	RMSE	88.67	96.59	55.02	34.27
	MAPE	99.39	116.56	77.89	66.56
9	RMSE	66.23	82.89	54.72	34.88
	MAPE	81.69	94.89	62.55	38.99
10	RMSE	55.79	83.86	42.78	23.68
	MAPE	66.33	99.97	53.60	44.78
11	RMSE	78.49	99.76	59.97	33.29
	MAPE	89.44	109.35	70.38	43.27
12	RMSE	70.22	88.98	60.97	31.87
	MAPE	88.28	99.65	72.67	23.99
13	RMSE	88.99	102.03	78.28	37.09
	MAPE	98.88	109.22	86.09	47.95
14	RMSE	80.44	95.89	74.00	43.98
	MAPE	87.48	101.67	77.80	48.00
15	RMSE	66.96	75.04	58.39	44.78
	MAPE	73.34	81.03	64.93	48.33
16	RMSE	78.56	80.23	69.76	60.06
	MAPE	83.37	88.08	70.39	64.59
17	RMSE	60.74	66.98	55.45	47.00
	MAPE	61.05	66.92	58.74	50.03
18	RMSE	89.77	96.66	81.36	52.99
	MAPE	93.88	101.92	77.68	53.55
19	RMSE	60.09	48.98	42.94	34.05
	MAPE	63.46	52.55	45.40	37.77
20	RMSE	31.03	35.99	29.11	20.61
	MAPE	34.46	38.78	32.08	23.06
21	RMSE	38.34	43.93	29.17	22.65
	MAPE	39.99	46.00	31.08	24.00

(Continued)

TABLE 2 (Continued)

Hour	Errors	ANN	NNAR	SVM	FAR(1)
22	RMSE	40.07	43.67	33.18	26.09
	MAPE	43.08	47.99	36.32	28.99
23	RMSE	35.07	39.49	28.71	22.55
	MAPE	39.86	42.08	32.86	25.06
24	RMSE	32.99	36.19	24.92	18.55
	MAPE	37.54	43.08	29.56	23.66

TABLE 3 Month-specific forecasting errors AT.

Hour	Errors	ANN	NNAR	SVM	FAR(1)
January	RMSE	29.97	42.88	23.85	20.63
	MAPE	33.93	46.38	27.23	23.05
February	RMSE	40.73	46.56	32.19	24.03
	MAPE	41.44	46.16	32.55	26.35
March	RMSE	51.09	55.97	43.80	35.74
	MAPE	55.99	59.75	47.06	36.03
April	RMSE	49.75	55.63	38.39	32.79
	MAPE	53.53	59.45	42.38	35.93
May	RMSE	55.59	65.28	47.59	38.88
	MAPE	59.49	69.58	51.80	41.47
June	RMSE	59.77	62.33	41.06	35.42
	MAPE	63.83	67.77	47.39	32.99
July	RMSE	51.44	58.83	43.77	40.64
	MAPE	54.82	63.02	48.35	42.47
August	RMSE	59.04	62.97	51.08	45.07
	MAPE	65.78	69.92	55.69	49.63
September	RMSE	60.84	65.98	57.54	35.87
	MAPE	65.29	69.28	61.99	40.35
October	RMSE	51.48	55.86	42.08	37.35
	MAPE	56.85	61.26	45.19	42.99
November	RMSE	62.48	67.87	53.56	43.74
	MAPE	65.88	72.28	58.98	49.50
December	RMSE	52.97	57.89	44.63	37.38
	MAPE	54.85	62.44	49.19	40.07

TABLE 4 Week-specific forecasting errors AT.

Hour	Errors	ANN	NNAR	SVM	FAR(1)
Monday	RMSE	59.76	67.04	45.05	29.04
	MAPE	63.33	70.77	51.38	35.58
Tuesday	RMSE	60.37	69.69	52.98	37.39
	MAPE	63.99	73.65	58.03	41.58
Wednesday	RMSE	56.97	63.27	53.09	33.05
	MAPE	59.19	65.45	57.86	36.77
Thursday	RMSE	59.55	62.65	48.77	37.79
	MAPE	62.75	64.89	51.99	40.92
Friday	RMSE	62.11	68.05	49.99	36.77
	MAPE	63.95	73.89	53.07	39.73
Saturday	RMSE	66.76	69.39	49.66	41.23
	MAPE	68.37	73.64	53.98	48.03
Sunday	RMSE	63.94	69.33	52.26	42.49
	MAPE	68.03	75.28	57.95	49.73

introduces a method for forecasting AT one day in advance using FDA. It effectively captures AT dynamics by applying a component estimation approach that separates the data into stochastic and deterministic components. The deterministic component is modeled using smoothing splines, while the stochastic component is forecasted using FAR, SVM, ANN, and NNAR models.

For empirical analysis, AT data from Berlin, Germany, was collected. The forecasting accuracy was assessed using RMSE and MAPE metrics for one-day-ahead predictions over a full year. The results indicate that the proposed component estimation method can effectively predict AT. Compared to SVM, ANN, and NNAR models, the suggested approach significantly reduces forecasting errors and improves accuracy. Among the tested models, SVM demonstrates the lowest overall forecasting error, followed by ANN and NNAR. Additionally, ANN outperforms NNAR in terms of prediction accuracy. Overall, FAR(1) is the most effective model, followed by SVM and ANN, while NNAR performs the worst, suggesting it is not suitable for AT forecasting.

Although this study offers valuable insights, it has certain limitations. The analysis is restricted to parametric (linear) models and is based on data from a single location. Therefore, the generalizability of the FAR(1) model to other geographic regions with different climatic patterns remains an open question and should be investigated in future work. Additionally, it does not consider the impact of external variables, raising questions about how these factors might influence AT predictions in future applications of this method.

second best, followed by the ANN model in third place, while the NNAR model shows the highest forecasting errors, making it the least effective.

4 Conclusion

Accurate forecasting is crucial because AT is a key element of weather and affects various aspects of life. However, predicting AT is challenging due to the complex nature of ts data, which includes both random fluctuations and predictable patterns. This study

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

Funding

Acknowledgments

We sincerely thank the referee for their valuable time, insightful comments, and constructive feedback, which have significantly enhanced this work. Their thoughtful suggestions have helped improve both the clarity and quality of our study. We truly appreciate their efforts in reviewing our manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Zahroh S, Hidayat Y, Pontoh RS, Santoso A, Sukono F, Bon A. Modeling and forecasting daily temperature in Bandung. In: *Proceedings of the International Conference on Industrial Engineering and Operations Management*. Riyadh (2019). p. 406–12.

2. Meshram SG, Kahya E, Meshram C, Ghorbani MA, Ambade B, Mirabbasi R. Long-term temperature trend analysis associated with agriculture crops. *Theor Appl Climatol*. (2020) 140:1139–59. doi: 10.1007/s00704-020-03137-z

3. Hoogenboom G. Contribution of agrometeorology to the simulation of crop production and its applications. *Agric Forest Meteorol*. (2000) 103:137–57. doi: 10.1016/S0168-1923(00)00108-8

4. Nag P, Nag A, Sekhar P, Pandit S. *Vulnerability to Heat Stress: Scenario in Western India*. Ahmedabad: National Institute of Occupational Health (2009). p. 380016.

5. Kumar S, Roshni T, Kahya E, Ghorbani MA. Climate change projections of rainfall and its impact on the cropland suitability for rice and wheat crops in the Sone river command, Bihar. *Theor Appl Climatol*. (2020) 142:433–51. doi: 10.1007/s00704-020-03319-9

6. Salman AG, Heryadi Y, Abdurahman E, Suparta W. Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Comput Sci*. (2018) 135:89–98. doi: 10.1016/j.procs.2018.08.153

7. Chaudhuri S, Middey A. Adaptive neuro-fuzzy inference system to forecast peak gust speed during thunderstorms. *Meteorol Atmos Phys*. (2011) 114:139–49. doi: 10.1007/s00703-011-0158-4

8. Liu H, He B, Qin P, Zhang X, Guo S, Mu X. Sea level anomaly intelligent inversion model based on LSTM-RBF network. *Meteorol Atmos Phys*. (2021) 133:245–59. doi: 10.1007/s00703-020-00745-2

9. Li X, Li Z, Huang W, Zhou P. Performance of statistical and machine learning ensembles for daily temperature downscaling. *Theor Appl Climatol*. (2020) 140:571–88. doi: 10.1007/s00704-020-03098-3

10. Liu H, Mi X, Li Y. Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network. *Energy Convers Manag*. (2018) 166:120–31. doi: 10.1016/j.enconman.2018.04.021

11. Li D, Robinson PM, Shang HL. Long-range dependent curve time series. *J Am Stat Assoc*. (2020) 115:957–71. doi: 10.1080/01621459.2019.1604362

12. Aue A, Norinho DD, Hörmann S. On the prediction of stationary functional time series *J Am Stat Assoc*. (2015) 110:378–92. doi: 10.1080/01621459.2014.909317

13. Bathia N, Yao Q, Ziegelmann F. Identifying the finite dimensionality of curve time series. *Ann Statist*. (2010) 38:3352–86 doi: 10.1214/10-AOS819

14. Bosq D. *Linear Processes in Function Spaces: Theory and Applications, Vol. 149*. Cham: Springer Science & Business Media (2000). doi: 10.1007/978-1-4612-1154-9

15. Hörmann S, Kidziński Ł, Hallin M. Dynamic functional principal components. *J R Stat Soc B: Stat Methodol*. (2015) 77:319–48. doi: 10.1111/rssb.12076

16. Cho H, Goude Y, Brossat X, Yao Q. Modeling and forecasting daily electricity load curves: a hybrid approach. *J Am Stat Assoc*. (2013) 108:7–21. doi: 10.1080/01621459.2012.722900

17. Horváth L, Kokoszka P, Rice G. Testing stationarity of functional time series. *J Econ*. (2014) 179:66–82. doi: 10.1016/j.jeconom.2013.11.002

18. Müller HG, Sen R, Stadtmüller U. Functional data analysis for volatility. *J Econ*. (2011) 165:233–45. doi: 10.1016/j.jeconom.2011.08.002

19. Zamani A, Haghbin H, Hashemi M, Hyndman RJ. Seasonal functional autoregressive models. *J Time Ser Anal*. (2022) 43:197–218. doi: 10.1111/jtsa.12608

20. Meintanis SG, Hušková M, Hlávka Z. Fourier-type tests of mutual independence between functional time series. *J Multivar Anal*. (2022) 189:104873. doi: 10.1016/j.jmva.2021.104873

21. Chang J, Chen C, Qiao X, Yao Q. An autocovariance-based learning framework for high-dimensional functional time series. *J Econom*. (2024) 239:105385. doi: 10.1016/j.jeconom.2023.01.007

22. Baek C, Kokoszka P, Meng X. Test of change point versus long-range dependence in functional time series. *J Time Ser Anal*. (2024) 45:497–512. doi: 10.1111/jtsa.12723

23. Ramsay JO. When the data are functions. *Psychometrika*. (1982) 47:379–96. doi: 10.1007/BF02293704

24. Ferraty F. *Nonparametric Functional Data Analysis*. Cham: Springer (2006).

25. Ramsay J, Silverman B. Principal components analysis for functional data. In: *Functional Data Analysis*. Cham: Springer (2005). p. 147–72. doi: 10.1007/b98888

26. Didericksen D, Kokoszka P, Zhang X. Empirical properties of forecasts with the functional autoregressive model. *Comput Stat*. (2012) 27:285–98. doi: 10.1007/s00180-011-0256-2

27. Power N. *Data Access Viewer*. Available online at: https://power.larc.nasa.gov/data-access-viewer (Accessed 11, 2022).