Arabian Journal of Geosciences (2021) 14: 622
https://doi.org/10.1007/s12517-021-06982-y

ORIGINAL PAPER

Prediction of 10-min, hourly, and daily atmospheric air temperature: comparison of LSTM, ANFIS-FCM, and ARMA

Arif Ozbek[1] · Aliihsan Sekertekin[2] · Mehmet Bilgili[1] · Niyazi Arslan[2]

Received: 7 September 2020 / Accepted: 17 March 2021 / Published online: 28 March 2021

Abstract

Prediction of atmospheric air temperature (AAT) time series is an important issue as it gives information to society and sustainability for future planning. In this study, a deep learning method, namely, long short-term memory (LSTM) network, based on one-step-ahead prediction approach was proposed to predict AAT using the actual time series data. For the proposed prediction method, a set of measurement data in 10-min, hourly, and daily intervals obtained from Mersin and Belen stations located in the Eastern Mediterranean Region of Turkey was used. Mean absolute percentage error (MAPE), root mean square error (RMSE), correlation coefficient (*R*), mean absolute error (MAE), and average bias were considered as evaluation criteria. According to the testing process, the RMSE, MAPE, MAE, *R*, and bias values for the 10-min interval AAT prediction were calculated as 0.35 °C, 1.40%, 0.25 °C, 0.995, and 0.074 °C, respectively. Considering the prediction results of the hourly AAT predication, the above statistical metrics with the same order were obtained as 0.61 °C, 1.85%, 0.43 °C, 0.945, and −0.013 °C. Concerning the daily AAT prediction results with LSTM, the above statistical metrics with the same order were computed as 1.33 °C, 3.27%, 0.99 °C, 0.97, and −0.116 °C. Compared to the hourly and daily AAT predictions, LSTM provided better accuracy results in predicting 10-min interval AAT. The prediction results from the three different time series data show that the prediction of AATs using LSTM can provide high accuracy results for short-term prediction using data with a long period time. On the other hand, adaptive neuro-fuzzy inference system with fuzzy C-means (ANFIS-FCM) method and autoregressive moving average (ARMA) model were also used to compare the results of LSTM method. Both LSTM and ANFIS-FCM network model showed high accuracy for the prediction of 10-min interval, hourly, and daily AAT data with RMSE values between 0.31 and 1.52 °C, while ARMA model failed to provide high accuracies for all predictions.

Keywords Deep neural network · Long short-term memory · Atmospheric air temperature · Mean absolute percentage error · 10-min · hourly · and daily intervals

Introduction

It is a known fact that many meteorological and environmental events, human life, and products in agricultural areas are very much affected by the physical condition of the environment, as well as by climatic conditions. Considering these climatic conditions, atmospheric air temperature (AAT), a crucial weather element whose value can change at any time, is among the most effective meteorological parameters on Earth (Cobaner et al. 2014). The AAT can influence the growth, development, and yield of crops in a particular region. Besides, many meteorological parameters, namely, solar energy, wind speed, soil temperature, humidity, rainfall, and atmospheric pressure are extremely all related to the AAT. Moreover, AAT is one of the most effective parameters in evapotranspiration, which is very important for water resource management and agricultural activities (Zahroh et al. 2019).

In recent years, drought, heavy snowfall, extreme rainfall, and cold weather have been among the outstanding anomalous weather events all over the world, which lead to health and environmental problems, material damage, human injury, and death. This means that irregular air change has an adverse

Responsible Editor: Zhihua Zhang

✉ Aliihsan Sekertekin
asekertekin@cu.edu.tr

[1] Department of Mechanical Engineering, Cukurova University, Ceyhan Campus, 01950, Ceyhan, Adana, Turkey

[2] Department of Geomatics Engineering, Cukurova University, Ceyhan Campus, 01950, Ceyhan, Adana, Turkey

effect on the quality of daily life (Park et al. 2019). In this sense, accurate estimation of AAT is essential for various decision-making fields, involving energy, tourism, transportation, and agricultural management (Venkadesh et al. 2013). Besides, accurate AAT estimation is of paramount importance in environmental studies which include the operational eco-environmental system. Considering the industrial perspective, accurate AAT prediction is an essential component of the energy management strategy to keep the indoor temperature within a certain comfort range and thus reduce energy consumption (Zhang et al. 2018).

Literature survey shows that accurate estimation of the AAT is an important indicator in many decision-making areas such as energy management, agricultural production, and environmental health (Radhika and Shashi 2009; Bilgili and Sahin 2009; Chevalier et al. 2011; Kisi and Shiri 2014; Zhang et al. 2014; Ramesh and Anitha 2014; Kisi and Sanikhani 2015; Salman et al. 2018; Sekula et al. 2019; Azad et al. 2020). In recent years, due to its impact on global human life, many researches on weather prediction have been carried out by various research communities (Venkadesh et al. 2013). For example, Kalogirou (Kalogirou 2001) depicted that the estimated variations in weather data, namely, solar radiation, air temperature, relative humidity, wind speed, and rainfall, are essential for the renewable energy industry to run the cost estimation, analyze the performance, and design the systems of renewable energy. Cobaner et al. (Cobaner et al. 2014) stated that the correct estimation of AAT is extremely important for various aspects of life. They also explained that the increase in air temperature dependent on global warming adversely affects all living things in the environment. Moreover, it can also give rise to many variations in steady natural phenomena, just as the average sea-level rise. Zahroh et al. (Zahroh et al. 2019) studied of predicting daily temperature with high accuracy contributing important information for the society and decision-makers. Zhang et al. (Zhang et al. 2014) suggested that the estimation of the AAT time series of an environment is necessary and essential for heating, ventilation and air conditioning (HVAC) systems to manage the optimum energy-saving operations in buildings.

Those above-mentioned literature studies indicate that the importance of AAT estimation and predicting methods is gradually increasing all over the world (Kisi and Shiri 2014). Data-driven method and physical process-driven method are the two main categories according to the AAT prediction methods. The physical process-driven method uses various observational parameters, such as solar radiation, clouds, pressure, and terrain data, for predicting (Radhika and Shashi 2009; Bilgili and Sahin 2009; Chevalier et al. 2011; Kisi and Shiri 2014; Zhang et al. 2014; Ramesh and Anitha 2014; Kisi and Sanikhani 2015; Sekula et al. 2019; Azad et al. 2020). Strong interpretability and high predictive accuracy are the most important advantages of this method. The disadvantages of this method are the complexity of the data modeling, difficulty of the data collection, and time-consuming solutions (Kalogirou 2001). Zhang et al. (Zhang et al. 2014) proposed the physical process-driven method for predicting of the AAT time series, which is dependent on the interrelationship among heat convection, clouds and ground heat radiation, direct solar radiation, and air temperature. Sekula et al. (Sekula et al. 2019) improved the accuracy of air temperature predictions using a physical process-driven method, namely, the ALADIN-HIRLAM numerical weather prediction (NWP) system.

The data-driven methods use statistical approaches including machine learning models, time series models, and hybrid models to predict AAT from past atmospheric air temperature data. Many researchers have developed models such as support vector regression (SVR) (Radhika and Shashi 2009; Chevalier et al. 2011; Ramesh and Anitha 2014; Kisi and Sanikhani 2015) and multivariate adaptive regression spline (MARS) (Ramesh and Anitha 2014) to predict AAT using time series. In addition to these time series models, some researchers have implemented machine learning methods such as artificial neural networks (ANNs) (Bilgili and Sahin 2009; Venkadesh et al. 2013; Kisi and Shiri 2014; Cobaner et al. 2014; Kisi and Sanikhani 2015; Fahimi Nezhad et al. 2019), adaptive neuro-fuzzy inference system (ANFIS) (Kisi and Shiri 2014; Cobaner et al. 2014), linear regression analysis (Meshram et al. 2020), and gene expression programming (GEP) (Kisi and Sanikhani 2015) to predict the AAT. Furthermore, some new hybrid methods are proposed in the literature to get better accuracy results in estimating AAT. ANFIS optimized with the particle swarm optimization (ANFIS-PSO), ANFIS optimized with the genetic algorithm (ANFIS-GA), ANFIS optimized with the ant colony optimization for continuous domains (ANFIS-ACO_R), ANFIS optimized with the differential evolution (ANFIS-DE) (Azad et al. 2020), ANFIS with subtractive clustering (ANFIS-SC), and ANFIS with grid partition (ANFIS-GP) (Kisi and Sanikhani 2015) are the examples introduced in previous studies.

In recent years, the deep learning approaches have been successfully applied to the prediction studies that need high-accuracy solutions (Sariturk et al. 2020; Benbahria et al. 2021). Therefore, they have received intense attention of researchers from variety of disciplines. In the deep learning process, recurrent neural networks (RNNs) are well suited in solving the sequence problems such as time series data. However, the extended sequence length causes long-term dependency problems (Kalogirou 2001). The deep learning concept originates from the study and structure of ANNs. A multilayered perceptron with multiple hidden layers can be thought of as a kind of deep learning structure. Low-level features can be connected with deep learning to create higher-level ones, referring to an attribute category or feature to explore the distributions of data features. Compared to the traditional

neural networks, deep learning is successful in solving the problems of slow training speed and over-adaptation. The long short-term memory (LSTM) network refers to a special form and development of the RNN. However, it has a much better universality than that of traditional RNN (Arslan and Sekertekin 2019). The LSTM has been successfully applied in various fields, especially time sequence problems, such as wind speed prediction (Balluff et al. 2015; Liu et al. 2018a, b; Chen et al. 2018, 2019a, b; Hu and Chen 2018; Huang et al. 2018; Wang and Li 2018; Yu et al. 2018; Liang et al. 2018; Zhang et al. 2019b, c, d), wind turbine power prediction (Qu et al. 2016; Wu et al. 2016; López et al. 2018; Dong et al. 2018; Shi et al. 2018), sea surface temperature prediction (Zhang et al. 2017), dam displacement prediction (Zhang et al. 2019a), electricity price prediction (Peng et al. 2018), solar irradiance prediction (Qing and Niu 2018), weather prediction (Zaytar and El Amrani 2016), electrical load prediction (Muzaffar and Afshari 2019), wind turbine signal forecasting (Qin et al. 2019), traffic speed and flow prediction (Ma et al. 2015; Tian and Pan 2015), prediction of $PM_{2.5}$ (particulate matter with a diameter of less than 2.5 μm) and PM_{10} concentration (Zhao et al. 2019; Tong et al. 2019; Kim et al. 2019), ozone concentration prediction (Pak et al. 2018), photovoltaic power generation prediction (Zhou et al. 2019a; Han et al. 2019), air pollutant concentration predictions (Li et al. 2017; Krishan et al. 2019; Xu et al. 2020), pixel-based land surface temperature prediction (Arslan and Sekertekin 2019), train speed prediction (Li et al. 2019), reservoir inflow prediction (Qi et al. 2019), remaining useful life prediction for super capacitor (Zhou et al. 2019b), and prediction interval of wind power (Yuan et al. 2019).

In this study, a data-driven method, namely the LSTM, which is a recent methodology of the time series prediction, is proposed for the time series prediction of AAT. The 10-min, hourly, and daily interval AAT time series data from the two measurement stations in the Mediterranean Region of Turkey are used as sample data. The main contribution and novelty of this study to the literature can be primarily expressed as follows: (i) An LSTM neural network is proposed to forecast the 10-min ahead, 1-h ahead, and 1-day ahead AATs since there are limited numbers of studies on the forecasting of AAT using the LSTM. (ii) In order to reveal the effectiveness of the LSTM, the LSTM results are compared with the findings of the adaptive neuro-fuzzy inference system (ANFIS) with fuzzy C-means (FCM) models and autoregressive moving average (ARMA) model. (iii) We also analyzed and presented the results of changing input variables (hidden layer number and epoch number) of LSTM that the previous works did not provided. (iv) Finally, the AAT prediction results of the 10-min ahead. the 1-h ahead, and the 1-day ahead are compared to show which data set is more feasible with the corresponding methods in terms of the short-term AAT prediction. The structure of this paper is organized as follows: The "MethodsS13" section describes the LSTM, ARMA, and ANFIS methodologies, and statistical metrics used to evaluate the efficiency of the methods. The "Results and discussion" section presents the experimental results including the description of the study area and material, and the analyses of LSTM results for both stations. Finally, the "Conclusion" section represents the conclusion part.

Methods

In this section, LSTM, ANFIS-FCM, and ARMA models were introduced. Concerning the software used for implementing the proposed models, LSTM and ANFIS-FCM models were applied using MATLAB, whereas ARMA was performed using the trial version of NUMXL, which is designed as an add-in for Microsoft Excel. Every method has its own advantages and disadvantages. ANFIS takes the advantages of two methods, namely, ANN and Takagi–Sugeno fuzzy inference systems, in one system that makes it a universal prediction method. On the other hand, LSTM uses its own gate-based deep learning abilities that make it better than ANNs and conventional RNNs. Concerning the ARMA model, it is combined version of auto-regressive (AR) or moving average (MA) models, and it is one of the widely used estimators for short-term predictions. The comparison of these three methods in AAT prediction will guide the researchers to choose the best of them for further identical studies.

Long short-term memory deep neural network model

LSTM, a special kind of RNNs, was firstly announced by Hochreiter and Schmidhuber (Hochreiter and Schmidhuber 1997) to solve vanishing and exploding issues in error signals of RNNs, which flow backwards in time. Therefore, it was constructed to model the sequences and their long-range dependencies better than the conventional RNNs (Zhang et al. 2017; Arslan and Sekertekin 2019). Figure 1 illustrates the architecture of the LSTM network, which is formed by the repeating and interconnected LSTM units, and the flow of data at time step t. Figure 1b depicts the structure of one of the LSTM units obtained from Fig. 1a.

In Fig. 1, the sequence data $x = (x_1, x_2, x_3, \ldots, x_t)$ refers to input variable to retrieve the hidden (output) state $h = (h_1, h_2, h_3, \ldots, h_t)$ and cell state $c = (c_1, c_2, c_3, \ldots, c_t)$. The first value of the sequence of x (x_1) is utilized in the first LSTM unit to obtain the first value of the hidden state (h_1) and the first updated cell state (c_1). At time step t, c_{t-1} and h_{t-1} fed the

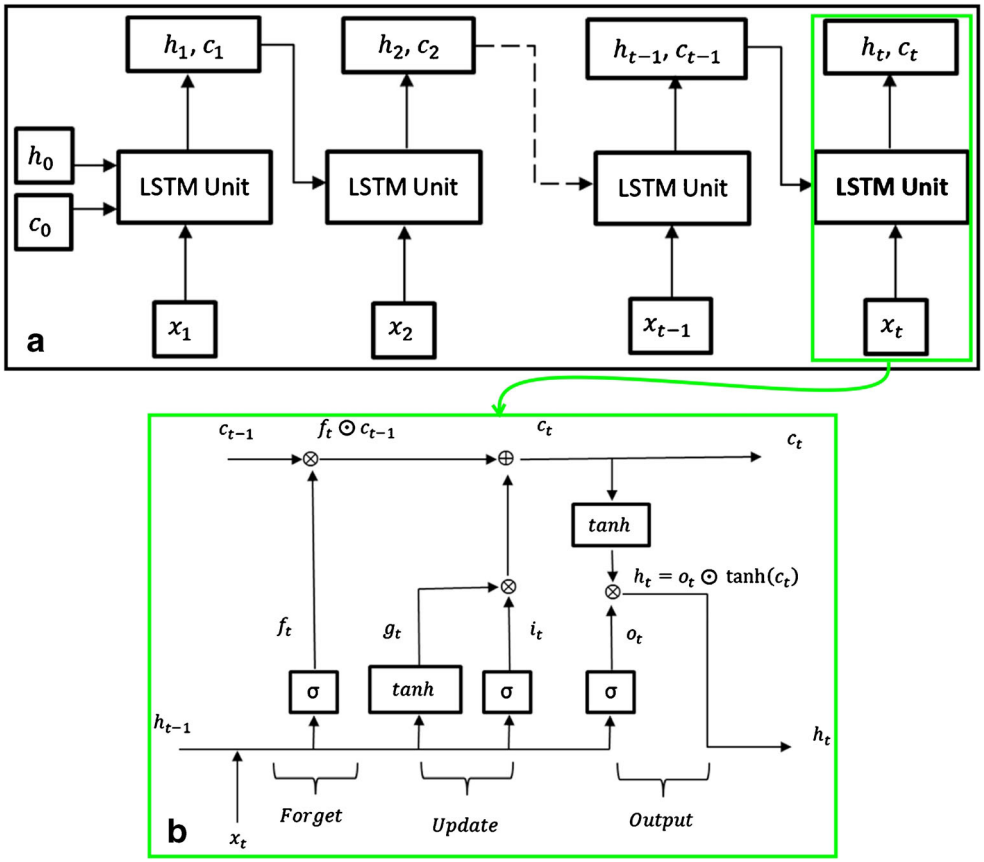

Fig. 1 The architecture of the LSTM network (a) and the flow of data at time step t (b)

LSTM unit to obtain h_t and c_t. The hidden state (h_t) at time t with cell state (c_t) can be calculated by:

$$h_t = o_t \odot \tanh(c_t) \tag{1}$$

where $\odot$ represents the Hadamard product (element-wise multiplication of vectors) and o_t refers to the output gate. The cell state manages the LSTM network by adding or removing information to it via the gates. The cell state (c_t) the time step t consists of information from the previous units and can be computed as follows (Hochreiter and Schmidhuber 1997; Mathworks 2019):

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{2}$$

In Eq. (2), input gate (i_t) controls the level of cell state update. Forget gate (f_t) manages the level of cell state reset. Cell candidate (g_t) adds information to the cell state, and output gate (o_t) controls the level of the cell state added to the hidden state. The variables i_t, f_t, g_t, and o_t can be computed using the following equations, respectively (Hochreiter and Schmidhuber 1997; Mathworks 2019).

$$i_t = \sigma(W_i x_t + R_i h_{t-1} + b_i) \tag{3}$$

$$f_t = \sigma(W_f x_t + R_f h_{t-1} + b_f) \tag{4}$$

$$g_t = \tanh(W_g x_t + R_g h_{t-1} + b_g) \tag{5}$$

$$o_t = \sigma(W_o x_t + R_o h_{t-1} + b_o) \tag{6}$$

where σ refers to the sigmoid function obtained by $\sigma(x) = (1 + e^{-x})^{-1}$. W is the input weights. R represents the recurrent weights and b is the bias term. All of these parameters can be given by Eq. (7).

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \tag{7}$$

In this study, the optimal number of hidden layers and epoch number (iteration size) in the training step of LSTM was determined by trial and error method as detailed in section "Determination of parameters."

Adaptive neuro-fuzzy inference system

Jang (1993) presented the learning structure and the architecture of ANFIS. According to Jang (1993), the ANFIS model functionally has the same structure as the Takagi-Sugeno type extraction model. An ANFIS model is a combination of two statistical systems, namely, FIS and ANN. ANFIS can model the dynamics of difficult systems in a best way. At first, the dataset is trained by ANFIS model in a similar manner as

ANN. The trained system is then operated as a FIS. ANFIS combines both ANN and FIS bases taking into account the advantages of both systems into a single system (Karakuş et al. 2017).

The neuro-fuzzy model in Tabari et al. (2012) is given in Fig. 2 representing a typical ANFIS architectural structure. It has a total of five layers as a fuzzy system based on multi-layered neural network. In the structure of this network, input and output nodes refer to input condition states and output response, respectively, as well as nodes in hidden layers that act as rules and MFs. Thus, this phenomenon removes the drawback of a traditional feed forward multi-layered network that is challenging for a supervisor to figure out or change. In this structure, a circle shape stands for a fixed node while a square shape represents an adaptive node. Simply, in Fig. 2, x and y are two inputs, and z refers to one the output. Thanks to its built-in optimal and adaptable techniques, computational efficiency, and high interpretability, the Sugeno model is the most widely applied among many fuzzy models in the literature. Two fuzzy if - then rules depending on the first order Sugeno fuzzy model can be given by:

$$\text{Rule1}: \text{if } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{then } z_1 = p_1 x + q_1 y + r_1 \quad (8)$$

$$\text{Rule2}: \text{if } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{then } z_2 = p_2 x + q_2 y + r_2 \quad (9)$$

where B_i and A_i express the fuzzy clusters in the first one, while r_i, q_i, and p_i define the architectural parameters in the training process. Figure 2 illustrates that five layers, detailed below, constitute the generalized ANFIS structure with two input and single output (Tabari et al. 2012):

Layer 1: This layer defines an input variable for each appropriate fuzzy set. The function by which membership degrees are created using MFs is indicated by the nodes in this layer. The node function of a node i is defined as follows:

$$O_i^1 = \mu_{A_i}(x), \quad i = 1, 2 \quad (10)$$

$$O_i^1 = \mu_{B_{i-2}}(y), \quad i = 3, 4 \quad (11)$$

where μ_{A_i} and μ_{B_i} are the MFs. In this layer, the parameters are defined as preliminary parameters.

Layer 2: In the nodes of layer 2, incoming signals are duplicated and the product out is sent. In this layer, all of the nodes compute the firing strength of a rule by multiplying the incoming signals.

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1, 2 \quad (12)$$

Layer 3: The ith node in layer 3 computes the ratio of the ith rule's firing strength to the sum of all rules firing strengths of all rules:

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \quad (13)$$

where $\overline{w}_i$ means the normalized firing strengths.

Layer 4: Node i in Layer 4 determines how the ith rule will contribute to the model output, via the node function in Eq. (14):

$$O_i^4 = \overline{w}_i z_i = \overline{w}_i (p_i x + q_i y + r_i), \quad i = 1, 2 \quad (14)$$

where $\{r_i, q_i, \text{and } p_i\}$ is the parameter set, and $\overline{w}_i$ is the output of the layer 3. In this layer, the parameters are called the consequent parameters.

Layer 5: This layer is named the output nodes where the single node calculates the overall output considering the sum of all incoming signals as in Eq. (15) (Tabari et al. 2012):

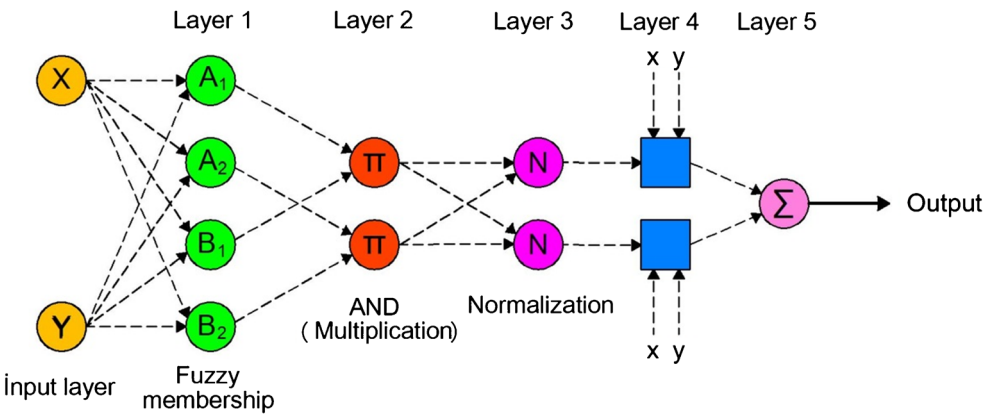

Fig. 2 ANFIS architecture structure (Tabari et al. 2012)

$$O_i^5 = \sum_{i=1}^{2} \overline{w}_i z_i = \frac{w_1 z_1 + w_2 z_2}{w_1 + w_2} \tag{15}$$

Consequently, the output z in Fig. 2 is obtained as:

$$Z = \left(\overline{w}_1 x\right) p_1 + \left(\overline{w}_1 y\right) q_1 + \left(\overline{w}_1\right) r_1 + \left(\overline{w}_2 x\right) p_2 + \left(\overline{w}_2 y\right) q_2 + \left(\overline{w}_2\right) r_2 \tag{16}$$

More information about ANFIS can be reached from the study of (Jang 1993). ANFIS-FCM model, used in this study, is obtained by combining ANFIS and fuzzy *C*-means (FCM) clustering method, which is an improvement and modification of *K*-means clustering (Bezdek 1981). Indeed, the principle of this model is minimizing an objective function that defines the distance from any given data point to a cluster center (Abdulshahed et al. 2015). This distance is weighted by the value of MFs of the data point (Karahoca and Karahoca 2011). Figure 3 represents the ANFIS structure with single input and single output that is utilized for time series forecasting, and this structure was applied in this study. As input and output, historical AAT and predicted AAT were considered, respectively.

Autoregressive moving average

Autoregressive moving average (ARMA) models are widely preferred in time series modeling. They provide practical linear model of stationary time series since they are capable of modeling the unknown process with the minimum number of parameters (Zhang and Moore 2015). These models are effective in very short-term forecasting because of their ability to extract the output power with an acceptable level of accuracy (Rajagopalan and Santoso 2009; Massaoudi et al. 2021). However, since the input data are not always stationary, these models provide low accuracy when modeling the weather parameters with high variability (Massaoudi et al. 2021). Compared to using only auto-regressive (AR) or moving average (MA) models, ARMA*(p, q)* models are used to describe the dynamic structure of the historical data with less parameter (Güldal and Tongal 2010). Concerning the mathematical notation of the ARMA*(p, q)*, in Eq. (17), S_i refers to the standardized observed time series, and p and q are the degree of the AR and MA models, respectively.

$$S_i = \sum_{j=1}^{p} \phi_j S_{i-j} + \varepsilon_i - \sum_{j=1}^{q} \theta_j \varepsilon_{i-j} \tag{17}$$

where the right part denotes autoregressive part and the left one is the moving average part, and ϕ_j and θ_j are the autoregressive and moving average parameters, respectively (Box and Jenkins 1976). In order to determine the p and q variables, the autocorrelation function plot (ACF) and partial autocorrelation plot (PACF) over several quarterly lags suggests which autoregressive and moving average terms should be included in the ARMA model. The ACF can be used as an indicator of the MA degree q, while PACF provides information about the AR degree p (Loganathan and Ibrahim 2010).

Error analysis

In our study, to determine the effectiveness of the LSTM in estimating AAT time series data, four statistical metrics such as mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), correlation coefficient (R), Nash–Sutcliffe efficiency coefficient (NSE) (Nash and Sutcliffe 1970), and average bias are considered. The relevant statistical metrics can be calculated as follows:

Mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |p(i) - o(i)| \tag{18}$$

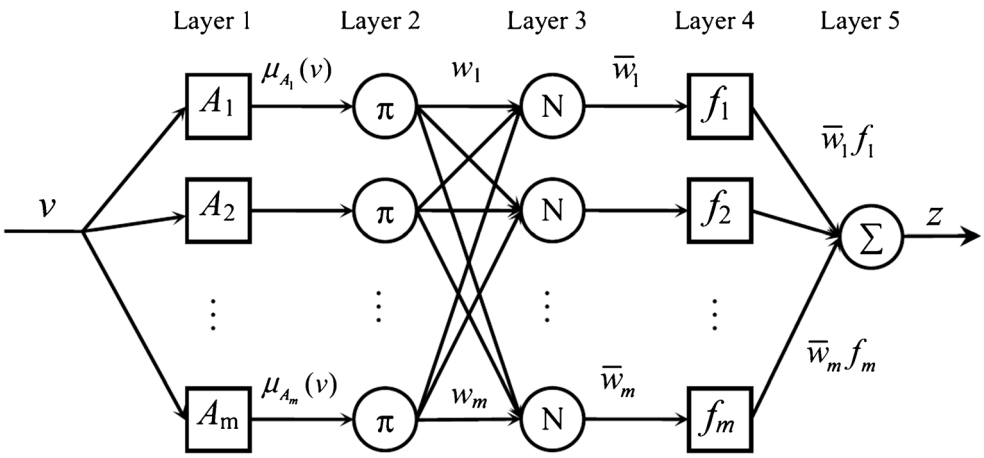

Fig. 3 The ANFIS structure with single input and single output used for time series data prediction (Rezaie et al. 2018)

Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[p(i)-o(i)]^2} \tag{19}$$

Mean absolute percentage error (MAPE):

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|p(i)-o(i)|}{o(i)}x100 \tag{20}$$

Correlation coefficient (R):

$$R = \left(\sum_{i=1}^{N}\left[p(i)-\overline{p}\right]\left[o(i)-\overline{o}\right]\right) / \left(\sqrt{\sum_{i=1}^{N}\left[p(i)-\overline{p}\right]^2}\sqrt{\sum_{i=1}^{N}\left[o(i)-\overline{o}\right]^2}\right) \tag{21}$$

Nash–Sutcliffe efficiency coefficient (NSE)

$$\mathrm{NSE} = 1-\frac{\sum_{i=1}^{N}(p(i)-o(i))^2}{\sum_{i=1}^{N}\left(o(i)-\overline{o}\right)^2} \tag{22}$$

Bias:

$$\mathrm{Bias} = \frac{1}{N}\sum_{i=1}^{N}[p(i)-o(i)] \tag{23}$$

where $p(i)$ and $o(i)$ represent the predicted and observed AAT at the time i, respectively. Besides, $\overline{p}$ and $\overline{o}$ stand for the mean of the predicted AAT and the observed AAT, respectively. N refers to the total number of AAT data.

Results and discussion

Study area and material

In this study, three different AAT time series data from two stations in the Mediterranean Region of Turkey, namely, Belen station and Mersin station, were utilized for evaluating the prediction performance of the LSTM network. Figure 4 represents the locations of the stations. Belen station and Mersin station are located in Hatay and Mersin provinces of Turkey, respectively. Besides, the stations are situated at geographical, economical, and cultural Çukurova region, which covers Adana, Mersin, Hatay, and Osmaniye provinces. The AAT data retrieved from the stations cover 10-min interval, hourly, and daily acquisitions, respectively. The AAT data from Belen station include 743 data in total (hourly data from 01 July 2016 to 31 July 2016), whereas the AAT data from Mersin station comprise of 37862 data in total (10-min interval data from 04 October 2014 to 24 June 2015). Furthermore, the daily AAT data series for Mersin station (730 daily data from January 1, 2018, to December 31, 2019) are utilized for 1-day ahead short-term AAT prediction.

Determination of parameters

In the experimental design, the percentage of the training and the testing/prediction data were defined by Pareto principle (Box and Meyer 1986) and assumed as 80% and 20%, respectively, for the AAT data of both stations. As stated in section "Study area and material," Belen station has 743 hourly data and Mersin station has 730 daily and 37862 10-min interval data. Concerning the Belen station, 594 hourly data were utilized in the training step of the methods and 149 hourly data were managed in the prediction step. On the other hand, considering the Mersin station, 30290 10-min interval data were included for training and 7572 10-min interval data were utilized for prediction. Lastly, the total of 730 daily data for the Mersin station were split into two parts, first as 80% training set and the other 20% testing set. To determine the optimal parameters of the LSTM network and ANFIS models for AAT prediction, the trial and error method was taken into consideration to retrieve minimum prediction errors. Therefore, the optimal parameters, namely, the number of the epoch and hidden layer were determined by the trial and error based on RMSE results. The optimal hidden layer and epoch number values for 10-min interval, hourly, and daily AAT predictions are highlighted in bold in Tables 1, 2, and 3, respectively. On the other hand, p and q values of ARMA were determined as 1 and 0, referring to ARMA(1,0), after interpretation of ACF and PCF plots. ARMA(1,1), ARMA(1,2), and ARMA(2,2) results were also obtained;

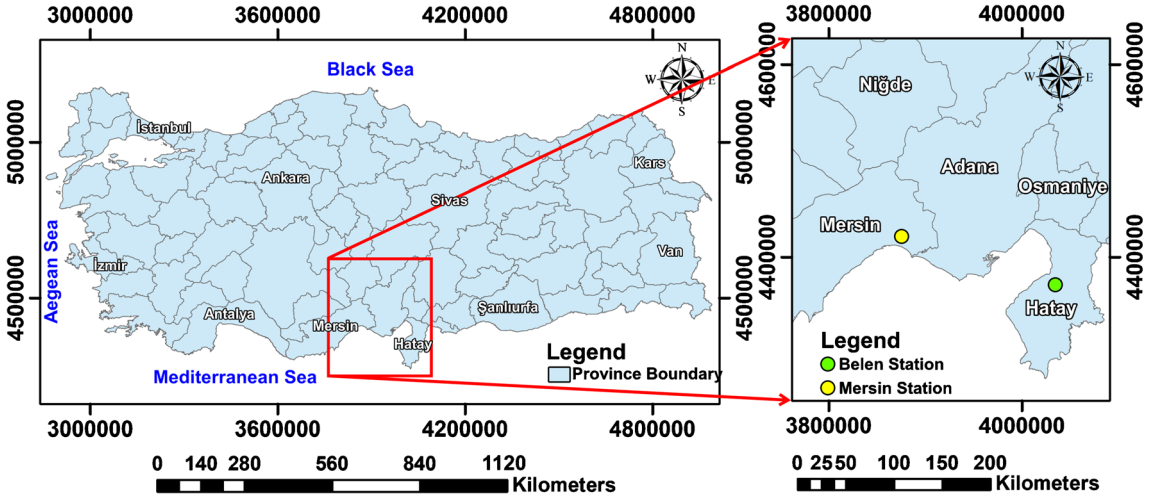

Fig. 4 Illustration of the stations' locations in the Coordinate System of World Geodetic System 1984 (WGS84)

however, none of them provided better results than ARMA(1,0). In all three models (LSTM, ANFIS-FCM, and ARMA), single input and single output data as AAT data were utilized.

LSTM prediction results for the 10-min interval AAT data of Mersin station

In Table 1, the selected hidden layer number and epoch number parameters for the 10-min interval AAT data of Mersin station range between 50 and 550. As seen from the table, the RMSE values were obtained between 0.3182 and 2.5752 °C. As the hidden layer size and iteration number are increased, the RMSE values decreased. It can be said that increasing of the hidden layer contributes to the prediction results with the improvement of the solution. The LSTM network model applied to 30290 samples with the hidden layer number 25 and iteration number (Epoch number) 600 gives a better result than the other selected parameters with the RMSE value of 0.3182 °C. As the number of samples for the input increased in the LSTM network model, it needs more hidden layer and iteration number to decrease the RMSE values. The number of hidden layer, epoch, and sample directly affects the quality of the solution and should be carefully taken into account in trial and error analysis stage.

Table 1 Determination of optimal parameters for 10-min interval AAT prediction from Mersin station using LSTM. The best results are shown in bold

Hidden layer number	Epoch number	RMSE (°C)	Hidden layer number	Epoch number	RMSE (°C)
5	100	2.5752	20	650	0.3194
5	200	1.0668	20	700	0.3195
5	300	0.7714	**25**	**600**	**0.3182**
5	400	0.7402	30	400	0.3289
5	500	0.7636	30	500	0.3225
10	200	0.4222	40	400	0.3615
10	300	0.3607	50	500	0.3330
10	400	0.3344	50	50	0.6302
10	500	0.3445	75	75	0.4883
10	600	0.3544	100	100	0.5332
15	500	0.3212	150	150	0.5774
20	300	0.3478	200	200	0.4562
20	400	0.3215	250	250	0.3790
20	450	0.3453	300	300	0.4301
20	500	0.3204	400	400	0.3703
20	550	0.3201	500	500	0.3492
20	600	0.3197	550	550	0.3585

Table 2 Determination of optimal parameters for hourly AAT prediction from the Belen station using LSTM. The best results are shown in bold

Hidden layer number	Epoch number	RMSE	Hidden layer number	Epoch number	RMSE
5	140	0.8414	140	140	0.6263
5	300	0.7201	140	150	0.6324
10	300	0.7119	150	150	0.6402
20	300	0.6971	160	160	0.7149
30	300	0.6502	175	175	0.6427
40	300	0.7434	200	140	0.6425
50	300	0.7845	200	200	0.8176
100	300	0.7370	250	140	0.6136
100	100	0.7094	250	250	0.8943
100	**140**	**0.6117**	300	140	0.6991
125	125	0.6431	300	300	0.7560
130	130	0.6502	400	400	0.7724
135	135	0.6377	500	500	0.9177

Figure 5a shows all observed data (blue) and predicted data (red) for the 10-min AAT data of Mersin station. The 10-min AAT values vary between −8 and 29 °C. Figure 5b illustrates the zoom in the original 10-min interval AAT values with blue and the predicted AAT values with red (top), and the difference between the observed and the predicted AAT values (bottom). The difference between the observed and predicted AAT values ranges between −2.0 and 2.2 °C. The top graph of Fig. 5b demonstrates how perfectly the 10-min observed and predicted AAT values are overlapping. The higher prediction errors are observed in extreme AAT values. The regression plot of the observed and predicted 10-min interval AAT are given in Fig. 5c to reveal the distribution of the data.

LSTM prediction results for the hourly AAT data of Belen station

In Table 2, the selected hidden layer number and epoch number parameters for Belen station range between 5 and 500 and RMSE value changes between 0.61 and 0.92 °C. The results are remarkable with these values. As the hidden layer size is increased, e.g., 300, the RMSE values are getting higher. It can be said that increasing of the hidden layer does not contribute to improvement of results for this number of sample. The LSTM network model applied to 594 samples with the hidden layer number 100 and iteration number (epoch number) 140 gives better result than for the other selected parameters. Therefore, it is important in the stage of LSTM network

Table 3 Determination of optimal parameters for daily AAT prediction from Mersin station using LSTM. The best results are shown in bold

Hidden layer number	Epoch number	RMSE	Hidden layer number	Epoch number	RMSE
5	100	1.6736	400	400	1.4652
5	150	1.3559	450	450	1.6371
5	**200**	**1.3343**	500	500	1.4636
5	250	1.3634	550	550	2.3575
5	300	1.3538	600	600	2.6079
5	400	1.3535	650	650	2.2541
5	500	1.3460	700	700	3.4160
5	600	1.3481	400	200	1.4364
10	200	1.3613	400	300	1.4186
10	300	1.3858	400	350	1.5204
50	50	1.9715	400	450	1.6364
100	100	1.4407	400	500	1.6253
150	150	1.4461	200	400	1.6372
200	200	1.7107	300	400	2.0635
250	250	1.8499	350	400	1.3885
300	300	1.5788	450	400	1.4274
350	350	2.0725	500	400	1.4549

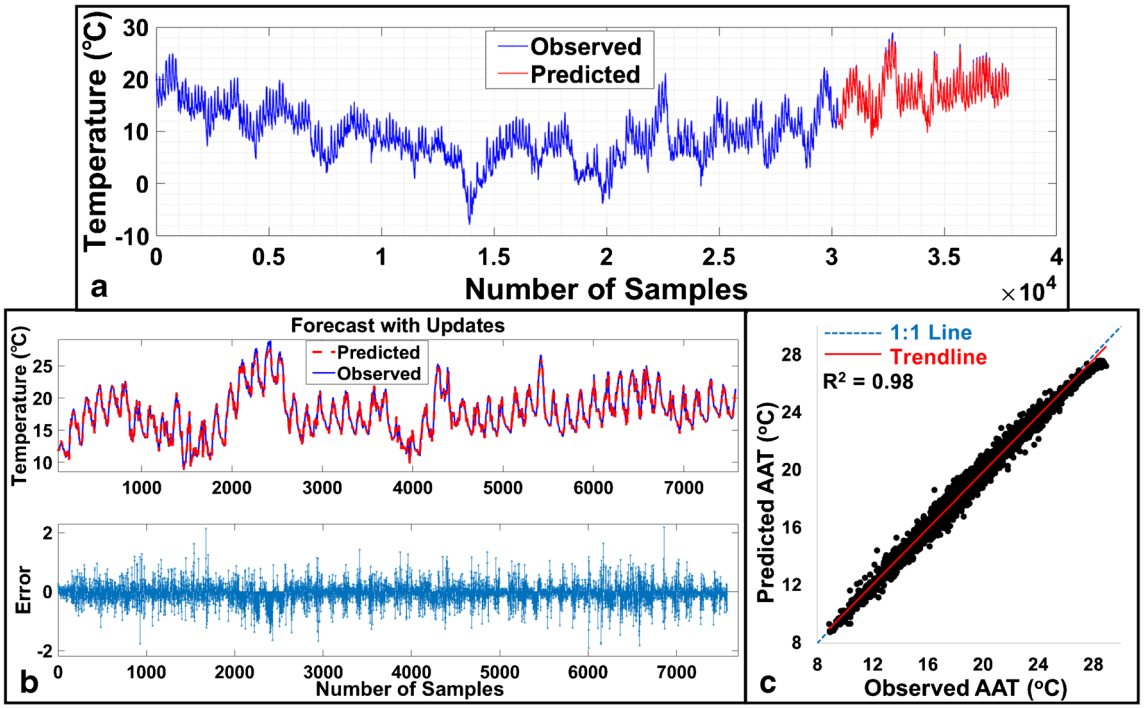

Fig. 5 The 10-min interval time series of AAT data of Mersin station with observed (blue) and predicted values (red) (a). Zoom in the observed and predicted data (top) and the difference between the observed and predicted data (bottom) (b). Regression plot of the observed and predicted data (c)

design to select the beginning and ending limits of these parameters.

Figure 6a shows all observed data (blue) and predicted data (red) for the hourly AAT data of Belen station. The hourly

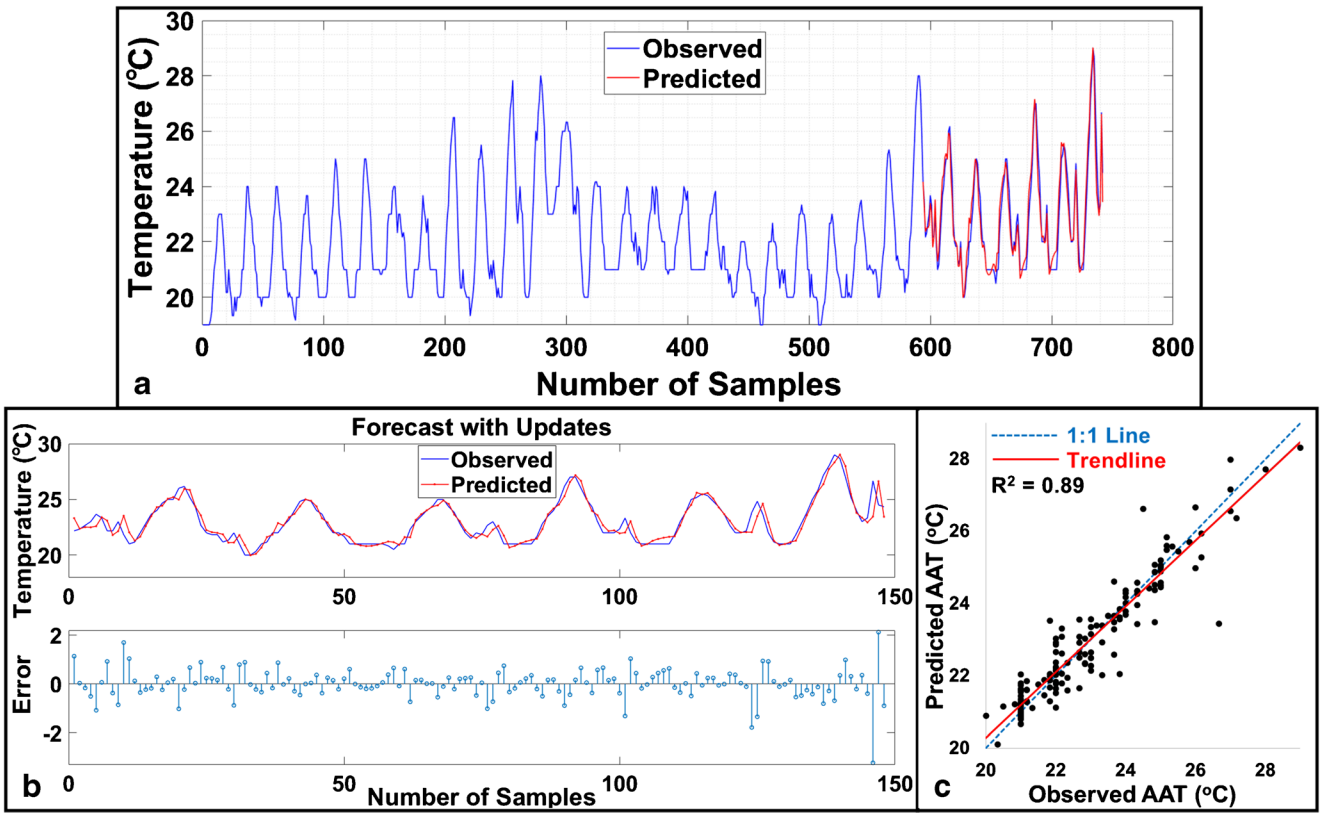

Fig. 6 The hourly time series of AAT data of Belen station with observed (blue) and predicted values (red) (a). Zoom in the observed and predicted data (top) and the difference between the observed and predicted data (bottom) (b). Regression plot of the observed and predicted data (c)

AAT values vary from 19 to 29 °C. Figure 6b reveals the zoom in the original AAT values with blue and the predicted AAT values with red (top), and the difference between the observed and predicted AAT values (bottom). The difference between the observed and predicted AAT values ranges between −2.8 and 2.1°C. It is clear from Fig. 6b that LSTM predicted hourly AAT values very successfully. The prediction errors are getting higher in the sudden changes in the time series. The regression plot of the observed and predicted hourly AAT is given in Fig. 6c to reveal the distribution of the data which supports the previous findings. Furthermore, the coefficient of determination (R^2) is presented in Fig. 6c. The observed data and the predicted values estimated by the LSTM network model almost coinciding with each other. The data has a pattern of periodical sinus wave every 200 h. Moreover, daily temperature changes also can be observed in the AAT data. Both of these patterns are well modeled with LSTM network model for future prediction. The R^2, equal to 0.89, shows the good fit of the observed and predicted data.

LSTM prediction results for the daily AAT data of the Mersin station

The selected hidden layer number and epoch number parameters for daily AAT data of Mersin station range between 5 and 700 as can be seen from Table 3. RMSE values were obtained between 1.3343 and 3.4160 °C. The LSTM network model applied to 730 samples with the hidden layer number 5 and iteration number (epoch number) 200 gives a better result than the other selected parameters with the RMSE value of 1.3343 °C. As seen, increasing the number of hidden layers and iterations does not decrease the error rate. To obtain the lowest RMSE value, 5 hidden layer and 200 epoch numbers were sufficient.

Figure 7a illustrates all observed (blue) and predicted (red) data of daily AAT for Mersin station. The daily AAT values vary between 17.8 and 42.2 °C. Figure 7b represents daily original AAT values with blue and the predicted AAT values with red (top), and the difference between the observed and the predicted AAT values (bottom). The difference between the observed and predicted AAT values ranges between −3.3 and 5.0 °C. The regression plot of the observed and predicted daily AAT are given in Fig. 7c to reveal the distribution of the data. The R^2, equal to 0.94, proves the good fit of the observed and predicted data.

Comparison of the 10-min, hourly, and daily AAT predictions

Considering different ANFIS-FCM architectures, the most suitable model structures, number of MFs, number of iterations (epoch number), and transfer functions were identified

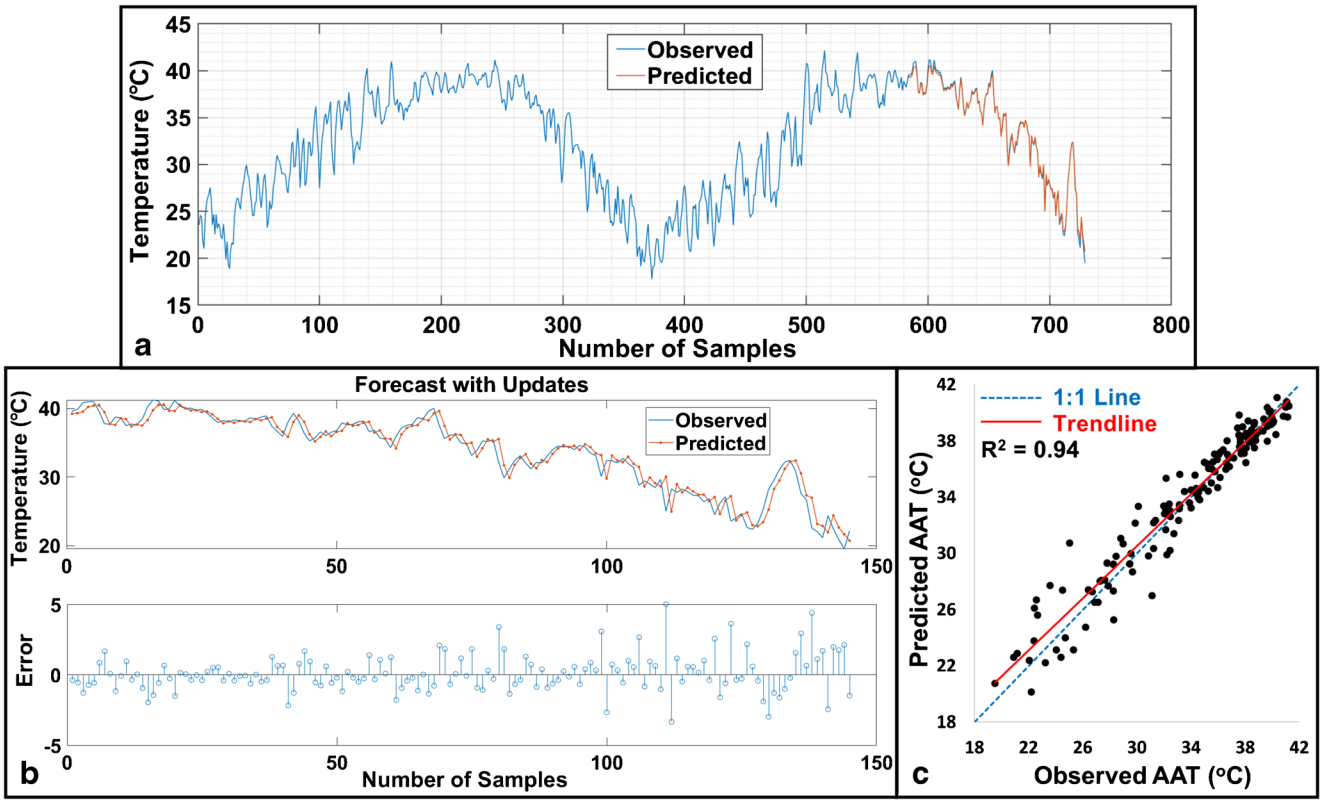

Fig. 7 The daily time series of AAT data of Mersin station with observed (blue) and predicted values (red) (a). Zoom in the observed and predicted data (top) and the difference between the observed and predicted data (bottom) (b). Regression plot of the observed and predicted data (c)

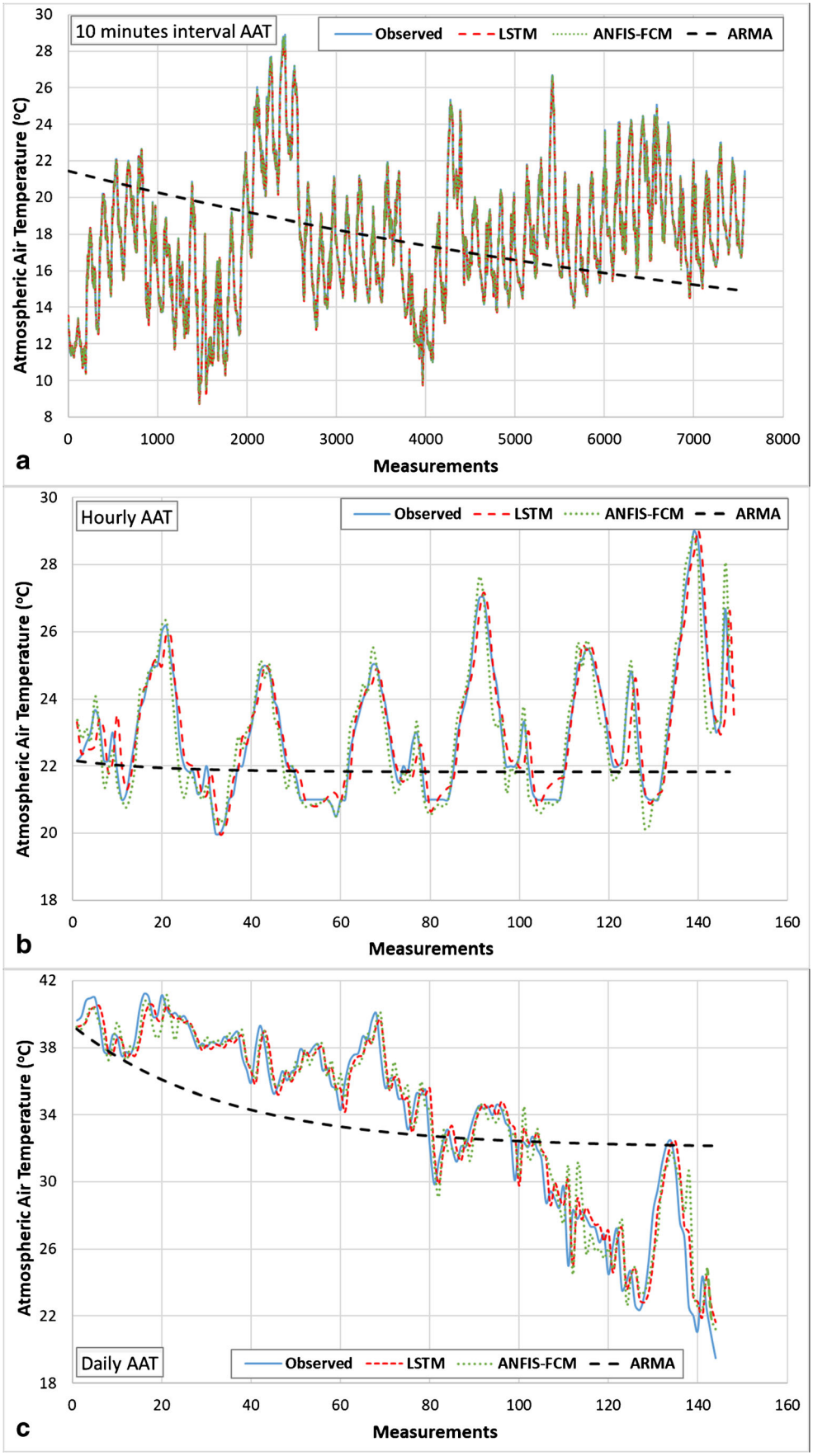

Fig. 8 Comparison of observed 10-min interval (a), hourly (b), and daily (c) atmospheric air temperature values with LSTM, ANFIS-FCM, and ARMA results

Table 4 The best statistical accuracy results of 10-min interval, hourly, and daily AAT predictions for LSTM and ANFIS-FCM models

Forecasting interval	Forecasting method	Error criteria					
		Bias (°C)	MAE (°)	RMSE (°)	MAPE (%)	*R*	NSE
10-min interval	LSTM	0.0297	0.2252	0.3182	1.2719	0.9958	0.9916
	ANFIS-FCM	**0.0176**	**0.2181**	**0.3104**	**1.2364**	**0.9959**	**0.9920**
	ARMA	−0.0567	3.5019	4.3342	21.2253	−0.2379	−0.5495
Hourly	LSTM	**−0.0130**	**0.4275**	**0.6117**	**1.8471**	**0.9454**	**0.8934**
	ANFIS-FCM	0.1205	0.5395	0.7564	2.3200	0.9257	0.8378
	ARMA	1.1436	1.6262	2.2050	6.6771	−0.0710	−0.3779
Daily	LSTM	**−0.1161**	**0.9870**	**1.3343**	**3.2673**	**0.9705**	**0.9411**
	ANFIS-FCM	−0.0803	1.0853	1.5203	3.6132	0.9608	0.9235
	ARMA	0.0104	3.4997	4.3953	11.7672	0.7026	0.3566

using the trial and error method. In the construction part, each ANFIS model was tested and the statistical metrics were utilized to choose the best results. Gaussmf and linear membership functions were used as input and output MFs, respectively. The number of MFs ranged from 2 to 6, and the number of iterations varied between 50 and 300 in the processing phase. Applying fuzzy technique alone has a disadvantage that the rules should be written by an expert. However, ANFIS has the advantage over fuzzy technique that the rules can be created automatically by the input–output relation (Akbari et al. 2018). Al-Hmouz et al. (Al-Hmouz et al. 2012) also stated that the ANFIS model performance is accurate, where some rules are covered by human expert training data set and the missing rules are detected by the ANFIS.

Figure 8 represents the comparison of the observed 10-min interval, hourly, and daily AAT values with LSTM, ANFIS-FCM, and ARMA results. Figure 8a contains 7572 10-min interval testing data, and 7b contains 149 hourly data and 146 daily data. It can be understood that LSTM gives more sensitive results than the other methods, especially in Fig. 8b and c. Besides, in Fig. 8, ANFIS_FCM model showed identical trends as LSTM; however, ARMA did not present satisfactory results in the prediction step due to considering the linear model approach. Some studies also presented identical results for forcasting with ARMA (Shafaei and Kisi 2016; Lai and Dzombak 2020).

Table 4 shows the best statistical accuracy results of 10-min interval, hourly, and daily AAT predictions in bold. As can be seen from the table, LSTM provided the best accuracies for both hourly and daily AAT prediction compared to ANFIS-FCM and ARMA models. On the other hand, LSTM and ANFIS-FCM provided close accuracy for 10-min interval results. ARMA model failed to predict the AAT time series for all three data sets, providing poor statistical results compared to LSTM and ANFIS-FCM models (Table 4). Considering the comparison of 10-min interval, hourly, and daily AAT predictions with LSTM method, it is clear from Table 4 that LSTM modeled and predicted the 10-min interval data better than the hourly and daily data. This is most probably due to the fact that 10-min interval data does not have sudden changes compared to hourly and daily data. Moreover, ANFIS model behaved in the same manner like LSTM with regard to 10-min interval, hourly, and daily data. ANFIS-FCM model provided solution that is slightly more sensitive compared to LSTM method for 10-min interval data. The prediction accuracies of all models decreased in 10-min interval AAT prediction; nevertheless, apart from ARMA, the best results for both LSTM and ANFIS models are still satisfactory. Concerning the NSE results, NSE=1 means that the model is perfectly fit with an estimation error variance equal to zero. On the other hand, NSE=0 shows the model is not meaningful for the relevant data set. Furthermore, if the NSE is negative, it means that the observed mean is a better predictor than the model as seen in 10-min and hourly ARMA results.

Conclusion

In this study, we proposed the LSTM network model, a deep learning method, to predict daily, hourly, and 10-min interval AAT values, which were originally obtained from two stations. For the proposed prediction model, a set of 10-min interval and daily AAT data obtained from Mersin station and hourly AAT data obtained from Belen station, located in the Eastern Mediterranean Region of Turkey, was utilized. The percentage of the training and testing/prediction data were set 80% and 20%, respectively. On the other hand, ANFIS-FCM and ARMA methods were also used to compare the results of the LSTM method. The optimal LSTM and ANFIS-FCM parameters such as the hidden layer number and epoch number were determined by trial and error method. Both LSTM and ANFIS-FCM network model showed high

accuracy for the prediction of 10-min interval, hourly, and daily AAT data with RMSE value between 0.31 and 1.52 °C. However, ARMA did not provide satisfactory results for all three data sets, presenting high RMSE values from 2.21 to 4.40 °C. It is clear from the results that the prediction error of the LSTM network decreases as the number of training data increases, since the LSTM produce better results for the 10-min interval AAT data than the hourly and daily one. LSTM provided the best accuracies for both hourly and daily AAT prediction compared to ANFIS-FCM model. At the same time, LSTM and ANFIS-FCM provided close accuracy for 10-min interval results.

Since AAT is one of the most important weather parameters for many climatological events, understanding the behavior of this parameter is crucial for a sustainable environment. Predicting and monitoring AAT variations make it possible for decision-makers to prepare an action plan in case of an emergency scenario. Therefore, as presented in this study, deep learning methods can help researchers to accurately model and predict AAT. Considering the experimental results, the LSTM network model is a powerful tool for 10-min interval, hourly, and daily AAT prediction.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Abdulshahed AM, Longstaff AP, Fletcher S (2015) The application of ANFIS prediction models for thermal error compensation on CNC machine tools. Appl Soft Comput 27:158–168. https://doi.org/10.1016/j.asoc.2014.11.012

Akbari S, Mahmood SM, Tan IM, Hematpour H (2018) Comparison of neuro-fuzzy network and response surface methodology pertaining to the viscosity of polymer solutions. J Pet Explor Prod Technol 8:887–900. https://doi.org/10.1007/s13202-017-0375-6

Al-Hmouz A, Shen J, Al-Hmouz R, Yan J (2012) Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning. IEEE Trans Learn Technol 5:226–237. https://doi.org/10.1109/TLT.2011.36

Arslan N, Sekertekin A (2019) Application of long short-term memory neural network model for the reconstruction of MODIS land surface temperature images. J Atmos Solar-Terrestrial Phys 105100:105100. https://doi.org/10.1016/j.jastp.2019.105100

Azad A, Kashi H, Farzin S, Singh VP, Kisi O, Karami H, Sanikhani H (2020) Novel approaches for air temperature prediction: a comparison of four hybrid evolutionary fuzzy models. Meteorol Appl 27. https://doi.org/10.1002/met.1817

Balluff S, Bendfeld J, Krauter S (2015) Short term wind and energy prediction for offshore wind farms using neural networks. In: 2015 International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, Palermo, pp 379–382. https://doi.org/10.1109/icrera.2015.7418440

Benbahria Z, Sebari İ, Hajji H, Smiej MF (2021) Intelligent mapping of irrigated areas from Landsat 8 images using transfer learning. Int J Eng Geosci. https://doi.org/10.26833/ijeg.681312

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Springer US, Boston, MA

Bilgili M, Sahin B (2009) Prediction of long-term monthly temperature and rainfall in Turkey. Energy Sources, Part A Recover Util Environ Eff 32:60–71. https://doi.org/10.1080/15567030802467522

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Holden–Day, San Francisco

Box GEP, Meyer RD (1986) An analysis for unreplicated fractional factorials. Technometrics 28:11–18. https://doi.org/10.1080/00401706.1986.10488093

Chen J, Zeng G-Q, Zhou W, du W, Lu KD (2018) Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy Convers Manag 165:681–695. https://doi.org/10.1016/j.enconman.2018.03.098

Chen M-R, Zeng G-Q, Lu K-D, Weng J (2019a) A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM. IEEE Internet Things J 6:6997–7010. https://doi.org/10.1109/JIOT.2019.2913176

Chen Y, Zhang S, Zhang W, Peng J, Cai Y (2019b) Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. Energy Convers Manag 185:783–799. https://doi.org/10.1016/j.enconman.2019.02.018

Chevalier RF, Hoogenboom G, McClendon RW, Paz JA (2011) Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks. Neural Comput Applic 20:151–159. https://doi.org/10.1007/s00521-010-0363-y

Cobaner M, Citakoglu H, Kisi O, Haktanir T (2014) Estimation of mean monthly air temperatures in Turkey. Comput Electron Agric 109:71–79. https://doi.org/10.1016/j.compag.2014.09.007

Dong D, Sheng Z, Yang T (2018) Wind power prediction based on recurrent neural network with long short-term memory units. In: 2018 International Conference on Renewable Energy and Power Engineering (REPE). IEEE, Toronto, pp 34–38. https://doi.org/10.1109/repe.2018.8657666

Fahimi Nezhad E, Fallah Ghalhari G, Bayatani F (2019) Forecasting maximum seasonal temperature using artificial neural networks "Tehran case study". Asia-Pac J Atmos Sci 55:145–153. https://doi.org/10.1007/s13143-018-0051-x

Güldal V, Tongal H (2010) Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in Eğirdir Lake level forecasting. Water Resour Manag 24:105–128. https://doi.org/10.1007/s11269-009-9439-9

Han S, Qiao Y, Yan J, Liu YQ, Li L, Wang Z (2019) Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl Energy 239:181–191. https://doi.org/10.1016/j.apenergy.2019.01.193

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu Y-L, Chen L (2018) A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm. Energy Convers Manag 173:123–142. https://doi.org/10.1016/j.enconman.2018.07.070

Huang Y, Liu S, Yang L (2018) Wind speed forecasting method using EEMD and the combination forecasting method based on GPR and LSTM. Sustainability 10:3693. https://doi.org/10.3390/su10103693

Jang JR (1993) ANFIS: adaptive network-based fuzzy inference system. IEEE Trans Syst Man Cybern Syst 23:665–685. https://doi.org/10.1109/21.256541

Kalogirou SA (2001) Artificial neural networks in renewable energy systems applications: a review. Renew Sust Energ Rev 5:373–401. https://doi.org/10.1016/S1364-0321(01)00006-5

Karahoca A, Karahoca D (2011) GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system. Expert Syst Appl 38:1814–1822. https://doi.org/10.1016/j.eswa.2010.07.110

Karakuş O, Kuruoğlu EE, Altinkaya MA (2017) One-day ahead wind speed/power prediction based on polynomial autoregressive model. IET Renew Power Gener 11:1430–1439. https://doi.org/10.1049/iet-rpg.2016.0972

Kim HS, Park I, Song CH, Lee K, Yun JW, Kim HK, Jeon M, Lee J, Han KM (2019) Development of a daily PM 10 and PM 2.5 prediction system using a deep long short-term memory neural network model. Atmos Chem Phys 19:12935–12951. https://doi.org/10.5194/acp-19-12935-2019

Kisi O, Sanikhani H (2015) Modelling long-term monthly temperatures by several data-driven methods using geographical inputs. Int J Climatol 35:3834–3846. https://doi.org/10.1002/joc.4249

Kisi O, Shiri J (2014) Prediction of long-term monthly air temperature using geographical inputs. Int J Climatol 34:179–186. https://doi.org/10.1002/joc.3676

Krishan M, Jha S, Das J, Singh A, Goyal MK, Sekar C (2019) Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. Air Qual Atmos Health 12:899–908. https://doi.org/10.1007/s11869-019-00696-7

Lai Y, Dzombak DA (2020) Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation. Weather Forecast 35:959–976. https://doi.org/10.1175/WAF-D-19-0158.1

Li X, Peng L, Yao X, Cui S, Hu Y, You C, Chi T (2017) Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environ Pollut 231:997–1004. https://doi.org/10.1016/j.envpol.2017.08.114

Li Z, Tang T, Gao C (2019) Long short-term memory neural network applied to train dynamic model and speed prediction. Algorithms 12:173. https://doi.org/10.3390/a12080173

Liang S, Nguyen L, Jin F (2018) A multi-variable stacked long-short term memory network for wind speed forecasting. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, Seattle, pp 4561–4564. https://doi.org/10.1109/BigData.2018.8622332

Liu H, Mi X, Li Y (2018a) Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manag 156:498–514. https://doi.org/10.1016/j.enconman.2017.11.053

Liu H, Mi X, Li Y (2018b) Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manag 159:54–64. https://doi.org/10.1016/j.enconman.2018.01.010

Loganathan N, Ibrahim Y (2010) Forecasting international tourism demand in Malaysia using Box Jenkins Sarima application. South Asian J Tour Herit 3:50–60

López E, Valle C, Allende H, Gil E, Madsen H (2018) Wind power forecasting based on echo state networks and long short-term memory. Energies 11:526. https://doi.org/10.3390/en11030526

Ma X, Tao Z, Wang Y, Yu H, Wang Y (2015) Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp Res Part C Emerg Technol 54:187–197. https://doi.org/10.1016/j.trc.2015.03.014

Massaoudi M, Chihi I, Sidhom L et al (2021) An effective hybrid NARX-LSTM model for point and interval PV power forecasting. IEEE Access 9:36571–36588. https://doi.org/10.1109/ACCESS.2021.3062776

Mathworks (2019) Long short-term memory networks. https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html. Accessed 22 Jul 2019

Meshram SG, Kahya E, Meshram C, Ghorbani MA, Ambade B, Mirabbasi R (2020) Long-term temperature trend analysis associated with agriculture crops. Theor Appl Climatol 140:1139–1159. https://doi.org/10.1007/s00704-020-03137-z

Muzaffar S, Afshari A (2019) Short-term load forecasts using LSTM networks. Energy Procedia 158:2922–2927. https://doi.org/10.1016/j.egypro.2019.01.952

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I — a discussion of principles. J Hydrol 10:282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Pak U, Kim C, Ryu U, Sok K, Pak S (2018) A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. Air Qual Atmos Health 11:883–895. https://doi.org/10.1007/s11869-018-0585-1

Park K, Lee et al (2019) Temperature prediction using the missing data refinement model based on a long short-term memory neural network. Atmosphere (Basel) 10:718. https://doi.org/10.3390/atmos10110718

Peng L, Liu S, Liu R, Wang L (2018) Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 162:1301–1314. https://doi.org/10.1016/j.energy.2018.05.052

Qi Y, Zhou Z, Yang L, Quan Y, Miao Q (2019) A decomposition-ensemble learning model based on LSTM neural network for daily reservoir inflow forecasting. Water Resour Manag 33:4123–4139. https://doi.org/10.1007/s11269-019-02345-1

Qin Y, Li K, Liang Z, Lee B, Zhang F, Gu Y, Zhang L, Wu F, Rodriguez D (2019) Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. Appl Energy 236:262–272. https://doi.org/10.1016/j.apenergy.2018.11.063

Qing X, Niu Y (2018) Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 148:461–468. https://doi.org/10.1016/j.energy.2018.01.177

Qu X, Xiaoning K, Chao Z et al (2016) Short-term prediction of wind power based on deep long short-term memory. In: 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). IEEE, pp 1148–1152

Radhika Y, Shashi M (2009) Atmospheric temperature prediction using support vector machines. Int J Comput Theory Eng:55–58. https://doi.org/10.7763/IJCTE.2009.V1.9

Rajagopalan S, Santoso S (2009) Wind power forecasting and error analysis using the autoregressive moving average modeling. In: 2009 IEEE Power & Energy Society General Meeting. IEEE, Calgary, pp 1–6. https://doi.org/10.1109/PES.2009.5276019

Ramesh K, Anitha R (2014) MARSpline model for lead seven-day maximum and minimum air temperature prediction in Chennai, India. J Earth Syst Sci 123:665–672. https://doi.org/10.1007/s12040-014-0434-z

Rezaie B, Nikoo SY, Rahmani Z (2018) A novel intelligent fast terminal sliding mode control for a class of nonlinear systems: application to atomic force microscope. Int J Dyn Control 6:1335–1350. https://doi.org/10.1007/s40435-017-0376-9

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

Sariturk B, Bayram B, Duran Z, Seker DZ (2020) Feature extraction from satellite images using segnet and fully convolutional networks (FCN). Int J Eng Geosci. https://doi.org/10.26833/ijeg.645426

Sekula P, Bokwa A, Bochenek B, Zimnoch M (2019) Prediction of air temperature in the Polish Western Carpathian mountains with the ALADIN-HIRLAM numerical weather prediction system. Atmosphere (Basel) 10:186. https://doi.org/10.3390/atmos10040186

Shafaei M, Kisi O (2016) Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. Water Resour Manag 30:79–97. https://doi.org/10.1007/s11269-015-1147-z

Shi X, Lei X, Huang Q, Huang S, Ren K, Hu Y (2018) Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short-term memory. Energies 11:3227. https://doi.org/10.3390/en11113227

Tabari H, Kisi O, Ezani A, Hosseinzadeh Talaee P (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. J Hydrol 444–445:78–89. https://doi.org/10.1016/j.jhydrol.2012.04.007

Tian Y, Pan L (2015) Predicting short-term traffic flow by long short-term memory recurrent neural network. In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity). IEEE, pp 153–158

Tong W, Li L, Zhou X, Hamilton A, Zhang K (2019) Deep learning PM2.5 concentrations with bidirectional LSTM RNN. Air Qual Atmos Health 12:411–423. https://doi.org/10.1007/s11869-018-0647-4

Venkadesh S, Hoogenboom G, Potter W, McClendon R (2013) A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks. Appl Soft Comput 13:2253–2260. https://doi.org/10.1016/j.asoc.2013.02.003

Wang J, Li Y (2018) Multi-step ahead wind speed prediction based on optimal feature extraction, long short term memory neural network and error correction strategy. Appl Energy 230:429–443. https://doi.org/10.1016/j.apenergy.2018.08.114

Wu W, Chen K, Qiao Y, Lu Z (2016) Probabilistic short-term wind power forecasting based on deep neural networks. In: 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, Beijing, pp 1–8. https://doi.org/10.1109/PMAPS.2016.7764155

Xu Y, Liu H, Duan Z (2020) A novel hybrid model for multi-step daily AQI forecasting driven by air pollution big data. Air Qual Atmos Health 13:197–207. https://doi.org/10.1007/s11869-020-00795-w

Yu C, Li Y, Bao Y, Tang H, Zhai G (2018) A novel framework for wind speed prediction based on recurrent neural networks and support vector machine. Energy Convers Manag 178:137–145. https://doi.org/10.1016/j.enconman.2018.10.008

Yuan X, Chen C, Jiang M, Yuan Y (2019) Prediction interval of wind power using parameter optimized Beta distribution based LSTM model. Appl Soft Comput 82:105550. https://doi.org/10.1016/j.asoc.2019.105550

Zahroh S, Hidayat Y, Pontoh RS (2019) Modeling and forecasting daily temperature in Bandung. In: Proceedings of the International Conference on Industrial Engineering and Operations Management, Riyadh, Saudi Arabia, pp 406–412

Zaytar MA, El Amrani C (2016) Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. Int J Comput Appl 143:7–11. https://doi.org/10.5120/ijca2016910497

Zhang Z, Moore JC (2015) Autoregressive moving average models. In: Mathematical and physical fundamentals of climate change. Elsevier Inc., Amsterdam, pp 239–290

Zhang X, Tan S-C, Li G (2014) Development of an ambient air temperature prediction model. Energy Build 73:166–170. https://doi.org/10.1016/j.enbuild.2014.01.006

Zhang Q, Wang H, Dong J, Zhong G, Sun X (2017) Prediction of sea surface temperature using long short-term memory. IEEE Geosci Remote Sens Lett 14:1745–1749. https://doi.org/10.1109/LGRS.2017.2733548

Zhang X, Zhang Q, Zhang G, Nie Z, Gui Z, Que H (2018) A novel hybrid data-driven model for daily land surface temperature forecasting using long short-term memory neural network based on ensemble empirical mode decomposition. Int J Environ Res Public Health 15:1032. https://doi.org/10.3390/ijerph15051032

Zhang J, Cao X, Xie J, Kou P (2019a) An improved long short-term memory model for dam displacement prediction. Math Probl Eng 2019:1–14. https://doi.org/10.1155/2019/6792189

Zhang J, Yan J, Infield D, Liu Y, Lien FS (2019b) Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Appl Energy 241:229–244. https://doi.org/10.1016/j.apenergy.2019.03.044

Zhang Z, Qin H, Liu Y, Wang Y, Yao L, Li Q, Li J, Pei S (2019c) Long short-term memory network based on neighborhood gates for processing complex causality in wind speed prediction. Energy Convers Manag 192:37–51. https://doi.org/10.1016/j.enconman.2019.04.006

Zhang Z, Ye L, Qin H, Liu Y, Wang C, Yu X, Yin X, Li J (2019d) Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression. Appl Energy 247:270–284. https://doi.org/10.1016/j.apenergy.2019.04.047

Zhao J, Deng F, Cai Y, Chen J (2019) Long short-term memory - fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. Chemosphere 220:486–492. https://doi.org/10.1016/j.chemosphere.2018.12.128

Zhou H, Zhang Y, Yang L, Liu Q, Yan K, du Y (2019a) Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. IEEE Access 7:78063–78074. https://doi.org/10.1109/ACCESS.2019.2923006

Zhou Y, Huang Y, Pang J, Wang K (2019b) Remaining useful life prediction for supercapacitor based on long short-term memory neural network. J Power Sources 440:227149. https://doi.org/10.1016/j.jpowsour.2019.227149