Daily average relative humidity forecasting with LSTM neural network and ANFIS approaches

Arif Ozbek (arozbek@cu.edu.tr)
Şaban Ünal
Mehmet Bilgili

Research Article

Keywords: Deep learning, Relative humidity (RH) forecasting, Long short-term memory (LSTM), Adaptive neuro-fuzzy inference system (ANFIS)

Posted Date: April 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1461025/v1

Abstract

Because hurricanes, droughts, floods, and heat waves are all important factors in measuring environmental changes, they can all result from changes in atmospheric air temperature and relative humidity (RH). Besides, climate, weather, industry, human health, and plant growth are all affected by RH. Accurately and consistently forecasting RH is a challenge due to its non-linear nature. The present study tried to predict one day ahead of RH in determined provinces from different climatic regions of Turkey (Ankara, Erzurum, Samsun, Diyarbakır, Antalya, and Bilecik) using long short-term memory (LSTM) and adaptive neuro-fuzzy inference system (ANFIS) with fuzzy c-means (FCM) based machine learning models. As evaluation criteria, root mean square error (RMSE), mean absolute error (MAE), and correlation coefficient (R) were employed. The outcomes from the forecasting models were also validated using observed data. During the testing stage, the smallest MAE and RMSE values were discovered to be 5.76% and 7.51%, respectively, in Erzurum province, with an R-value of 0.892 when using the LSTM method. Moreover, the smallest MAE and RMSE values were obtained to be 5.95% and 7.67% respectively, in Erzurum province with an R-value of 0.887 using the ANFIS model according to the hourly RH prediction. The results indicate that both the LSTM and ANFIS approaches performed well in daily RH prediction, with the LSTM and ANFIS approaches producing nearly identical results.

1. Introduction

Relative humidity (RH) is the ratio of the amount of water vapor available in the air to the maximum amount of water vapor that can be held in the air at the same temperature. In other words, RH is defined as the ratio of moisture partial pressure in the air to water vapor saturation pressure at the same temperature. If the relative humidity of the air at a given temperature is 100%, it indicates that the air has reached saturation, and if more moisture is added, condensation will form in the air. In addition to humans, plants, animals, and even inanimate objects are affected by the relative humidity in the air. At a certain temperature and humidity, people feel very comfortable both indoors and outdoors. Efforts are being made to provide the necessary comfort conditions using a variety of methods. Relative humidity is critical in providing the necessary comfort, particularly for livestock, as well as for human comfort. To maximize the efficiency of livestock, the relative humidity of the environment in which the animals are housed must be kept under control (Zhou et al., 2021). Relative humidity governs evaporation at the earth's surface as well as plant leaf transpiration. Furthermore, it prevents solar radiation from reaching the earth, thereby preventing overheating or cooling. The importance of relative humidity in food storage cannot be overstated. To avoid food damage, precautions must be taken depending on whether the relative humidity is low or high. Furthermore, relative humidity affects roads, buildings, machinery, electrical transmission lines, and various mechanical and electronic devices and tools we use. All of these can be harmed by both low and high relative humidity. Relative humidity must be kept under control at all times during the manufacturing and use processes. As a result, predicting relative humidity is critical for design and manufacturing processes.

The study conducted by Parishwad et al. emphasized that relative humidity, air temperature, and wind speed were the most important parameters in the sizing of air conditioning systems, and that optimum sizing could be accomplished by accurately estimating these parameters. Using meteorological data from various cities in India, a model was developed to estimate RH, air temperature, and wind speed. The error rates for air temperature, RH, and wind speed in the developed model were reported to be 10.5%, 14.6%, and 26.7%, respectively (Parishwad et al., 1998). Mass and energy balance equations are utilized in the physical modeling of a modern building for automatic control of the HVAC system. To obtain accurate models using this method, the physical properties of the building must be well understood. Inadequate information can have an impact on HVAC facility design and control, resulting in increased energy consumption and poor indoor air quality. The artificial neural network model proposed by Mustafaraj et al. was used to predict RH and indoor air temperature. For a set period of time, indoor and outdoor temperature and RH values were recorded, and the resulting data was used in the artificial neural network model. Very accurate results can be obtained and complex problems can be solved much more easily and quickly using this method, which does not require as detailed data entry as physical models (Mustafaraj et al., 2011). Because of rising energy costs and current environmental regulations, it is possible to argue that dynamic control is required for HVAC systems. The relative humidity value, as well as other meteorological data, is required for the dynamic control of HVAC systems. Significant energy savings will be possible thanks to the estimation of meteorological data using various methods and the dynamic control of HVAC systems (Mustafaraj et al., 2010). Martin et al. proposed a model to estimate the city's temperature and RH values. Although there are many studies for predicting temperature, it has been noted that there are only a few studies for predicting relative humidity. It has been stated that by predicting a city's relative humidity value, the energy management systems of the buildings can be designed accordingly, resulting in energy savings in the buildings (Martin et al., 2015). Also, relative humidity and temperature have a substantial effect on the performance of air-source heat pumps operating in different climatic conditions. Icing in the heat pump may occur depending on the relative humidity value in the outdoor environment, resulting in performance loss. Knowing how relative humidity varies by climatic region and taking appropriate measures will also help to prevent heat pump performance loss (Wu et al., 2020).

There are other areas where relative humidity is effective. For example, high RH levels combined with high air temperatures can cause more thermal stress and death in humans. According to research, low humidity levels play a significant role in severe influenza epidemics, particularly in autumn and winter in temperate climates. For these reasons, the significance of forecasting humidity fluctuations and extremes and taking appropriate measures was emphasized (Langendijk et al., 2021). Liu et al. looked at how temperature and relative humidity affected learning performance. It was determined that relative humidity is a more important factor in learning performance than temperature (Liu et al., 2021). In the study conducted by Pei et al., it was investigated how temperature and relative humidity affect the filters used to provide the necessary air standards for laboratories. In accordance with the findings of the study, RH, rather than temperature, influences the performance of the filters (Pei et al., 2021). Also, relative humidity

can affect the evaporation rate and equilibrium size of particles in the air, which can affect particle removal rates and virus viability.

The amount of energy consumed in buildings is affected by indoor thermal comfort conditions. It is generally determined by the RH and air temperature. For the estimation of comfort parameters, simulation programs have been developed. However, in order to use them, extremely detailed data must be provided. According to Mba et al., predicting temperature and relative humidity values and transferring this information to the building energy management system can save a significant amount of energy (Mba et al., 2016). Using building energy simulation tools is an efficient way to assess indoor air temperature and RH parameters. To make the simulation tools work properly, all kinds of detailed parameters about the building must be collected, such as changes in weather conditions, the structure of the building, its geographical location, the lights, and the energy generated by the equipment. It takes time and a lot of effort to collect these parameters. Most of the time, accessing all of this data may be impossible. Furthermore, the development of these simulation models necessitates expert knowledge, and the computation process is time-consuming. On the other hand, it is well known that some nonlinear models, such as artificial neural networks, outperform linear models. Because neural networks can infinitely approach any continuous function, they are widely utilized in various fields. Although many temperature prediction models have been developed, studies on relative humidity are limited (Shi et al., 2018).

The thermal conductivity of building components is an important parameter that directly affects building heating and cooling loads. The effects of RH and air temperature on the thermal conductivity of building insulation materials have been studied in the literature. The thermal conductivity of insulation materials is measured at room temperature and in dry conditions. In practice, building walls are subjected to changing outdoor conditions such as RH and air temperature. When the temperature rises from 20°C to 60°C, the heat transmission coefficients of the insulation materials used in buildings rise by 8.8–21.4%, and the thermal conductivity coefficient rises by 14.4% when the relative humidity rises from 0–100%. It was discovered that there was an increase ranging from 14.8–186.7%. As a result, the amount of heat transferred from the building walls was estimated to have increased by 1.5 to 1.9 times (Wang et al., 2022). In a study conducted by Karyono et al., it was discovered that estimating the relative humidity value and using this data in the building energy management system could save 2% of the energy consumed in buildings (Karyono et al., 2022).

In a study, Csavina et al. demonstrated the effects of RH and wind speed on atmospheric dust concentrations in semi-arid climates. It has been emphasized that estimating the harmful effects of dust and aerosol emissions in the air is possible, and that estimation can be performed with the help of wind speed and relative humidity parameters (Csavina et al., 2014). Humidity is a significant component of the hydrological cycle, influencing both the air and the climate. The concept of relative humidity is widely applied in many fields, including hydrology, ecology, agriculture, and medicine (Gunawardhana et al., 2017). Temperature and relative humidity both have a significant impact on air quality and human health. As a result, mapping relative humidity and temperature extremes during the hot summer months is critical for public service and environmental policy (L. Li & Zha, 2018).

The relative humidity of the air must be taken into consideration in the food and agricultural industries as well. The effect of RH on water loss during pear storage was studied. The finite element method was used to analyze a three-dimensional model for this. Increasing the RH from 95 to 97 percent reduces water loss in pears by 3%, according to the findings. As a result, the significance of controlling indoor relative humidity by estimating relative humidity change during storage was emphasized (Nguyen et al., 2007). Biaobrzewski's study highlighted the importance of relative humidity in grain storage. The artificial neural network method has been claimed to be precise in estimating relative humidity (Białobrzewski, 2008). Because temperature and humidity control have a direct impact on product quality, it is critical in industries such as tobacco processing, food production, and biological product production (Shi et al., 2018). Kim et al. studied the effects of RH on egg quality and laying hen physiological stress. Based on what we learned from the research, high relative humidity reduces egg quality and increases stress in chickens (Kim et al., 2021). The importance of relative humidity during mushroom drying was emphasized in a study conducted by (X. Li et al., 2022).

The importance of RH and air temperature control in the production of electronic devices was emphasized by Conseil-Gudla et al. It has been stated that condensation may occur as a result of humidity fluctuations, which will have a negative impact on electronic devices. It has been stated that the necessary precautions must be taken by predicting the relative humidity (Conseil-Gudla et al., 2018). According to Shi et al., accurate, reliable, and stable RH and air temperature control in industrial production is critical for efficiency. Simultaneously, the importance of accurately predicting air temperature and RH in climate control systems in terms of system stability was emphasized. The artificial neural network method was shown in the study to be very sensitive for estimating relative humidity (Shi et al., 2018). Sohani et al. studied the effects of absolute and RH on photovoltaic performance using the artificial neural network method. It has been discovered that as absolute and relative humidity levels rise, so does the performance of photovoltaics (Sohani et al., 2020). The relationship between relative humidity and metallurgical engineering is also very close. The effect of RH on the production process of aluminum alloys was highlighted in the study conducted by Safyari et al. As a result of the research, cracks can form in aluminum alloys when relative humidity is high, and relative humidity should be kept under control during the manufacturing process (Safyari et al., 2021). Temperature and relative humidity effects on high voltage power transmission lines were investigated. It has been stated that high voltage line initial investment costs can be reduced by taking the necessary measures in high voltage lines based on seasonal and regional changes by predicting the relative humidity value (Ma et al., 2021).

In another investigation, the researchers indicate that, the curing times and strengths of concrete are determined based on the assumption that the relative humidity value is 100%. However, in field studies, this is frequently not possible to provide, and the concrete strength values calculated based on the 100% relative humidity are higher. Indeed, it has been emphasized that the relative humidity value of the outdoor environment shapes concrete strengths, so the relative humidity value should also be taken into account in field studies (Liao et al., 2008). As a result, it is critical to know the RH value of the environment before pouring concrete and to take appropriate precautions in concrete construction

projects that are exposed to the outside environment. Kwon et al. proposed a method for estimating the compressive strength of concrete used in the construction of nuclear power plants. The study revealed that, in addition to temperature, RH is an important parameter in concrete strength, and the importance of predicting air temperature and RH parameters for the concrete strength of nuclear power plants to be built in different climatic regions and seasons was emphasized (Kwon et al., 2014). Shen et al. investigated the impact of RH on the curing time of concrete and crack formation in concrete (Shen et al., 2017). Temperature and humidity are known to cause corrosion in reinforced concrete structures. Chauhan and Sharma investigated the effects of air temperature and RH changes on corrosion in reinforced concrete, concluding that both parameters should be estimated for each climate zone and reinforced concrete reinforcement calculations should be made accordingly (Chauhan & Sharma, 2019). Xi et al. studied the effect of RH on the linear viscoelastic properties of asphalt mixtures (Xi et al., 2021).

Interpolation, simulation, and regression approaches are commonly applied to predict RH and air temperature. The interpolation method is the simplest approach, but it introduces significant errors and uncertainties (Vincent & Mekis, 2006; Yang et al., 2004). For more precise estimations, regression methods can be used (L. Li & Zha, 2018). Gunawardhana et al. proposed an alternative model for estimating relative humidity. Meteorological data from many years was statistically converted into a smaller data set in the proposed method, and a relative humidity forecasting model was developed as a result. It was stated that the obtained results were very good (Gunawardhana et al., 2017). The artificial neural networks method was used in the study conducted by Kuzugudenli to estimate relative humidity using meteorological data from 177 different points in Turkey. According to the results, the method used for estimating relative humidity produced very sensitive (R2 = 0.84) results (Kuzugudenli, 2018). Accurate outdoor temperature and RH estimation have a crucial impact on agricultural and industrial production, as well as environmental and economic policies. It is possible to predict the outdoor air temperature and RH very precisely using models developed using the artificial neural network method (Al-Shawwa et al., 2018). Hutapea et al. conducted research on the estimation of RH using the long short-term memory network method (Hutapea et al., 2020). Adnan et al. emphasized the importance of accurate and reliable estimation of relative humidity in all areas related to global climate change. A machine learning method was used to estimate RH in the Hunza river basin in Pakistan. It has been determined that the error rates in the results obtained using the method are quite low, and thus the MARS method can be easily used in the estimation of relative humidity (Adnan et al., 2021). Due to its highly complex and non-linear nature, research on relative humidity estimation is limited. In the study by Qadeer et al., a machine learning method was used to estimate relative humidity. In accordance with reports, the machine learning method can predict relative humidity very precisely, and this data can be used in various applications, including the operation and design of HVAC systems, cooling towers, and thermal power plants (Qadeer et al., 2021).

2. Research Significance And Novelty Of The Study

As can be seen in the literature research described, the effects of relative humidity (RH) on various fields, such as humans, animals, foodstuffs, air-conditioning applications, electronic devices, and building

materials, have been investigated, but there have been few studies on its forecasting using a machine learning approach. For one-day ahead RH estimate utilizing historical meteorological RH data, we offer two alternative machine learning algorithms, namely long short-term memory (LSTM) and adaptive neurofuzzy inference system (ANFIS) using fuzzy c-means (FCM). The findings from all of the strategies utilized in this research were compared using identical data sets in order to identify which strategy gave more accurate results. Machine learning methods have been demonstrated to be effective in data forecasting, particularly for predicting relative humidity. The calculations were carried out using the average daily relative humidity values obtained by the Turkish State Meteorological Service during an eight-year period in six provinces selected from six distinct climatic areas of Turkey. The most significant benefit of these suggested methodologies is the ability to estimate relative humidity levels at any station with acceptable accuracy without the requirement for a variety of meteorological data or complicated computations. ANFIS and LSTM neural networks, among other techniques, may be used to model univariate time series data as independent variables. The followings are the study's most significant contributions to the literature:

- The use of deep learning techniques in the estimate of RH values across time series, which has only been researched in a small number of studies till now.
- Determining the most accurate technique by contrasting the best configurations of ANFIS and LSTM neural network models against one another.
- Comparison of the time series estimation for relative humidity (RH) values of meteorological stations from different climate types, such as continental, Mediterranean, Black Sea, and Marmara (transitional) climates.

Because of this, it is anticipated that this study will fill a gap in the literature and make major contributions to the field.

3. Method

3.1. Long-short term memory (LSTM) neural network

A Long-Short Term Memory (LSTM) neural network is expressed as a type of recurrent neural network (RNN) that could learn long-term dependencies between time steps of series data. Unlike RNNs, machine learning techniques based on LSTM are developed to prevent the long-term dependency problem. The LSTM neural network performs the training process using back propagation over time, thus eliminating the vanishing gradient problem. Conventional artificial neural networks have neurons, whereas LSTM neural networks have memory blocks connected via sequential layers (Abdel-Nasser & Mahmoud, 2019). The basic elements of an LSTM neural network consist of a set of the input layer, an LSTM layer, a fully connected layer, and a regression output layer. This network structure functions first with input layers and then with an LSTM layer. With a series of the input layer, sequence or time-series data is entered into the network architecture. An LSTM layer gets to know long-term dependencies between time steps of

sequence data. Next, the network completes the function with regression and fully connected output layers (Bilgili et al., 2021; Ozbek et al., 2021; Sekertekin et al., 2021).

Figure 1(a) shows an LSTM layer structure. This graph shows the flow of a time series X with C channels of length S through an LSTM layer. In the graph, c_t and h_t indicate the cell state and the hidden state at time step t, respectively. The first LSTM block utilizes the initial state of the network and the first time step of the sequence to calculate the first output and the updated cell state. At time step t, the LSTM block utilizes the current state of the network (c_{t-1}, h_{t-1}) and the next time step of the sequence to calculate the output and the updated cell state c_t (Mathworks, 2021; Zolfaghari & Golabi, 2021).

Figure 1(b) shows the data flow at time step t for the LSTM block. It is clear from the figure that the gates forget, update, and output the cell and hidden states. The state of a layer is composed of the hidden state or output state and the cell state. The output of the LSTM layer is derived from the hidden state at this time step t. The cell state includes knowledge learned from the previous time steps. The layer adds knowledge to or extracts knowledge from the cell state at each step. The layer manages these updates using the gates. As seen in the LSTM layer architecture, input gate (i), forget gate (f), cell candidate (g), and output gate (o) components are included. These elements check the hidden state and cell state of the layer. The input gate checks the level of cell state update. The forget gate checks the level of cell state reset (forget). The cell candidate adds knowledge to the cell state. The output gate checks the level of cell state added to the hidden state (Hochreiter & Schmidhuber, 1997).

The input weights W, the recurrent weights R, and the bias b form the learnable weights of an LSTM layer. The matrices W, R, and b can be given as follows (Mathworks, 2021):

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \quad (1)$$

The cell state at time step t is expressed by

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

where $\odot$ demonstrates the element-wise multiplication of vectors. The hidden state at time step t is expressed by

$$h_t = o_t \odot \sigma_c\left(c_t\right) \quad (3)$$

where σ_c demonstrates the state activation function. The hyperbolic tangent function (tanh) is utilized to calculate the state activation function, and it is expressed by

$$\sigma(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (4)$$

At time step t, the components of the input, forget, cell, and output gates can be identified as follows, respectively (Mathworks, 2021):

$$i_t = \sigma_g\left(W_i x_t + R_i h_{t-1} + b_i\right) \quad (5)$$

$$f_t = \sigma_g\left(W_f x_t + R_f h_{t-1} + b_f\right) \quad (6)$$

$$g_t = \sigma_c\left(W_g x_t + R_g h_{t-1} + b_g\right) \quad (7)$$

$$o_t = \sigma_g\left(W_o x_t + R_o h_{t-1} + b_o\right) \quad (8)$$

In these calculations, σ_g demonstrates the gate activation function. The sigmoid function is utilized to calculate the gate activation function, and it is expressed by

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (9)$$

3.2. Adaptive neuro-fuzzy inference system (ANFIS)

Although artificial neural networks (ANNs) are effective for modeling different natural processes, they also have their shortcomings (Tabari et al., 2012). The Adaptive Neuro-Fuzzy Inference System (ANFIS), first presented by (Jang, 1993), is a universal predictive approach and thus can approximate any natural continuous procedure on a consolidated set with any grade of accuracy (Goyal et al., 2014). ANFIS, a Sugeno-type fuzzy systems network expression, is equipped with neural learning abilities. This technique operates similarly to neural networks and creates a method for fuzzy modeling procedures to learn information about a dataset. An ANFIS provides a strategy with the support of a fuzzy rule or rules. Fuzzy rules are statements that present the relationship between the inputs and outputs of the system in the form of if-then statements depending on linguistic variables (Ay & Kisi, 2014).

It can be considered that the system of fuzzy inference consists of two *x* and *y* inputs and one *z* output. Assuming the rule base includes two fuzzy if-then rules of Takagi and Sugeno type:

$$Rule1: if x is A_1 and y is B_1,\ then f_1 = p_1 x + q_1 y + r_1 \quad (10)$$

$$Rule2: if x is A_2 and y is B_2,\ then f_2 = p_2 x + q_2 y + r_2 \quad (11)$$

where A_i and B_i describe the fuzzy clusters in the antecedent, while r_i, q_i and p_i depict the design factors in the training procedure. As can be seen in Fig. 2, the ANFIS structure consists of five layers.

Layer 1: This layer defines an input variable for every suitable fuzzy set. The nodes in this layer indicate the function by which membership degrees are formed using MFs. The node function of a node *i* is described as (Jang, 1993)

$$O_i^1 = \mu_{A_i}(x), i = 1, 2 \text{ (12)}$$

$$O_i^1 = \mu_{B_{i-2}}(y), i = 3, 4 \text{ (13)}$$

where μ_{A_i} and μ_{B_i} are the MFs.

Layer 2: In this layer, the product output is created by multiplying the incoming signals. All nodes determine the firing strength of a rule by multiplying the incoming signals.

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), i = 1, 2 \text{ (14)}$$

Layer 3: The i^{th} node in Layer 3 calculates the ratio of the i^{th} rule's firing strength to the sum of all rules firing strengths of all rules

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2 \text{ (15)}$$

where $\bar{w}_i$ is the normalized firing strengths.

Layer 4: Node *i* in Layer 4 identifies how the i^{th} rule will support the model output through the node function

$$O_i^4 = \bar{w}_i z_i = \bar{w}_i \left(p_i x + q_i y + r_i\right), i = 1, 2 \text{ (16)}$$

where {r_i, q_i, *and* p_i} is the parameter set, and $\bar{w}_i$ is the output of the layer 3.

Layer 5: This layer is called the output nodes. Here, the single node computes the output considering the sum of all incoming signals (Tabari et al., 2012)

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i z_i = \frac{w_1 z_1 + w_2 z_2}{w_1 + w_2} \text{ (17)}$$

Accordingly, the output *z* in Fig. 2 is found as:

$$Z = \left(\bar{w}_1 x\right) p_1 + \left(\bar{w}_1 y\right) q_1 + \left(\bar{w}_1\right) r_1 + \left(\bar{w}_2 x\right) p_2 + \left(\bar{w}_2 y\right) q_2 + \left(\bar{w}_2\right) r_2 \quad (18)$$

3.3. Error analysis

In our study, four statistical error criteria such as mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and correlation coefficient (R) are considered. Their calculation methods are as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |p(i) - o(i)| \quad (19)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} [p(i) - o(i)]^2} \quad (20)$$

$$R = \frac{\left(\sum_{i=1}^{N}\left[p(i) - \bar{p}\right]\left[o(i) - \bar{o}\right]\right)}{\left(\sqrt{\sum_{i=1}^{N}\left[p(i) - \bar{p}\right]^2}\sqrt{\sum_{i=1}^{N}\left[o(i) - \bar{o}\right]^2}\right)} \quad (21)$$

where *p(i)* and *o(i)* describe the predicted and observed RH at the time *i*, respectively. Besides, $\bar{p}$ and $\bar{o}$ stand for the mean of the predicted RH and the observed RH, respectively. N represents the total number of RH data.

4. Results And Discussion

4.1. Case study and data preparation

The calculations were performed using the average daily relative humidity values measured by the General Directorate of State Meteorology Affairs between 2012 and 2019 in 6 different provinces (Ankara, Erzurum, Samsun, Diyarbakır, Antalya and Bilecik) selected from 6 different climatic regions of Turkey as shown in Fig. 3. Table 1 gives the geographic coordinates of the stations as well as a statistical summary of the dataset. As seen from the table, the Diyarbakır station has the highest RH fluctuation and the lowest average RH value. Furthermore, Samsun station has the smallest standard deviation value of 12.31 percent.

Table 1
Geographic coordinates of the provinces and statistical summary of the dataset

Station	Location (DD)		Altitude (m)	Maximum RH (%)	Minimum RH (%)	Average RH (%)	Standard Deviation (%)
	Latitude	Longitude					
ANKARA	39.9727	32.8637	891	99.00	17.17	58.44	17.95
ERZURUM	39.9529	41.1897	1758	98.67	25.04	66.54	16.76
SAMSUN	41.3441	36.2563	4	95.40	21.58	67.67	12.31
DIYARBAKIR	37.8973	40.2027	674	99.67	11.75	53.34	25.32
ANTALYA	36.9063	30.7990	64	95.71	15.83	62.64	16.51
BILECIK	40.1414	29.9772	539	99.46	27.38	68.62	13.82
DD: Degree Decimal							

The main steps of the suggested machine-learning techniques shown in Fig. 4 can be expressed as follows (Adnan et al., 2021):

1. The data was pre-processed and revised for the next step by interpolating missing values after it was collected from the stations.

2. The original data were then normalized and the proposed models trained, validated, and tested, respectively.
3. Each model was applied to these data separately with the same input data combinations for RH prediction.
4. To examine the precision of the developed model, various statistical criteria such as MAE, RMSE, and R were used.
5. Eventually, RH was forecasted using the designed machine-learning approaches.
MATLAB R2018a (Trail Version) was used to perform the models.

4.2. Fuzzy C-Means clustering (ANFIS-FCM)

Initially, the ANFIS-FCM clustering model was chosen, and numerous clusters, such as epoch number, input number, and Membership Functions (MFs) were used in the simulation. The method's performance was then evaluated by computing MAE, RMSE, and R values. In the analysis, the influence of the epoch number, input number, and MFs parameters were investigated. While the best results are obtained in Ankara, Erzurum, and Diyarbakir when 2 MFs and 5 input numbers are used against 10 epochs; it was obtained when 2 MFs and 5 input numbers were used against 100 epoch numbers in Samsun, Antalya, and Bilecik. It was seen that only the epoch number had an effect among the clusters. Table 2 displays the best statistical accuracy results of daily RH predictions for ANFIS-FCM and LSTM methods. As a result of the simulation, the closest results to the observed values were obtained in Diyarbakır with MAE

5.98%, RMSE 8.21%, and R-value 0.948; while the farthest results were obtained in Antalya with MAE 8.52%, RMSE 10.92%.and R-value 0.721.

4.3. Long short-term memory (LSTM)

Although 300 epoch numbers are used in all provinces with this method; the best results were obtained when 5 hidden layers were used in Samsun, 10 hidden layers in Erzurum, Diyarbakir, Antalya, and Bilecik, and 25 hidden layers in Ankara. In this method, while the effect of the hidden layer was observed, it was seen that the epoch number was not effective. As shown in Table 2, the lowest MAE value is equal to 5.76% and RMSE is 7.51% with an R-value of 0.892 in Erzurum. In all analyzes made for all provinces, it has been observed that increasing the number of hidden layers and epochs has no effect on decreasing the MAE and RMSE values. The highest MAE and RMSE values were found in Antalya province as 8.37% and 10.84%, respectively, with an R-value of 0.726.

Table 2
The best statistical accuracy results of daily RH predictions for ANFIS-FCM and LSTM

Station	Forecasting Method	MAE (%)	RMSE (%)	R
ANKARA	ANFIS FCM	7.28	9.63	0.841
	LSTM	**7.02**	**9.31**	**0.852**
ERZURUM	ANFIS FCM	5.95	7.67	0.887
	LSTM	**5.76**	**7.51**	**0.892**
SAMSUN	ANFIS FCM	7.27	9.95	0.606
	LSTM	**7.27**	**9.91**	**0.608**
DIYARBAKIR	ANFIS FCM	5.98	8.21	0.948
	LSTM	**5.83**	**8.09**	**0.950**
ANTALYA	ANFIS FCM	8.52	10.92	0.721
	LSTM	**8.37**	**10.84**	**0.726**
BILECIK	ANFIS FCM	7.27	9.13	0.723
	LSTM	**7.05**	**8.93**	**0.736**

Figure 5 depicts the RH values observed with the training (7 years) and testing (1 year) data for the provinces to be used in the ANFIS FCM and LSTM algorithms. 2920 daily observed data points were used in the training and testing processes. Within these measurement values, 2555 pieces of data (87.5% of the entire dataset) from 01 January 2012 to 31 December 2018 were used to train the models, and the last 365 daily observed data (12.5% of the entire dataset) from 01 January 2019 to 31 December 2019

were employed to test them. In all provinces, the same training and testing ratios were applied. The X-axis in this figure displays the number of observed values while the Y-axis represents the RH in %. As can be seen from the figure, the prediction of the RH time series by LSTM (red-colored dataset) corresponds to the actual values in the testing part in all provinces.

The results show that the ANFIS FCM model and the LSTM neural network perform well with this type of sinusoidal data. To examine the prediction results in greater depth, Fig. 6 is a close-up of only the testing values. As can be seen from this figure, the estimations obtained by both ANFIS FCM and LSTM methods in Diyarbakir are very close to the real values. While the most accurate estimates were obtained in Diyarbakir, the most remote values were obtained in Samsun.

The variation between the observed and predicted RH values is shown in Fig. 7. While the error value between observed and predicted RH data ranged between − 30% and + 30% in Ankara, Samsun, and Antalya provinces, the least error was obtained in the province of Diyarbakir. It is clear from the figures that, especially in the spring and summer months, error rates were determined to be much lower in Diyarbakir province. The reason for this is that the RH values are very low on these dates. Figure 7 also shows a plot of observed and predicted RH on a daily basis to reveal the data distribution, which supports the previous results. Moreover, the correlation coefficient (R) is presented in this figure. The observed and forecasted data obtained by the LSTM approach almost coincide with each other in Diyarbakir province with an R-value of 0.9496. The smallest R-value was obtained in Samsun with a 0.6082.

Figure 8 shows histograms for all provinces that show the frequency distribution of RH differences between predicted and observed RH for ANFIS FCM and LSTM methods. In particular, results obtained from Diyarbakır tend to show the closest thing to a normal distribution, with an almost symmetrical distribution of about zero. In that province, the error distribution follows a Gaussian curve. The results in Samsun and Antalya appear to be skewed to the positive, while the result in Ankara appears to be skewed to the negative for the models. In both the ANFIS FCM and LSTM methods, more than 85% of the model predictions have an error $\leq \pm 10\%$ for Diyarbakır as can be seen in this figure. In Antalya, on the other hand, more than 72% of the model predictions have an error of $\leq \pm 10\%$. For all the methods under consideration, greater than 90%, the error values are $\leq \pm 16\%$ and the high errors of about 24% and greater in prediction occurred only in Samsun and Antalya provinces.

The Taylor diagrams (Taylor, 2001) in Fig. 9 were constructed for evaluating the results. As is known, a Taylor diagram allows for determining the accuracy of the proposed models in a variety of ways. The distance between the points representing the models and the point representing the observed value is inversely proportional to the overall similarity of the models to the measured values. In comparison to the ANFIS FCM model, Fig. 9 clearly shows that the LSTM model predictions are closer to the observations. The lowest RMSE value of 7.51% was obtained in Erzurum when the LSTM method was used. In Diyarbakir, a correlation coefficient (R) close to 95% was obtained when both methods were used. While the province with the lowest standard deviation was Samsun, the highest was obtained in Diyarbakır.

5. Conclusion

This research proposes two distinct machine learning algorithms for one-day ahead relative humidity (RH) estimates based on historical meteorological RH data spanning eight years. The approaches are described in detail in the following sections. The following are the findings of the research, as described below:

- Machine learning techniques have been demonstrated to be beneficial in data forecasting, particularly in the prediction of RH levels. The ANFIS FCM and LSTM have shown satisfactory estimate performance when using time series histories.
- Both models demonstrated adequate accuracy in the prediction of daily RH data, with MAE values ranging between 5.76% and 8.52%.
- When compared to the ANFIS FCM model, the LSTM model estimates produced results that were closer to the observations in all provinces.
- The best results were found in Diyarbakır, where the dry climatic area predominates, using both the ANFIS FCM and the LSTM approaches. The province of Samsun, on the other hand, had the lowest outcomes when it came to calculations.
- In this study, it was verified that the LSTM neural network, which was developed using a deep learning technique, is an extremely effective tool for short-term daily RH forecasting.
- The time-series method based on ANFIS and LSTM neural networks provides predictions by taking into account the underlying periodicities in the data as it moves through time. In this sense, the most significant advantage of the univariate modeling approach proposed in this study is the ability to use data collected in the past as an input variable.
- The developed model can be used to forecast the one-day ahead RH value of any station without the need for complex calculations or extensive meteorological data.

Declarations

Acknowledgement The Turkish State Meteorological Service released data, which the authors extremely grateful recognize.

Author contributions ***Arif Ozbek*** did all the calculations and analysis and wrote the manuscript by taking inputs from all the co-authors. ***Şaban Ünal*** and ***Mehmet Bilgili*** provided the relative humidity dataset and co-wrote the manuscript.

Data availability All the data produced in this study are available on request from the corresponding author.

Code availability All codes used in this study are available on request from the corresponding author.

Ethics approval Not applicable.

Consent to participate The authors express their consent to participate in research and review.

Consent for publication The authors express their consent for publication of research work.

Conflict of interest The authors declare no competing interests.

References

1. Abdel-Nasser M, Mahmoud K (2019) Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural Comput Appl 31(7):2727–2740. https://doi.org/10.1007/s00521-017-3225-z
2. Adnan M, Muhammad Adnan R, Liu S, Saifullah M, Latif Y, Iqbal M (2021) Prediction of Relative Humidity in a High Elevated Basin of Western Karakoram by Using Different Machine Learning Models. Weather Forecast 1–20. https://doi.org/10.5772/intechopen.98226
3. Al-Shawwa M, Al-Absi AA, Hassanein SA, Baraka KA, Abu-Naser SS (2018) Predicting Temperature and Humidity in the Surrounding Environment Using Artificial Neural Network. Int J Acad Pedagogical Res (IJAPR) 2(9):1–6
4. Ay M, Kisi O (2014) Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. J Hydrol 511(January):279–289. https://doi.org/10.1016/j.jhydrol.2014.01.054
5. Białobrzewski I (2008) Neural modeling of relative air humidity. Comput Electron Agric 60(1):1–7. https://doi.org/10.1016/j.compag.2007.02.009
6. Bilgili M, Yildirim A, Ozbek A, Celebi K, Ekinci F (2021) Long short-term memory (LSTM) neural network and adaptive neuro-fuzzy inference system (ANFIS) approach in modeling renewable electricity generation forecasting. Int J Green Energy 18(6):578–594. https://doi.org/10.1080/15435075.2020.1865375
7. Chauhan A, Sharma UK (2019) Influence of temperature and relative humidity variations on non-uniform corrosion of reinforced concrete. *Structures*, *19*(December 2018), 296–308. https://doi.org/10.1016/j.istruc.2019.01.016
8. Conseil-Gudla H, Staliulionis Z, Mohanty S, Jellesen MS, Hattel JH, Ambat R (2018) Humidity build-up in electronic enclosures exposed to different geographical locations by RC modelling and reliability prediction. *Microelectronics Reliability*, *82*(November 2017), 136–146. https://doi.org/10.1016/j.microrel.2018.01.013
9. Csavina J, Field J, Félix O, Corral-Avitia AY, Sáez AE, Betterton EA (2014) Effect of wind speed and relative humidity on atmospheric dust concentrations in semi-arid climates. Sci Total Environ 487(1):82–90. https://doi.org/10.1016/j.scitotenv.2014.03.138
10. Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A (2014) Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. Expert Syst Appl 41(11):5267–5276. https://doi.org/10.1016/j.eswa.2014.02.047

11. Gunawardhana LN, Al-Rawas GA, Kazama S (2017) An alternative method for predicting relative humidity for climate change studies. Meteorol Appl 24(4):551–559. https://doi.org/10.1002/met.1641
12. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
13. Hutapea MI, Pratiwi YY, Sarkis IM, Jaya IK, Sinambela M (2020) Prediction of relative humidity based on long short-term memory network. *AIP Conference Proceedings*, *2221*(March). https://doi.org/10.1063/5.0003171
14. Jang JSR (1993) ANFIS: Adaptive-Network-Based Fuzzy Inference System. IEEE Trans Syst Man Cybernetics 23(3):665–685. https://doi.org/10.1109/21.256541
15. Karyono K, Romano A, Abdullah BM, Cullen J, Bras A (2022) The role of hygrothermal modelling for different housing typologies by estimating indoor relative humidity, energy usage and anticipation of fuel poverty. Build Environ 207(PB):108468. https://doi.org/10.1016/j.buildenv.2021.108468
16. Kim D-H, Lee Y-K, Lee S-D, Lee K-W (2021) Impact of relative humidity on the laying performance, egg quality, and physiological stress responses of laying hens exposed to high ambient temperature. *Journal of Thermal Biology*, *103*(July 2021), 103167. https://doi.org/10.1016/j.jtherbio.2021.103167
17. Kuzugudenli E (2018) Relative humidity modeling with artificial neural networks. Appl Ecol Environ Res 16(4):5227–5235. https://doi.org/10.15666/aeer/1604_52275235
18. Kwon SH, Jang KP, Bang JW, Lee JH, Kim YY (2014) Prediction of concrete compressive strength considering humidity and temperature in the construction of nuclear power plants. Nucl Eng Des 275:23–29. https://doi.org/10.1016/j.nucengdes.2014.04.025
19. Langendijk GS, Rechid D, Sieck K, Jacob D (2021) Added value of convection-permitting simulations for understanding future urban humidity extremes: case studies for Berlin and its surroundings. Weather and Climate Extremes 33(February):100367. https://doi.org/10.1016/j.wace.2021.100367
20. Li L, Zha Y (2018) Mapping relative humidity, average and extreme temperature in hot summer over China. Sci Total Environ 615:875–881. https://doi.org/10.1016/j.scitotenv.2017.10.022
21. Li X, Zhang Y, Zhang Y, Liu Y, Gao Z, Zhu G, Xie Y, Mowafy S (2022) Relative humidity control during shiitake mushroom (Lentinus edodes) hot air drying based on appearance quality. *Journal of Food Engineering*, *315*(December 2020), 110814. https://doi.org/10.1016/j.jfoodeng.2021.110814
22. Liao WC, Lee BJ, Kang CW (2008) A humidity-adjusted maturity function for the early age strength prediction of concrete. Cem Concr Compos 30(6):515–523. https://doi.org/10.1016/j.cemconcomp.2008.02.006
23. Liu C, Zhang Y, Sun L, Gao W, Jing X, Ye W (2021) Influence of indoor air temperature and relative humidity on learning performance of undergraduates. Case Stud Therm Eng 28(July):101458. https://doi.org/10.1016/j.csite.2021.101458
24. Ma XQ, He K, Lu JY, Xie L, Ju Y, Zhao LX, Gao WS (2021) Effects of temperature and humidity on ground total electric field under HVDC lines. Electr Power Syst Res 190(August 2020):1–8. https://doi.org/10.1016/j.epsr.2020.106840

25. Martin M, Afshari A, Armstrong PR, Norford LK (2015) Estimation of urban temperature and humidity using a lumped parameter model coupled with an EnergyPlus model. Energy Build 96:221–235. https://doi.org/10.1016/j.enbuild.2015.02.047
26. Mathworks (2021) *Long short-term memory networks*. https://www.mathworks.com/help/deeplearning/ug/long-shortterm-memory-networks.html
27. Mba L, Meukam P, Kemajou A (2016) Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region. Energy Build 121:32–42. https://doi.org/10.1016/j.enbuild.2016.03.046
28. Mustafaraj G, Chen J, Lowry G (2010) Development of room temperature and relative humidity linear parametric models for an open office using BMS data. Energy Build 42(3):348–356. https://doi.org/10.1016/j.enbuild.2009.10.001
29. Mustafaraj G, Lowry G, Chen J (2011) Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. Energy Build 43(6):1452–1460. https://doi.org/10.1016/j.enbuild.2011.02.007
30. Nguyen TA, Verboven P, Schenk A, Nicolaï BM (2007) Prediction of water loss from pears (Pyrus communis cv. Conference) during controlled atmosphere storage as affected by relative humidity. *Journal of Food Engineering*, *83*(2), 149–155. https://doi.org/10.1016/j.jfoodeng.2007.02.015
31. Ozbek A, Sekertekin A, Bilgili M, Arslan N (2021) Prediction of 10-min, hourly, and daily atmospheric air temperature: comparison of LSTM, ANFIS-FCM, and ARMA. Arab J Geosci 14(7). https://doi.org/10.1007/s12517-021-06982-y
32. Parishwad GV, Bhardwaj RK, Nema VK (1998) Prediction of monthly-mean hourly relative humidity, ambient temperature, and wind velocity for India. Renewable Energy 13(3):363–380. https://doi.org/10.1016/S0960-1481(98)00010-X
33. Pei C, Ou Q, Pui DYH (2021) Effects of temperature and relative humidity on laboratory air filter loading test by hygroscopic salts. *Separation and Purification Technology*, *255*(August 2020), 117679. https://doi.org/10.1016/j.seppur.2020.117679
34. Qadeer K, Ahmad A, Qyyum MA, Nizami AS, Lee M (2021) Developing machine learning models for relative humidity prediction in air-based energy systems and environmental management applications. J Environ Manage 292(April):112736. https://doi.org/10.1016/j.jenvman.2021.112736
35. Safyari M, Hojo T, Moshtaghi M (2021) Effect of environmental relative humidity on hydrogen-induced mechanical degradation in an Al–Zn–Mg–Cu alloy. Vacuum 192(July):110489. https://doi.org/10.1016/j.vacuum.2021.110489
36. Sekertekin A, Bilgili M, Arslan N, Yildirim A, Celebi K, Ozbek A (2021) Short-term air temperature prediction by adaptive neuro-fuzzy inference system (ANFIS) and long short-term memory (LSTM) network. Meteorol Atmos Phys 133(3):943–959. https://doi.org/10.1007/s00703-021-00791-4
37. Shen D, Wang M, Chen Y, Wang W, Zhang J (2017) Prediction of internal relative humidity in concrete modified with super absorbent polymers at early age. Constr Build Mater 149:543–552. https://doi.org/10.1016/j.conbuildmat.2017.05.121

38. Shi X, Lu W, Zhao Y, Qin P (2018) Prediction of Indoor Temperature and Relative Humidity Based on Cloud Database by Using an Improved BP Neural Network in Chongqing. IEEE Access 6:30559–30566. https://doi.org/10.1109/ACCESS.2018.2844299
39. Sohani A, Shahverdian MH, Sayyaadi H, Garcia DA (2020) Impact of absolute and relative humidity on the performance of mono and poly crystalline silicon photovoltaics; applying artificial neural network. J Clean Prod 276:123016. https://doi.org/10.1016/j.jclepro.2020.123016
40. 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
41. Vincent LA, Mekis É (2006) Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. Atmos - Ocean 44(2):177–193. https://doi.org/10.3137/ao.440205
42. Wang Y, Liu K, Liu Y, Wang D, Liu J (2022) The impact of temperature and relative humidity dependent thermal conductivity of insulation materials on heat transfer through the building envelope. J Building Eng 46(13):103700. https://doi.org/10.1016/j.jobe.2021.103700
43. Wu Y, Wang W, Sun Y, Cui Y, Duan D, Deng S (2020) An equivalent temperature drop method for evaluating the operating performances of ASHP units jointly affected by ambient air temperature and relative humidity. Energy Build 224:110211. https://doi.org/10.1016/j.enbuild.2020.110211
44. Xi L, Luo R, Liu H (2021) Effect of relative humidity on the linear viscoelastic properties of asphalt mixtures. Constr Build Mater 307(July):124956. https://doi.org/10.1016/j.conbuildmat.2021.124956
45. Yang JS, Wang YQ, August PV (2004) Estimation of Land Surface Temperature Using Spatial Interpolation and Satellite-Derived Surface Emissivity. J Environ Inf 4(1):40–47. https://doi.org/10.3808/jei.200400035
46. Zhou M, Aarnink AJA, Huynh TTT, van Dixhoorn IDE, Koerkamp G, P. W. G (2021) Effects of increasing air temperature on physiological and productive responses of dairy cows at different relative humidity and air velocity levels. J Dairy Sci. https://doi.org/10.3168/jds.2021-21164
47. Zolfaghari M, Golabi MR (2021) Modeling and predicting the electricity production in hydropower using conjunction of wavelet transform, long short-term memory and random forest models. Renewable Energy 170:1367–1381. https://doi.org/10.1016/j.renene.2021.02.017

Figures

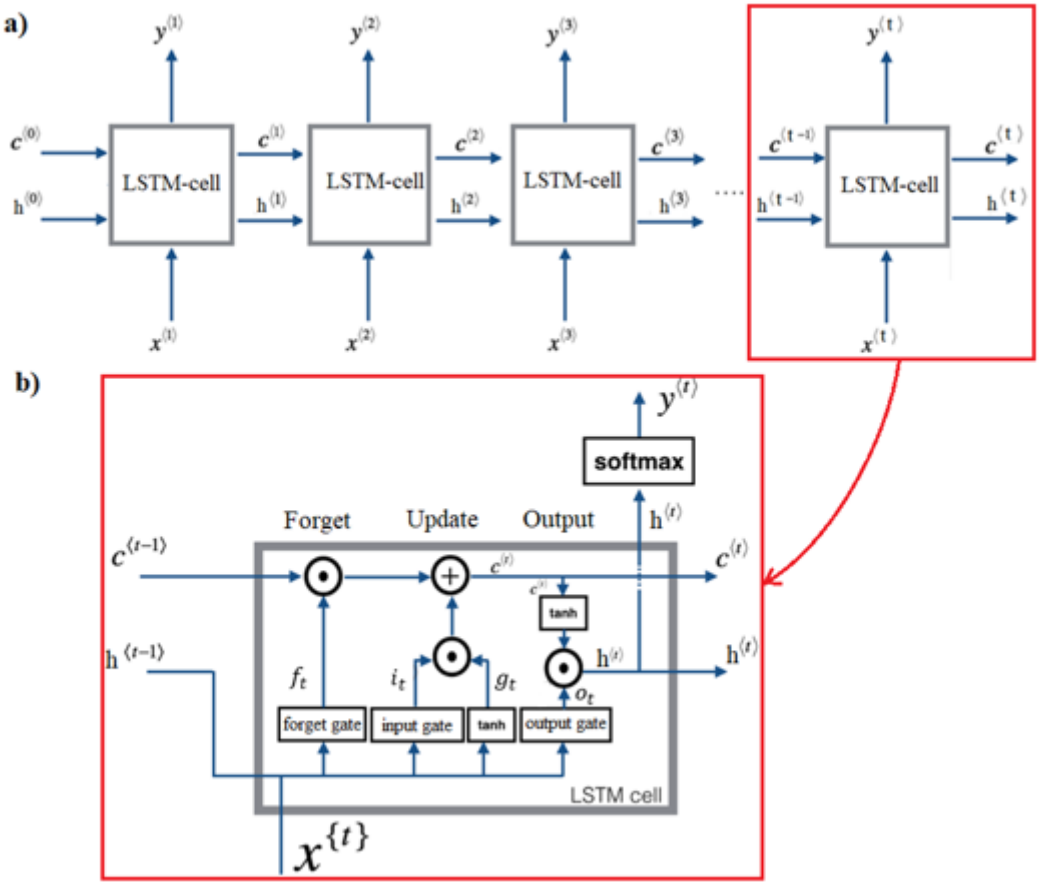

Figure 1

(a) LSTM layer architecture, (b) the flow of data at time step *t*

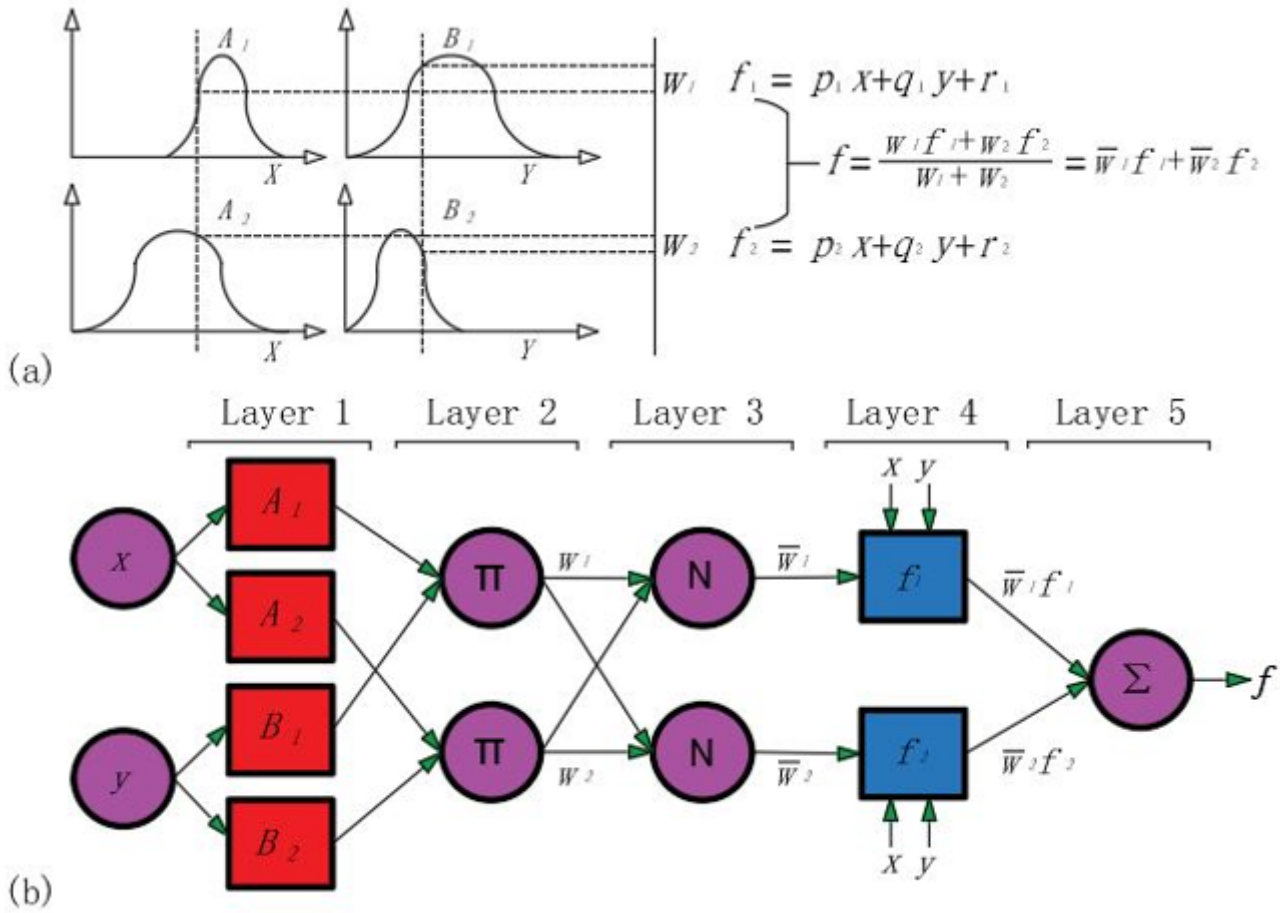

Figure 2

ANFIS architecture structure (a) Type-3 fuzzy reasoning; (b) equivalent ANFIS (type-3 ANFIS) (Jang, 1993)

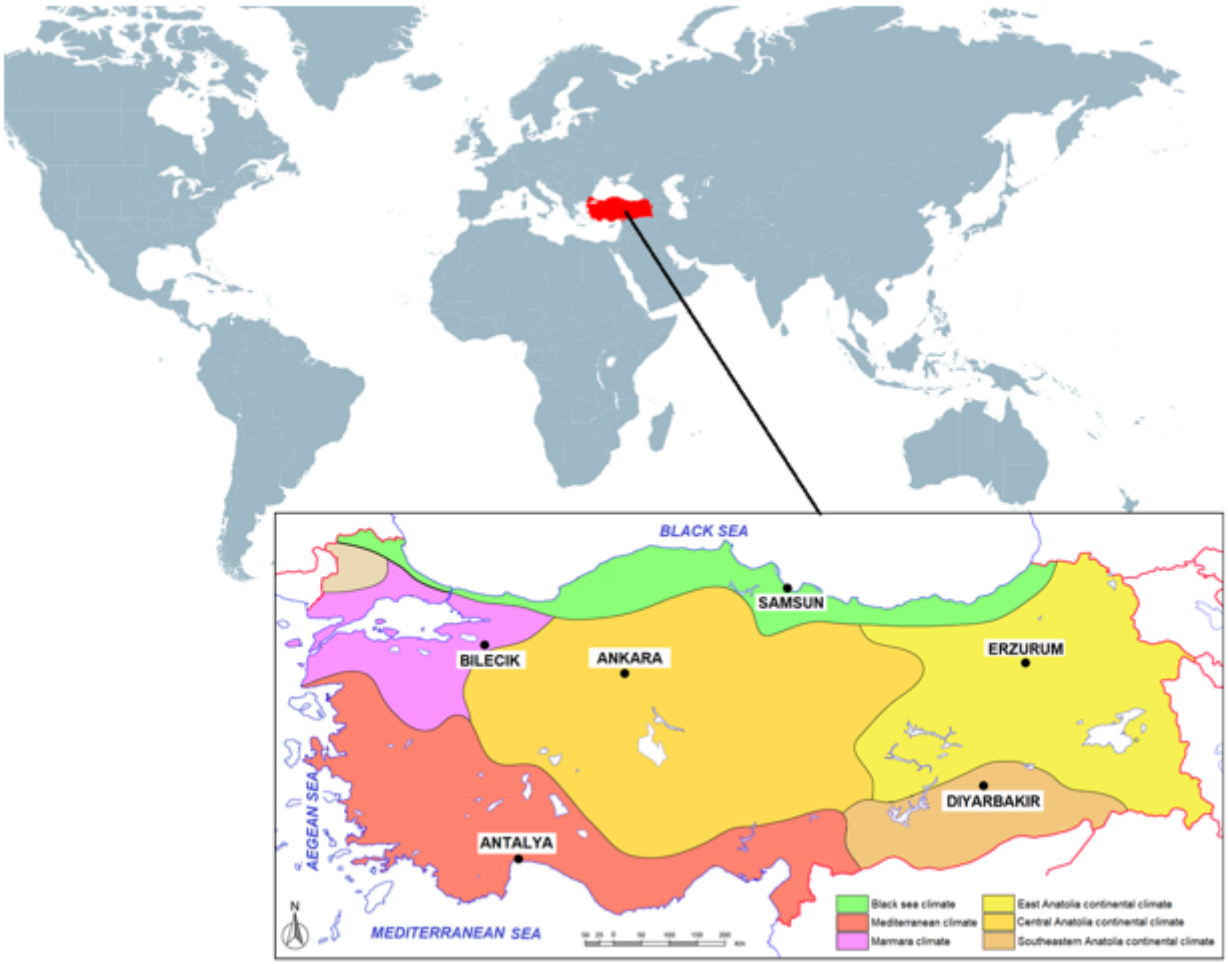

Figure 3

Locations of measuring stations

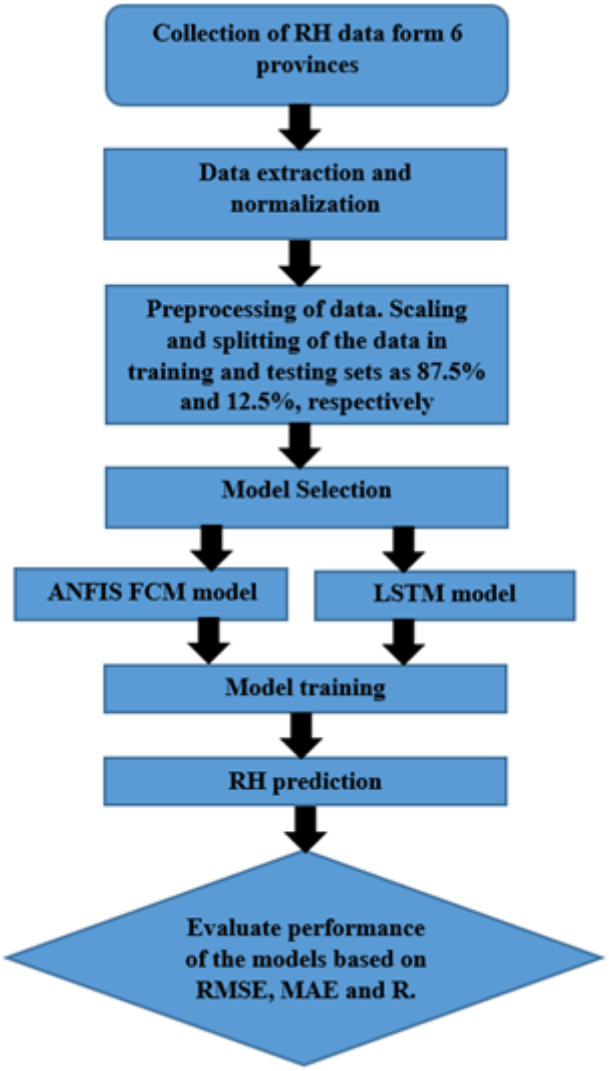

Figure 4

Flowchart for models proposed in RH prediction

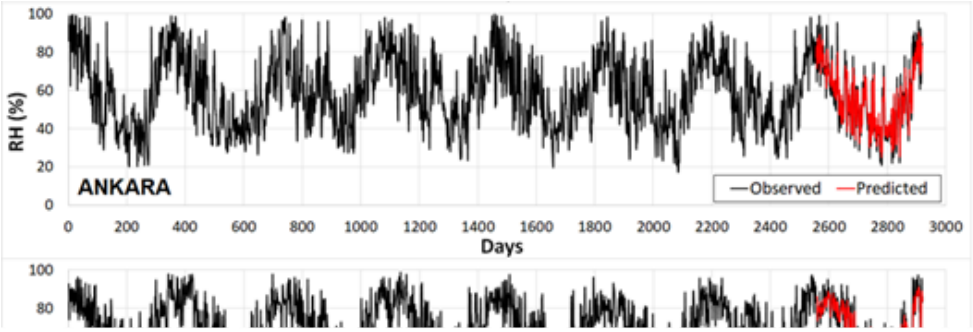

Figure 5

The daily RH data series for the stations with observed (black) and predicted (red) values.

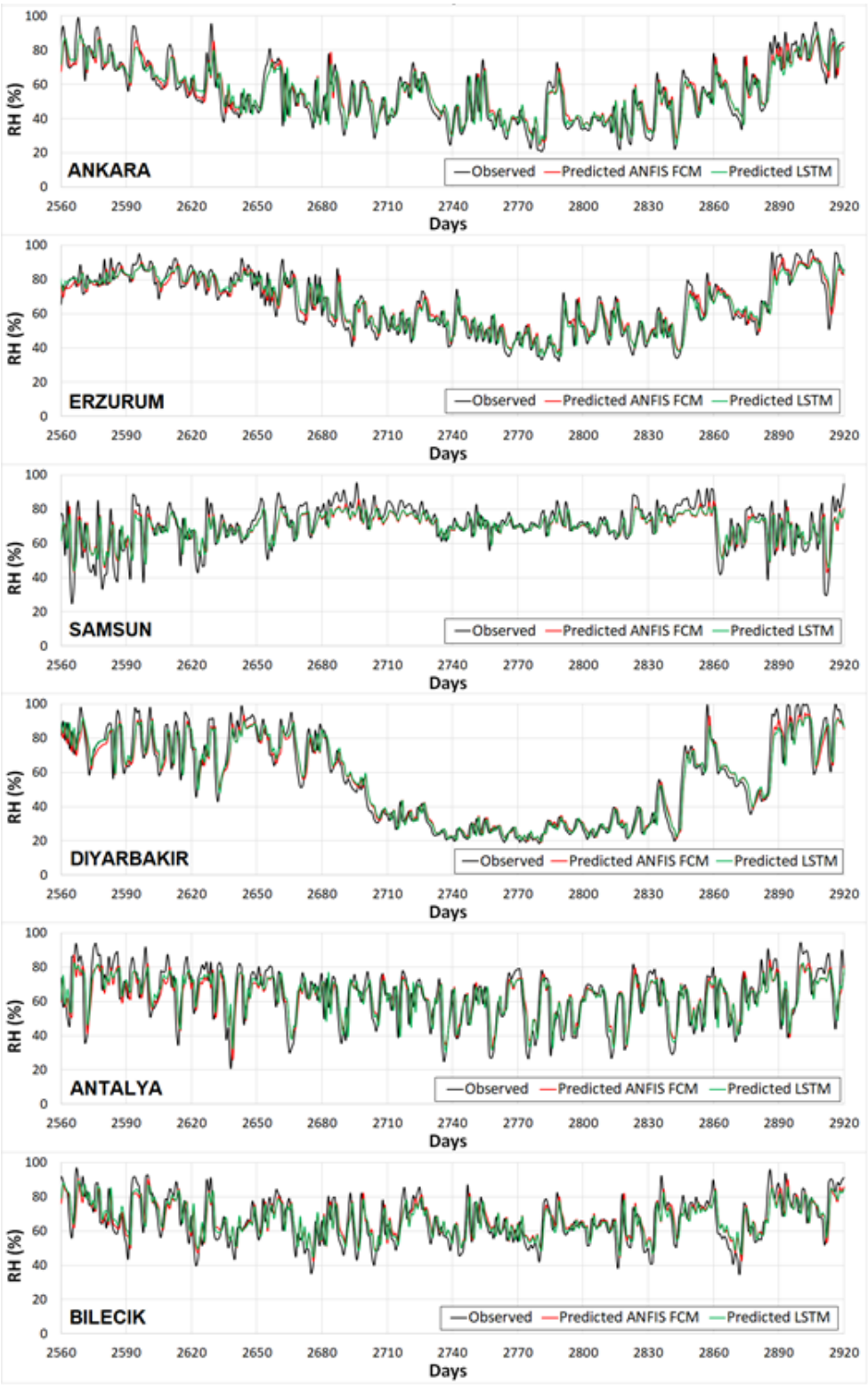

Figure 6

The hourly testing RH data observed (black) and predicted values for ANFIS FCM (red) and LSTM (green) methods

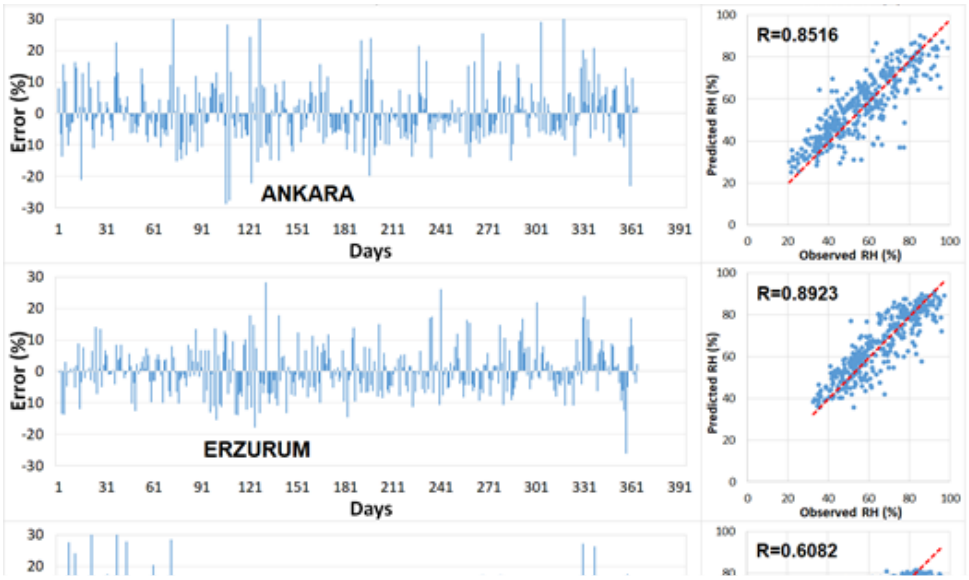

Figure 7

The error value and regression plot of the observed (real) and predicted RH values for LSTM method

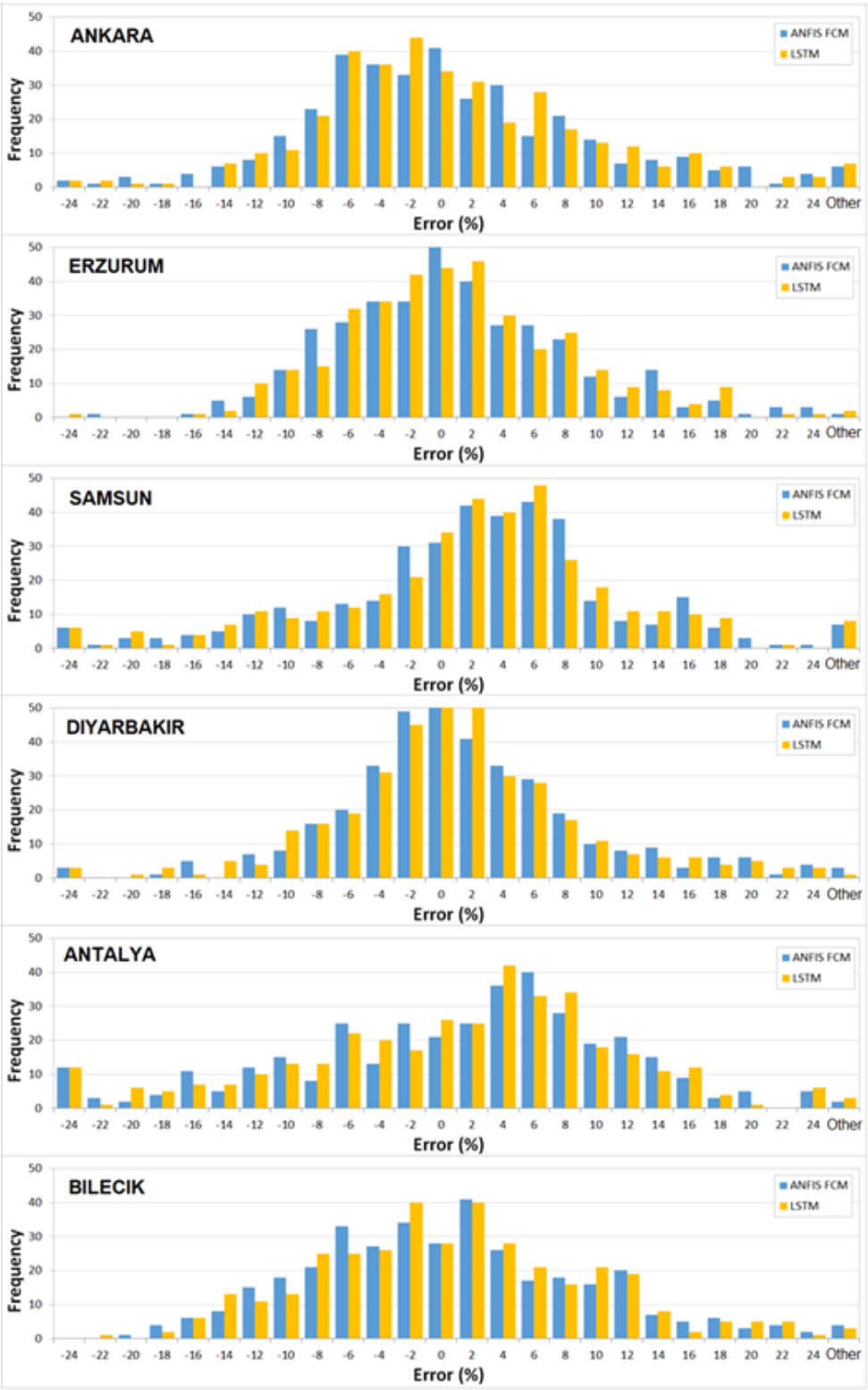

Figure 8

Histogram of prediction errors for the models at all provinces

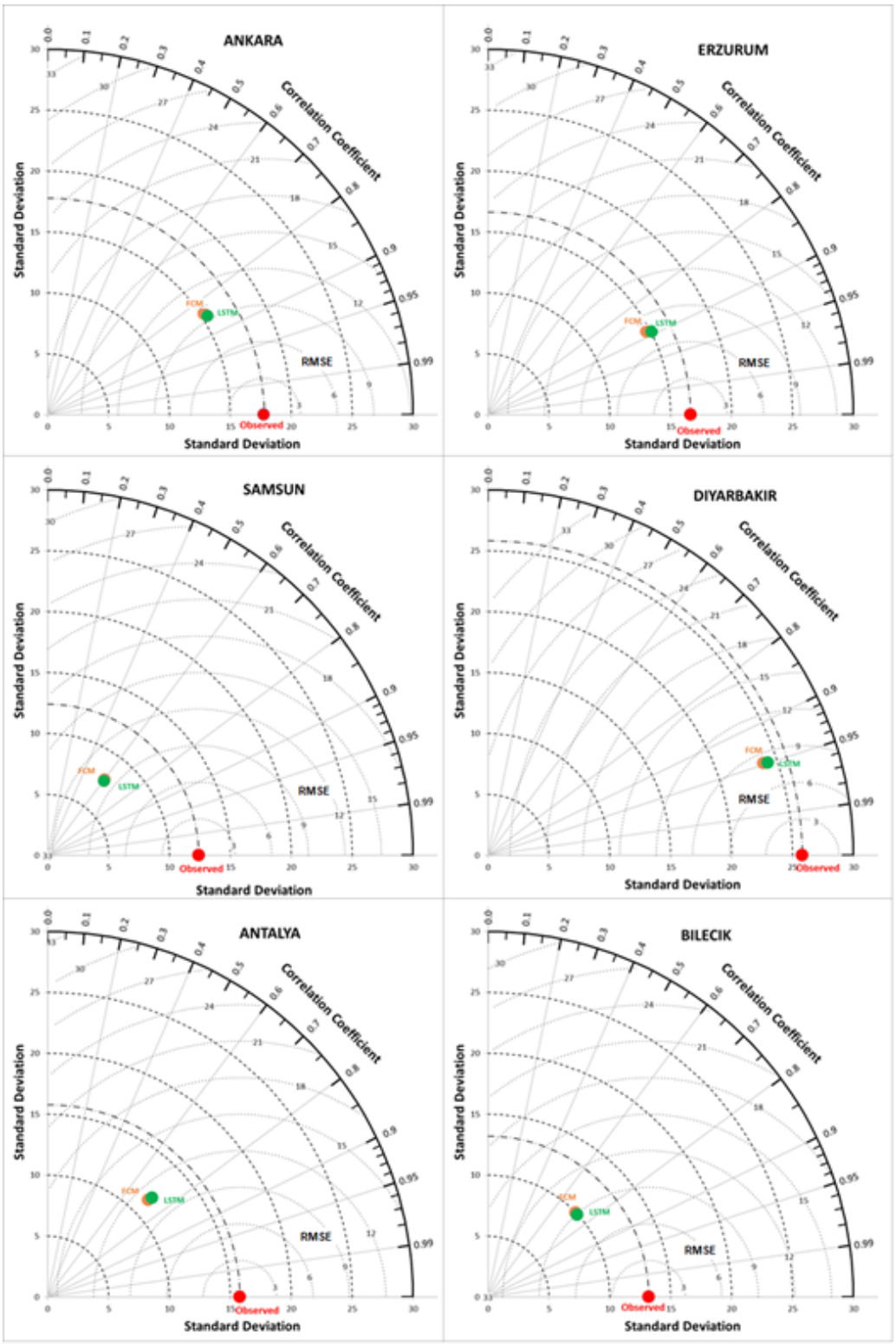

Figure 9

Taylor diagrams of the proposed models' errors for the test data