



Deep learning-based effective fine-grained weather forecasting model

Pradeep Hewage¹ · Marcello Trovati¹ · Ella Pereira¹ · Ardhendu Behera¹

Received: 11 September 2019 / Accepted: 9 June 2020 / Published online: 22 June 2020
© The Author(s) 2020

Abstract

It is well-known that numerical weather prediction (NWP) models require considerable computer power to solve complex mathematical equations to obtain a forecast based on current weather conditions. In this article, we propose a novel light-weight data-driven weather forecasting model by exploring temporal modelling approaches of long short-term memory (LSTM) and temporal convolutional networks (TCN) and compare its performance with the existing classical machine learning approaches, statistical forecasting approaches, and a dynamic ensemble method, as well as the well-established weather research and forecasting (WRF) NWP model. More specifically Standard Regression (SR), Support Vector Regression (SVR), and Random Forest (RF) are implemented as the classical machine learning approaches, and Autoregressive Integrated Moving Average (ARIMA), Vector Auto Regression (VAR), and Vector Error Correction Model (VECM) are implemented as the statistical forecasting approaches. Furthermore, Arbitrage of Forecasting Expert (AFE) is implemented as the dynamic ensemble method in this article. Weather information is captured by time-series data and thus, we explore the state-of-art LSTM and TCN models, which is a specialised form of neural network for weather prediction. The proposed deep model consists of a number of layers that use surface weather parameters over a given period of time for weather forecasting. The proposed deep learning networks with LSTM and TCN layers are assessed in two different regressions, namely multi-input multi-output and multi-input single-output. Our experiment shows that the proposed lightweight model produces better results compared to the well-known and complex WRF model, demonstrating its potential for efficient and accurate weather forecasting up to 12 h.

Keywords Long short-term memory · Temporal convolutional networks · Weather prediction · WRF · Neural network · Time-series data analysis

1 Introduction

Weather forecasting refers to the scientific process of predicting the state of the atmosphere based on specific time frames and locations [1]. Numerical weather prediction (NWP) utilises computer algorithms to provide a forecast based on current weather conditions by solving a large system of nonlinear mathematical equations, which are based on specific mathematical models. More specifically, these models define a coordinate system, which divides the earth into a 3-dimensional grid. The weather parameters such as winds, solar radiation, the phase change of water, heat transfer, relative humidity, and surface hydrology are measured

within each grid and their interaction with neighbouring grids to predict atmospheric properties for the future [2].

Meteorology adopted a more quantitative approach with the advancement of technology and computer science, and forecast models became more accessible to researchers, forecasters, and other stakeholders. Many NWP systems were developed in recent years, such as Weather Research and Forecasting (WRF) model, where increasing high-performance computing power has facilitated the enhancement and the introduction of regional or limited area models [3]. As a consequence, the WRF model became the world's most-used atmospheric NWP model due to its higher resolution rate, accuracy, open-source nature, community support, and a wide variety of usability within different domains [4, 5].

According to [1], data-driven computer modelling systems can be utilised to reduce the computational power of NWPs. In particular, artificial neural network (ANN) can be used for this purpose due to their adaptive nature and

✉ Pradeep Hewage
23147458@go.edgehill.ac.uk

¹ Department of Computer Science, Edge Hill University, Ormskirk, Lancashire L39 4QP, UK

learning capabilities based on prior knowledge. This feature makes the ANN techniques very appealing in application domains for solving highly nonlinear phenomena. Deep models for multivariate time-series forecasting often use Recurrent Neural Networks (RNN) and Temporal Convolutional Networks (TCN). Recently, a variant of RNN called Long Short-Term Memory (LSTM) has attracted considerable attention due to its superior performance. Such models have attracted considerable attention due to their superior performance [6–8]. Deep networks often use stacked neural networks and include several layers as part of the overall composition known as nodes. The computation takes place at the node level since it allows the combination of data input through a set of coefficients. Subsequently, the activation function gets established on the basis of input-weight products while signal progresses through the network [9]. Regression technique is often employed to develop and evaluate neural network models for accurate weather prediction as the weather information is captured by time-series data consisting of real numbers [10].

This article presents developing and evaluating a light-weight and novel weather forecasting system using modern neural networks. Figure 1 depicts a general overview of the research discussed in this article. More specifically, a suitable machine learning model is proposed by exploring temporal modelling approaches of LSTM and TCN, and compare its performance with classical machine learning approaches, statistical forecasting models, and a dynamic ensemble method. Secondly, we use the proposed model for short-term weather prediction and compare the model accuracy with the well-established WRF model. Finally, we reform the model for long-term weather forecasting, and analyse the model accuracy and compared the performance to the state-of-art WRF model.

In this study, we investigate LSTM and TCN over RNN since there is an inherent issue of the vanishing gradient problem with the RNN [6]. The LSTM and TCN can overcome this vanishing gradient issue, but it can easily use up the high capacity of memory [8, 11]. The rest of the article is organised as follows: Sect. 2 focuses on related work,

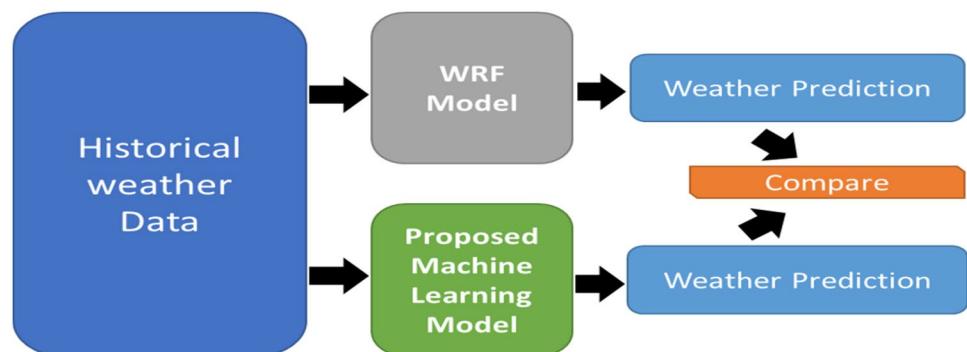
and Sect. 3 discusses the research aims and objectives. In Sect. 4, we present the WRF model and its challenges, and Sect. 5 discusses the sequence modelling and prediction. In Sects. 6 and 7, we discuss the methodology and results. Finally, Sect. 8 concludes the article.

2 Related work

Numerical weather prediction (NWP) concept was proposed by Lewis Fry Richardson in 1922, and practical use of NWP began in 1955 after the development of programmable computers [1]. Neural networks-based weather forecasting has been evolved significantly in the last three decades. Before the year 2000, the model output statistics (MOS) was the most widely used approach to improve the numerical models' ability to forecast by relating model outputs to observational data [12–14]. A mixed statistical or dynamic technique for the weather forecasting was introduced by [15] in 1983. The work in [16] added a new perception to dynamic modelling in 1991. These approaches have limitations and challenges such as massive computational requirements, lack of design methodologies for selecting the model architecture and parameters, and time-consuming to prediction resulting less reliability as the difference between the current time and the forecast time increases [13, 16, 17].

Artificial neural network-based minimum temperature prediction system was introduced in 1991 using the back-propagation algorithms [18, 19]. This concept considerably reduced the computational requirements of MOS directing an effective forecast [16]. A snowfall and rainfall forecasting model was introduced in 1995 from weather radar images with ANN [20]. The results show that the ANN is more effective than the traditional cross-correlation method, and the persistence prediction method is producing a substantial reduction in prediction error. In 1998, Oishi et al. developed a severe rainfall prediction method using AI [21]. The development method was unique as it is introduced inference (i.e. knowledge-based) rather than using numerical simulations. A multi-polynomial high order neural network

Fig. 1 Overview of the research



(M-PHONN)-based rainfall prediction model was developed by Hui Qi and Ming Zhang in 2001 [22]. This new model has features such as increasing the speed, accuracy, and the robustness of the rainfall estimate. Therefore, this model could be used to complement the already established auto-estimator algorithms.

A multilayer perceptron network was trained with the backpropagation algorithm with momentum for temperature forecasting in 2002 [23]. The results were very encouraging and clearly demonstrated the potential for future weather forecasting applications. In the same year, a comparative was carried out analysing different neural network models for daily maximum and minimum temperature, and wind speed [24]. The results show that the radial basis function network (RBFN) produced the most accurate forecast compared to the Elman recurrent neural network (ELNN) and multi-layered perceptron (MLP) networks. In 2005, a rough set of fuzzy neural network was introduced to forecast weather parameters; dew temperature, wind speed, temperature, and visibility [25]. This model has several fuzzy rules, and their initial weights were estimated with a deeper network for weather forecasting. Moreover, Hayati and Mohebi proposed a successful model for temperature forecasting based on MLP.

A feature-based neural network model was introduced in 2008 to predict maximum temperature, minimum temperature, and relative humidity [26]. Neural network features are extracted over different periods as well as from the time-series weather parameter itself. In particular, feedforward ANN is utilised in this approach with backpropagation for supervised learning. The prediction results have a high degree of accuracy, and this modelling is recommended as an alternative to traditional meteorological approaches by [27–29]. In 2012, a backpropagation neural network (BPN) was implemented for temperature forecasting [27, 30]. This network has successfully identified the nonlinear structural relationship between various input weather parameters. Furthermore, a new hybrid model was introduced in 2014 to forecast the temperature which is based on an ensemble of neural networks (ENN) [31], and the results suggested that including image data would improve the prediction results. In the same year, a deep neural network-based feature representation for weather prediction model was developed for the temperature and dew point prediction [32].

In 2015, eight different novel regression tree structures were applied to short-term wind speed prediction [33]. The author also compared the best regression tree approach against other AI approaches such as support vector regression (SVR), MLP, extreme learning machines, and multi-linear regression approach. The best regression tree yields the best results for wind speed prediction. In the same year, a deep neural network was introduced for ultra-short-term wind forecasting with success [34]. Deep

learning with LSTM layers has been introduced to precipitation nowcasting by Shi et al. [11]. The experimental results show that the LSTM network has the ability to capture spatiotemporal correlations and can be used to precipitation nowcasting. In the same year, a model was developed to predict the temperate in Nevada using a deep neural network with stacked denoising auto-encoders with higher accuracy of 97.97% compared to traditional neural networks (94.92%) [35]. In 2016, the multi-stacked deep learning LSTM approach was utilised to forecasting weather parameters temperature, humidity, and wind speed [36]. The author suggested that the model could be used to predict other weather parameters based on the effectiveness and accuracy of the results.

Traditional machine learning methods were analysed for radiation forecasting in 2017 [37]. The author concluded that the SVR, regression trees, and forests have produced a promising outcome for radiation forecasting. In 2018, the backpropagation neural (BPN) network's performance compared with linear regression and regression tree for temperature forecasting [38]. As a result, a significant better temperature yields the BPN. In 2018, a short-term local rain and temperature forecasting model was developed using deep neural network [39]. The author concluded that the deep neural networks yield the highest accuracy for rain prediction among several machine learning methods. In the same year, the neural network approach is utilised to create models to predict sea surface temperature and soil moisture [40, 41].

The selected state-of-the-art deep learning approaches for weather forecasting and their contributions and differences with the previous approaches are discussed in Table 1.

The above existing weather forecasting models are able to predict up to maximum three weather parameters. Besides, weather forecasting is an entirely nonlinear process, and each parameter often depends upon one or more other parameters [13, 42, 43]. These larger numbers of interrelated parameters work together, aiming for an accurate weather forecast in a more reliable NWP such as met office and WRF models [4, 44]. A maximum of up to four input weather parameters is considered in the existing AI-based forecasting models.

Based on the related work, it is evident that:

- There is no identified attempt to compare an AI-based weather prediction with a well-established and existing weather forecasting model such as WRF;
- There has been little or no attempt to compare traditional machine learning approaches with cutting-edge deep learning technologies for weather forecasting;
- Most of the existing approaches use less than four interrelated input parameters for neural network-based weather forecasting model;

Table 1 Existing deep learning approaches and their contributions

Deep learning approach for weather forecasting	Contribution and difference with the previous approaches
Deep neural networks for ultra-short-term wind forecasting [34]	Results show that carefully selection of deep neural networks outperforms shallow ones. The model accepts a single input parameter and predicts a single parameter, and the model is limited to very short-term forecasting (less than an hour)
Weather forecasting using deep learning techniques [61]	Recurrent neural network is used for prediction of the rainfall with adequate accuracy level. The model uses a single input single output and is used for short-term forecasting
Short-term local weather forecast using dense weather station by deep neural network [39]	Deep neural network is used to predict rain and temperature. The researchers use four input parameters and predict one parameter at a given time. This model is able to predict data accurately up to an hour
Convolutional LSTM network: a machine learning approach for precipitation nowcasting [11]	Formulated precipitation nowcasting as a spatiotemporal sequence forecasting problem. The proposed model is a Single-input single-output and able to produce a state-of-the-art performance for up to 6 h
Forecasting the weather of Nevada: a deep learning approach [35]	This model accepts four input parameters and predicts one output as temperature. Results indicated that stacked denoising auto-encoder deep learning model predicts accurate long-term temperature
Sequence to sequence weather forecasting with long short-term memory recurrent neural networks [36]	Multi-stacked LSTMs are used to map sequences of weather values of the same length. Use three input parameters and predict one parameter at a time
A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction [62]	Contributed the bidirectional gated recurrent network for wind power forecasting. The model used wind direction and wind speed as inputs and predicted the results more accurately up to 6 h

- A complete AI-based weather forecasting model with up to 10 input/output weather parameters is yet to be explored.

3 Research aim and objectives

The work presented in this article aimed to develop a weather forecasting model to address the above-mentioned drawbacks using state-of-the-art deep models by establishing the following objectives.

1. To propose an efficient neural network-based weather forecasting model by exploring temporal modelling approaches of LSTM and TCN, and compare its performance with the existing approaches;
2. Use the proposed neural network model for short-term weather prediction and compare the results with WRF model prediction;
3. Fine-tune the proposed model for long-term weather forecasting;
4. Compare the model performances for long-term forecasting with the WRF model prediction.

Our approach is targeted to develop deep neural networks to solve the regression problem of weather forecasting. We propose two different regression models to assess proposed deep learning models, namely multi-input multi-output (MIMO) and multi-input single-output (MISO). In this

article, we addressed the above objectives in detail in various sections. Objective 1, an effective neural network-based weather forecasting model is proposed and compared its performance with existing approaches in Sect. 7.1. Objective 2, the proposed model is used to short-term weather forecasting and compared its performance with the WRF model predictions in Sect. 7.2. For Objective 3 and Objective 4, the proposed model is fine-tuned for long-term forecasting and compared the results with the WRF model predictions in Sect. 7.3.

4 Weather research and forecasting (WRF) model

The WRF model was developed by Norwegian physicist Vilhelm Bjerknes in the latter part of the 1990s as part of a collaborative partnership with many environmental and meteorology organisations. The model involves solving of various thermodynamic equations so that numerical weather-based predictions can be made mainly through different vertical levels [45, 46]. The primary role of the WRF is to carry out analysis focusing on climate time scale via linking physics data between land, atmosphere and ocean. The WRF model is currently the world's most-used atmospheric model since its initial public release in the year 2000 [5].

In order to investigate the model for real cases, it is necessary to install and configure WPS (WRF pre-processing system), WRF ARW (advanced research WRF model), and

post-processing software. The WRF post-processing is not described in this article, as the main objective is to collect historical weather data for prediction and analyses. Interested researchers can refer to [47] for further details. The WRF ARW and the WPS share common routines, like WRF I/O API. Therefore, the successful compilation of the WPS depends upon the successful compilation of the WRF ARW model [4].

The WRF model needs to run in two different modes to extract time-series data. Firstly, historical weather data are collected and subsequently, predicted weather data is identified for evaluation purposes. For each instance, the model runs in a single domain mode and utilises different “*namelist.wps*” and “*namelist.input*” files to configure the WPS and WRF-ARW components [17]. GRIdded binary or general regularly distributed information in binary, often used as GRIB data, which is a concise data format commonly used in meteorology to store historical and forecast weather data [17, 48]. According to [49], Global Forecast System (GFS) GRIB data provides 0.25 degrees resolution and available to download every 3 h freely. Therefore, the GFS 3-hourly data are selected for this project, with a horizontal resolution set to 10 km.

One of the primary challenges in the WRF is its requirement for massive computational power to solve the equations that describe the atmosphere. Furthermore, atmospheric processes are associated with highly chaotic dynamical systems, which cause a limited model’s accuracy. As a consequence, the model forecast capabilities are less reliable as the difference between the current time and the forecast time increases [1, 50]. In addition, the WRF is a large and complex model with different versions and applications, which lead to the need for greater understanding of the model, its implementation and the different option associated with its execution [5]. The GFS 0.25 degrees dataset is the freely available highest resolution dataset for the WRF model. This allows the user to forecast weather data at a horizontal resolution about 27 km [48, 49]. This implies that the user can predict data with increased accuracy up to 27 km. The model calculates the lesser resolution data based on results obtained. Thus, the model obtains better results for long-range forecast and not for a selected geographical region, such as a farm, school, places of interest, and so on [5, 17, 51].

Based on the above discussion, we propose a novel lightweight weather prediction model that could run on a standalone PC for accurate weather prediction and could easily be deployed in a selected geographical region.

Table 2 Surface weather parameters (10 identified parameters used by our model)

Variable	Description	Measuring unit
TSK	Skin temperature or surface temperature	°K
PSFC	Surface pressure	Pa
U10	X component of wind at 10 m	m/s
V10	Y component of wind at 10 m	m/s
Q2	2-m specific humidity	Kg/Kg
Rainc	Convective rain (accumulated precipitation)	mm
Rainnc	Non-convective rain	mm
Snow	Snow water equivalent	Kg/m ²
TSLB	Soil temperature	°K
SMOIS	Soil moisture	m ³ /m ³

5 Sequence modelling and prediction

The modelling task has been highlighted before defining a network structure which involves time-series weather data sequence x_0, \dots, x_T and wish to predict some corresponding outputs y_0, \dots, y_T at each time. As presented in Table 2, there are 10 different weather parameters in data at a given time t , $x_t = [p_1, \dots, p_{10}]$. The aim is to predict the value y_t at time t , which is constrained to only previously observed inputs: x_0, \dots, x_{t-1} . Therefore, the sequence modelling network can be defined as a function $\mathcal{F} : \mathcal{X}^{T+1} \rightarrow \mathcal{Y}^{T+1}$ that produces the mapping $\hat{y}_0, \dots, \hat{y}_T = \mathcal{F}(x_0, \dots, x_T)$, if it satisfies the causal constraints, i.e. y_t only depends on x_0, \dots, x_t and not on any future inputs x_{t+1}, \dots, x_T . The main idea of learning in the sequence modelling is to find a network \mathcal{F} which minimises the loss (ℓ) between the actual outputs and the predictions, $\ell(y_0, \dots, y_T, \mathcal{F}(x_0, \dots, x_T))$ in which the sequences and predictions are drawn according to some distribution.

The WRF model with GFS-GRIB data can produce a large amount of historical weather data. Recurrent neural networks (RNN), LSTM, and TCN are extremely expressive models which are appropriate in such a scenario. These networks have attracted considerable attention due to their superior performance based on ability to learn highly complex vector-to-vector mapping [52, 53]. The LSTM is a specialised form of RNN that is designed for sequence modelling [52, 54]. Highly dimensional hidden states \mathbf{H} are the basic building blocks of RNN which are updated with nonlinear activation function \mathcal{F} . At a given time t , the hidden state \mathbf{H}_t is updated by $\mathbf{H}_t = \mathcal{F}(\mathbf{H}_{t-1}, x_t)$. The structure of \mathbf{H} works as the memory of the network.

The state of the hidden layer at a given time is conditioned on its previous state. The RNN is extremely

deep as they are maintained a vector activation through time at each timestep. This will result in high training time-consuming due to the exploding and the vanishing gradient problems [6]. The development of LSTM and TCN architectures have been addressed the gradient vanishing issue with RNN [55]. Therefore, we use state-of-art LSTM and TCN architecture to minimise the loss $\ell(y_0, \dots, y_T, \mathcal{F}(x_0, \dots, x_T))$ for effective modelling and prediction of time-series weather data.

5.1 Proposed deep model with long short-term memory (LSTM) layers

The proposed model is based on LSTM networks and uses temporal weather data to identify the patterns and produces weather predictions. As discussed in Sect. 5, we experiment with the state-of-the-art LSTM, which is a specialised form of RNN, and it is widely applied to handle temporal data. The key concepts of the LSTM have the ability to learn long-term dependencies by incorporating memory units. These memory units allow the network to learn, forget previously hidden states, and update hidden states [6, 9, 56]. Figure 2a shows the deep learning model consisting of stacked LSTM layers for weather forecasting using surface weather parameters. Table 2 describes the surface weather parameters, which are used as the input parameters. The model provides outputs, which are the predicted weather parameters.

Figure 2b shows the LSTM memory architecture used in our model. More specifically, the proposed model has the input vector $X_t = [p_1, p_2, \dots, p_9, p_{10}]$ at a given time step t , which consists of 10 different ($p_1 \dots p_{10}$) weather parameters. In a given time t , the model updates the memory cells for

long-term C_{t-1} and short-term H_{t-1} recall from the previous timestep $t - 1$ via:

$$I_t = \tanh(w_{xi}X_t + w_{hi}H_{t-1} + b_i)$$

$$J_t = \text{sigm}(w_{xj}X_t + w_{hj}H_{t-1} + b_j)$$

$$F_t = \text{sigm}(w_{xf}X_t + w_{hf}H_{t-1} + b_f)$$

$$O_t = \tanh(w_{xo}X_t + w_{ho}H_{t-1} + b_o)$$

$$C_t = C_{t-1} \odot F_t + I_t \odot J_t$$

$$H_t = \tanh(C_t) \odot O_t \quad (1)$$

The notations of Eq. (1) are: w_* —weight matrices, b_* —biases, \odot —element-wise vector product, I_t —input gate and J_t —input moderation gate contributing to memory, F_t —forget gate, and O_t —output gate as a multiplier between memory gates. To allow the LSTM to make complex decisions over a short period of time, there are two types of hidden states, namely C_t and H_t [6, 57]. The LSTM has the ability to selectively consider its current inputs or forgets its previous memory by switching the gates I_t and F_t . Similarly, the output gate O_t learns how much memory cell C_t needs to be transferred to the hidden state H_t . Compared to the RNN, these additional memory cells give the ability to learn enormously complex and long-term temporal dynamics with the LSTM.

In this work, we propose two types of deep models to solve the regression problem involving weather forecasting,

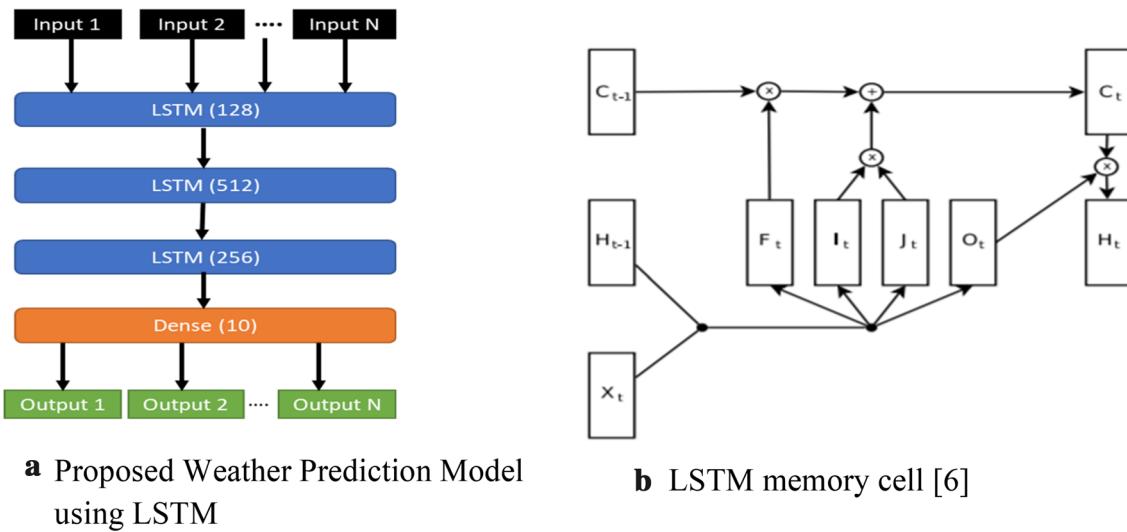


Fig. 2 a Proposed layered LSTM and b LSTM memory cell used for this research

namely multi-input multi-output (MIMO) and multi-input single-output (MISO).

5.1.1 MIMO-LSTM

In the MIMO, all the weather parameters (i.e. 10 surface weather parameters in this study) are fed into the network, which is expected to predict the same number of parameters (i.e. 10 parameters in this study) as the output. Therefore, only one model is required for weather forecasting. Figure 3a depicts the basic arrangement of the MIMO.

5.1.2 MISO-LSTM

In MISO approach, all of the weather parameters (i.e. 10 surface weather parameters in this study) are fed into the network, which is expected to predict a single parameter. 10 different models are required as each of them is trained to predict a particular weather parameter. Figure 3b depicts the basic arrangement of the MIMO and the MISO.

5.2 Proposed deep model with temporal convolutional network (TCN) layers

The main characteristic of the TCN is that the network can take a sequence of any length as inputs and map it to an output sequence of the same length, just similar to the RNN categories. These networks involve causal convolutions and initially developed to examine long-range patterns using a hierarchy of temporal convolutional filters [8, 58, 59]. TCN architecture is quite simple and is informed by recent generic convolutional architectures for sequential data. This architecture has no skip connections across layers, conditioning, context stacking or gated activations, and autoregressive prediction and a very long memory.

The TCNs use dilated convolutions that enable an exponentially large receptive field, allowing very deep networks and very long effective history [60]. For instance, the dilation convolution operation F for a 1-D sequence of a given weather parameter p^1 , i.e. $p = (p_0^1, \dots, p_t^1)$ and a filter $f : \{0, \dots, k-1\}$, on element $s = p_{\hat{i}}^1$ (where $\hat{i} = 0, \dots, t$) of the sequence is defined as:

$$F(s) = (p *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot p_{s-d,i} \quad (2)$$

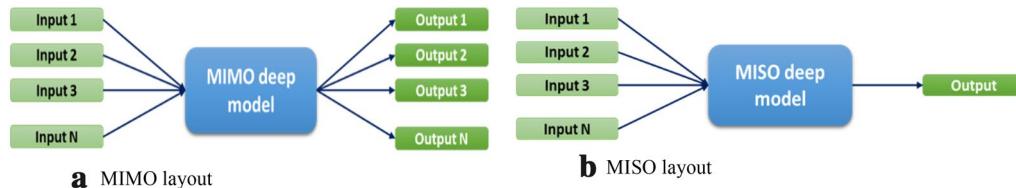


Fig. 3 The proposed MIMO and MISO deep architecture for weather forecasting

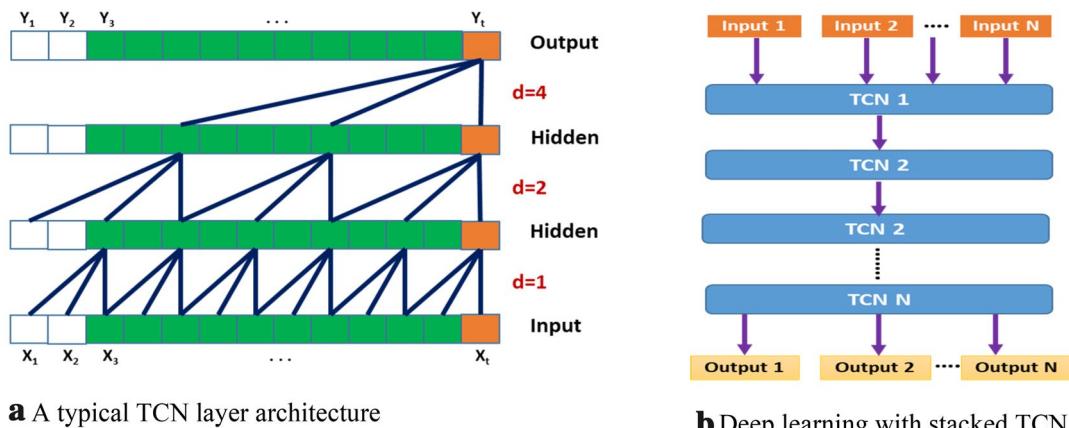


Fig. 4 Architectural elements in a TCN with causal convolution and different dilation factors. The input to the TCN is x_t and output y_t . The x_t contains 10-dimensional weather parameter

The notations of Eq. (2) are: d —dilation factor, k —filter size, and $s - d.i$ accounts for the direction of the past. The TCN consists of stacked units of one dimensional convolution with activation functions [7]. The architectural elements in a TCN with configurations dilations dilation factors $d = 1, 2$, and 4 are shown in Fig. 4. The main purpose of the dilation to introduce a fixed step between every adjacent filter taps, and larger dilations and larger filter sizes k enable effectively expanding the receptive field [8, 59]. The increment of d exponentially increase the depth of the network in these convolutions and this guarantees that there is some filters that hits each input within the effective history [59].

5.2.1 MIMO-TCN and MISO-TCN

Similar to LSTM in Sects. 5.1.1 and 5.1.2, we also use the TCN in our proposed MIMO and MISO models.

5.3 Proposed model for weather forecasting

As discussed in Sects. 5.1 and 5.2, the LSTM and TCN deep learning approaches are proposed for weather forecasting. The MIMO and MISO are the two types of deep models to solve the regression problem. Therefore, proposed models for weather forecasting are MIMO-LSTM, MISO-LSTM, MIMO-TCN, and MISO-TCN. Deep learning models are discussed in [11, 34, 39, 61, 62] are single input single output models. The MISO are experimented in [35, 36] and a MIMO is discussed in [63]. All these models can be accepted up to four input parameters at a given time. Increased number of input parameters will increase the forecasting accuracy of an NWP model by distinguishing interrelationships among parameters [17, 47]. Our proposed model uses ten input parameters which has not been explored in the past for neural network-based weather forecasting. Subsequently, the research discusses in this article is explored for both MIMO and MISO.

Moreover, [62] uses the bidirectional recurrent network with weather-related input parameters successfully to predict the wind power up to 6 h. Therefore, bidirectional LSTM experiments in long-term forecasting and compare with the proposed model. Most of the researches discussed in Table 1 are attempted to forecasting a single or few parameters for a specific purpose rather developing a complete weather forecasting model. Our proposed model explores to complete AI-based fine-grained weather forecasting model.

We use Keras as a tool to implement both LSTM and TCN deep learning networks [56, 64–66].

6 Methodology

This is an empirical-based study and is focused on analysing the quantitative temporal weather data. There are 10 surface weather parameters utilised in this research for weather prediction. These weather parameters are identified by considering their usefulness in precision farming. Moreover, these surface parameters can be captured at a chosen location using various sensors using a local weather station.

6.1 Surface weather parameters

The surface weather parameters are observed and reported in for monitoring and forecasting purposes [67]. In our previous study, we defined 10 surface weather parameters for the forecasting, which can be extruded from GRIB data using the WRF model [66]. Those 10 surface parameters, as shown in Table 2.

The surface parameters of wind direction and wind speed can be calculated from the WRF surface variables U_{10} and V_{10} [4]. Table 2 shows the surface weather parameters which are utilised in this research. The *XLAT*—reference latitude and *XLONG*—reference longitude parameters are used with each data point for the location identification.

6.2 Data collection and preparation

As described in Sect. 4, the GRIB data is used to run the WRF model. A total of 12 weather parameters are extracted from the period of January 2018 to May 2018. This is used as the training dataset to train the proposed models. Similarly, the parameters in June 2018 data are used to test the network. This is to test different trained deep models to identify the best model for forecasting. The parameters in July 2018 are considered as the validation dataset, which is used as the ground truth to compare prediction from the best model. The WRF model is being run in forecast mode using the same format GRIB data for the month of July 2018

Table 3 Different LSTM layer configurations for MIMO-LSTM and MISO-LSTM

Configuration	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Config 1	128	512	512	256	
Config 2	256	2048	2048	1024	256
Config 3	256	512	1024	512	256
Config 4	256	1024	1024	512	
Config 5	64	256	512	128	
Config 6	128	512	256		

to evaluate the overall prediction performance of the WRF model.

The training data set has been normalised to keep each value in between –1 and 1, and the same maximum and minimum variable values are used to normalise the testing and the evaluation data set. We apply a sliding window of 7 days temporal resolution on each dataset as input to the model and the temporal resolution of next 3 h data as the model's output. By using this sliding window method, the size of our training dataset is ~6.5 GB with a sample size of 675,924, and the testing dataset is ~1.19 GB with a sample size of 114,450.

6.3 Model details

As shown in Table 3, six different configurations are considered for both MIMO-LSTM and MISO-LSTM models. Figure 2a depicts the general architecture of the proposed model. Each configuration has a different number of layers, and each layer consists of a different number of nodes. Each configuration is experimented with:

- Fixed learning rate (LR) and adaptive learning rate [68]. In the fixed learning rate, we set LR = 0.01. In the adaptive learning rate method, the LR (initial LR = 0.1) is reduced to half of the current LR in every 20 epochs to find the optimal model with best LR.
- Adam [69] and SGD [70] optimizers to minimise a given cost function [56, 64].

The MIMO-TCN and MISO-TCN approaches have experimented with different configurations and controls, such as;

- Filter sizes: 32, 64, 128, 256, and 512
- Stacked TCN layers: 1, 2, 3, and 4 and
- With different activation functions such as ‘linear’ and ‘tanh’

Table 4 Comparison of machine learning approaches for MISO

Parameter	SR	SVR	ARIMA	VAR	VECM	AFE	RF	LSTM	TCN
TSK	0.002401549	0.002254852	0.002284599	0.002276563	0.002121585	0.002117985	0.002095814	0.002041361	0.001738656
PSFC	9.359E–05	8.90012E–05	9.2467E–05	9.08745E–05	8.8452E–05	8.79532E–05	8.76859E–05	8.16E–05	8.74041E–05
U10	0.005820971	0.005620015	0.00568859	0.005689786	0.005425125	0.005325658	0.00486305	0.002748407	0.004384032
V10	0.009827752	0.008465238	0.00920003	0.008902459	0.008625459	0.008612125	0.007865233	0.003732091	0.007427616
Q2	0.00698015	0.006901244	0.006976667	0.006885608	0.006841126	0.006827854	0.006795488	0.006379222	0.006752483
Rainc	0.004125379	0.003956207	0.004072365	0.004006562	0.003756566	0.003654545	0.003465855	0.002799961	0.003260107
Rainnc	0.021597916	0.019257844	0.016020204	0.015784525	0.013299866	0.013198986	0.009548721	0.000502061	0.001895714
Snow	1.65547E–06	9.87E–07	1.18746E–06	9.98926E–07	5.98926E–07	5.13656E–07	3.72155E–07	1.74E–07	1.34E–07
TSLB	0.000934762	0.000847989	0.000904632	0.000869562	0.000853657	0.000796566	0.000762486	0.000724035	0.000376134
SMOIS	0.000359895	0.000285655	0.000327851	0.000302515	0.000273252	0.000271652	0.000249451	0.00024636	9.98907E–05

Lower MSE is better and is shown in bold

According to [8, 59, 71], the following controls are kept constant within these experiments as these do not impact on final results significantly in the regression model for time-series data; kernel size: 2, dilations: 7, where dilation values are: 1, 2, 4, 8, 16, 32, 64, batch size-64, and dropout rate-0, learning rate-0.01.

6.4 Evaluation metric

The proposed deep regression models are evaluated using the most common metrics of mean squared error (MSE), which is calculated as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_a - y_b)^2 \quad (3)$$

where y_a is the actual expected output, y_b is the model's prediction, and n is number of samples.

6.5 Baseline approaches

Performances of the proposed LSTM and TCN models are compared with the following three types of baseline approaches. These approaches do not consider the temporal information rather count as another dimension in multivariate weather data.

- Classic machine learning approaches

Standard regression (SR), support vector regression (SVR), and random forest (RF).

- Statistical machine learning approaches

Autoregressive integrated moving average (ARIMA), vector auto regression (VAR), and vector error correction model (VECM).



Fig. 5 MISO analysis of different approaches to predicting different weather parameters (SR standard regression, ARIMA autoregressive integrated moving average, VAR vector autoregression, SVR support

vector regression, VECM vector error correction model, AFE arbitrage of forecasting experts, RF random forest, LSTM long short-term memory, TCN temporal convolutional network)

- A dynamic ensemble method

Arbitrage of forecasting expert (AFE).

We use both linear and RBF (radial basis function) kernels for SVR in our experiments and use the grid search algorithm technique to optimise both C and γ parameters.

In linear kernel, the parameter C is selected among the range [0.01–10,000] with multiples of 10. In RGB kernel, the parameters C is selected as above but γ is selected among the range [0.0001, 0.001, 0.01, 0.1, 0.2, 0.5, 0.6, 0.9]. For RF [71], we select number of trees as [100, 250, 500]. For ARIMA model, we use the parameters $p=2$,

Table 5 Comparison of machine learning approaches for MIMO

Parameter	SR	SVR	RF	LSTM	TCN
TSK	0.003701561	0.003652545	0.003612458	0.003271054	0.003578392
PSFC	0.005358824	0.00325658	0.002720456	0.002112675	0.000279068
U10	0.008420962	0.006994518	0.006890641	0.005394089	0.00632667
V10	0.015627757	0.012985601	0.012056545	0.006311009	0.010195208
Q2	0.009980163	0.009979542	0.009929825	0.009881492	0.006578324
Rainc	0.006125415	0.005231452	0.005095621	0.002878811	0.004785024
Rainnc	0.021599896	0.016958456	0.016332367	0.003070845	0.015204848
Snow	1.65518E–05	9.72462E–06	7.27815E–06	2.39E–06	1.60078E–06
TSLB	0.004349349	0.003756588	0.003665241	0.003427306	0.000485899
SMOIS	0.000979024	0.000782515	0.000771265	0.000648767	0.000756974

Lower MSE is better and is shown in bold

$d=0$, and $q=1$ [72]. For VAR and VECM, the auto option is selected for weather forecasting [73, 74]. The given software package is used for the AFE [75].

The baseline performances are compared with the proposed LSTM and TCN networks. These models are evaluated using the testing dataset to select the optimal model or a model with the least MSE, which can be used as a tool for future forecasting. The selected optimal is used to forecast the weather parameters for the validation dataset (model prediction), and the model predicted values are evaluated with respect to the ground truth. Similarly, the WRF model has been run in forecast mode using the same format GRIB data for the month of July 2018 (WRF Prediction). These WRF predicted values are evaluated with respect to the ground truth. Then, we compare the model prediction and WRF prediction to determine the possibility to use the proposed model for short-term weather forecasting (i.e. 3-h prediction). Then, the optimal model is re-tuned for long-term weather forecasting, such as 6, 9, 12, 24, and 48 h. Similar to the short-term forecasting, we compare the model predictions and WRF predictions to determine up to what extent the proposed model can be used for weather forecasting.

7 Results and discussion

There are three types of results, namely: (1) a comparison of various machine learning techniques, statistical forecasting approaches, and a dynamic ensemble method with the proposed approach for weather forecasting, (2) performance of short-term weather forecasting, and (3) performance of long-term weather forecasting using the proposed model. More specifically, the short-term weather forecasting refers

to 3-h weather prediction, and long-term weather forecasting refers to 6-h, 9-h, 12-h, 24-h, and 48-h weather predictions.

7.1 Comparison of machine learning techniques for short-term weather forecasting

As described in Sect. 6.5, we examine the classic machine learning approaches (i.e. SR, SVR, RF), statistical forecasting approaches (i.e. ARIMA, VAR, and VECM), and a dynamic ensemble method (i.e. AFE). Finally, we compare these performances with the proposed deep models (i.e. MISO-LSTM, MISO-TCN, MIMO-LSTM, MIMO-TCN) consisting of cutting-edge networks such as LSTM and TCN layers. As described in Sects. 5.1 and 5.2, these models are evaluated using two different regression types, namely MISO and MIMO.

We evaluate the MISO models to determine the MISO-optimal with the least MSE for weather prediction. Table 4 and Fig. 5 represent the comparison of machine learning approaches for MISO. As per Table 4 and Fig. 5, the MISO-LSTM provides better performance with the least MSE for 6 parameters out of 10. Thus, the LSTM combined model with 10 parameters (i.e. MISO-LSTM) has been selected as the MISO proposed model.

Similarly, we evaluate the MIMO models to determine the MIMO-optimal with the least MSE for weather prediction. Table 5 and Fig. 6 represent the comparison of machine learning approaches for MIMO. We do not consider the approaches ARIMA, VAR, VECM, and AFE in MIMO. Therefore, we compare SR, multi-output SVR [76], and RF with the proposed deep models MIMO-LSTM and MIMO-TCN. The results are subsequently evaluated via the mean squared error. This is used to assess the best model (i.e. least MSE) after comparing the performance of all models.

As per Table 5 and Fig. 6, the MIMO-LSTM provides high accuracy output with least MSE for 6 parameters out

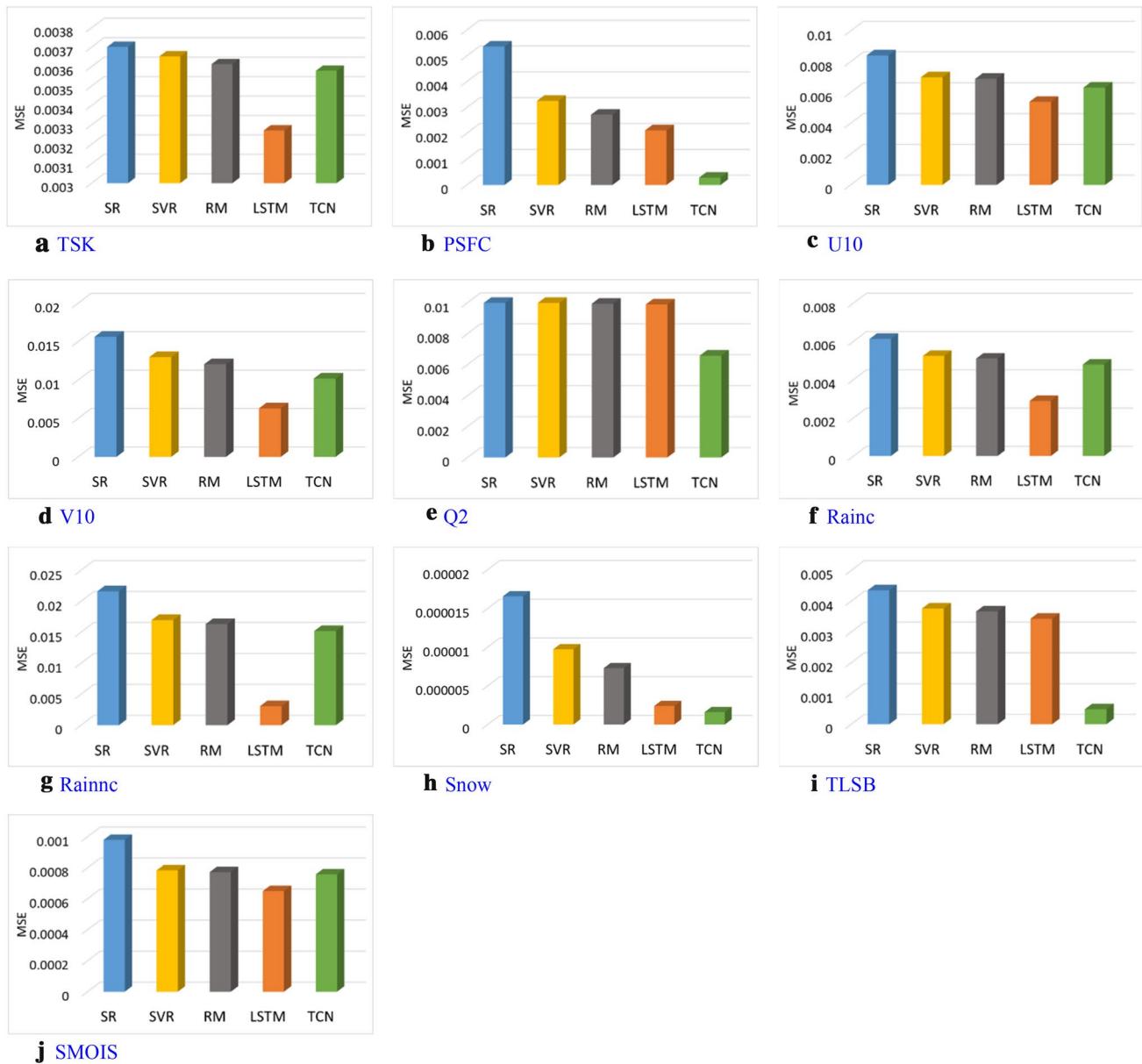


Fig. 6 MIMO analysis of different approaches to predicting different weather parameters (SR standard regression, SVR support vector regression, RM random forest, LSTM long short-term memory, and TCN temporal convolutional network)

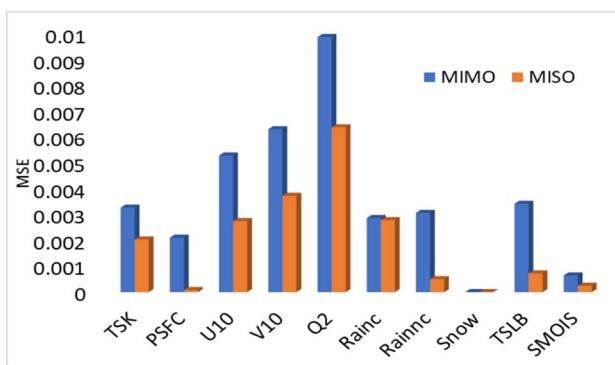
of 10. Therefore, the MIMO-LSTM has been selected as the proposed model (i.e. MIMO-optimal).

In both MIMO and MISO, the LSTM and the TCN produce high performance with smaller errors compared to the classic machine learning approaches and statistical forecasting approaches as presented in Figs. 5 and 6. The reason is that the selected parameters do not follow a linear path within selected sequential timeslots [77, 78] and there is a nonlinear interrelationship among parameters [6, 53, 79]. Besides, the sequential information is not encoded by the

classic machine learning approaches and statistical forecasting models. The LSTM and TCN encode both multivariate and sequential information by taking them into another dimension in the input data [6, 59, 80].

Table 6 MSE comparison

Parameter	MIMO-optimal	MISO-optimal
TSK	3.27E–03	2.04E–03
PSFC	2.11E–03	8.16E–05
U10	5.39E–03	2.75E–03
V10	6.31E–03	3.73E–03
Q2	9.88E–03	6.38E–03
Rainc	2.87E–03	2.79E–03
Rainnc	3.07E–03	5.02E–04
Snow	2.39E–06	1.74E–07
TSLB	3.42E–03	7.24E–04
SMOIS	6.48E–04	2.46E–04

**Fig. 7** Comparison of MIMO and MISO

7.2 Proposed models for short-term weather prediction

The least MSE for the MIMO is identified in the configuration with three LSTM layers, with 128, 512, and 256 number of nodes, respectively (i.e. MIMO-optimal model). We use the SGD optimiser with a fixed learning rate of 0.01 to optimise the MSE regression loss function. The model is trained for 230 epochs. In MISO, all these 10 models have different configurations with a different number of LSTM layers and nodes, activation functions, and optimisers (i.e. MISO-optimal). Table 6 and Fig. 7 graphically represent the comparison of MSE in each variable for both MIMO-optimal and MISO-optimal.

Table 6 shows the comparison of MSE in each variable for both MIMO and MISO. Figure 7 graphically represents these values to get an idea of whether to use the MIMO model or the MISO combined model to use as the best model for future predictions.

According to Fig. 7, there is no major gap between MSE values for each variable when compare the MIMO-optimal and MISO-optimal. These differences are less than 0.04 for each variable. These error figures are significantly smaller.

Table 7 Comparison of the proposed deep model with the WRF forecasting model for 3-h prediction

Parameter	Mean squared Error	
	WRF model	Proposed model (MIMO-LSTM)
TSK	4.0209727	2.788245
PSFC	227869.02	123881.22
U10	10.540705	5.327054
V10	12.0824	4.6248293
Q2	1.1117266e ⁻⁶	7.716598e⁻⁷
Rainc	15.942339	0.11341145
Rainnc	18.627722	0.83847433
Snow	0.0	0.016857434
TSLB	8.140333	2.6088953
SMOIS	8.523197e⁻⁵	0.000024246839

Moreover, the MISO-optimal requires 10 different models for the prediction of 10 different weather parameters. Therefore, we consider the MIMO-optimal (i.e. MIMO-LSTM) model as a tool for future forecasting since it is easier to handle and less time and power consumption (only one model to run) than running 10 different models of MISO-optimal.

As described in Sect. 6.5, the validation dataset is utilised to get weather prediction using the proposed model. Similarly, the WRF model is run in forecast mode using the July 2018 data to compare results. Both WRF and model predicted values are compared with respect to the ground truth and calculated the MSE. Table 7 and Fig. 8 represent the MSE comparison values for each variable.

When comparing Table 7 and Fig. 8, the proposed deep model (i.e. MIMO-LSTM) provides comparatively best results (bolded in the table) on eight occasions out of 10. The WRF model provides the best results for the snow and soil moisture (SMOIS) variables. On both occasions, these error figures are quite small. For example, MSE for the variable snow is 0.0168574 kg/m². This is quite a small and therefore, negligible. Similarly, the SMOIS has got a minimal and negligible error value. Figure 8k shows an overall comparison of both models.

As there are 125,373 samples in the July 2018 evaluation data, the proposed deep model and the WRF model will produce a similar number of samples as the predicted data. It is difficult to visualise all of these predictions because of the large sample size and therefore, a random sample of the 100 samples has been taken from the test set to compare with the respective ground truth. Figure 9 shows a comparison of the proposed deep model's predictions versus the WRF model predictions. For each graph, the ground truth, WRF prediction, and the proposed deep model's predictions are represented by each line with blue, green, and red colours, respectively.

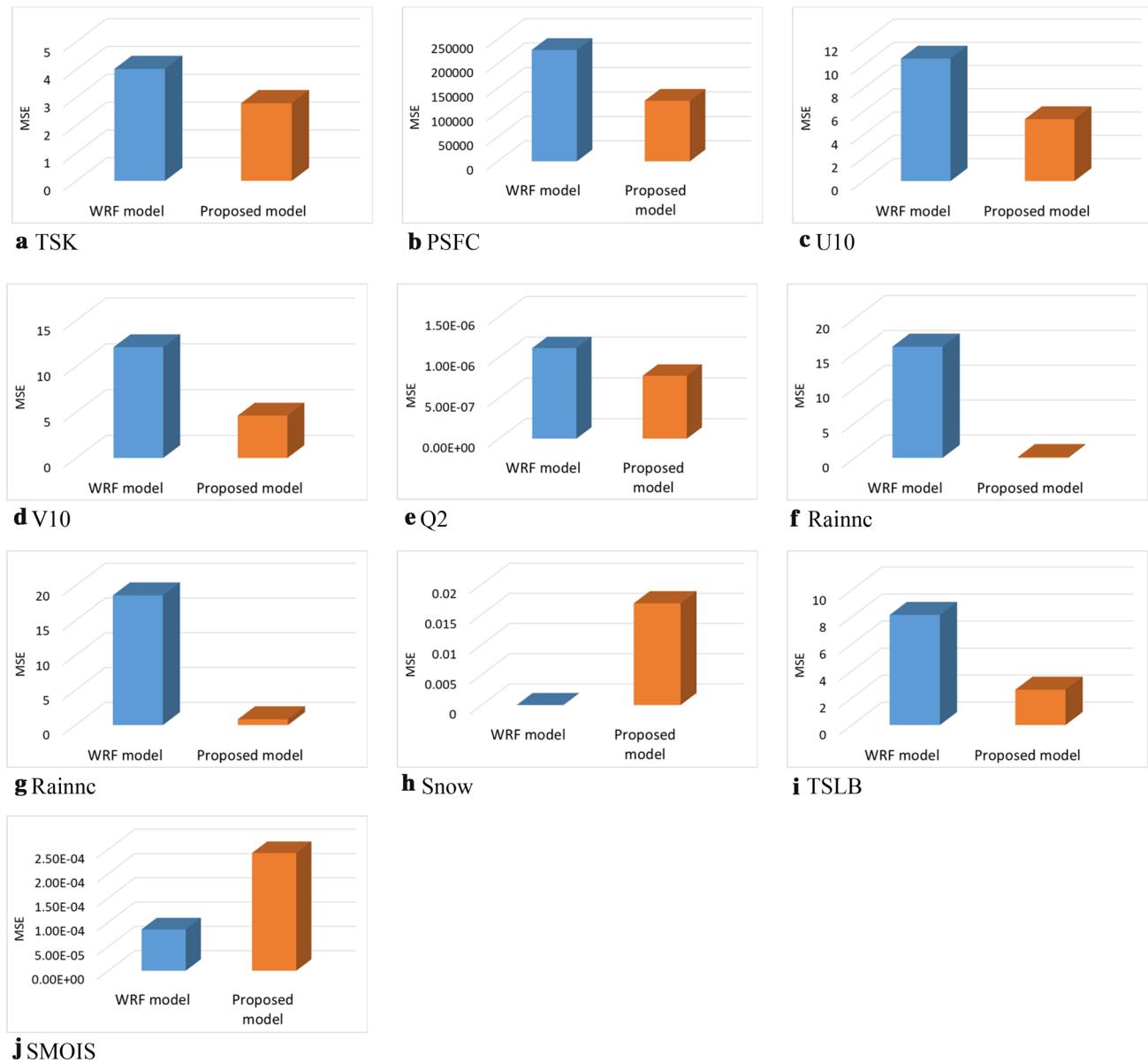


Fig. 8 Analysis of weather prediction of the WRF model and proposed deep learning LSTM model

As per Fig. 9, the red line-chart (deep model prediction) follows closely to the blue line-chart (ground truth) compared to the green-chart (WRF prediction). The WRF prediction is widely diverted in the parameters Rainnc and Rainnc compared to the actual values. The deep model prediction is diverted in the parameter snow compared to the actual values. According to Fig. 7h, the highest snow prediction is 0.24 kg/m^2 . This is quite a small figure and can be negligible. Overall, the deep learning model provides a better short-term (up to 3 h) prediction compared to the WRF model.

7.3 Proposed model for long-term weather forecasting

As described in Sect. 7.2, the proposed model (i.e. MIMO-LSTM) can be utilised for short-term weather forecasting, and it yields more accurate results compare to the well-known WRF model. In this section, our study is focused on exploring long-term weather prediction using the same historical weather data with 10 surface weather parameters.

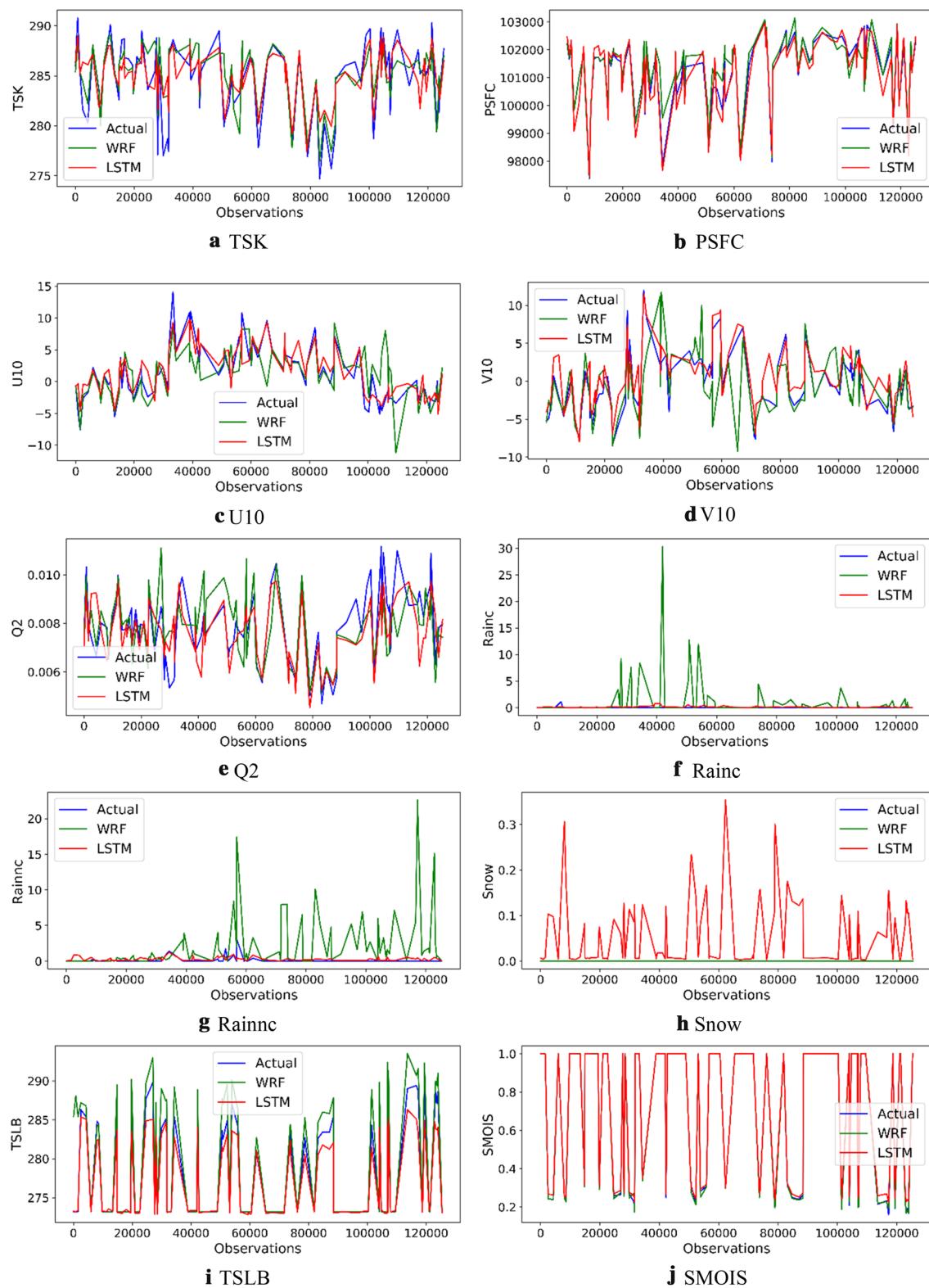


Fig. 9 Comparison of WRF prediction versus the MIMO-LSTM model prediction for 100 random data samples with respect to the ground truth (colour figure online)

7.3.1 Selection of an appropriate technique

As discussed in Sect. 7.1, the proposed model provides better performance compared to other machine learning techniques. Therefore, we use the same deep learning model with the LSTM layers for the long-term weather forecasting with the following variations. All these three variants use the same configuration and controls, which are comparable to the proposed MIMO-LSTM model.

- (a) Load the MIMO-LSTM optimal model weights (3-h) and fine-tune models for the long-term forecasting (shortened form: LSTM LW)
- (b) Train models for each time frame without loading the optimal model weights (shortened form: LSTM WL). That is train the model at the beginning of the training dataset and new labels.
- (c) We have also experimented with Bidirectional LSTM (Bi-LSTM). Compared to the LSTM, the Bi-LSTM has used two layers; one layer performs the operations following the forward direction (time-series data) of the data sequence, and the other layer applies its operations on in the reverse direction of the data sequence [81].

The following Table 8 shows the comparison of these three variations for each timeslot. As shown in Table 8, the Bi-LSTM provides slightly better results compared to the LSTM LW except for the timeslot 3-h. The LSTM WL produces weaker results compared to the both LSTM LW. The reason is that the LSTM LW used its optimal weight, which is already configured to re-train and yield a prediction. Moreover, this is re-tune the model which is matched to the new dataset [55]. The Bi-LSTM is also trained the model at the beginning similar to the LSTM WL. However, the Bi-LSTM provides more accurate results due to the ability to preserve the past and future values [81].

The only drawback of the Bi-LSTM is that time taken to training, testing, and predicting data [82]. This is less efficient compared to the LSTM LW. Moreover, as can be observed in Table 8, there is a slight gap in the overall figures of MSE in both LSTM LW and Bi-LSTM. Therefore, we have selected the LSTM LW method for long-term forecasting for an effective and efficient outcome.

7.3.2 Long-term weather forecasting

The proposed model (i.e. MIMO-LSTM) consists of three LSTM layers with other controls. As described in Sect. 7.3.1 the LSTM with loading the optimal weight method is used for the long-term weather prediction. Therefore, the optimal model is re-tuned (i.e. load optimal model weight and re-train models) for timeslots 3-h, 6-h, 9-h, 12-h, 24-h, and 48-h. While re-tuning, the optimal

Table 8 Comparison of LSTM LW, LSTM WL, and Bi-LSTM

Parameter	LSTM LW	LSTM WL	Bi-LSTM
(a) 3 h			
TSK	N/A	0.003271054	0.002371392
PSFC	N/A	0.002112675	0.001007641
U10	N/A	0.005394089	0.008889356
V10	N/A	0.006311009	0.010825
Q2	N/A	0.009881492	0.00885295
Rainc	N/A	0.002878811	0.004197211
Rainnc	N/A	0.003070845	0.025307791
Snow	N/A	2.39E–06	1.06E–06
TSLB	N/A	0.003427306	0.001056143
SMOIS	N/A	0.000648767	0.000677912
Overall	N/A	0.003699844	0.006318646
(b) 9 h			
TSK	0.004679656	0.003785324	0.002954833
PSFC	0.00337704	0.005435103	0.002543765
U10	0.016287696	0.01789222	0.015199178
V10	0.022693845	0.032980144	0.026619522
Q2	0.016228491	0.017330563	0.014454748
Rainc	0.007961646	0.007261488	0.006792006
Rainnc	0.08320849	0.087723635	0.0691833
Snow	2.13871E–06	1.39469E–05	1.86E–06
TSLB	0.002113115	0.002125454	0.002216928
SMOIS	0.001027248	0.001156121	0.00075886
Overall	0.015757935	0.017570399	0.014072499
(c) 24 h			
TSK	0.003225559	0.003989982	0.003520491
PSFC	0.012482793	0.010315491	0.007714262
U10	0.026440082	0.026202237	0.024926782
V10	0.03660787	0.042136274	0.036013693
Q2	0.026067492	0.030222168	0.02755576
Rainc	0.08263268	0.07865509	0.078575564
Rainnc	0.15932418	0.16492906	0.158401
Snow	4.65178E–07	0.000137658	0.000442552
TSLB	0.004401047	0.004503616	0.005910429
SMOIS	0.001600785	0.001434334	0.001202147
Overall	0.035278295	0.03625259	0.034456268
(d) 48 h			
TSK	0.004480547	0.005869389	0.003708232
PSFC	0.018504778	0.013365718	0.016115312
U10	0.045134	0.037737582	0.03475978
V10	0.04253545	0.04715329	0.042574175
Q2	0.050479617	0.04151997	0.038551033
Rainc	0.061815947	0.068089165	0.059418406
Rainnc	0.16204703	0.16313162	0.15197921
Snow	3.72323E–06	0.000231712	6.40E–05
TSLB	0.007845704	0.012153346	0.005880864
SMOIS	0.00158867	0.001342647	0.001104945
Overall	0.039443548	0.038059445	0.03541559

Only included the results for 3, 9, 24, and 48 h. The other tables are included in the supplementary section

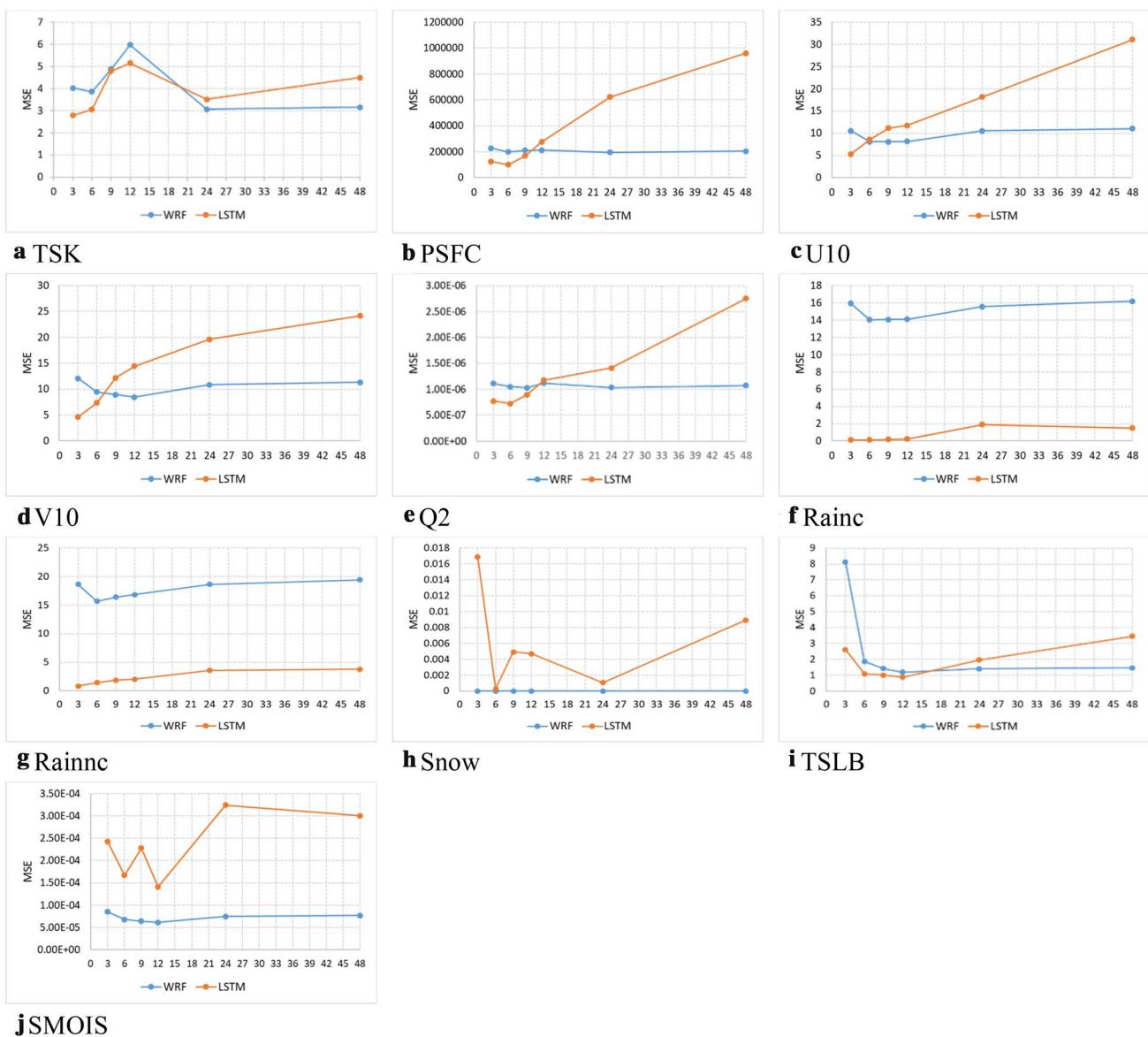


Fig. 10 Compare proposed MIMO-LSTM model prediction with WRF prediction for long-term forecasting. The MSE values are calculated with respect to the ground truth in both WRF and LSTM models

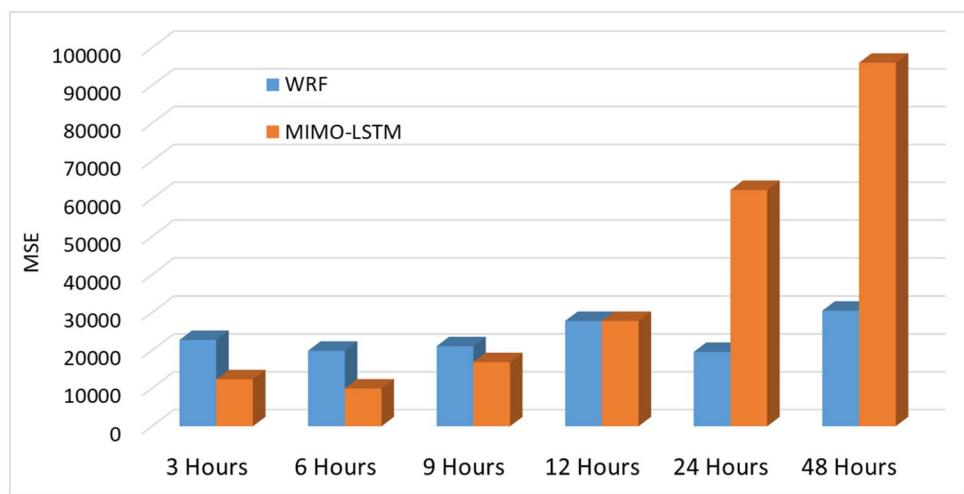
models are found in different epochs such as 80, 10, 10, 10, and 10 for timeslots 6, 9, 12, 24, and 48 h, respectively.

Similar to the short-term weather forecasting, the optimal model for each timeslot is used to forecast the weather parameters for the July 2018 data (model prediction), and the model predicted values are evaluated with respect to the ground truth. The WRF model has been run in forecast mode using the same format GRIB data for the month of July 2018 (WRF prediction) based on the same conditions as model prediction (i.e. input 7 days data and predict weather parameters for timeslot 6, 9, 12, 24 and 48). The WRF predicted values are evaluated with respect to the ground truth. Finally, compare the model prediction and WRF prediction

to determine what extent the deep learning model can be used for weather forecasting. Figure 10 shows a comparison of MSE values related to the proposed model and the WRF model for each time slot.

According to the results presented in Fig. 10, it is obvious that the WRF model produces better forecasting results for the very long-term compared to the deep learning model. The reason is that the WRF model is combined with many other climate models [4, 83, 84] and data is coming to the system globally [4, 49]. The deep learning model has predicted these outputs based on 5 months of training data. We could receive better results if we increase the size of the training dataset [56]. The Rainc and Rainnc parameters show

Fig. 11 Comparison of overall MSE for each timeslot



much better results in the deep learning model compared to the WRF model for long-term forecasting. The experiments of [39] already proved that the deep learning neural networks yield the highest accuracy for rain prediction.

Contrarily, the SMOIS and snow parameters show weak results in deep learning compared to the WRF model at all timeslots. Simply, these error patterns are rather low (maximum error: Snow— 0.016 kg/m^2 , SMOIS— $0.00035 \text{ m}^3/\text{m}^3$) and can be negligible. This could be resolved by increasing the size of the sample data. All other occasions, the deep learning model provide more accurate prediction compared to the WRF model up to some extent, than the WRF model produces better prediction compared to the deep learning model. Figure 11 shows the comparison of overall error values of the WRF model and proposed deep learning model.

As indicated in Fig. 11, the deep learning model produces better predictions compared to the WRF model prediction up to 12 h overall. Therefore, we can use deep learning with LSTM model up to 12 h of weather forecasting much accurately compared to the well-recognised WRF model. The comparison of WRF prediction versus the LSTM model prediction for 50 random data samples with respect to the ground truth is shown in Fig. 12. For each graph, the ground truth, WRF prediction, and the proposed deep model's predictions are represented by each line with blue, green, and red colours, respectively.

As per Fig. 12, the red line-chart (deep model prediction) followed closely to the blue line-chart (ground truth) up to some extent and diverted when time increases in many parameters. The green line-chart (WRF model prediction) also diverted from the blue line-chart when time increased, but this diversion is relatively small compared with the red line-chart. As shown in Fig. 12vi, vii, the rainc and rainnc values are accurate in the deep learning model compared to the WRF model for up to 48 h. As discussed earlier, the WRF model produces a better

prediction for the Snow and SMOIS parameters. As shown in Fig. 12x, the difference is negligible for the parameter SMOIS. As shown in Fig. 12viii, the maximum snow values are shown in the 3 h line-chart. This value is equal to 0.24 kg/m^2 , and this is a relatively negligible figure. Overall, the deep learning model delivers a better forecasting prediction compared to the WRF model for up to 12 h.

7.4 Applicability of the new model

As described in Sect. 7.3, the proposed model can be used for weather prediction. Even, this model generates more accurate predictions compared to the well-recognised WRF model for up to 12 h. We use historical weather data to evaluate and validate these models. The only issue is we still use the WRF model to extract GRIB data to use as input for the new model (we use GFS GRIB data). On the other hand, it requires a minimum of 3 h of access GFS data after taking the atmospheric measurements. This includes the time taken to upload data to the website [4, 85]. In addition, the WRF model also taken the time to extract the GFS data depends on the computer system. Hence, the input data which are used in the new model are not the current atmospheric measurement data (i.e. older more than 3 h). Therefore, it is not practicable to use WRF data with the new model, and it will be highly beneficial to consider the use of local weather station data for weather forecasting.

8 Conclusion and future work

In this article, we demonstrate that the proposed lightweight deep model can be utilised for weather forecasting up to 12 h for 10 surface weather parameters. The model outperformed the state-of-the-art WRF model for up to 12 h. The proposed model could run on a standalone computer, and it could easily

Fig. 12 Comparison of WRF prediction versus the LSTM model prediction for 50 random data samples with respect to the ground truth (colour figure online)

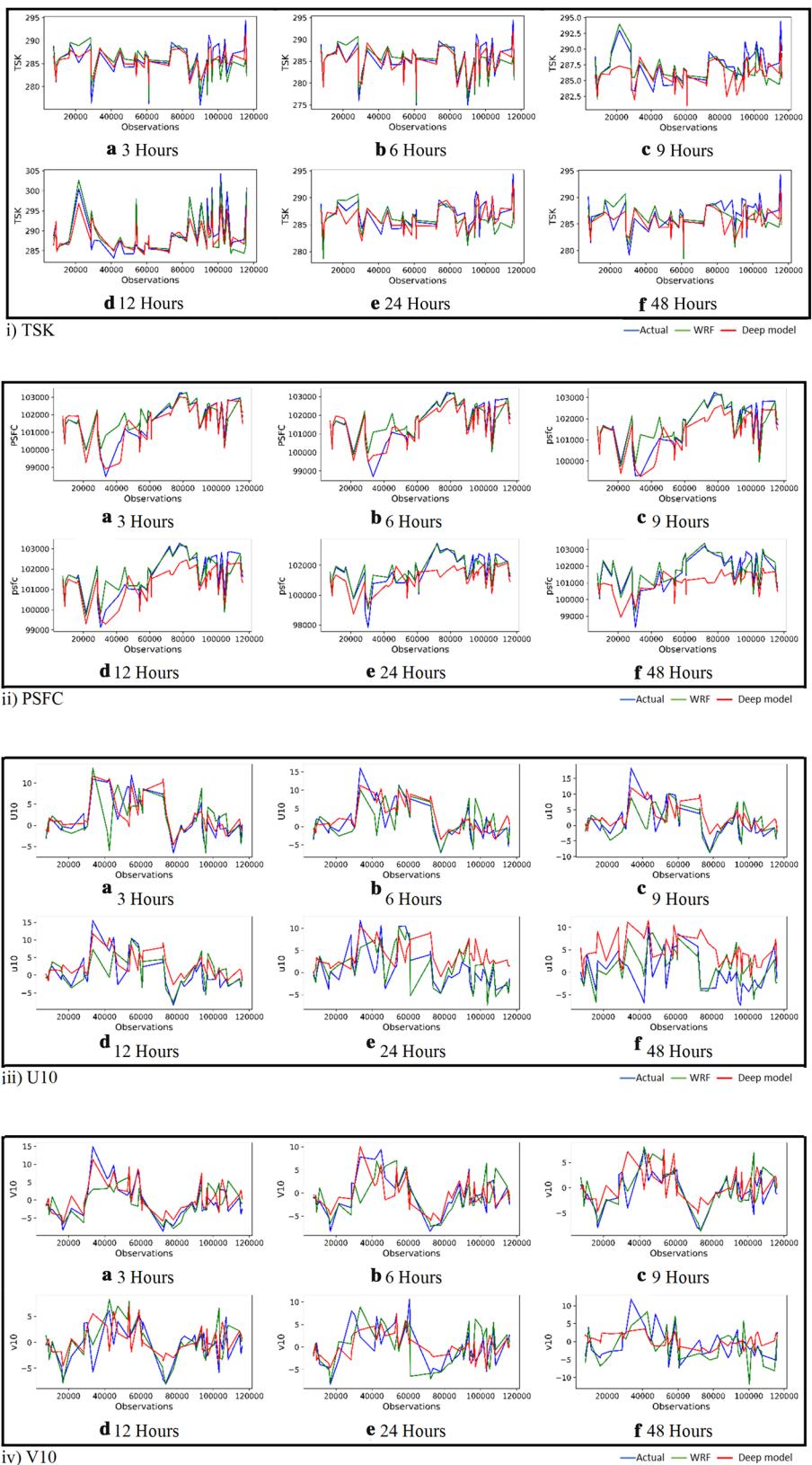
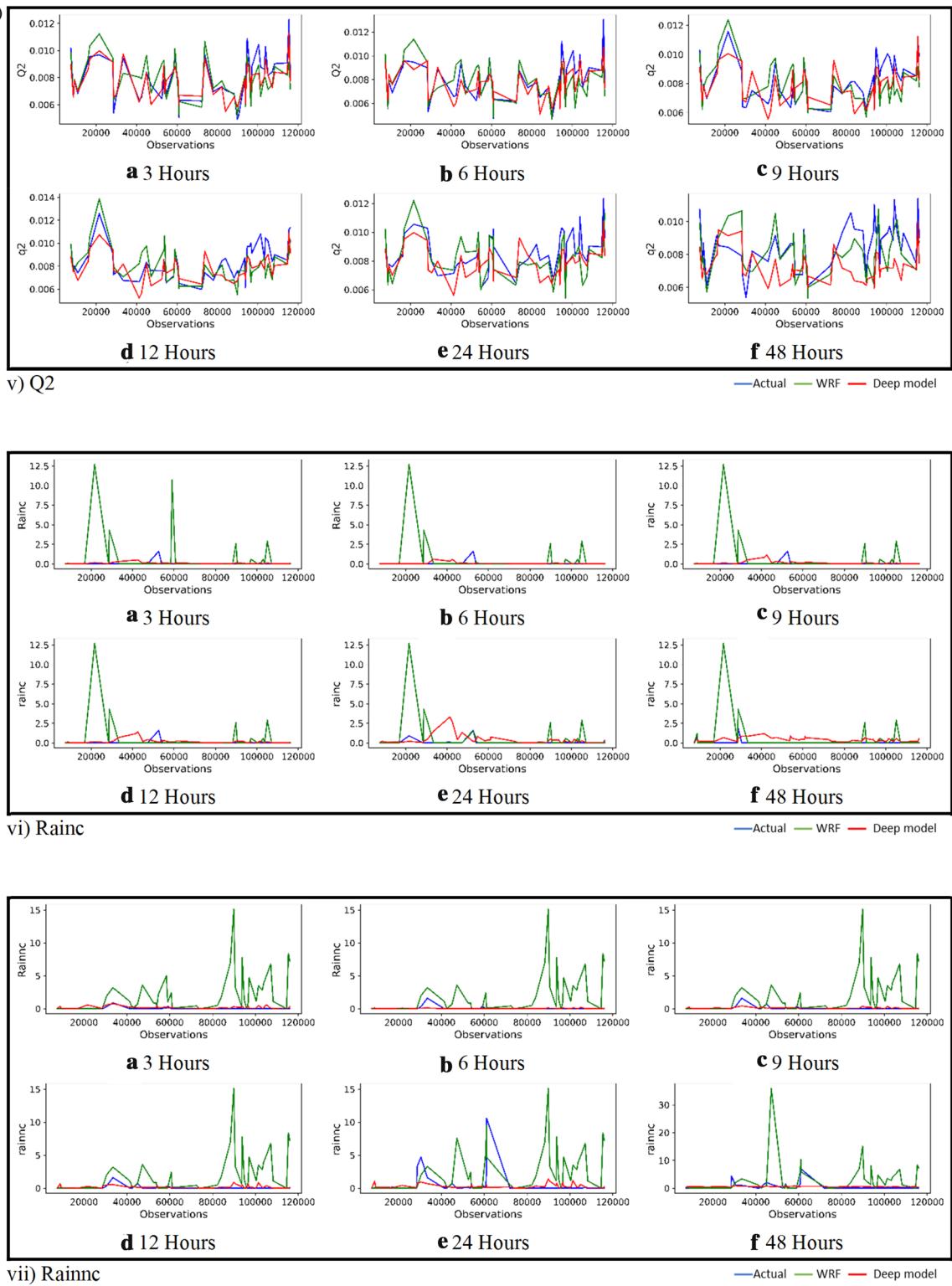
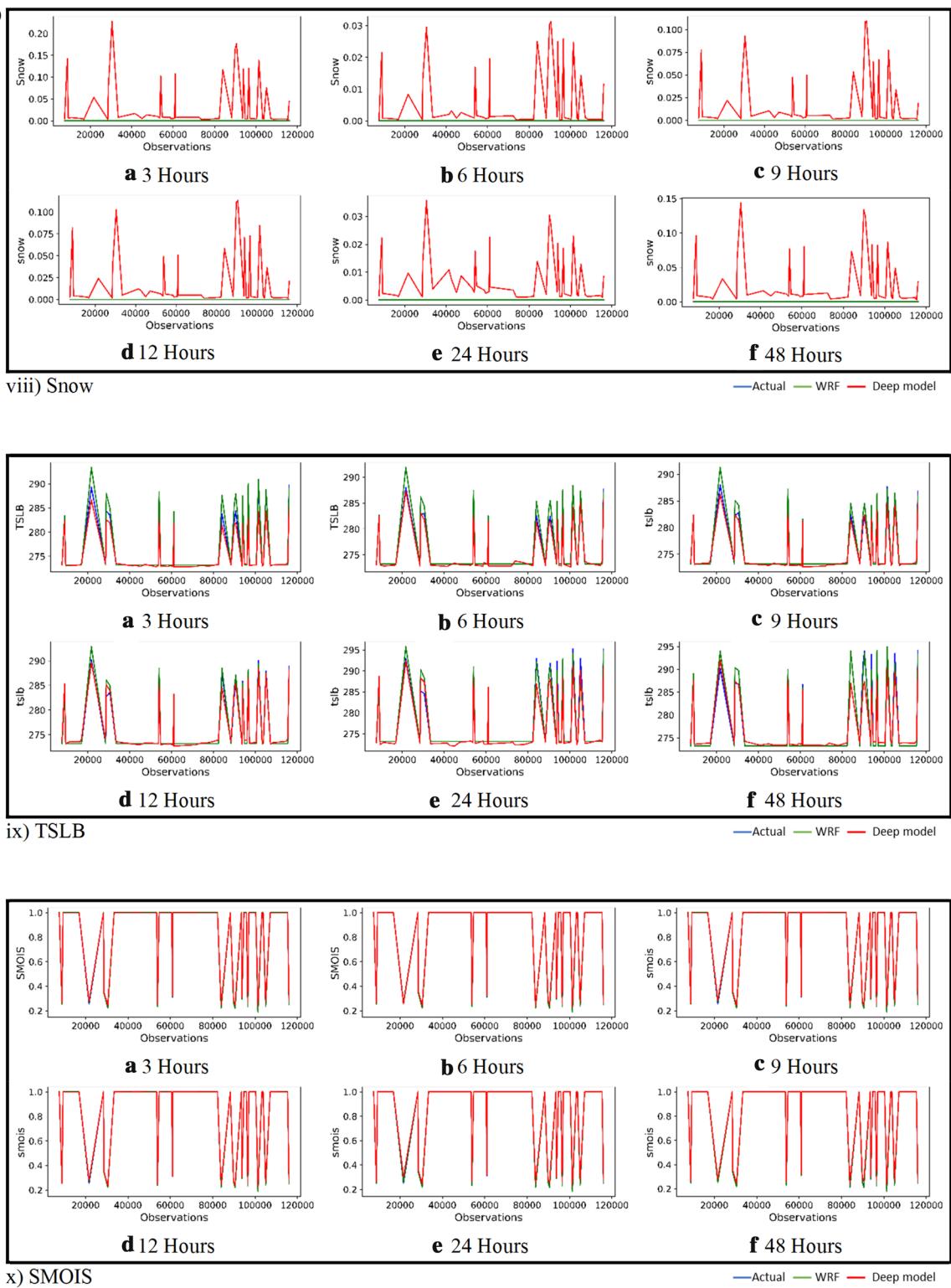


Fig. 12 (continued)

be deployed in a selected geographical region for fine-grained short to medium-term weather prediction. Furthermore, the proposed model is able to overcome some challenges within the WRF model, such as the understanding of the model and its installation, as well as its execution and portability.

In particular, the deep model is portable and can be easily installed into a Python environment for effective results [17, 56]. This process is highly efficient compared to the WRF model.

Fig. 12 (continued)



This research is carried out using ten different surface weather parameters, and an increased number of inputs would probably lead to enhanced results. For example, there are 36 different pressure levels defined in the WRF model [17]. Only the pressure at two meters is considered within

this research. There is a possibility to increase the accuracy of the results if we introduce all 36 possible pressure levels to the proposed model. However, it will increase the model complexity requiring a large number of parameters to estimate. Furthermore, January to May weather data is utilised

for training the deep model, and the increase in the size of training dataset could help towards improved results in a deep learning network [56, 86].

Besides, we used the MIMO approach within this research to predict weather data. Table 5 and Fig. 7 show that the MISO approach produces better MSE values compared to the MIMO. Therefore, there is a huge potential that the MIMO approach will increase the accuracy of the results; even this method is less efficient compared to the MIMO. Besides, the Bi-LSTM yields high accuracy long-term prediction compared to the LSTM, as presented in Table 7. Therefore, we could get more accurate results if we use Bi-LSTM; even this method is not efficient due to high time-consumption.

These experiments show that we can apply the neural network approach for weather prediction. Based on the geographical appearance of location (such as the top of a mountain, land covered by several mountains, the slope of the land, etc.) the regional weather forecasting may not be accurate. As a solution, we could develop a lightweight (neural network-based) short-term weather forecasting system for the community of users utilising weather station data. These are our future experimentation.

Acknowledgements This research work is partly supported by Clive Blaker and Rich Kavanagh (Precision Decisions Ltd). We are grateful to Prof Mark Anderson (Former Professor at Edge Hill University) for his valuable inputs and support at the initial stage of this study. We are also grateful to Dr Alan Gadian (National Centre for Atmospheric Sciences, University of Leeds) for his valuable support to identify weather parameters.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Hayati M, Mohebi Z (2007) Application of artificial neural networks for temperature forecasting. *Int J Electr Comput Eng* 1(4):5
2. Lynch P (2006) The emergence of numerical weather prediction: Richardson's dream. Cambridge University Press, Cambridge
3. Oana L, Spataru A (2016) Use of genetic algorithms in numerical weather prediction. In: 2016 18th international symposium on symbolic and numeric algorithms for scientific computing (SYNASC), Timisoara, Romania, September 2016, pp 456–461. <https://doi.org/10.1109/synasc.2016.075>
4. NCAR/UCAR (2019) WRF model users site. <http://www2.mmm.ucar.edu/wrf/users/>. Accessed 21 Jan 2019
5. Powers JG et al (2017) The weather research and forecasting model: overview, system efforts, and future directions. *Bull Am Meteorol Soc* 98(8):1717–1737. <https://doi.org/10.1175/BAMS-D-15-00308.1>
6. Jozefowicz R, Zaremba W, Sutskever I (2015) An empirical exploration of recurrent network architectures. In: International conference on machine learning, pp 2342–2350
7. Kim TS, Reiter A (2017) Interpretable 3D human action analysis with temporal convolutional networks, April 2017. <http://arxiv.org/abs/1704.04516>. Accessed 22 Feb 2019
8. Lea C, Flynn MD, Vidal R, Reiter A, Hager GD (2017) Temporal convolutional networks for action segmentation and detection. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, July 2017, pp 1003–1012. <https://doi.org/10.1109/cvpr.2017.113>
9. Hinton GE (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507. <https://doi.org/10.1126/science.1127647>
10. Choi T, Hui C, Yu Y (2011) Intelligent time series fast forecasting for fashion sales: a research agenda. In: 2011 international conference on machine learning and cybernetics, July 2011, vol 3, pp 1010–1014. <https://doi.org/10.1109/icmlc.2011.6016870>
11. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W, Woo W (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R (eds) Advances in neural information processing systems, Curran Associates, Inc., vol 28, pp 802–810
12. N. US Department of Commerce (2019) Model output statistics (MOS). https://www.weather.gov/mdl/mos_home. Accessed 05 Jul 2019
13. Glahn HR, Lowry DA (1972) The use of model output statistics (MOS) in objective weather forecasting. *J Appl Meteorol* 11(8):1203–1211. [https://doi.org/10.1175/1520-0450\(1972\)011%3c1203:TUOMOS%3e2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011%3c1203:TUOMOS%3e2.0.CO;2)
14. Woodcock F (1984) Australian experimental model output statistics forecasts of daily maximum and minimum temperature. *Mon Weather Rev* 112(10):2112–2121. [https://doi.org/10.1175/1520-0493\(1984\)112%3c2112:AEMOSF%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112%3c2112:AEMOSF%3e2.0.CO;2)
15. Kruizinga S, Murphy AH (1983) Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. *Mon Weather Rev* 111(11):2244–2254. [https://doi.org/10.1175/1520-0493\(1983\)111%3c2244:UOAAPT%3e2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111%3c2244:UOAAPT%3e2.0.CO;2)
16. Abdel-Aal RE, Elhadidy MA (1995) Modeling and forecasting the daily maximum temperature using abductive machine learning. *Weather Forecast* 10(2):310–325. [https://doi.org/10.1175/1520-0434\(1995\)010%3c0310:MAFTDM%3e2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010%3c0310:MAFTDM%3e2.0.CO;2)
17. Skamarock C et al (2008) A description of the advanced research WRF version 3. <https://doi.org/10.5065/d68s4mvh>
18. Schizas CN, Michaelides S, Pattichis CS, Livesay RR (1991) Artificial neural networks in forecasting minimum temperature (weather). In: 1991 second international conference on artificial neural networks, November 1991, pp 112–114
19. Rumelhart DE, McClelland JL (1987) The PDP perspective. In: Parallel distributed processing: explorations in the microstructure of cognition: foundations, MITP
20. Ochiai K, Suzuki H, Shinozawa K, Fujii M, Sonehara N (1995) Snowfall and rainfall forecasting from weather radar images with artificial neural networks. In: Proceedings of ICNN'95—international conference on neural networks, November 1995, vol 2, pp 1182–1187. <https://doi.org/10.1109/icnn.1995.487781>
21. Oishi S, Ikeuchi S, Kojiri T (1998) Severe rainfall prediction method using artificial intelligence. In: SMC'98 conference proceedings. 1998 IEEE international conference on systems, man, and cybernetics (Cat. No. 98CH36218), October 1998, vol 5, pp 4820–4825. <https://doi.org/10.1109/icsmc.1998.727615>

22. Qi H, Zhang M (2001) Rainfall estimation using M-PHONN model. In: IJCNN'01. International joint conference on neural networks. Proceedings (Cat. No. 01CH37222), July 2001, vol 3, pp 1620–1624. <https://doi.org/10.1109/ijcnn.2001.938403>
23. Jaruszewicz M, Mandziuk J (2002) Application of PCA method to weather prediction task. In: Proceedings of the 9th international conference on neural information processing. ICONIP'02, November 2002, vol 5, pp 2359–2363. <https://doi.org/10.1109/iconip.2002.1201916>
24. Maqsood I, Khan MR, Abraham A (2002) Intelligent weather monitoring systems using connectionist models, vol 10, p 21
25. Li K, Liu YS (2005) A rough set based fuzzy neural network algorithm for weather prediction. In: 2005 international conference on machine learning and cybernetics, August 2005, vol 3, pp 1888–1892. <https://doi.org/10.1109/icmlc.2005.1527253>
26. Mathur S, Kumar A, Ch M (2008) A feature based neural network model for weather forecasting. *Int J Comput Intell* 4:209–216
27. Abhishek K, Singh MP, Ghosh S, Anand A (2012) Weather forecasting model using artificial neural network. *Procedia Technol* 4:311–318. <https://doi.org/10.1016/j.protcy.2012.05.047>
28. Reddy BSP, Kumar KV, Reddy BM, Raja N (2012) ANN approach for weather prediction using back propagation. *Int J Eng Trends Technol* 3:19–23
29. Shrivastava G, Karmakar S, Kumar Kowar M, Guhathakurta P (2012) Application of artificial neural networks in weather forecasting: a comprehensive literature review. *Int J Comput Appl* 51(18):17–29. <https://doi.org/10.5120/8142-1867>
30. Nayak R, Patheja PS, Waoo AA (2012) An artificial neural network model for weather forecasting in Bhopal. In: IEEE-international conference on advances in engineering, science and management (ICAESM-2012), March 2012, pp 747–749
31. Ahmadi A, Zargaran Z, Mohebi A, Taghavi F (2014) Hybrid model for weather forecasting using ensemble of neural networks and mutual information. In: 2014 IEEE geoscience and remote sensing symposium, July 2014, pp 3774–3777. <https://doi.org/10.1109/igassr.2014.6947305>
32. Liu JNK, Hu Y, You JJ, Chan PW (2014) Deep neural network based feature representation for weather forecasting. In: Proceedings of the international conference on artificial intelligence, ICAI, p 7
33. Troncoso A, Salcedo-Sanz S, Casanova-Mateo C, Riquelme JC, Prieto L (2015) Local models-based regression trees for very short-term wind speed prediction. *Renew Energy* 81:589–598. <https://doi.org/10.1016/j.renene.2015.03.071>
34. Dalto M, Matuško J, Vašák M (2015) Deep neural networks for ultra-short-term wind forecasting. In: 2015 IEEE international conference on industrial technology (ICIT), March 2015, pp 1657–1663. <https://doi.org/10.1109/icit.2015.7125335>
35. Hossain M, Rekabdar B, Louis SJ, Dascalu S (2015) Forecasting the weather of Nevada: a deep learning approach. In: 2015 international joint conference on neural networks (IJCNN), July 2015, pp 1–6. <https://doi.org/10.1109/ijcnn.2015.7280812>
36. Akram M, El C (2016) Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. *Int J Comput Appl* 143(11):7–11. <https://doi.org/10.5120/ijca2016910497>
37. Voyant C et al (2017) Machine learning methods for solar radiation forecasting: a review. *Renew Energy* 105:569–582. <https://doi.org/10.1016/j.renene.2016.12.095>
38. Sharaff A, Roy SR (2018) Comparative analysis of temperature prediction using regression methods and back propagation neural network. In: 2018 2nd international conference on trends in electronics and informatics (ICOEI), May 2018, pp 739–742. <https://doi.org/10.1109/icoei.2018.8553803>
39. Yonekura K, Hattori H, Suzuki T (2018) Short-term local weather forecast using dense weather station by deep neural network. In: 2018 IEEE international conference on big data (big data), December 2018, pp 1683–1690. <https://doi.org/10.1109/bigdata.2018.8622195>
40. Patil K, Deo MC (2018) Basin-scale prediction of sea surface temperature with artificial neural networks. In: 2018 OCEANS—MTS/IEEE Kobe Techno-Oceans (OTO), May 2018, pp 1–5. <https://doi.org/10.1109/oceanskobe.2018.85558780>
41. Rodríguez-Fernández NJ et al (2018) SMOS neural network soil moisture data assimilation. In: IGARSS 2018—2018 IEEE international geoscience and remote sensing symposium, July 2018, pp 5548–5551. <https://doi.org/10.1109/igarss.2018.8519377>
42. Elsner JB, Tsonis AA (1992) Nonlinear prediction, chaos, and noise. *Bull Am Meteorol Soc* 73(1):49–60. [https://doi.org/10.1175/1520-0477\(1992\)073%3c0049:NPCAN%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073%3c0049:NPCAN%3e2.0.CO;2)
43. Taylor JW, Buizza R (2002) Neural network load forecasting with weather ensemble predictions. *IEEE Trans Power Syst* 17(3):626–632. <https://doi.org/10.1109/TPWRS.2002.800906>
44. Met Office (2019) How weather forecasts are created. Met Office. https://www.metoffice.gov.uk/weather/learn-about/how-forecasts-are-made/_index. Accessed 21 Feb 2019
45. Stensrud DJ (2007) Parameterization schemes: keys to understanding numerical weather prediction models. Cambridge University Press, Cambridge
46. Weather Research and Forecasting Model MMM: Mesoscale & Microscale Meteorology Laboratory (2019) <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>. Accessed 14 May 2019
47. UCAR (2019) WRF model updates. <http://www2.mmm.ucar.edu/wrf/users/graphics/WRF-post-processing.htm>. Accessed 01 Feb 2019
48. NOAA (2017) Reading GRIB files. http://www.cpc.ncep.noaa.gov/products/wesley/reading_grib.html. Accessed 23 Jan 2019
49. N. C. for E. P. W. S. S. D. of Commerce (2015) NCEP GFS 0.25 degree global forecast grids historical archive, Jan 26. <https://doi.org/10.5065/D65D8PWK>
50. Baboo SS, Shereef IK (2010) An efficient weather forecasting system using artificial neural network. *Int J Environ Sci Dev*. <https://doi.org/10.7763/ijesd.2010.v1.63>
51. Routray A, Mohanty UC, Osuri KK, Kar SC, Niyogi D (2016) Impact of satellite radiance data on simulations of Bay of Bengal tropical cyclones using the WRF-3DVAR modeling system. *IEEE Trans Geosci Remote Sens* 54(4):2285–2303. <https://doi.org/10.1109/TGRS.2015.2498971>
52. Elman JL (1990) Finding structure in time. *Cogn Sci* 14(2):179–211. https://doi.org/10.1207/s15516709cog1402_1
53. Graves A (2012) Supervised sequence labelling. In: Graves A (ed) Supervised sequence labelling with recurrent neural networks. Springer, Berlin, pp 5–13
54. Chung J, Gulcehre C, Cho K, Bengio Y (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling, December 2014. <http://arxiv.org/abs/1412.3555>. Accessed 04 Jul 2019
55. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
56. Gulli A, Pal S (2017) Deep learning with Keras. Packt Publishing Ltd, Birmingham
57. Behera A, Keidel A, Debnath B (2018) Context-driven multi-stream LSTM (M-LSTM) for recognizing fine-grained activity of drivers. In: Brox T, Bruhn A, Fritz M (eds) Pattern recognition. Springer, Cham, p 16
58. Yang J, Nguyen MN, San PP, Li XL, Krishnaswamy S (2015) Deep convolutional neural networks on multichannel time series for human activity recognition. In: Twenty-fourth international joint conference on artificial intelligence, June 2015. <https://doi.org/10.1109/ijcai.2015.100>

- [www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/10710.](http://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/view/10710) Accessed 04 Jul 2019
59. Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, March 2018. <http://arxiv.org/abs/1803.01271>. Accessed 27 Feb 2019
60. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions, November 2015. <http://arxiv.org/abs/1511.07122>. Accessed 03 May 2019
61. Salman AG, Kanigoro B, Heryadi Y (2015) Weather forecasting using deep learning techniques. Presented at the 2015 international conference on advanced computer science and information systems (ICACSIS), 2015, pp 281–285
62. Deng Y, Jia H, Li P, Tong X, Qiu X, Li F (2019) A deep learning methodology based on bidirectional gated recurrent unit for wind power prediction. Presented at the 2019 14th IEEE conference on industrial electronics and applications (ICIEA), pp 591–595
63. Fente DN, Singh DK (2018) Weather forecasting using artificial neural network. In: 2018 second international conference on inventive communication and computational technologies (ICICCT), April 2018, pp 1757–1761. <https://doi.org/10.1109/icicct.2018.8473167>
64. Keras (2019) Home—Keras documentation. <https://keras.io/>. Accessed 28 Jan 2019
65. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in neural information processing systems, vol 25. Curran Associates Inc., Red Hook, pp 1097–1105
66. Hewage P, Behera A, Trovati M, Pereira E (2019) Long-short term memory for an effective short-term weather forecasting model using surface weather data. In: Artificial intelligence applications and innovations, pp 382–390
67. Federal Meteorological (2017) Surface weather observations and reports, November 2017, p 98
68. Brownlee J (2016) Using learning rate schedules for deep learning models in Python with Keras. Machine learning mastery, June 21. <https://machinelearningmastery.com/using-learning-rate-schedules-deep-learning-models-python-keras/>. Accessed 28 Jan 2019
69. Brownlee J (2017) Gentle introduction to the adam optimization algorithm for deep learning. Machine learning mastery, July 02. <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>. Accessed 28 Jan 2019
70. Ruder S (2016) An overview of gradient descent optimization algorithms, September 2016. <http://arxiv.org/abs/1609.04747>. Accessed 28 Jan 2019
71. Wright MN, Ziegler A (2015) ranger: a fast implementation of random forests for high dimensional data in C++ and R. <http://arxiv.org/abs/1508.04409>
72. Rahman M, Islam AS, Nadví SYM, Rahman RM (2013) Comparative study of ANFIS and ARIMA model for weather forecasting in Dhaka. Presented at the 2013 international conference on informatics, electronics and vision (ICIEV), pp 1–6
73. Sun C-S, Wang Y-N, Li X-R (2008) A vector autoregression model of hourly wind speed and its applications in hourly wind speed forecasting. Proc Chin Soc Electr Eng 28(14):112
74. Jaupllari S, Zoto O (2013) An assessment of demand for imports through the VECM model. J Knowl Manag Econ Inf Technol 3(6):1–17
75. Cerqueira V, Torgo L, Pinto F, Soares C (2019) Arbitrage of forecasting experts. Mach Learn 108(6):913–944. <https://doi.org/10.1007/s10994-018-05774-y>
76. Sanchez-Fernandez M, de-Prado-Cumplido M, Arenas-Garcia J, Perez-Cruz F (2004) SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems. IEEE Trans Signal Process 52(8):2298–2307. <https://doi.org/10.1109/tsp.2004.831028>
77. Bishop CM (2006) Pattern recognition and machine learning. Springer, Berlin
78. McCREA M et al (2005) Standard regression-based methods for measuring recovery after sport-related concussion. J Int Neuropsychol Soc 11(1):58–69. <https://doi.org/10.1017/S135561770500083>
79. Kavitha S, Varuna S, Ramya R (2016) A comparative analysis on linear regression and support vector regression. In: 2016 online international conference on green engineering and technologies (IC-GET), November 2016, pp 1–5. <https://doi.org/10.1109/get.2016.7916627>
80. Basak D, Pal S, Patranabis D (2007) Support vector regression. Neural Inf Process Lett Rev 11(10):203–224
81. Althelaya KA, El-Alfy EM, Mohammed S (2018) Evaluation of bidirectional LSTM for short-and long-term stock market prediction. In: 2018 9th international conference on information and communication systems (ICICS), April 2018, pp 151–156. <https://doi.org/10.1109/icaics.2018.8355458>
82. Salehinejad H, Sankar S, Barfett J, Colak E, Valaei S (2017) Recent advances in recurrent neural networks, December 2017. <http://arxiv.org/abs/1801.01078>. Accessed 04 Jul 2019
83. Liu Z, Liu S, Hu F, Li J, Ma Y, Liu H (2012) A comparison study of the simulation accuracy between WRF and MM5 in simulating local atmospheric circulations over Greater Beijing. Sci China Earth Sci 55(3):418–427. <https://doi.org/10.1007/s11430-011-4310-2>
84. Hernández-Ceballos MA, Skjøth CA, García-Mozo H, Bolívar JP, Galán C (2014) Improvement in the accuracy of back trajectories using WRF to identify pollen sources in southern Iberian Peninsula. Int J Biometeorol 58(10):2031–2043. <https://doi.org/10.1007/s00484-014-0804-x>
85. National Centers for Environmental Information (2019) Global forecast system (GFS). <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>. Accessed 19 Jul 2019
86. Hewage P et al (2020) Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Comput. <https://doi.org/10.1007/s00500-020-04954-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.