







Long-Short Term Memory for an Effective Short-Term Weather Forecasting Model Using Surface Weather Data

Pradeep Hewage^(✉) , Ardhendu Behera , Marcello Trovati ,
and Ella Pereira 

Department of Computer Science, Edge Hill University, Ormskirk L39 4QP, UK
{hewagep, beheraa, Marcello.Trovati,
pereirae}@edgehill.ac.uk,
<https://www.edgehill.ac.uk/computerscience>

Abstract. Numerical Weather Prediction (NWP) requires considerable computer power to solve complex mathematical equations to obtain a forecast based on current weather conditions. In this article, we propose a lightweight data-driven weather forecasting model by exploring state-of-the-art deep learning techniques based on Artificial Neural Network (ANN). Weather information is captured by time-series data and thus, we explore the latest Long Short-Term Memory (LSTM) layered model, which is a specialised form of Recurrent Neural Network (RNN) for weather prediction. The aim of this research is to develop and evaluate a short-term weather forecasting model using the LSTM and evaluate the accuracy compared to the well-established Weather Research and Forecasting (WRF) NWP model. The proposed deep model consists of stacked LSTM layers that uses surface weather parameters over a given period of time for weather forecasting. The model is experimented with different number of LSTM layers, optimisers, and learning rates and optimised for effective short-term weather predictions. 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 short-term weather forecasting.

Keywords: Long Short-Term Memory · Numerical Weather Prediction · WRF · Surface weather parameters · 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 large systems of non-linear mathematical equations, which are based on specific mathematical models [2]. Meteorology adopted a more quantitative approach with the advance 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 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 NWP. In particular, Artificial Neural Network (ANN) can be used for this purpose due to their adaptive nature and learning capabilities based on prior knowledge. This feature makes the ANN techniques very appealing in application domains for solving highly nonlinear phenomena. Recurrent Neural Networks (RNN) and deep learning have attracted considerable attention due to their superior performance [6]. Deep learning allows stacked neural networks and includes several layers as part of 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 weights. Subsequently, the activation function gets established on the basis of input-weight products while signal progresses take place in the network [7].

Figure 1 depicts a general overview of the proposed research. More specifically, the proposed weather prediction model is defined on the Long Short-Term memory (LSTM), a specialised form of the RNN, for weather prediction. The performance of the proposed model is subsequently compared to the well-established WRF model to assess and evaluate its computational efficiency and accuracy.

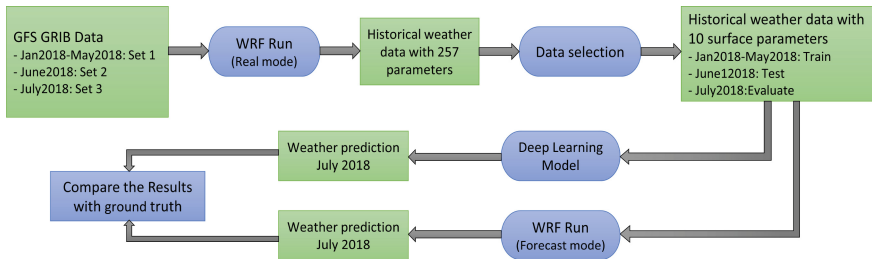


Fig. 1. Overview of the research (Data, processes, and dataflow show in green rectangles, blue rounded rectangles, and arrows respectively) (Color figure online)

2 Weather Research and Forecasting Model

The Weather Research and Forecasting (WRF) model was developed by Norwegian physicist Vilhelm Bjerknes and applied various thermodynamic equations so that numerical weather-based predictions can be made mainly through different vertical levels [8]. The primary role of the WRF was to carry out analysis focusing on climate time scale via linking physics data between land, atmosphere and ocean. The WRF model became the world’s most-used atmospheric model since its initial public release was in the year 2000 [5].

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. GRIdded Binary (GRIB) data are a concise data format commonly used in meteorology to store historical and forecast weather data [9, 10]. Global Forecast System (GFS) GRIB data provides 0.25° resolution and available to download every three hours freely [11]. Therefore, the GFS three hourly data are selected for this project, with a horizontal resolution set to 27 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 causes a limited model accuracy. As a consequence, the model forecast capabilities are less reliable as the difference between current time and the forecast time increases [1, 12]. 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° 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 [10, 11]. 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 provides 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, 9, 13].

This article proposes a novel lightweight weather prediction model that could run on a standalone PC for accurate short-term weather prediction and could easily be deployed in a selected geographical region. As depicted in Fig. 2, the proposed model is based on state-of-the-art deep learning techniques that use Artificial Neural Network (ANN) and modern LSTM layers.

3 Proposed Deep Model Using Long Short-Term Memory (LSTM) Network

The approach discussed in this work is based on temporal weather data to identify the patterns and produces weather predictions. As discussed above, we use the state-of-the-art Long Short-Term Memory (LSTM), which is a specialised form of Recurrent Neural Network (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, 7, 14].

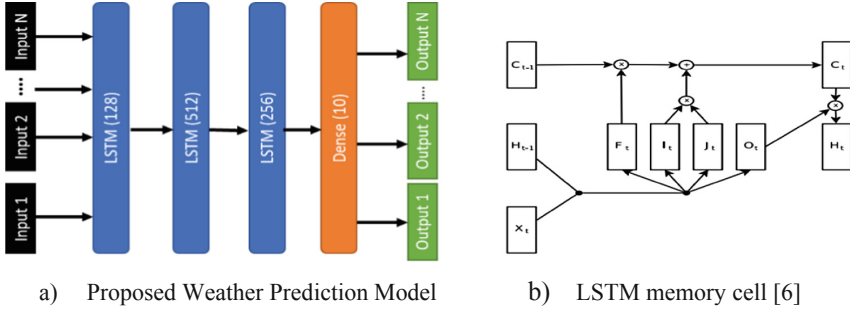


Fig. 2. (a) Proposed layered LSTM and (b) LSTM memory cell used for this research

Figure 2a shows the proposed lightweight model consisting of stacked LSTM layers for weather forecasting using surface weather parameters. Table 1 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 $x_t = [p_1, p_2, p_3 \dots p_{10}]$ at a given timestep t , which consists of 10 different ($p_1, p_2 \dots p_{10}$) weather parameters. In the given time t , the model is updated the memory cells for long-term c_{t-1} and short-term h_{t-1} recall from the previous timestep $t - 1$ via:

$$\begin{aligned}
 i_t &= \tanh(w_{xi}x_t + w_{hi}h_{t-1} + b_i) \\
 j_t &= \text{sigm}(w_{xj}x_t + w_{hj}h_{t-1} + b_j) \\
 f_t &= \text{sigm}(w_{xf}x_t + w_{hf}h_{t-1} + b_f) \\
 o_t &= \tanh(w_{xo}x_t + w_{ho}h_{t-1} + b_o) \\
 c_t &= c_{t-1} \odot f_t + i_t \odot j_t \\
 h_t &= \tanh(c_t) \odot o_t
 \end{aligned} \tag{1}$$

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, 15]. 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. We use Keras to implement the proposed model using LSTM [14, 16, 17].

Table 1. Surface weather parameters (10 identified variables 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

4 Methodology

4.1 Surface Weather Parameters

For monitoring and forecasting purposes, the surface weather parameters are observed and reported in [18]. The surface parameters of wind direction and wind speed can be calculated from the WRF surface variables U_{10} and V_{10} [4]. Table 1 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.

4.2 Data Collection and Preparation

As described in Sects. 2 and 4.1, the GRIB data is used to run the WRF model and extract data for the 12 weather parameters for the period of January 2018 to May 2018. This is used as the training dataset to train the proposed models. Similarly, the June 2018 data is extracted to test the network. This is used to test different trained AI models to identify the best model for forecasting. The July 2018 dataset is also extracted as the evaluation dataset set 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 to evaluate the overall model (WRF prediction).

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 data set and the evaluation data set. Then apply the sliding window method for each dataset based on seven days data to train and next 3 h data as the model's output. The capacity of the final training dataset became 6.5 GB and the testing dataset 0.27 GB.

4.3 Proposed Model with Optimal Number of LSTM Layers

Several different configurations have been utilised to train and test the proposed models. Figure 2a depicts the general architecture of the proposed model. Each layer consists of a number of nodes. Optimisers are generally used in deep learning networks to minimise a given cost function by updating the model parameters such as weights and bias values. For better learning, it is often useful to reduce learning rates as the training progresses when training a deep network. This can be achieved by using adaptive learning rate methods [14]. Therefore, the fixed learning rate and adaptive learning rate methods have been explored to train the proposed deep models.

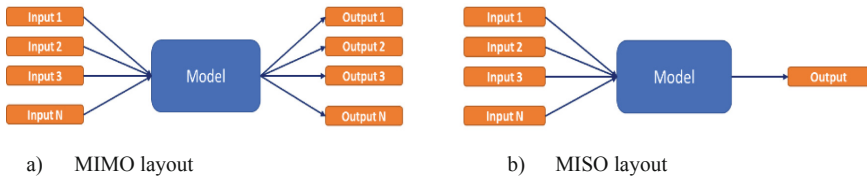


Fig. 3. Further classifications - MIMO and MISO

The proposed model is evaluated using two different types, namely multi-input multi-output (MIMO) and multi-input single-output (MISO). In the MIMO, all 10 variables are fed into the network, which are expected to predict the same 10 variables as the output. In contrast, in the MISO approach, all 10 variables are fed into the network with a single variable output. In the MIMO, only one model is required for weather forecasting involving 10 different variables. Whereas, in the MISO, 10 different models are required as each of them is trained to predict a particular weather parameter with all 10 variables as input. Figure 3 depicts the basic arrangement of the MIMO and the MISO. The both MIMO and MISO classifications are utilised within these experiments.

The most common metrics for evaluating regression models include Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Explained Variance (EV) [19]. In this work, we use the MSE to evaluate the models. As described in Sect. 4.3, there are different deep learning models, which are trained with different optimisers, different learning rates and regressions (MIMO and MISO). These models are evaluated using the June 2018 data to select the best model or a model with the least MSE, which can be used as a tool for future forecasting. The selected best model 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. 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. Finally, we compare the model prediction and WRF prediction to determine the possibility to use a deep learning model for weather forecasting.

5 Results and Discussion

As described in Sect. 4.3, deep learning models are trained with different configurations and controls. The results are subsequently evaluated via the Mean Squared Error (MSE). This is used to assess the least error model and after comparing all evaluation reports. The least MSE for the MIMO is identified in the configuration with three LSTM layers, with the respective number of nodes 128, 512, and 256. We use the SGD optimizer with fixed learning rate of 0.01 to optimize the MSE regression loss function. The model is trained for 230 epochs for optimum performance.

The best values for each variable for the MISO are found in different configurations with different optimisers. Figure 4 graphically represents the comparison of MSE in each variable for both MIMO and MISO. As per Fig. 4, there is no significant difference between MSE values for each variable between the MIMO and the MISO. These differences are less than 0.04 for each variable. These error figures are significantly smaller, and the MISO requires 10 different models for the prediction of 10 different weather parameters. Therefore, we consider the MIMO model since it is much more efficient (only one model to run) than running 10 different models of MISO.

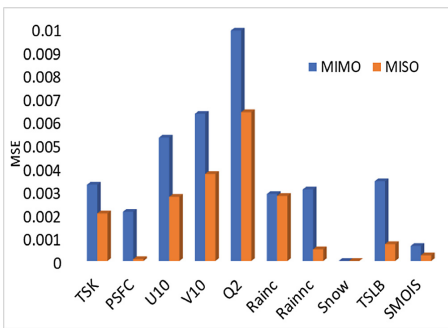


Fig. 4. Comparison of MIMO and MISO

Table 2. Comparison of models (Less MSE is better)

Parameter	MSE	
	WRF model	Deep model
TSK	4.0209727	2.7882845
PSFC	227869.02	123881.22
U10	10.540705	5.327054
V10	12.0824	4.6248293
Q2	1.11172e ⁻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	8.52319e⁻⁵	2.4246839e ⁻⁵

As described in Sect. 4.3, the July 2018 weather data are utilised to get weather prediction using the same format described in Sect. 4.2. 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 2 represents the MSE comparison values for each variable.

When comparing the Table 2 figures, the proposed deep model provides comparatively best results in eight occasions out of 10. The WRF model provides the best results for the Snow and Soil Moisture (SMOIS) variables. In 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.

6 Conclusion and Future Work

In this paper, we demonstrate that the proposed lightweight deep model can be utilised for short-term weather forecasting. The model outperformed the state-of-the-art WRF model. The proposed model could run on a standalone computer unit and it could easily be deployed in a selected geographical region. 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 running and portability. In particular, the deep model is portable and can be easily installed into a Python environment for an effective results [9, 14]. This process is highly efficient compared to the WRF model.

This research is carried out using ten different surface weather parameters and an increased number of inputs would probably lead to enhanced results. However, it will increase the model complexity requiring a large number of parameters to estimate. Furthermore, January to May weather data is utilised to train the deep model and the increase in the size of training dataset could help towards an improved results in a deep learning network [14]. Besides, we used the MIMO approach within this research to predict weather data. Figure 4 shows that the MISO approach produces better MSE values compared to the MIMO. Therefore, there is a huge potential that the MISO approach will increase the accuracy of the results even this method is less efficient compared to the MIMO. These are our future experimentation.

Acknowledgement. The research work is partly supported by Clive Blacker and Rich Kavanagh (Precision Decisions Ltd). We are also grateful to Dr Alan Gadian (National Centre for Atmospheric Sciences, University of Leeds) and Mr John Collins (SimCon software validation Ltd) for their valuable inputs and support at the initial stage of this study.

References

1. Hayati, M., Mohebi, Z.: Application of artificial neural networks for temperature forecasting. *Int. J. Electr. Comput. Eng.* **1**(4), 5 (2007)
2. Lynch, P.: *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge University Press, Cambridge (2006)
3. Oana, L., Spataru, A.: Use of genetic algorithms in numerical weather prediction. In: 2016 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, pp. 456–461 (2016)
4. WRF Model Users Site (2019). <http://www2.mmm.ucar.edu/wrf/users/>. Accessed 21 Jan 2019
5. Powers, J.G., et al.: The weather research and forecasting model: overview, system efforts, and future directions. *Bull. Am. Meteorol. Soc.* **98**(8), 1717–1737 (2017)
6. Jozefowicz, R., Zaremba, W., Sutskever, I.: An empirical exploration of recurrent network architectures. In: *ICML*, pp. 2342–2350 (2015)
7. Hinton, G.E.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006)
8. Stensrud, D.J.: *Parameterization schemes: keys to understanding numerical weather prediction models* (2007)
9. Skamarock, C., et al.: *A Description of the Advanced Research WRF Version 3* (2008)

10. Noaa, Reading GRIB Files (2017). http://www.cpc.ncep.noaa.gov/products/wesley/reading_grib.html. Accessed 23 Jan 2019
11. National Centers for Environmental Prediction/National Weather Service/NOAA/U.S. Department of Commerce, NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive, 26 January 2015. <https://rda.ucar.edu/datasets/ds084.1/>
12. Baboo, S.S., Shereef, I.K.: An efficient weather forecasting system using artificial neural network. *Int. J. Environ. Sci. Dev.* **1**, 321–326 (2010)
13. Routray, A., Mohanty, U.C., Osuri, K.K., Kar, S.C., Niyogi, D.: 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**, 2285–2303 (2016)
14. Gulli, A., Pal, S.: *Deep Learning with Keras*. Packt Publishing Ltd., Birmingham (2017)
15. Behera, A., Keidel, A., Debnath, B.: Context-driven multi-stream LSTM (M-LSTM) for recognizing fine-grained activity of drivers. In: Brox, T., Bruhn, A., Fritz, M. (eds.) *Pattern Recognition. GCPR 2018. LNCS*, vol. 11269, pp. 298–314. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12939-2_21
16. Keras, Home - Keras Documentation (2019). <https://keras.io/>. Accessed 28 Jan 2019
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc. (2012)
18. Federal Meteorological Handbook Number 1: Surface Weather Observations and Reports, p. 98, November 2017
19. Metrics - Keras Documentation. <https://keras.io/metrics/>. Accessed 28 Jan 2019