

Prediction of temporal atmospheric boundary layer height using long short-term memory network

By NISHANT KUMAR¹, KIRTI SONI^{2*}, and RAVINDER AGARWAL¹, ¹Thapar Institute of Engineering and Technology, Patiala, India; ²CSIR-National Physical Laboratory, New Delhi, India

(Manuscript Received 29 November 2020; in final form 2 May 2021)

ABSTRACT

Nowadays, the city's rapid growth of industrialisation, population, human activities, vehicular traffic density, unplanned urbanisation with poor ventilation contributes to increasing large amount of pollutants concentration. Atmospheric Boundary Layer (ABL) height is a basic parameter to define the pollution carrying capacity of any area in a big city. In the time series analysis and prediction of ABL height, the existing models use linear (AR, ARMA, ARIMA etc.) and non-linear (ANN, ANFIS etc) algorithms, but these models less capable of identifying the hidden pattern and underlying dynamics of ABL patterns. This paper presents a Long Short-Term Memory (LSTM) model using deep learning-based algorithms for temporal/seasonal and annual ABL height prediction and identified the latent dynamics of the ABL height pattern. The results of the model have been compared with the measurements made by SONIC Detection And Ranging (SODAR) system. LSTM model is used for prediction and to analyse their performance affected by the model. The observed ABL height data and model data are used to predict the ABL height by applying the neural network of LSTM. It is observed from the analysis that the optimal results can be achieved when the number of neurons is equal to 32, an epoch is equal to 500. To obtain the acceptable accuracy of prediction, various error-based performance indices have been calculated. Mean Absolute Percentage Error (MAPE) and relative Root Mean Square Error (rRMSE) have been calculated for the updated network with predicted values 17.3% and 7.33%, and, for the updated network with observed values 10.62% and 5.95%, respectively. Also, the performance of the proposed model has been estimated for the annual and seasonal prediction of ABL height. The results depict rRMSE values (7.49% and 5.59%) as lowest during post-monsoon for seasonal prediction and (10.29% and 5.86%) highest for annual prediction.

Keywords: atmospheric boundary layer height, LSTM network, SODAR, deep learning

1. Introduction

Atmospheric Boundary Layer (ABL) is a base of life on the Earth. The ABL includes a fair-weather cloud related to thermals and a stratocumulus cloud that fills the upper portion of the well-mixed moist boundary layer (Garratt, 1994). Fog is a form of stratocumulus cloud that touches the ground, also reported as a special feature of ABL (Stull, 2012). Continuously increasing atmospheric pollution and its long-term exposure pose many challenges for global public health which includes cardiovascular and cardio-respiratory diseases. It has become a latent need to accurately monitor the atmospheric variables that allow for the prediction of air pollution behaviour, to issue early alarms for the protection of the population. One of the variables of greatest interest is the dynamical height

of the ABL, the lowest layer of the troposphere is directly influenced by the Earth surface by means of both natural and anthropogenic emissions (Koffi et al., 2016).

Many authors discussed soft computing and artificial intelligence-based model to study the effect of ABL on the atmosphere and human beings simultaneously (Terradellas et al., 2005). These models utilised different techniques i.e. Neural Network, Fuzzy Logic Technique, Deep Learning, Adaptive Neuro-Fuzzy Inference System (ANFIS), etc. (Niska et al., 2004; Freeman et al., 2018). These techniques are used to investigate, simulate, and analyse complex issues and phenomena in an attempt to solve real-world problems. The accurate modelling of ABL height and the realisation of prediction techniques are of great significance to the pollution control board, meteorological department, atmospheric scientists to study the climatic condition for local or remote areas

*Corresponding author. e-mail: 2006.kirti@gmail.com

(Caughey, 1984; Singal et al., 1994; Levi et al., 2020; Liu et al., 2020). For example, ABL height is significant in calculating the pollution carrying capacity in a specific area. Air quality assessments at the local or regional scale require for a variety of purposes such as emission control, air quality forecasts, and implementation of legislation. A key input to this model is the meteorological data that require computing the transport, dispersion, and removal of pollutants. ABL height determines the volume available for the diffusion of contaminants and also an essential parameter, i.e. Wind Profile in atmospheric flow models (Koffi et al., 2016). Besides, ABL height is vital to air-space study, wind structure of the area, atmospheric stability class, etc. To calculate/measure the ABL height, there are two methods, viz. in-situ or direct method (such as tethered balloons, masts, rawinsonde etc.) and remote sensing method (such as LIDAR, SODAR, wind profilers) (Bradley, 2007; Lee and Pal, 2017). SONIC Detection And Ranging (SODAR) is one of the remote sensing instruments (Beyrich, 1997; Emeis et al., 2008). This instrument works with the acoustic wave, where the acoustic wave is transmitted into the atmosphere and gets reflected on to the antenna due to the inhomogeneous structure in the atmosphere (Gilman et al., 1946). The output of the SODAR data is plotted in the form of an echogram, which represents the reflection of the signal from the atmosphere. This data has been used for the prediction of ABL height using time series prediction models. The time series models can be classified as Linear and Non-linear models. Auto Regressive (AR), Auto Regressive Moving-Average (ARMA), Auto Regressive Integrated Moving-Average (ARIMA) and its variation are linear models. They work on predefined equations to fit a mathematical model to a univariate time series (Selvin et al., 2017). These models fail to cope up with latent dynamics in the data. The advanced soft computing techniques such as Artificial Neural Network (ANN), Adaptive Network-based Fuzzy Inference (ANFIS) system, Genetic Algorithm, and Fuzzy Inference System, are non-linear models and being successfully applied for modelling of different parameters such as ABL Height, pollutant, temperature, etc. (Chelani, 2005; Rehman and Mohandes, 2008; Ettouney et al., 2009; Paoli et al., 2010; Kumar and Jha, 2013; Vivas et al., 2020). However, these models are less capable of identifying the hidden pattern and underlying dynamics of data. Deep learning algorithms are capable of identifying the hidden patterns and underlying dynamics in the data through a self-learning process (Selvin et al., 2017; Zaidi et al., 2020). In the case of ABL height, the data generated is enormous and is highly non-linear (Vivas et al., 2020). Unlike other algorithms, deep learning models can provide effectively good predictions by analysing the

interactions and hidden patterns within the data. Long Short-Term Memory (LSTM) is the special type of a Recurrent Neural Network (RNN) (Hochreiter and Schmidhuber, 1997; Kawakami, 2008). LSTM is used in the deep learning methods, that works on the gradient-based learning algorithm and designed to minimise the backflow of the error.

In the area of meteorological parameters, time series analysis using neural network models and deep learning algorithms have used different input variables for predicting the weather data. The data for time series has been used for heterogeneous weather information. Vivas et al. (2020) has applied deep learning for the detection of ABL height from atmospheric LiDAR signals in Colombia. Rehman and Mohandes (2008) reported that an Artificial neural network has used for the prediction of global solar radiation with air temperature and relative humidity for Saudi Arabia from 1998 to 2002. Zhao et al. (2019) has applied the LSTM method to predict RADAR sea clutter and the performance of a particular method has found superior over neural network. However, for long term prediction, the performance of model is not as good as short-term prediction.

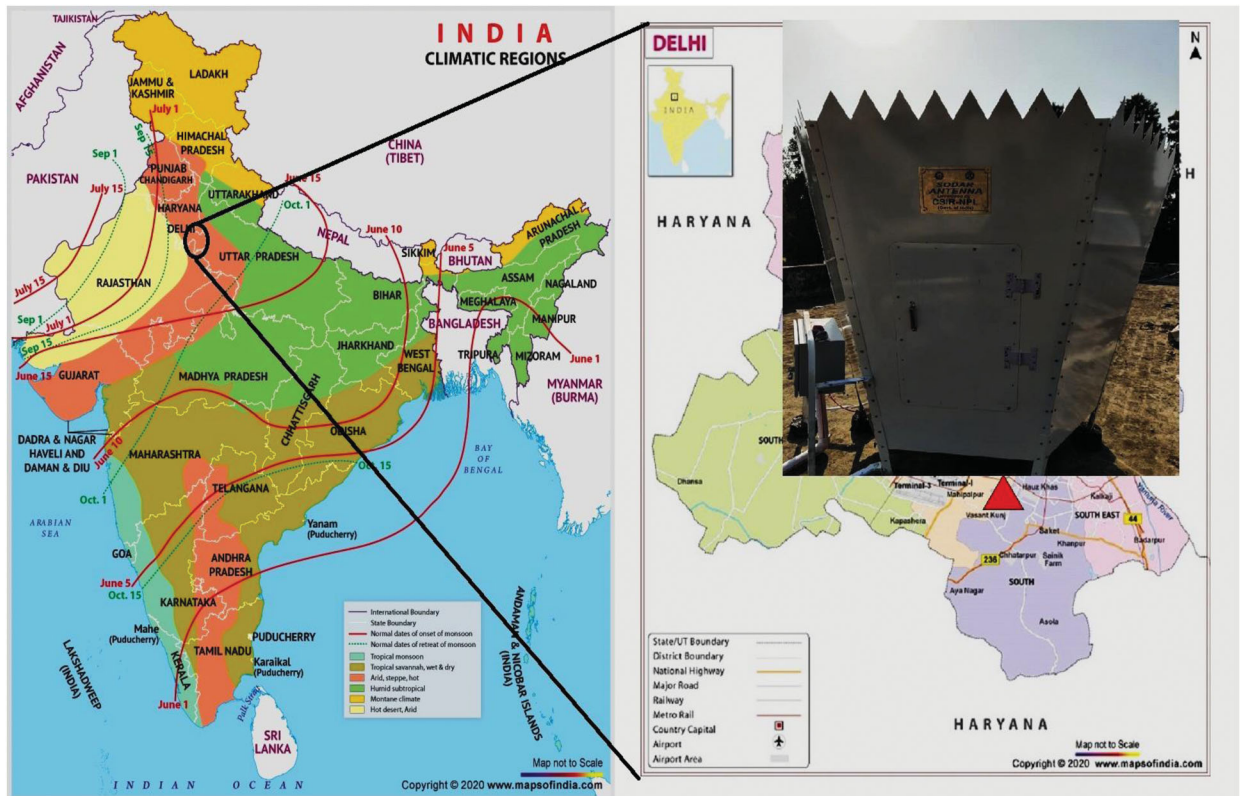
In the present work, the LSTM model has been implemented to obtain the future value of ABL height using SODAR data. The seasonal and annual variability have been studied for significant difference in ABL height. This model incorporates the factor and principle, which affect performance. Also, the performance of the model has been quantified for the data of ABL height using the SODAR system. Further, to estimate the performance of the model in picture, a time-series comparison of annual and seasonal ABL height variations is presented. This work will help to air quality study, seasonal variability for the Delhi region. Section 2 briefly describes the Delhi region, SODAR ABL height data. This section also describes the ABL height LSTM model and architecture. Section 3 begins with the general result of LSTM network then describes the annual and seasonal ABL height analysis and prediction. Section 4 and 5 discuss and conclusions of annual and seasonal ABL height prediction.

2. Data and methodology

Delhi is a site in the northern region of India at 715 feet above sea level, where it has a semi-arid or steppe climate, with scorching summers, heavy rainfalls in the monsoon months, and cold winter (Ramachandran et al., 2012; Kumar et al., 2017b). There are dust storms in summer and foggy mornings in winter. The temperature gradually rises to 46°C in the summer and falls to 2°C in winter (Roy et al., 2011). In the winter months, temperature inversion and low wind speed are the leading causes

Table 1. Designing specifications of CSIR-NPL monostatic SODAR.

Transmitted power (electrical)	90 Watts
Transmitted power (acoustical)	15 Watts
Pulse width	100 ms
Pulse repetition period	4 sec
Operational range	1000 m
Receiver bandwidth	50 Hz
Frequency of operation	2250 Hz
Acoustic velocity	340 m/s (average)
Receiver Gain	80 dB
Transmit-receive antenna	Parabolic reflector dish surrounded by conical acoustic cuff
Receiver area	2.5 sq. m
Preamplifier sensitivity	The fraction of a micro-Volt

Fig. 1. Monsoon Map of India and SODAR system installed at Delhi (Courtesy: www.mapindia.com).

of the accumulation of airborne pollutants in Delhi. Delhi, being the capital of India has been chosen as the site for the installation of the SODAR as it faces extensive change in the atmosphere (Figure 1).

SODAR system was developed at CSIR-NPL in 1973 and modified from time to time. The system is installed on the roof of the main building, CSIR-NPL, New Delhi. Table 1 shown the design specifications of SODAR system which are used for measurement of ABL height (Gera et al., 2011; Kumar et al., 2019). The data of ABL height is obtained in the echogram, which is shown in

Fig. 2 and measured by visualisation (Kumar et al., 2017a, 2017b). In this study, the data set has a temporal value (1 hour = one value) for one year (1 Dec 2018–30 Nov 2019).

2.1. Basic of LSTM network

LSTM network predicts the next moment state based on the data at the previous moment state (Zhao et al., 2019). From Fig. 3, it has been observed that the most

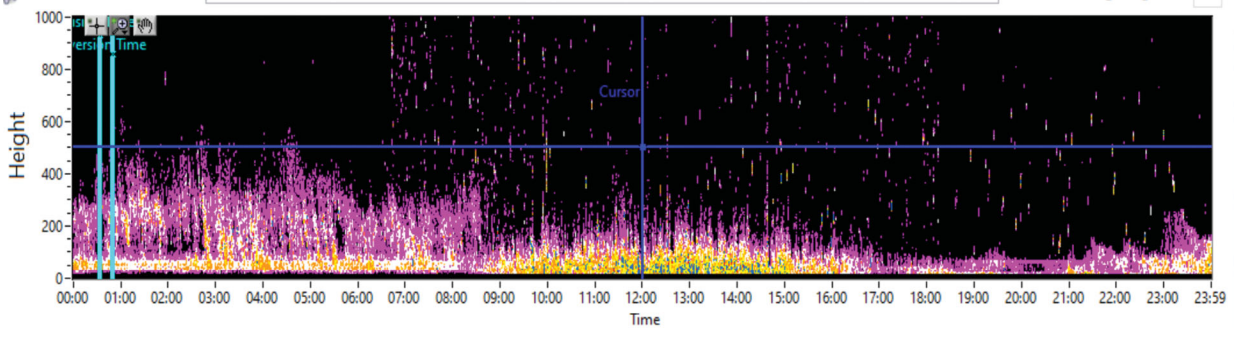


Fig. 2. 24 hours echogram of SODAR system.

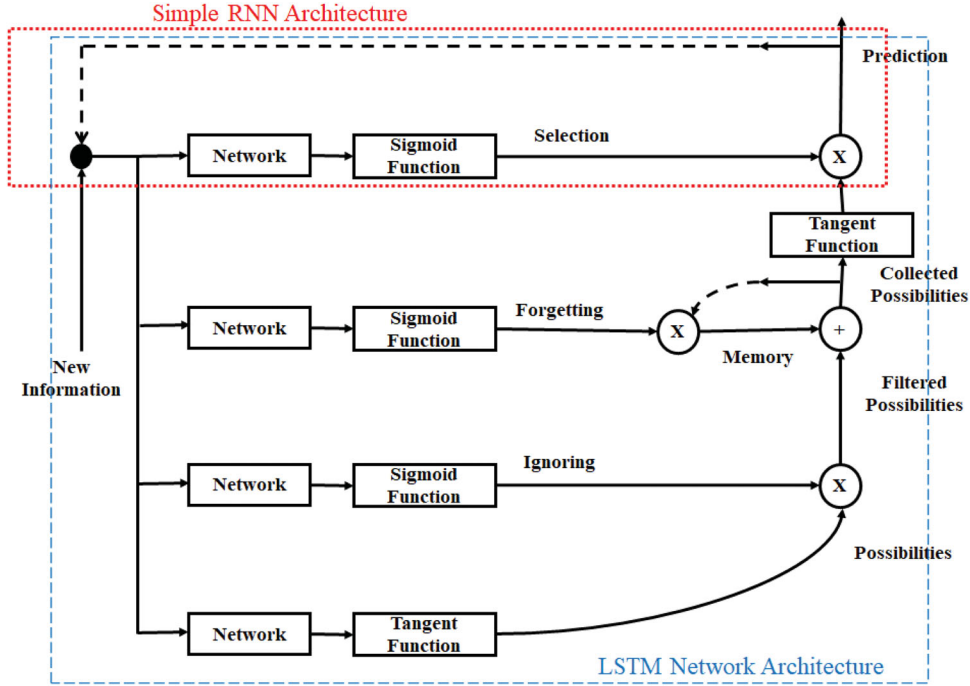


Fig. 3. Network architecture.

significant difference between ordinary RNN and LSTM is the hidden unit. In LSTM network, the hidden layer does self-looping, which can be seen as multiple copies of the same neural network, and each neural network module passes the information to the next level. It is also recompensed for the incapability of RNN to predict long-distance relationships (Hochreiter and Schmidhuber, 1997). It is used due to the famous part of the current deep learning field to solve the time-series prediction problem. A neural network contains one input layer, one output layer, and many intermediate layers, which are called the hidden layer (Soni and Parmar, 2020). The output of the input layer appearances, the input of the first hidden layer, and the output of each hidden layer constitutes the input of each subsequent hidden layer (Kumar et al., 2015). LSTM is a type of RNN but the usually

hidden layers are replaced with LSTM cells and these cells are composed of various gates that control the input flow (Selvin et al., 2017; Zhao et al., 2019). An LSTM cell consists of input gate (consist of input), cell state (runs through the entire network and has the ability to add or remove information with help of gates), forget gate (control level of cell state reset/forget), and output gate (control level of cell state added to hidden state). The details of architecture of LSTM network have been described in Hochreiter and Schmidhuber (1997) and Kawakami (2008).

2.2. Sodar ABL height prediction model

Figure 4 represents the block diagram of development of the ABL height LSTM network prediction model for the

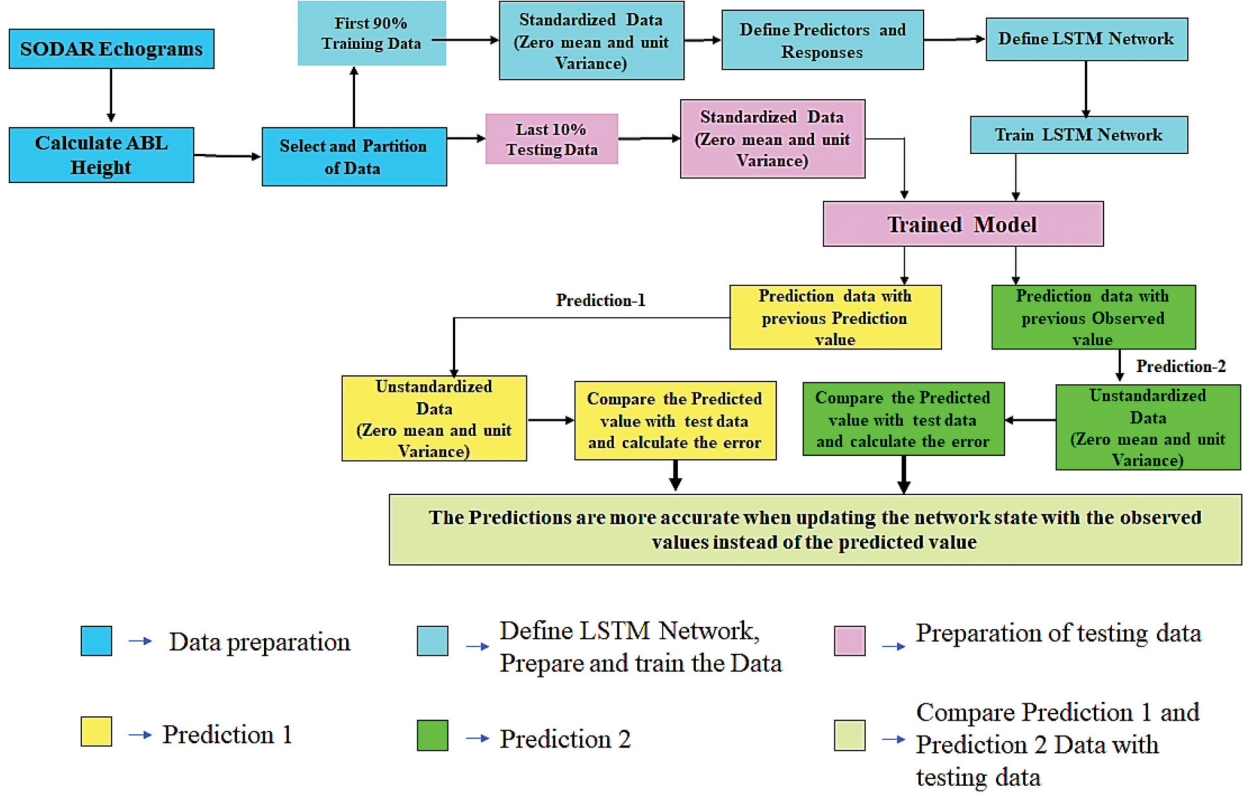


Fig. 4. Block diagram of the LSTM ABL height model.

training and testing of the network. Block diagram depicts that the LSTM prediction network could uninterruptedly pass the ABL height characteristics, i.e. from the current step to the next step, and predicate the next point from the previous point. Thus, it predicts the temporal ABL height. While creating the LSTM prediction model, the input data are a set of temporal ABL height, that is measured by the SODAR system.

For the solving of ABL height prediction problematic, the input data are a set of ABL height temporal series. As the series progresses, the hidden layer of the previous point will affect the hidden layer of the next point. This feature is of great help to the ABL height that has a specific nonlinear relationship between the previous data and the last data. Therefore, the configuration of LSTM network was a sequence-to-sequence regression and trained from Back Propagation Through Time (BPTT). It is observed that as time increases, the values are also changing with a specific time, and there is some periodic change. From the above observation, LSTM neural network model is good for nonlinear parameter.

To check the generalised capacity and accuracy of model, first 90% of the input data are selected as the training set, and the last 10% as the test set, due to checking the pattern of convection and inversion period

of ABL on temporal basis and below functions are selected as performance measures of prediction model:

Root Mean Square Error, RMSE

$$= \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Predicated}_i - \text{Measured}_i)^2}$$

Relative Root Mean Square Error, rRMSE

$$= \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\text{Predicated}_i - \text{Measured}_i}{\text{Measured}_i} \right)^2}$$

Mean Absolute Error, MAE

$$= \frac{1}{n} \sum_{i=1}^n |\text{Predicated}_i - \text{Measured}_i|$$

Mean Absolute Percentage Error, MAPE

$$= \frac{100}{n} \sum_{i=1}^n \left| \frac{\text{Predicated}_i - \text{Measured}_i}{\text{Measured}_i} \right|$$

In this paper, the pre-defined Deep Learning Models available in the Deep Learning toolbox of MATLAB R2019a software is used for deep learning training. These pre-defined networks are retrained and fine-tuned on augmented training and validation sets with SODAR ABL height data. For the better fit, to prevent the training and testing from diverging, the data are standardised with

zero mean and unit variance (Unal et al., 2003). The networks are trained using ‘adam’ function as an optimiser (Kingma and Ba, 2014). LSTM network architecture parameters are defined in Table 2.

Table 2. LSTM architecture parameters.

Network parameters	Value
Gradient decay factor	0.9
Squared gradient decay factor	0.9990
Initial learn rate	0.005
Learn rate schedule	Piecewise
Learn rate drop factor	0.20
Learn rate drop period	125 ^a
Gradient threshold method	I2 norm
Gradient threshold	1
Verbose frequency	50
Validation frequency	50
Shuffle	once
Sequence length	Longest
State activation function	tanh
Gate activation function	sigmoid
Input weight initializer	glorot
Recurrent weight learn initializer	orthogonal
Bias initializer	Unit-forget-gate
Bias learns rate factor	1

^aDrop the learning rate after 125 epochs by a factor of 0.2.

The performance of the prediction model is quantified for the test set and used for estimating the RMSE, rRMSE, MAE, and MAPE. The calculated error is lower than to gets a higher accuracy of the prediction model (Adnan et al., 2020; Kumar et al., 2021).

The input data set has a nearly one-year data point. When the complete data set is used in each training process, the gradient cannot be corrected, and the network cannot converge to the global optimum. The size of the mini-batch is selected with care such that the whole data-set is to be passed through the network every epoch during training and no data is discarded. The experiment is done for ABL height prediction models using the 3 months data (1 April 2019 to 30 June 2019). During this period, available data is continuous and more dynamic in nature. Also, the change in ABL height is maximum during May month. In this paper, two types of predictions are presented. Firstly, the network is updated with previously predicted values as input to the function, which is called Prediction-1 and secondly, the network is updated with observed values instead of the predicted values, this is called Prediction-2. When the LSTM network completes the one loop, i.e. the input ABL height data set passes through the network once and returns once, the process is called an epoch.

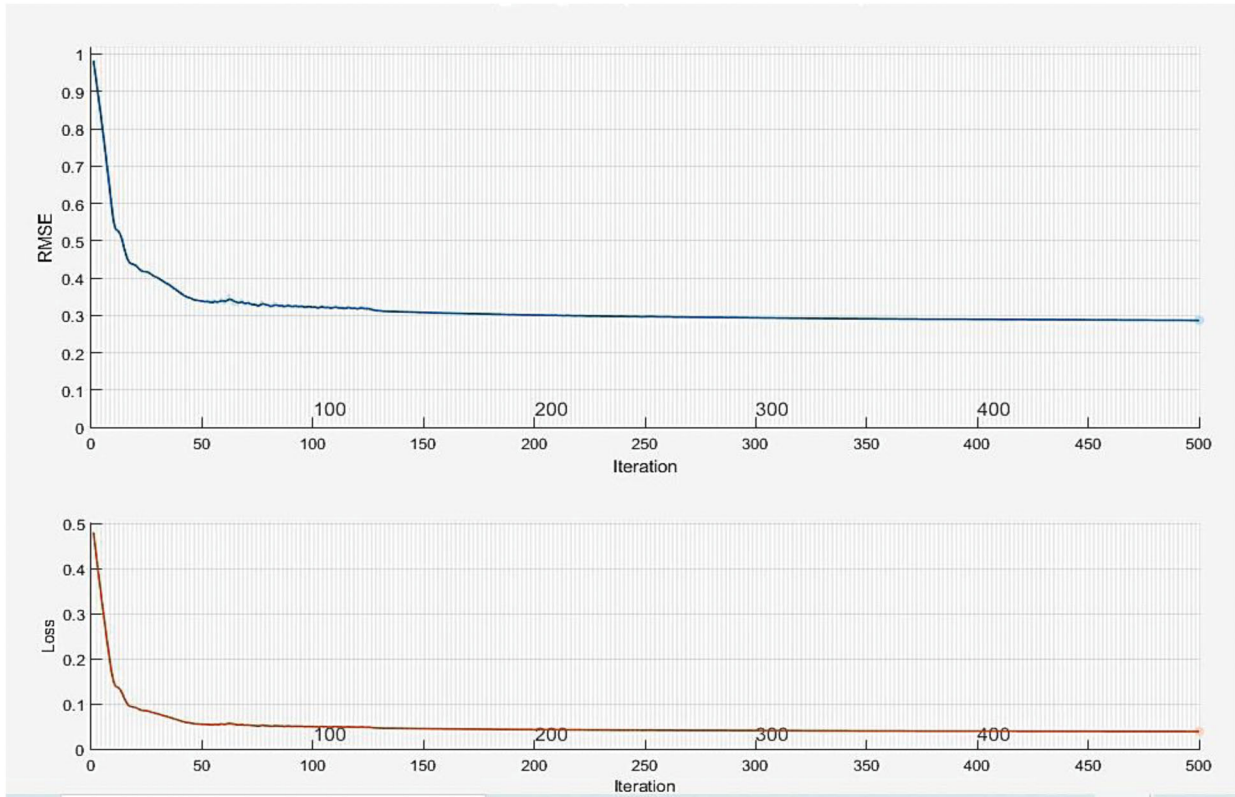


Fig. 5. Training progress with hidden layer 32 and Epoch 500.

Table 3. Comparison of accuracy of LSTM model.

Hidden layer	Prediction-1				Prediction-2			
	RMSE1	rRMSE1	MAE1	MAPE1	RMSE2	rRMSE2	MAE2	MAPE2
Max epochs – 500								
2	790.3	24.69	617.5	72.37	315.69	9.86	197.41	17.11
5	317.83	9.93	199.61	20.22	212.48	6.64	133.99	12.74
10	270.58	8.45	181	20.1	214.71	6.7	130.24	12.29
20	250.93	7.84	175.36	18.27	202.13	6.31	121.86	11.43
25	289.34	9.04	197.85	20.47	204.76	6.39	120.59	11.43
28	290.05	9.06	198.89	22.33	205.31	6.41	118.31	11.25
30	235.4	7.35	166.91	17.31	192.01	6	114.01	10.96
32	220.31	7.33	167.12	17.3	193.81	5.95	113.72	10.62
35	272.51	8.51	194.47	19.56	197.15	6.16	117.52	10.87
50	264.61	8.26	186.84	19.27	192.48	6.01	112.21	10.54
100	291.14	9.09	205.52	21.92	198.86	6.21	116.06	11.11
Max epochs – 250								
32	373.89	11.68	250.20	24.37	193.66	6.05	115.99	10.90
Max epochs – 750								
32	271.66	8.48	189.62	20.68	193.35	6.04	112.08	10.77

^aRed colour row shows best prediction result.

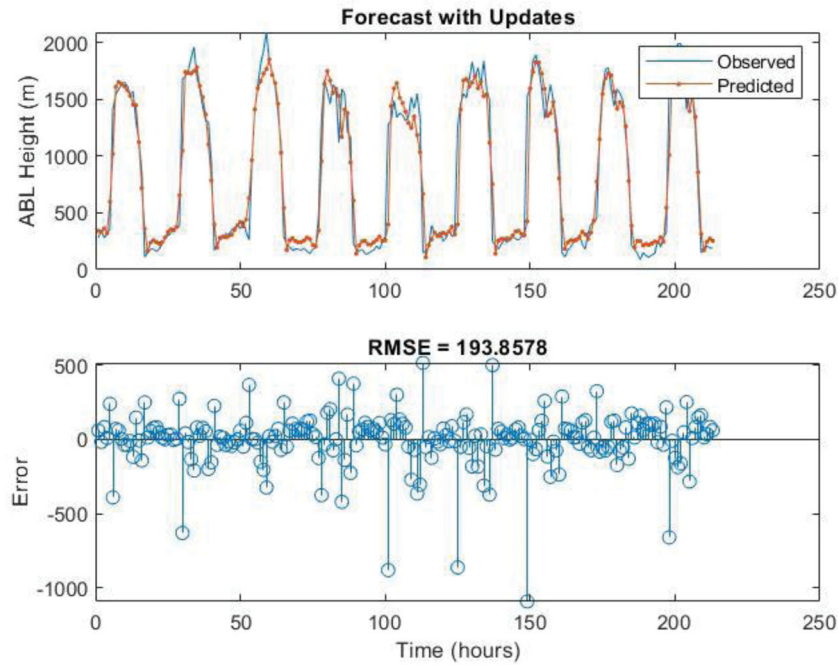


Fig. 6. Prediction result LSTM Network update with observed values (Prediction-2), hidden layer 32 and maximum epochs 500.

Table 4. Statistical analysis of LSTM model with hidden layer 32 and epoch 500.

Parameter	Training data	Test data	Prediction-1 data	Prediction-2 data
Mean	737	837	849	852
Median	430	551	547	400
Kurtosis	-1.22	-1.67	-1.66	-1.71
Skewness	0.56	0.35	0.32	0.26
Coefficient of variation	0.80	0.68	0.72	0.78

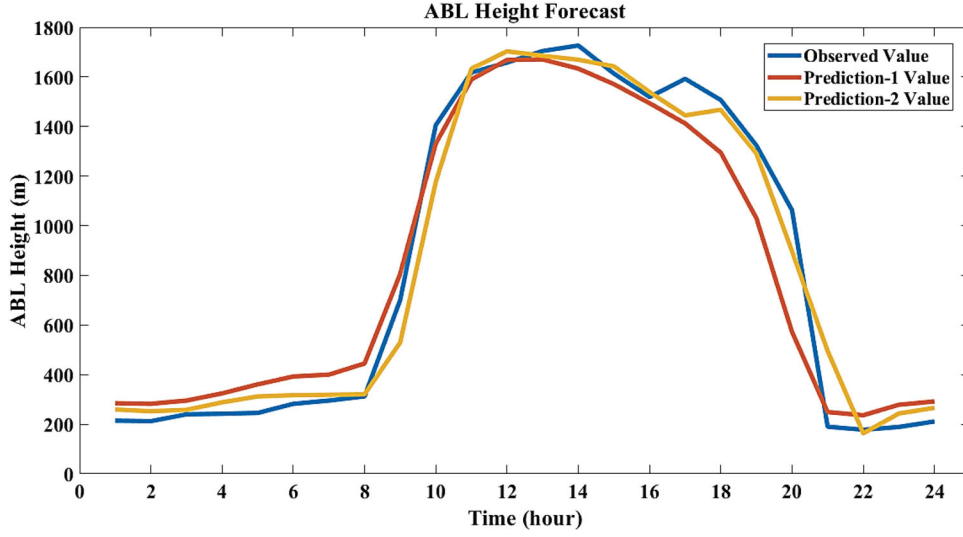


Fig. 7. Comparison of Line Plot of Observed ABL Height and Predication ABL Height.

Table 5. Comparison between NAR and LSTM Model.

	NAR	Prediction-1	Prediction-2
RMSE	259.65	220.31	193.85
MAE	164.97	167.12	113.72

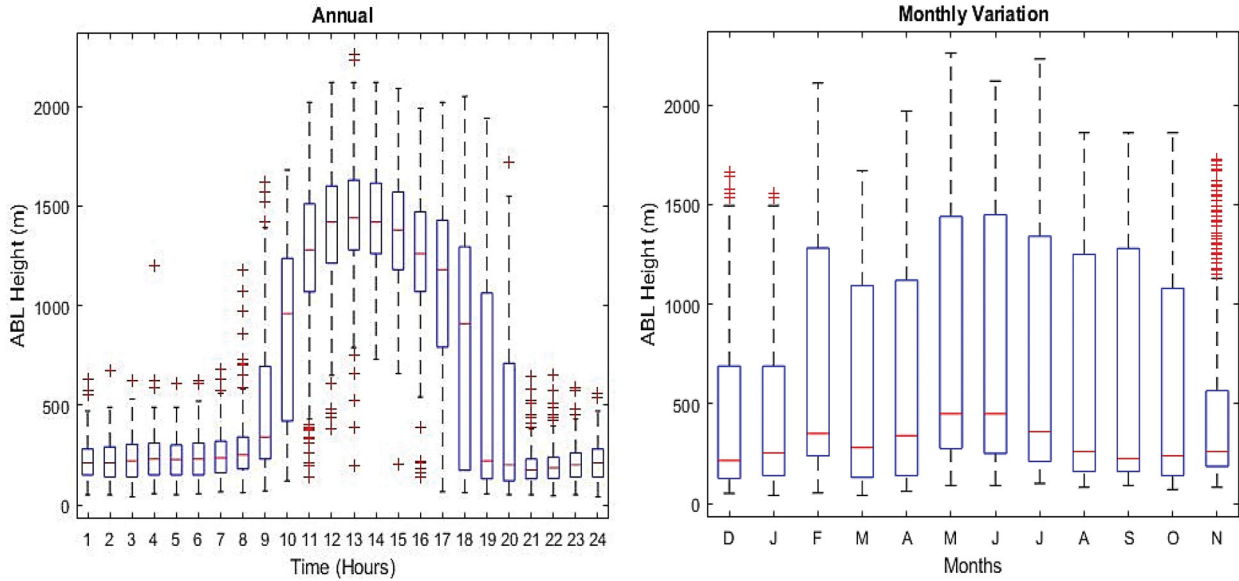


Fig. 8. Annual ABL height Temporal and Monthly variation.

3. Results

Performance evaluation of the experiment for ABL height LSTM model has been obtained for the optimal parameters. The testing is done for two parameters, i.e. hidden neuron and epochs with constant data set (2160 data point). To avoid overfitting, the networks are

trained using different hidden layer varying in the range (2, 5, 10, 20, 30, 50, 100, 35, 28, 32, respectively) with different epochs (250, 500, 750). The network (32 hidden layer and 500 epochs) training progress is shown in Fig. 5, which provide the best result from among other combination.

Table 6. Monthly ABL height variation during different hours.

Month	Maximum ABL height (m)	Corresponding hour of maximum ABL height (hrs)	Average ABL height (m) during		
			Day time (09:00–18:00 hr)	Remaining hours (19:00–08:00 hr)	Diurnal average (m)
December	1255	13:00	795	170	435
January	1120	13:00	710	200	415
February	1565	13:00	1125	265	620
March	1360	12:00	1035	185	545
April	1435	13:00	1145	225	615
May	1745	12:00	1480	350	810
June	1635	13:00	1475	325	795
July	1520	12:00	1380	325	760
August	1410	12:00	1305	260	690
September	1485	14:00	1280	190	645
October	1445	12:00	995	225	550
November	1405	11:00	850	230	490

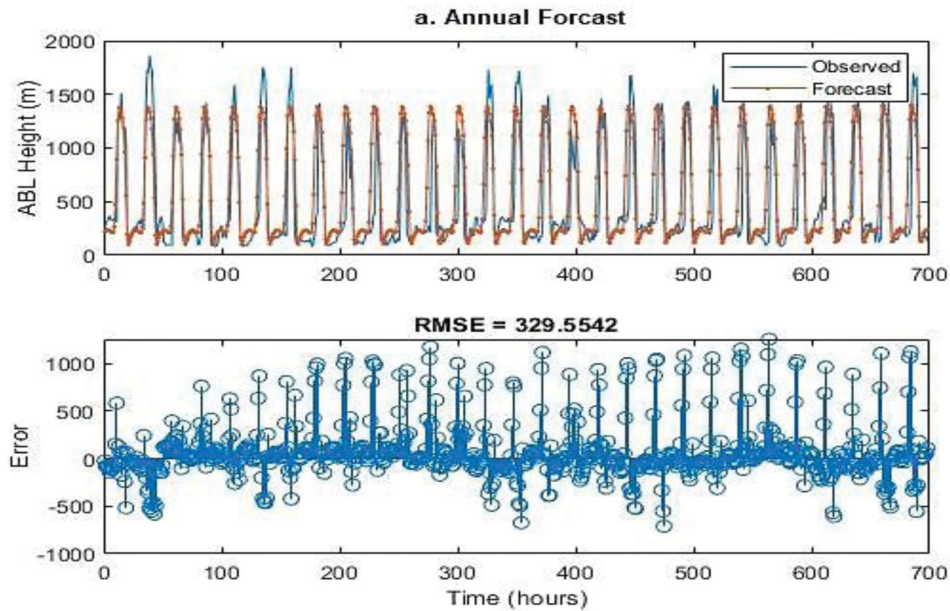


Fig. 9. Prediction result LSTM Network update with predicted values (Prediction-1), hidden layer 32 and maximum epochs 500.

Results obtained after testing is shown in Table 3 and Fig. 6. It has been observed that as the epochs and hidden layer (neuron) increases, the error value on the test set and training set both decreases. Table 3 shows that after certain epochs and the number of hidden layers and the error value increases, i.e. the prediction accuracy is improving.

To check the uncertainties in the LSTM model, different parameters have been calculated and represented in Table 4 and Fig. 7. It has been observed that the testing values and predicted values have followed the same

pattern. Figure 7 shows the line plot of temporal average of testing ABL height data and prediction ABL height data. In the LSTM model, Prediction-1 represents the updated network with predicted value (1650 m maximum) and Prediction-2 represents the updated network with observed values (1700 m maximum), with respect to the highest value of ABL height from SODAR 1775 m.

Also, Non-linear Auto Regressive (NAR) model is used to predict the ABL height using the predefined model of MATLAB 2019a. The maximum value of errors is obtained for each model is shown in Table 5. It has

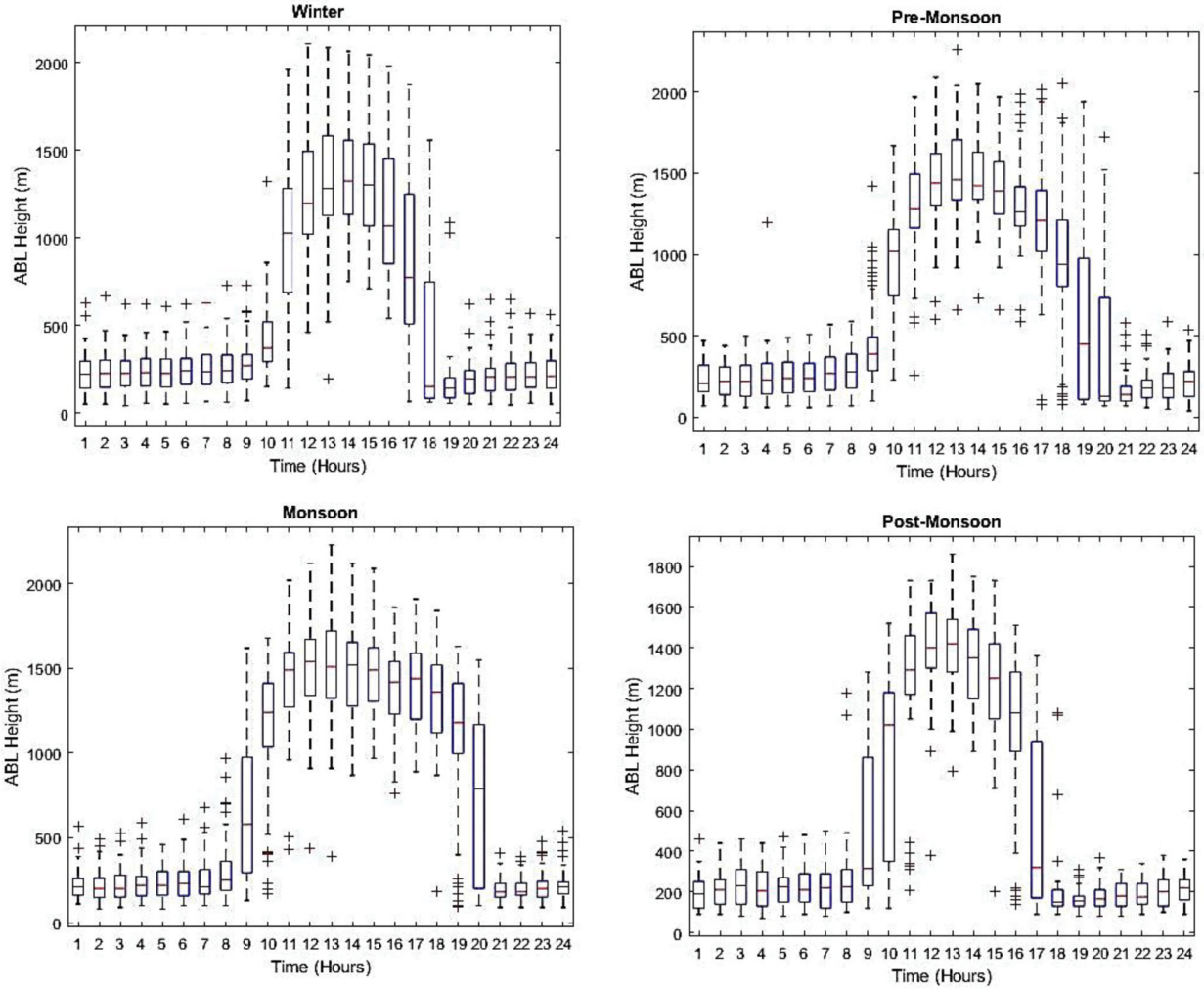


Fig. 10. Box plot of temporal seasonal ABL height.

been observed that Prediction-2 model values are more accurate than the NAR model. However, NAR model is used for the information from previous lags to predict future instances. The ABL height is highly dynamical in nature for day-to-day atmosphere. This causes a learning problem to NAR architecture and fails to capture the dynamical changes accurately.

3.1. Analysis and prediction of annual ABL height

The ABL is a zone having nearly constant potential temperature and specific relative humidity with height. ABL height determines the volume available for dispersion of pollutants and characterises the structure of the lower atmosphere. Higher the value of mixing height, the greater is dispersion rate and vice versa. Based on the observation of SODAR echograms, the ABL height is changing continuously. So, the box plot is used to interpret the data (Frigge et al., 1989; Williamson et al.,

1989). The box plot uses the median, the approximate quartiles, and the lowest and highest data points to convey the level, spread, and symmetry of a distribution of data values. Every box has a central mark, which indicates median value and whereas the bottom-line represents the 25th percentiles, and top-line indicates 75th percentiles of the data. The whiskers cover the most extreme data points and the outliers have plotted individually using the '+' symbol. Figure 8 represents the annual variation of temporal ABL height and Month average. The vertical bars denote the $\pm \sigma$ standard deviation from the temporal average. Figure 8 and Table 6, presents the temporal average SODAR data for about one year (December 2018 to November 2019).

The annual ABL height data have been used to retained LSTM model to predict the annual temporal ABL height. The result obtained shows the Prediction-2 model has the lowest RSME (187.71 m). Figure 9, Prediction-1 model provides a good result for 30 days

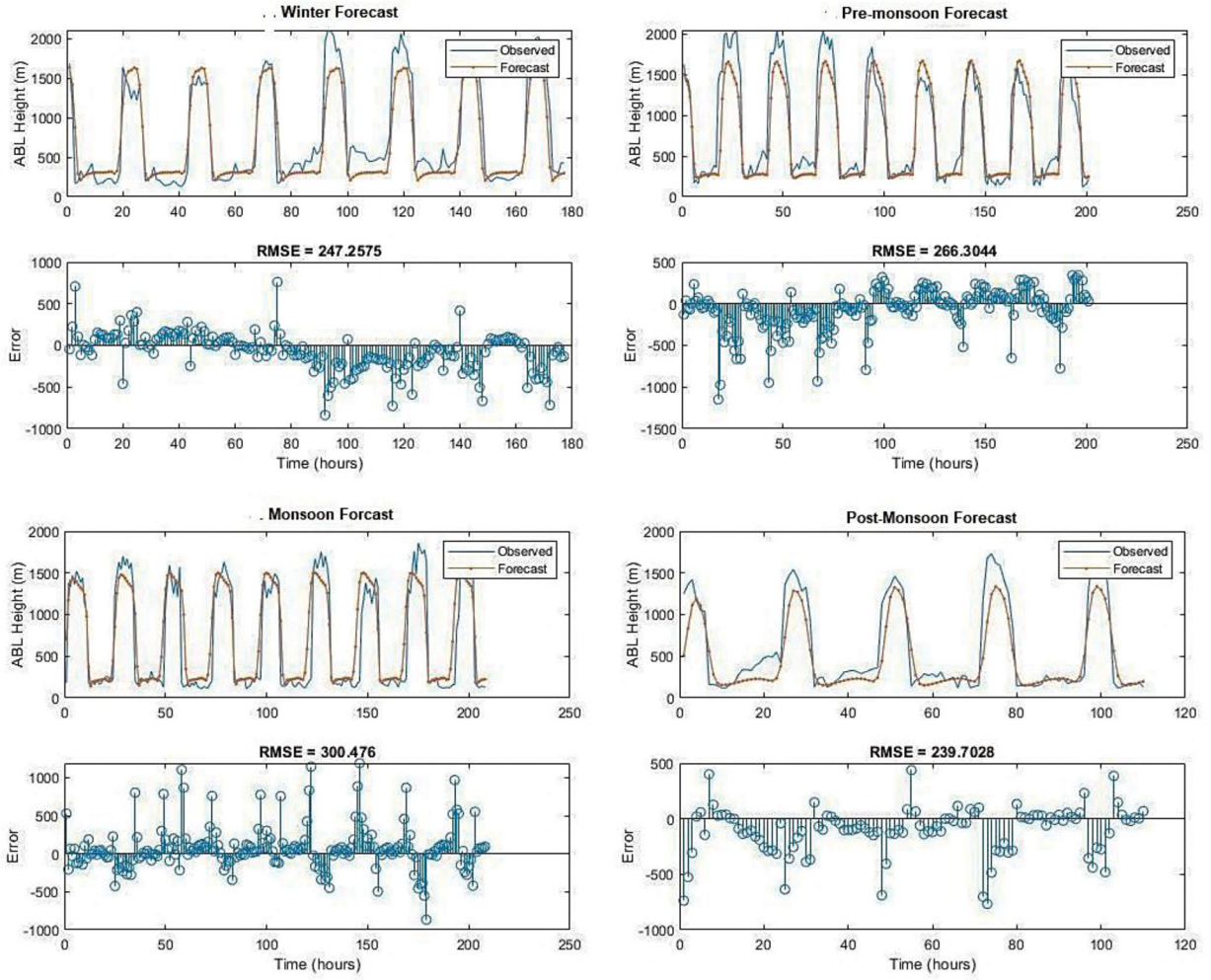


Fig. 11. Seasonal ABL height Prediction result from the update network state with predicted values (Prediction-1).

with RSME value (329.55 m). During convection period, the predicted ABL height is lower than observed, which increases the error. Also, the annual data set has been applied on NAR model and obtained a higher RSME value (261.80 m) compared to Prediction-2 model.

3.2. Comparison and prediction of seasonal ABL height

Total data set has been classified into four seasons based on meteorology over northern India (Soni et al., 2011; Ramachandran et al., 2012; Kumar et al., 2017b) namely, winter (December-January-February), Pre-monsoon (March-April-May), Monsoon (June-July-August-September) and Post-monsoon (October-November) for the analysis of seasonal ABL height. The temporal seasonal variation of mixing height during the whole year of observation is shown in Fig. 10. It is observed that the convection period is most extended during monsoon and

lowest during the post-monsoon. The maximum value (about 1510 m) of mixing height has been observed in the pre-monsoon season, while the minimum value of mixing height (around 1315 m) is found during the winter season. The monthly average ABL Height during different hours has been represented in Table 6 and found that in May month (Pre-Monsoon season), ABL height is the highest and lowest in January (winter season).

ABL height has a positive correlation with temperature and wind speed, whereas a negative relationship with relative humidity (Kumar et al., 2017b). Temperature and wind speed influences positively to the ABL, while relative humidity influences negatively to the ABL heights, during all the seasons due to an increase or decrease in solar heating. The convective boundary layer height increases and decreases during the day time due to the change in surface temperature. The variation in surface temperature carbon monoxide mixing ratio the existence of atmospheric convection; therefore, it strongly affects

Table 7. Comparison of seasonal prediction of ABL height.

		Hidden layers – 32; Max Epochs – 500								NAR models
		Prediction-1				Prediction-2				
	Data point	RMSE1	rRMSE1	MAE1	MAPE1	RMSE2	rRMSE2	MAE2	MAPE2	
Annual	6984	329.55	10.29	200.13	23.35	187.71	5.86	118.78	17.04	261.80
Winter	1776	247.25	7.72	181.83	19.59	187.77	5.86	125.81	13.19	307.33
Pre-Monsoon	2016	266.30	8.32	182.99	20.02	181.18	5.66	116.24	11.04	245.91
Monsoon	2088	300.47	9.38	187.14	15.62	255.45	7.98	151.71	12.74	281.25
Post-Monsoon	1104	239.70	7.49	163.73	25.15	178.92	5.59	118.57	15.17	250.94

the height. Generally, it is high between 11:00 to 14:00 during all seasons. Mixing height starts decreasing due to the decrease in solar heat during the evening. The ABL height during the convection period is found 1510m, 1485m, 1395m, and 1315m in pre-monsoon, monsoon, post-monsoon, and winter seasons, respectively.

All the season data is used for the prediction of seasonal ABL height. Both LSTM models are retrained and tested with seasonal data and to obtain the result for the prediction of seasonal ABL height. Figure 11 represents the seasonal ABL height prediction result from the Prediction-1 model. It is observed that when the prediction days increases the accuracy of Prediction-1 model decrease. Seasonal data is used to retrain and test the NAR model. It has been observed that the NAR model gives a higher RSME value.

4. Discussion

The four season's data sets of the ABL height in different atmospheric conditions are selected to train the LSTM model. Then, the trained prediction model is used to predict the ABL height of each season to compare with the annual prediction (Figures 9 and 11). The predictions ABL height shows a comparable result with the SODAR data, especially over transition period (from inversion to convection or vice versa). Table 7 shows the error of ABL height prediction using the LSTM network, which is lower in winter season and higher in annual data. Whereas error in all conditions of Prediction-2 is lower as compare to Prediction-1 and NAR model. But the related error generated by Prediction-1 is not much higher as compared to Prediction-2. The LSTM network model is providing high prediction accuracy as compared to NAR model. The LSTM models simulate the seasonal ABL height reasonably well, i.e. inferred from the periodicity in the ABL height time series. It also provides reliable ABL height simulations and predictions at any sites where the yearly ABL height pattern remains somewhat similar. It is observed that ABL height is highest in the convection period (daytime), do not follow any

periodicity and somewhat errors are highest during this period. These LSTM models are useful to the pollution regulatory body to control atmospheric pollution.

5. Conclusion

A neural network based temporal ABL height prediction is presented in this paper. It has been observed that LSTM neural network architecture is capable of capturing the hidden dynamics of temporal ABL and able to make predictions. The critical point of this approach is the decomposition of the temporal data into seasonal and the annual prediction of ABL height. Application of LSTM instead of classical neural network has enabled to obtain better accuracy of the seasonal prediction and, aggregate annual data. The proposed model is applied to ABL height data measurement using the SODAR system situated at CSIR-NPL, Delhi. The obtained results of prediction are in good agreement with the actual measurements. It is found that the optimal results achieved in short training time, when the number of neurons is equal to 32 and epoch are equal to 500, which provide accurate prediction result to obtain a long-term prediction. The LSTM model has foundation for the next step of accurate prediction of ABL height. Two types of prediction LSTM model have been applied for prediction of annual and seasonal ABL height. Both the models have provided good accuracy as compared to NAR model. These LSTM networks will apply to a different environmental parameter to find out the atmospheric condition in future.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adnan, R. M., Liang, Z., Parmar, K. S., Soni, K. and Kisi, O. 2021 . Modeling monthly streamflow in mountainous basin

- by MARS, GMDH-NN and DENFIS using hydroclimatic data. *Neural Comput. Appl.* 33(7), 2853–2871.
- Beyrich, F. 1997. Mixing height estimation from SODAR data—a critical discussion. *Atmos. Environ.* 31, 3941–3953. doi:10.1016/S1352-2310(97)00231-8
- Bradley, S. 2007. *Atmospheric Acoustic Remote Sensing: Principles and Applications*. Taylor & Francis Group, USA: CRC Press.
- Caughey, S. J. 1984. Observed characteristics of the atmospheric boundary layer. In: *Atmospheric Turbulence and Air Pollution Modelling*. Springer, Dordrecht, pp. 107–158.
- Chelani, A. B. 2005. Predicting chaotic time series of PM10 concentration using artificial neural network. *Int. J. Environ. Stud.* 62, 181–191. doi:10.1080/0020723042000285906
- Emeis, S., Schäfer, K. and Munkel, C. 2008. Surface-based remote sensing of the mixing-layer height—a review. *Metz.* 17, 621–630. doi:10.1127/0941-2948/2008/0312
- Ettouney, R. S., Farouq, S. M., John, G. Z., Mahmoud, A. E. and Hisham, M. E. 2009. Forecasting of ozone pollution using artificial neural networks. *Manag. Environ. Qual.* 20, 668–683. doi:10.1108/14777830910990843
- Freeman, B. S., Taylor, G., Gharabaghi, B. and Thé, J. 2018. Forecasting air quality time series using deep learning. *J. Air Waste Manag. Assoc.* 68, 866–886. doi:10.1080/10962247.2018.1459956
- Frigge, M., Hoaglin, D. C. and Iglewicz, B. 1989. Some implementations of the boxplot. *Am. Statist.* 43, 50–54.
- Garratt, J. R. 1994. The atmospheric boundary layers. *Earth. Sci. Rev.* 37, 89–134. doi:10.1016/0012-8252(94)90026-4
- Gera, B. S., Raghavendra, T., Singh, G., Ojha, V. K., Malik, J. and co-authors. 2011. Instrumentation and computer capabilities for improving SODAR data acquisition. *Int. J. Remote Sens.* 32, 4807–4817. doi:10.1080/01431161.2010.489072
- Gilman, G. W., Coxhead, W. B. and Willis, F. H. 1946. Reflection of sound signals in the troposphere. *J. Acoust. Soc. Am.* 18, 274–283. doi:10.1121/1.1916364
- Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. *Neural Comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Kawakami, K. 2008. Supervised sequence labelling with recurrent neural networks. Ph.D. dissertation.
- Kingma, D. P. and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Koffi, E. N., Bergamaschi, P., Karstens, U., Krol, M., Segers, A. and co-authors. 2016. Evaluation of the boundary layer dynamics of the TM5 model over Europe. *Geosci. Model Dev.* 9, 3137–3160. doi:10.5194/gmd-9-3137-2016
- Kumar, N. and Jha, G. K. 2013. A time series ANN approach for weather forecasting. *Int. J. Control Theory Comput. Model* 3, 19–25. doi:10.5121/ijctcm.2013.3102
- Kumar, N., Parmar, K. S., Soni, K., Garg, N. and Agarwal, R. 2015. Prediction of ventilation coefficient, using a conjunction model of wavelet-neuro-fuzzy model: A case study Delhi, India. *Acad. J. Sci. Res.* 3, 184–191.
- Kumar, N., Soni, K. and Agarwal, R. 2021. A comprehensive study of different feature selection methods and machine-learning techniques for SODAR structure classification. *Model. Earth Syst. Environ.* 7, 209–212. doi:10.1007/s40808-020-00872-0
- Kumar, N., Soni, K., Agarwal, R. and Singh, M. 2017a. SODAR as a diagnostics tool for urban air-quality and health care system. *J. Acoust. Soc. India* 44, 213–222.
- Kumar, N., Soni, K., Agarwal, R. and Singh, M. 2019. Preliminary analysis of operating conditions of acoustic sounder (SODAR) before the installation. *J. Acoust. Soc. Am.* 146, 2997–2998. doi:10.1121/1.5137378
- Kumar, N., Soni, K., Garg, N., Agarwal, R., Saha, D. and co-authors. 2017b. SODAR pattern classification and its dependence on meteorological parameters over a semiarid region of India. *Int. J. Remote Sens.* 38, 3466–3482. doi:10.1080/01431161.2017.1294774
- Lee, T. R. and Pal, S. 2017. On the potential of 25 years (1991–2015) of rawinsonde measurements for elucidating climatological and spatiotemporal patterns of afternoon boundary layer depths over the contiguous US. *Adv. Meteorol.* 2017, 1–20.
- Levi, Y., Dayan, U., Levy, I. and Broday, D. M. 2020. On the association between characteristics of the atmospheric boundary layer and air pollution concentrations. *Atmos. Res.* 231, 1–16.
- Liu, C., Huang, J., Wang, Y., Tao, X., Hu, C. and co-authors. 2020. Vertical distribution of PM2.5 and interactions with the atmospheric boundary layer during the development stage of a heavy haze pollution event. *Sci. Total Environ.* 704, 135329–135348. doi:10.1016/j.scitotenv.2019.135329
- Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J. and Kolehmainen, M. 2004. Evolving the neural network model for forecasting air pollution time series. *Eng. Appl. Artif. Intell.* 17, 159–167. doi:10.1016/j.engappai.2004.02.002
- Paoli, C., Cyril, V., Marc, M. and Marie-Laure, N. 2010. Forecasting of preprocessed daily solar radiation time series using neural networks. *Sol. Energy* 84, 2146–2160. doi:10.1016/j.solener.2010.08.011
- Ramachandran, S., Kedia, S. and Srivastava, R. 2012. Aerosol optical depth trends over different regions of India. *Atmos. Environ.* 49, 338–347. doi:10.1016/j.atmosenv.2011.11.017
- Rehman, S. and Mohandes, M. 2008. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* 36, 571–576. doi:10.1016/j.enpol.2007.09.033
- Roy, S. S., Singh, R. B. and Kumar, M. 2011. An analysis of local spatial temperature patterns in the Delhi Metropolitan Area. *Phys. Geogr.* 32, 114–138. doi:10.2747/0272-3646.32.2.114
- Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K. and Soman, K. P. 2017. Stock price prediction using LSTM, RNN and CNN-sliding window model. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, pp. 1643–1647.
- Singal, S. P., Gera, B. S. and Pahwa, D. R. 1994. Application of SODAR to air pollution meteorology. *Remote Sens.* 15, 427–441. doi:10.1080/01431169408954084

- Soni, K. and Parmar, K. S. 2020. Soft computing applications in air pollution meteorology. In: *Intelligent Data Analytics for Decision-Support Systems in Hazard Mitigation*, Springer, Singapore, pp. 441–469.
- Soni, K., Singh, S., Bano, T., Tanwar, R. S. and Nath, S. 2011. Wavelength dependence of the aerosol angstrom exponent and its implications over Delhi, India. *Aerosol Sci. Technol.* 45, 1488–1498. doi:10.1080/02786826.2011.601774
- Stull, R. B. 2012. *An Introduction to Boundary Layer Meteorology*. Vol. 13. Springer Business Media.
- Terradellas, E., Soler, M. R., Ferreres, E. and Bravo, M. 2005. Analysis of oscillations in the stable atmospheric boundary layer using wavelet methods. *Boundary Layer Meteorol.* 114, 489–518. doi:10.1007/s10546-004-1293-y
- Unal, Y., Kindap, T. and Karaca, M. 2003. Redefining the climate zones of Turkey using cluster analysis. *Int. J. Climatol.* 23, 1045–1055. doi:10.1002/joc.910
- Vivas, D. R., Sánchez, E. and Reina, J. H. 2020. Deep learning the atmospheric boundary layer height. *arXiv preprint arXiv:2004.04353*. <https://arxiv.org/pdf/2004.04353.pdf>
- Williamson, D. F., Parker, R. A. and Kendrick, J. S. 1989. The box plot: a simple visual method to interpret data. *Ann. Intern. Med.* 110, 916–921. doi:10.7326/0003-4819-110-11-916
- Zaidi, S., Zela, A., Elsken, T., Holmes, C., Hutter, F. and co-authors. 2020. Neural ensemble search for performant and calibrated predictions. *arXiv preprint arXiv:2006.08573*.
- Zhao, J., Wu, J., Guo, X., Han, J., Yang, K. and co-authors. 2019. Prediction of radar sea clutter based on LSTM. *J. Ambient Intell. Hum. Comput.* 1–8.