



A Hybrid Deep Learning Model for Multi-step Ahead Prediction of PM_{2.5} Concentration Across India

Pranjol Goswami¹ · Manoj Prakash² · Rakesh Kumar Ranjan³ · Amit Prakash¹

Received: 25 September 2022 / Accepted: 10 May 2023 / Published online: 3 June 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Fine particulate matter (PM_{2.5}) concentration in ambient air has become a major concern across the globe. All major cities of India have reported an elevated concentration of PM_{2.5} that has severe consequences to the health, economy, and ecosystem of the region. As a result, it becomes imperative to develop adequate tools for forecasting particulate matter concentration. Most of the research works mostly focused on single-step prediction horizon, thereby limiting their use. In the present work, a hybrid model has been proposed to forecast multi-step ahead concentrations of PM_{2.5} in ambient air across India covering different agroclimatic zones. The hybrid model architecture was an encoder-decoder-based sequence to sequence model framework that was built with convolutional long short-term memory (LSTM), bidirectional LSTM and 3D convolution neural network. The model was tested across 26 Indian cities covering 13 major agroclimatic zones of India. The performance of the model was also analysed for consecutive hour sequential prediction taking last 24-h data as input to the model. The model output was also compared with signal to noise ratio to explore the reason for variations in model performance. A distinct trend was found between signal to noise ratio and model output. As noise increases, the model performances suffer. Overall, the model was found to be stable as its performance errors across different time horizon has little variations. The proposed model has the potential to be used for long-term forecasting by incorporating other predictor variables series.

Keywords Hybrid deep learning model · Encoder-decoder · LSTM · Bidirectional LSTM · Convolutional LSTM · 3D convolution neural network · PM_{2.5} · Multistep ahead prediction · SNR

1 Introduction

Human-induced activities have increased the level of air pollution, and its adverse consequences were suffered by all. Researchers across the world have been concerned with the

challenging problem of high ambient aerosol concentrations that directly affect our health, economy and the climates [1, 2]. A World Health Organization (WHO) report [3] suggests that 90% of the population of the planet breathes air that crosses the WHO air quality recommendations, and every year, around 7 million people have lost their lives from exposure to ambient air pollution. A survey conducted by WHO in the year 2016 reported that out of the world's top twenty polluted cities, fourteen cities belong to India. Therefore, reliable forecasting of PM_{2.5} concentration is imperative to forewarn the public as well as policy planners to take corrective measures. In the last couple of decades, different forecasting methods such as the deterministic model [4–8], statistical model [9] and artificial neural networks (ANN)-based models were explored and attempted by various researchers across the world. In recent years, ANN-based models gained prominence owing to its ability to handle linear and nonlinear variability present in the environmental data with limited set of variables. The general regression neural network [10] and neuro-fuzzy models [11], feed forward network, radial basis function network, multilayer

✉ Amit Prakash
ap@tezu.ernet.in; amitprakash.amit@gmail.com

Pranjol Goswami
goswamipranjol@gmail.com

Manoj Prakash
manoj.prakash@leadics.com; manoj.prakash@gmail.com

Rakesh Kumar Ranjan
rkrnanjan@cus.ac.in; rakeshranjan.jnu@gmail.com

¹ Department of Environmental Science, Tezpur University, Napaam 784028, Assam, India

² Leadics, Bengaluru 560038, Karnataka, India

³ Department of Geology, Sikkim University, Gangtok 737102, Sikkim, India

perceptron (MLP) model [12], back-propagation neural network [13, 14], and recurrent neural network (RNN) [15, 16] with error back-propagation learning techniques are some of the widely used ANN architecture for air pollution forecasting [17–20]. With the advancement in computing technology, an emerging field of machine learning and artificial intelligence has attracted a lot of researchers to apply deep learning techniques (a subset of machine learning) for diverse problems of societal relevance.

Among many deep learning architectures, long short-term memory (LSTM) network was mostly used for air quality time series forecasting in the recent past because of its ability to capture long- and short-term dependencies [21–26]. However, different hybrid network architectures were also employed for different aspects of air quality studies [27–29]. The LSTM network was further used to ascertain the long-term and short-term dependencies [30] and effectiveness of encoder-decoder networks for building prediction machines with time-series data [31, 32]. A hybrid model consisting of convolutional LSTM and CNN was also attempted to predict the concentration of particulate matter [33, 34] where convolutional LSTM was used for sequential spatiotemporal information and CNN for extracting temporal features in parallel. Similarly, transfer learning BiLSTM model was also examined for hourly, daily and weekly prediction of air quality [35].

Most of these studies were focused on predicting air quality at single (or few) monitoring stations making use of such modelling techniques rather limited. India being a very diverse (7th largest country by area with 2nd most populated in the world) country spreading from $8^{\circ}4'$ to $37^{\circ}6'$ latitude and $68^{\circ}7'$ to $97^{\circ}25'$ longitude, models trained on selected cities may not be efficient for use in other cities. The model architecture needs constant alteration for use in other cities. Therefore, a uniform simpler model is essentially required which is less data intensive and can be applicable across India without the need for structural changes. The present study aims to develop a hybrid deep learning network with uniform model architecture that can be applicable for all monitoring stations across India. Besides developing and testing the model for multi-step ahead forecasting, it is also ascertained the dependence of model performance on the relative variance measure in terms of signal-to-noise ratio (SNR) of the input data. Section 2 details site description of data pre-processing; model development and architecture are presented in Sect. 3. Result and discussions are given in Sect. 4 and conclusion in Sect. 5.

2 Site and Data Description

For the present study, the air pollution data were acquired for 26 different cities across the country from the Central Pollution Control Board (CPCB), Government of India

(<http://www.cpcb.nic.in/>) (Fig. 1). For the ease of analysis, India was further subdivided into 15 Agroclimatic zones (Table 1) as per Indian Meteorological Department (IMD), Government of India (GOI) classification [36]. No data was available for two regions, namely, WH (U.T. of Jammu and Kashmir and Union Territory of Ladakh) and the IR of India (Andaman and Nicobar Island, Lakshadweep Island), hence were not included in the present study. The data was collected for the duration from 1 January 2015, to 31 May 2020 depending upon the data availability. Details of the data used in this work have been presented in Table 1.

2.1 Data Pre-Processing

The data acquired from the secondary sources were often infected with outliers and missing values. Therefore, pre-processing of the data to eliminate and minimise such errors is highly imperative. In the present study, the unreasonably high values were considered as outliers and were replaced with the help of linear interpolation method [37]. Similar techniques were followed for filling of missing values present in the data.

3 Model Development

3.1 Network Architecture

In the present study, an ANN architecture with deep learning framework was adopted for multi (Eight) step ahead forecasting. The model architecture is an encoder-decoder-based (Fig. 2) [38] sequence-to-sequence hybrid model, which has three main components, namely,

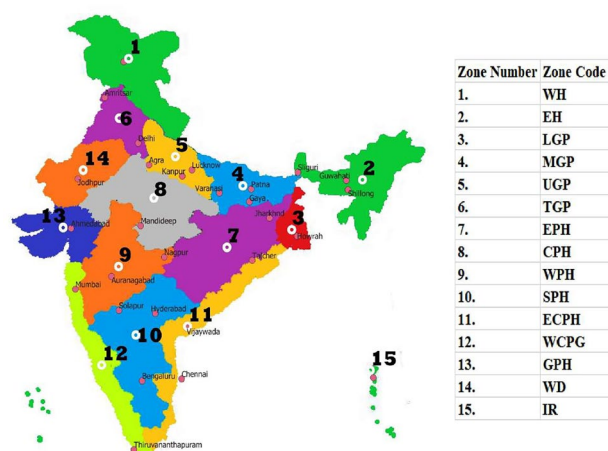


Fig. 1 Agroclimatic zones of India

Table 1 Agroclimatic zones and data description

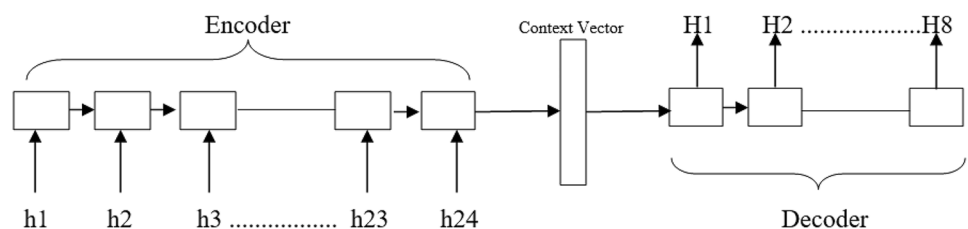
Sl. no.	Agroclimatic zone	Zone code	Station name	Data period	
				From	To
1	Eastern Himalayan Region	EH	Shillong	27 August 2019	31 May 2020
			Guwahati	16 February 2019	31 May 2020
			Siliguri	1 February 2018	31 May 2020
2	Lower Gangetic Plain Region	LGP	Howrah	19 January 2018	31 May 2020
3	Middle Gangetic Plain Region	MGP	Gaya	1 January 2016	31 May 2020
			Patna	21 October 2017	10 April 2020
			Varanasi	1 January 2015	31 May 2020
4	Upper Gangetic Plains Region	UGP	Agra	11 May 2015	31 May 2020
			Kanpur	12 May 2015	31 May 2020
			Lucknow	28 March 2015	31 May 2020
5	Trans-Ganga Plains Region	TGP	Amritsar	27 February 2017	31 May 2020
			Delhi	3 November 2016	31 May 2020
6	Eastern Plateau and Hills	EPH	Jamshedpur	9 January 2019	31 May 2020
			Talcher	7 February 2018	31 May 2020
7	Central Plateau and Hills	CPH	Mandideep	2 January 2018	31 May 2020
			Nagpur	30 March 2016	31 May 2020
8	Western Plateau and Hills	WPH	Aurangabad	1 October 2017	31 May 2020
			Mumbai	6 May 2018	31 May 2020
			Solapur	1 January 2016	31 May 2020
9	Southern Plateau and Hills	SPH	Bengaluru	23 March 2015	31 May 2020
			Hyderabad	1 October 2015	31 May 2020
10	Eastern Coastal Plains and Hills	ECPH	Chennai	23 March 2015	31 May 2020
			Vijayawada	26 April 2017	27 October 2019
11	Western Coastal Plains and Ghats	WCPC	Thiruvananthapuram	21 June 2017	31 May 2020
12	Gujarat Plains and Hills	GPH	Ahmedabad	1 January 2018	31 May 2020
13	Western Dry Region	WD	Jodhpur	21 September 2017	31 May 2020
14	Western Himalayan Division	WH	No data available during the period	–	–
15	Island Region	IR	No data available during the period	–	–

1. 3D-CNN: 3-dimensional convolutional neural network model
2. ConvLSTM: convolutional long short-term memory
3. BiLSTM: bidirectional long short-term memory

Essentially, 3D-CNN and LSTM networks are the backbone of this architecture. The LSTM model was widely applied for time series prediction, because of its ability to store the information in self-recurrent cells that can be retrieved at different time steps. The LSTM network performs exceedingly well in reducing the Gaussian

noise present in the data [39], but unable to filter out non-Gaussian noise, which was inherently present in the data set. To address these shortcomings, BiLSTM network was applied to reduce the overfitting of noisy data. Besides, the ConvLSTM model performed better in datasets having long-duration sequential features with multiple temporal information [40]. Furthermore, the 3D-CNN model was advancement over 2D-CNN model that has better processing ability for large contextual data helpful in extracting the spatiotemporal features. The ability of the 3D-CNN to

Fig. 2 Encoder-decoder-based sequence-to-sequence model



extract features from large sequential data into different time–frequency domains was exploited to reduce noise present in the data as well as to abstract features that can be stored and further fed into the next fully connected layer. The schematic diagram of hybrid model architecture is presented in Fig. 3.

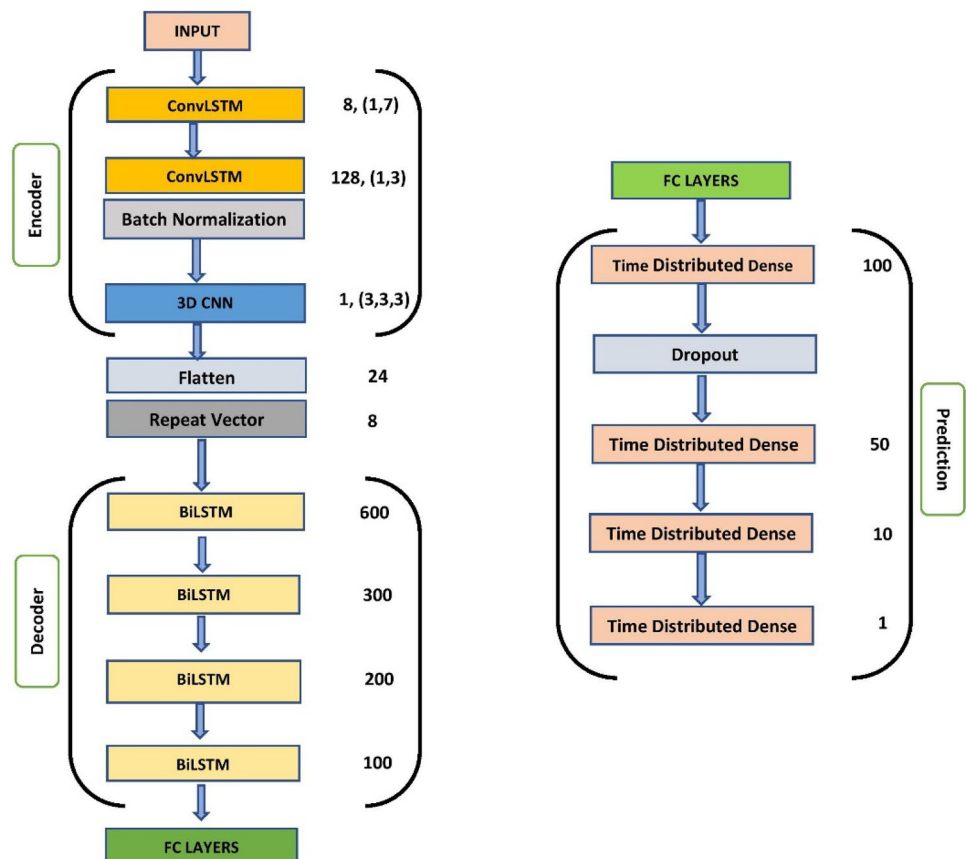
In the model architecture, ConvLSTM encoder layer generates a feature map that was further refined and filtered by the second ConvLSTM network with Batch normalisation layer. The output is fed into 3D-CNN to extract spatiotemporal patterns from the state matrix. The output of the 3D-CNN layer then feeds into the decoder layer having four BiLSTM networks. BiLSTM will generate a string of the entire sequence containing values for 8 h. The first, second and third fully connected layers act as an interpretation layer for each time step of the output sequence, and the last fully connected layer is the final output layer of the model that generates the final predicted value of 8 steps ahead prediction. Concurrently, a dropout layer was used after the first BiLSTM to minimize the overfitting. Each layer of filter is a CNN model abstract feature. Since initial network layers receive the noisy raw data, fewer filters were used to capture the basic features only. In the subsequent layers, the number of filters was increased to capture deeper abstraction of features. A smaller filter size

or kernel size can capture more features than a larger kernel size. We applied 64 numbers of filters of size (1,7) in the first ConvLSTM layer. In the second ConvLSTM layer, the number of filters was increased to 128, and kernel size was decreased to (1,3). Odd numbers of kernel size were used to maintain symmetry around the centre or origin of the abstraction layer.

The BiLSTM layer, used in the model, acts as a decoder and generates output of multiple values in a sequence. Cross validation and out of sample testing techniques were employed to evaluate the model performance. A similar model framework was earlier applied by [41] for learning smart manufacturing problems using time series data. However, they used stacked ConvLSTM as an encoder and stacked BiLSTM as a decoder layer for an auto encoder model framework. In the present study, Stacked ConvLSTM layer outputs were fed into the 3D-CNN layer. The air pollution time series data are the net outcome of the complex interplay between different stochastic and dynamic processes having different characteristic frequencies [16]. Therefore, 3D-CNN was used to take into account the characteristic features, enhancing the ability of the network for better prediction.

To forecast $PM_{2.5}$ value for the next 8 h, we used 3 sequences of 8-h durations, i.e., 24 h of data as input sequence

Fig. 3 Model architecture



with the next 8 h of data as target. But this number of instances would be rather limited for training a deep learning model. Therefore, an overlapping moving window method was used during training of the time series data for generating more training instances. This method is a modified rolling window method as proposed in [42] and later adopted for air pollution studies by [35]. Here, a large training dataset was generated by shifting the entire sequence by one step (Fig. 4) as discussed as follows.

Let us consider a time series $u(t) = \{u_1, u_2, u_3, \dots, u_t\}$. In order to forecast the next k values of the sequence $\hat{s} = (\hat{s}_1, \hat{s}_2, \dots, \hat{s}_k)$ equivalent to $(u_{t+1}, u_{t+2}, \dots, u_{t+k})$ with the help of last observation and a moving window of fixed size w , it would be

$$\hat{s} = (\hat{s}_1, \hat{s}_2, \dots, \hat{s}_k) = f(u_{t-w}, u_{t-w+1}, u_{t-w+2}, \dots, u_t)$$

When the above operations are applied to a univariate time series of length N , it generates a sequence to sequence prediction with an input set $U \in R^{n \times w}$ and output set $S \in R^{n \times k}$. Here, n is the size of training data given by

$$n = (N - w - k + 1).$$

As evident from the above description, the entire sequence of 24-h time series data (h1 to h24) was converted into 3×8 sequence internally and mapped to the next 8-h values (h25 to h32). In the next step, the data were shifted by one value, and now, the data (h2 to h25) were mapped to the next 8-h sequence (h26 to h33) and so on.

Since the model was used in many stations situated at different geographical locations of India, the model parameters had been generalized in such a way that it could result in optimum value for most of the stations.

3.2 Hyperparameters

Hyperparameters in machine learning are special kinds of parameters that play a significant role in determining the performance of a deep learning model. The hyperparameters used in this paper are listed in Table 2.

The most widely used activation function for deep learning was the rectified linear unit (ReLU), which is $f(x) = \max(0, x)$. A new activation function was proposed by the Google Brain team [43], named ‘swish’ which is $f(x) = x \cdot \text{sigmoid}(x)$,

Table 2 Optimal architecture of the parameters used in the study

Layer	Hyperparameters	Value
ConvLSTM-1	Filter	64
	Kernel size	(1,7)
	Activation function	Swish
ConvLSTM-2	Filter	128
	Kernel size	(1,3)
	Activation function	NA
3D CNN	Filter	1
	Kernel size	(3,3,3)
	Activation function	Swish
BiLSTM-1	Units, Activation Function	600, tanh
BiLSTM-2	Units, activation function	300, tanh
BiLSTM-3	Units, activation function	200, tanh
BiLSTM-4	Units, activation function	100, tanh
Dense-1	Units, activation function	100, swish
Dense-2	Units, activation function	50, swish
Dense-3	Units, activation function	10, swish
Dense-4	Units, activation function	1, linear
	Dropout	0.3
	Learning rate	10e-4
	Optimizer	Adam

which performs better in a deeper network. Hence, in the present study, the swish activation function was used, and for BiLSTM part, ‘tanh’ activation function was applied.

There exist different types of optimizing algorithms such as Gradient Descent, Stochastic Gradient Descent, Momentum Based Gradient Descent, Adaptive Moment Estimation (Adam), Nesterov Accelerated Gradient (NAG) and Root Mean Square Propagation (RMSProp) to minimize the loss function during the training of a machine learning model. In the present work, Adaptive Moment Estimation (Adam) optimizer was used due to its adaptive nature and combined momentum component [44].

3.3 Model Evaluation

Furthermore, the effectiveness of the model was tested following a walk forward validation method. Statistical error metrics like root mean square error (RMSE), mean absolute

Fig. 4 Overlapping data by one step

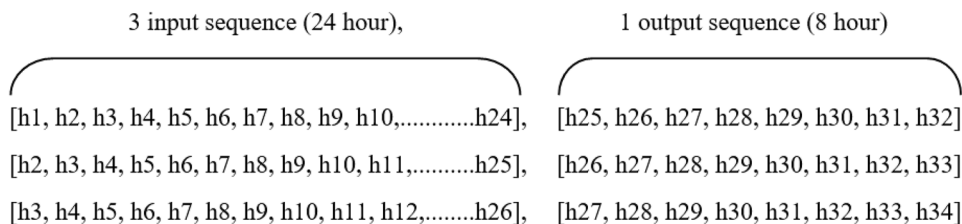


Table 3 Best fit Statistical distribution parameters values

Agro climatic zone	Site	Best fit distribution	Location parameter	Scale parameter	Scale/location
Eastern Himalayan Region	Guwahati	Gamma	1.00E+01	5.55E+01	5.55E+00
	Shillong	Lognorm	9.18E+00	1.05E+01	1.14E+00
	Siliguri	Gamma	1.00E+01	4.79E+01	4.79E+00
Lower Gangetic Plain Region	Howrah	Beta	1.00E+01	5.34E+03	5.34E+02
Middle Gangetic Plain Region	Gaya	Gamma	9.88E+00	7.62E+01	7.71E+00
	Patna	Lognorm	4.89E+00	8.32E+01	1.70E+01
	Varanasi	Beta	9.93E+00	2.85E+13	2.87E+12
Upper Gangetic Plain Region	Agra	Gamma	9.94E+00	6.55E+01	6.59E+00
	Kanpur	Gamma	9.97E+00	8.55E+01	8.57E+00
	Lucknow	Gamma	1.00E+01	7.81E+01	7.81E+00
Trans-Ganga Plain Region	Amritsar	Gamma	9.95E+00	2.84E+01	2.85E+00
	Delhi	Gamma	9.96E+00	7.18E+01	7.21E+00
Eastern Plateau and Hills	Jamshedpur	Gamma	3.94E+00	2.20E+01	5.58E+00
	Talcher	Gamma	1.00E+01	5.01E+01	5.01E+00
Central Plateau and Hills	Mandideep	Beta	9.95E+00	5.30E+03	5.33E+02
	Nagpur	Gamma	1.00E+01	2.88E+01	2.88E+00
Western Plateau and Hills	Aurangabad	Beta	9.67E+00	4.45E+02	4.60E+01
	Mumbai	Gamma	1.00E+01	2.43E+01	2.43E+00
	Solapur	Beta	1.00E+01	4.53E+08	4.54E+07
Southern Plateau and Hills	Bengaluru	Gamma	9.92E+00	2.19E+01	2.20E+00
	Hyderabad	Beta	9.78E+00	1.91E+02	1.95E+01
Eastern Coastal Plain and Ghats	Chennai	Beta	1.00E+01	2.63E+03	2.63E+02
	Vijayawada	Beta	9.82E+00	1.73E+02	1.76E+01
Western Coastal Plain and Ghats	Thiruvananthapuram	Gamma	1.00E+01	2.03E+01	2.03E+00
Gujrat Plain and Hills	Ahmedabad	Beta	9.67E+00	6.43E+13	6.64E+12
Western Dry Region	Jodhpur	Beta	8.59E+00	1.62E+12	1.88E+11

India has higher concentration of PM_{2.5} especially during monsoon and post monsoon season. In winter and pre-monsoon seasons, the highest concentration was confined to Indo Gangetic plain region.

A detailed statistical distribution analysis was carried out to ascertain the nature of distribution prevalent in PM_{2.5} data in different agroclimatic zones. It is imperative to understand the nature of the statistical distribution in any data as it determines the effectiveness of the model performance measures used in forecasting problems. To ascertain the best fit model, each dataset was tested against 7 common distributions applied in case of air pollution studies, namely, Normal, Log Normal, Logistic, Laplace, Weibull, Gamma and Beta. The best fitted distribution (Supplementary Fig. F1(A-Z)) was selected based on the minimum sum of square error criteria. The results (Table 3) reveal the predominance of Gamma and Beta distribution at 24 out of 26 sites. Overall, at 14 sites, PM_{2.5} concentration follows Gamma distribution, Beta distribution at 10 sites and lognormal at the remaining two sites. It is pertinent to mention that all the sites at UGP, TGP and EPH have Gamma distribution as best fit distribution. Similarly, LGP, ECPH, GPH follow Beta distribution

only. Log normal distribution was observed only at Shillong in EH region and at Patna in MGP. At the rest of the regions, mixed distribution fitting results were obtained.

4.2 Model Performance Evaluation Results

The model performance evaluations were carried out through commonly used error functions such as RMSE, MAE and MAPE. The error functions RMSE and MAE are scale dependent and not an optimum measure to compare different data sets where mean differences are larger. However, MAPE is a unitless function that is scale independent and more suitable for model comparisons even with data that have large variance and are infected with extreme values. The model's performances (Fig. 6) in terms of RMSE values for 8-consecutive-hour advance predictions ranged from the minimum of 7.09 in Shillong at EH region to the maximum of 53.81 in Patna at MGP. Similarly, in terms of MAE, minimum of 5.41 and maximum of 34.09 were obtained at Shillong and Patna respectively. In terms of MAPE, the minimum value of 18.6% and a maximum of 52.7% were observed in Hyderabad

and Chennai, respectively. The first step prediction errors in terms of RMSE were found to be less than $10 \mu\text{g}/\text{m}^3$ at nine cities whereas Shillong in EH, Howrah in LGP and Mandideep and Nagpur in CPH were found to be in between 10 and $15 \mu\text{g}/\text{m}^3$. In terms of MAE, 15 out of 26 cities have less than $10 \mu\text{g}/\text{m}^3$ MAE value, and overall, 24 out of 26 cities have $\leq 15 \mu\text{g}/\text{m}^3$ MAE values. Only Patna (MAE = 20.52) in the MGP and Talcher (MAE = 20.29) in EPH have MAE values more than $15 \mu\text{g}/\text{m}^3$. Similarly, 11 out of 26 sites have MAPE values less than 30%, and 20 sites show MAPE values $< 35\%$. Apart from Jodhpur in WD and Chennai in ECPH zone, the rest of the sites have MAPE values less than 40%. (The detailed results are presented in Supplementary Table T2). Overall, for predicting up to 8-h ahead concentrations, the model performance was found to be relatively better in the central, southern and western regions of India in comparison to northern and eastern regions. The robustness of the prediction ability of the proposed model framework is also evident from the 1-h ahead and 8-h average $\text{PM}_{2.5}$ concentration estimates. The performance evaluation for 1-h ahead and 8-h average prediction horizon is essential as the regulatory air quality was reported mostly in this temporal range. The minimum values of RMSE, MAE and MAPE for 1-h ahead concentration was found to be 5.81 (at Shillong), 3.92 (at Aurangabad) and 10.8 (Howrah), respectively. The maximum RMSE of 41.384 and 29.04 was obtained respectively for 1-h and 8-h average periods at Patna and Talcher, respectively. Overall, less than 20 RMSE values were observed at 18 sites and 13 sites for 1-h ahead and 8-h average forecasting horizon, respectively. In case of MAE, again, minimum and maximum values for 1-h and

8-h average were found at Shillong and Patna respectively. The results were found to be more uniform across India in terms of MAE as only 3 sites namely Delhi, Talcher and Patna, have MAE values more than $20 \mu\text{g}/\text{m}^3$.

Across agroclimatic zones (Table 4), the best model performance for 1-h ahead prediction was achieved in terms of RMSE for SPH (7.6) followed by CPH (8.4) and WPH (9.4). Overall, 7 zones have RMSE values less than 20 and 3 zones each have RMSE values in the range 20 to 30 and 30 to $40 \mu\text{g}/\text{m}^3$, respectively. In terms of MAE, similar trend in the error was observed with the minimum value (5.4) obtained at SPH and maximum (16.3) at MGP. In case of 8-consecutive-hour advance predictions, the same pattern was obtained with minimum RMSE and MAE values of 11.1 and 8.0, respectively, observed at SPH, and maximum RMSE (40.4) and MAE (25.8) were obtained for MGP. However, the results for MAPE values were slightly different with minimum error value observed at LGP (10.8%) and EH (21.3%) for 1-h ahead and cumulative 8-consecutive-hour ahead prediction, respectively.

The observed forecasting results exhibit spatial variability in model performance. As evident from the heatmap (Fig. 7) for multi-step hourly forecast, the regions mostly along the southeast, south, central and southwestern of India have better model performance in terms of MAE values. In most parts of North India, relatively poor model performance was observed except for Jamshedpur, Agra, Amritsar and Shillong. It is worth mentioning that, for the first step, model performance is best across India except at Patna and Talcher. Till date, cross-country analysis of model performance was not attempted in India, although ANN with deep learning architecture was attempted for

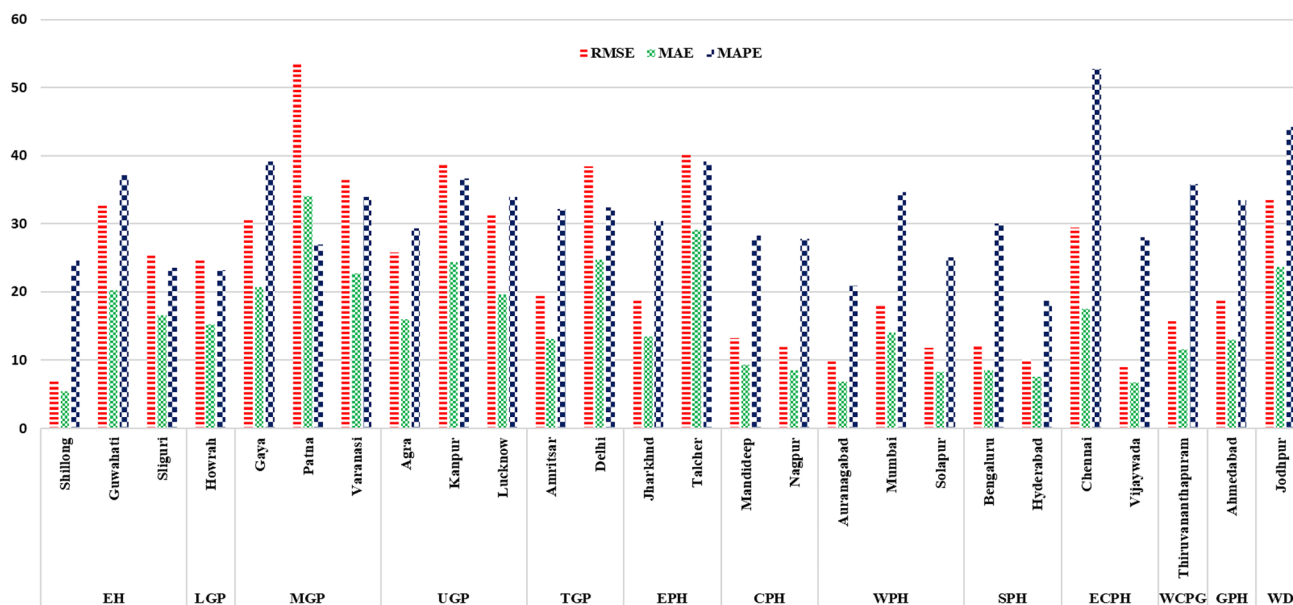


Fig. 6 Model performance results for 8-h cumulative forecasting at different cities in various agroclimatic zones of India

Table 4 Model performance agroclimatic zone wise

Zone	Error Metrics	1 h	2 h	3 h	4 h	5 h	6 h	7 h	8 h	Cumulative error
CPH	RMSE	8.4	10.7	12.1	12.9	13.5	13.8	14.0	14.3	12.6
	MAE	5.8	7.6	8.6	9.3	9.7	9.9	10.0	10.2	8.9
	MAPE	18.0	23.5	26.8	29.1	30.6	31.5	32.0	33.0	28.1
ECPH	RMSE	15.0	18.0	19.3	19.8	20.0	20.2	20.4	20.5	19.2
	MAE	8.9	11.1	12.1	12.5	12.7	12.9	13.2	13.4	12.1
	MAPE	28.4	36.4	39.9	41.4	42.2	43.3	44.9	46.4	40.4
EH	RMSE	12.9	17.7	20.6	22.3	23.3	24.2	24.7	25.3	21.8
	MAE	8.3	11.3	13.4	14.7	15.4	16.1	16.5	17.1	14.1
	MAPE	17.9	22.8	26.3	29.1	30.8	32.1	33.3	35.0	28.4
EPH	RMSE	20.4	27.1	29.9	31.2	31.4	31.4	31.4	31.9	29.6
	MAE	14.0	19.3	21.7	22.6	22.8	23.1	23.2	23.7	21.3
	MAPE	21.0	30.1	34.5	36.6	37.8	38.5	39.0	40.4	34.7
GPH	RMSE	10.6	15.4	18.3	19.8	20.6	21.4	21.8	22.1	19.1
	MAE	7.3	10.5	12.5	13.6	14.4	14.9	15.3	15.6	13.0
	MAPE	18.7	26.0	31.2	34.7	37.2	38.9	40.0	41.5	33.5
LGP	RMSE	9.6	15.4	20.2	23.9	26.8	29.2	31.3	33.0	24.8
	MAE	6.4	9.8	12.9	15.4	17.3	18.9	20.3	21.6	15.3
	MAPE	10.8	15.2	19.3	22.7	25.6	28.2	30.6	33.0	23.2
MGP	RMSE	30.3	35.5	38.8	40.7	42.3	43.4	44.4	45.5	40.4
	MAE	16.3	21.5	24.7	26.7	28.1	29.1	29.8	30.6	25.8
	MAPE	20.5	26.9	31.3	34.3	36.3	37.8	39.0	40.6	33.3
SPH	RMSE	7.6	10.1	11.1	11.5	11.8	11.9	12.0	12.2	11.1
	MAE	5.4	7.2	8.0	8.4	8.7	8.8	8.9	9.0	8.0
	MAPE	15.8	21.3	23.8	25.4	26.3	26.9	27.3	27.9	24.3
TGP	RMSE	14.4	21.6	26.5	29.7	31.8	33.1	34.1	35.2	29.1
	MAE	9.3	14.2	17.7	19.9	21.4	22.3	23.0	23.8	19.0
	MAPE	15.9	23.6	29.8	34.1	36.5	37.9	39.2	41.3	32.3
UGP	RMSE	19.8	26.8	30.6	32.9	34.4	35.3	36.0	37.0	32.1
	MAE	11.5	16.3	19.1	20.9	22.0	22.7	23.3	24.1	20.0
	MAPE	19.3	26.0	31.0	34.4	36.7	38.2	39.6	41.5	33.3
WCPG	RMSE	12.4	15.3	16.0	16.3	16.8	16.9	17.1	17.6	16.1
	MAE	8.6	10.9	11.6	11.8	12.2	12.5	12.6	12.9	11.6
	MAPE	27.2	33.6	35.3	35.2	36.8	37.8	39.5	41.2	35.8
WD	RMSE	21.6	28.6	32.1	33.9	35.7	36.9	37.6	38.0	33.5
	MAE	14.9	20.2	23.0	24.4	25.5	26.6	27.2	27.6	23.7
	MAPE	26.1	36.4	42.8	46.4	48.5	50.2	51.5	52.4	44.3
WPH	RMSE	9.4	12.4	13.8	14.3	14.2	13.9	13.7	13.8	13.3
	MAE	6.7	8.9	10.0	10.5	10.6	10.4	10.3	10.3	9.7
	MAPE	18.1	24.0	27.1	28.8	29.3	29.0	29.0	29.9	26.9

selected pollution hotspots in India by many researchers [45, 46]. The results obtained at different locations and the corresponding observed values were further subjected to statistical distribution analysis. The best fit distribution was found to be the same as the original training data at each location in India. (The statistical distribution analysis plots for test results are not displayed here as it is same as that of the observed training dataset).

4.3 Effects of Data Length and SNR on Model Performance

The variability in the forecasting results across the different agroecosystems of India prompts us to examine the effect of data length and the nature of deterministic signal and random components present in the PM_{2.5} time series using correlation analysis and signal-to-noise ratio (SNR)

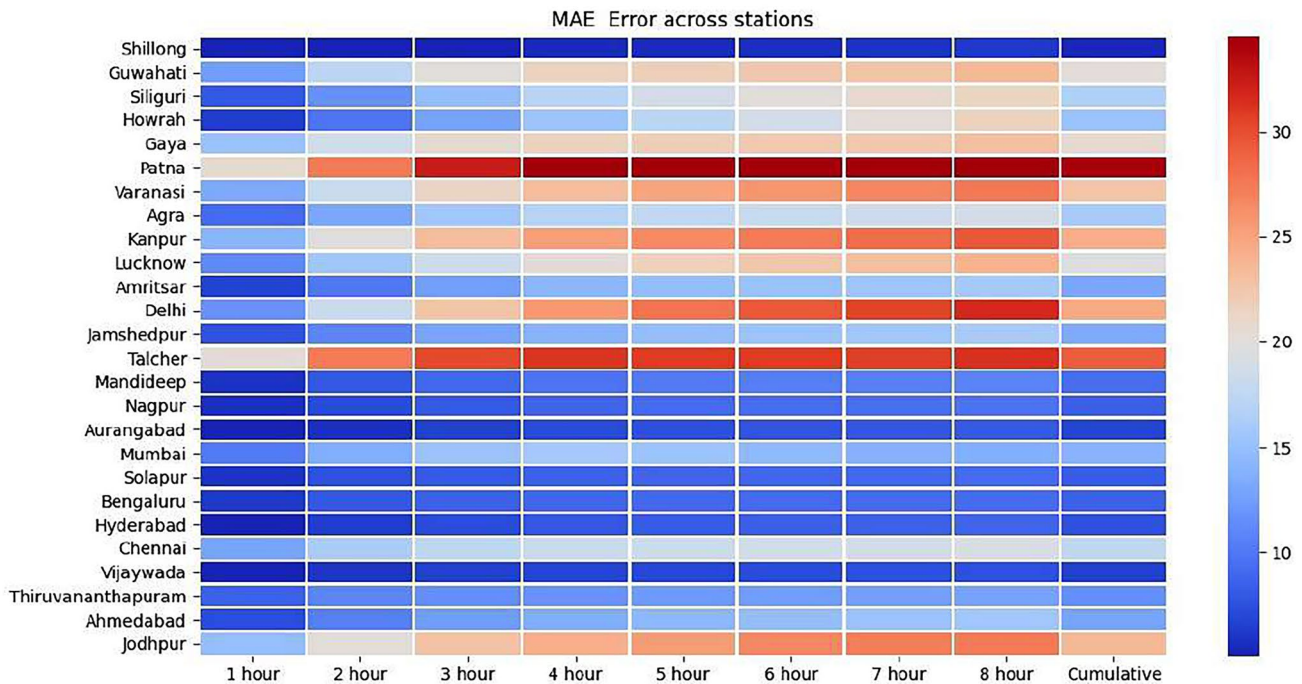


Fig. 7 Heat map of model performance errors in terms of MAE values

measurements. SNR quantifies the fraction of desired or good information with respect to unwanted or false information in each data series. In the present study, SNR was calculated [47] (Table 5) for each pre-processed dataset through the following equation:

$$SNR = \frac{\mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of the time series data. Such equations are used in situations where all values are non-negative. The scatter plots and trend line between the data length and model performance error (Supplementary Fig. F2) does not show any relationship between them indicating minimal or no effect of data length on model performance. A scatter plot (Fig. 8) of MAE vs. SNR for 1-h ahead and 8-h cumulative forecasting reveals a sharply decreasing trend. It is evident that as the noise component reduces, the model error also declines significantly for both the forecasting horizons. It is pertinent to mention that when SNR is greater than ~ 1.5 , error variance reduces significantly, i.e. model performance improves significantly. The variability in the results across the sites in India may be attributed to the level of noise present in the data series. In northern India, the pollution sources vary significantly because of large population density and traffic loads, thereby increasing the relative variance in the data set. The Indo Gangetic plain (IGP) region is known for large-scale farming and agricultural waste burning. It is to be noted that westerly winds are dominant in this region throughout the year except for monsoon season when

easterlies bring monsoon rains. Westerlies wind-driven dust storms and agricultural burning bring large uncertainty in the dust load over the area. The poor performances for multistep ahead forecast of model in this region may be attributed to the weather-induced uncertainties.

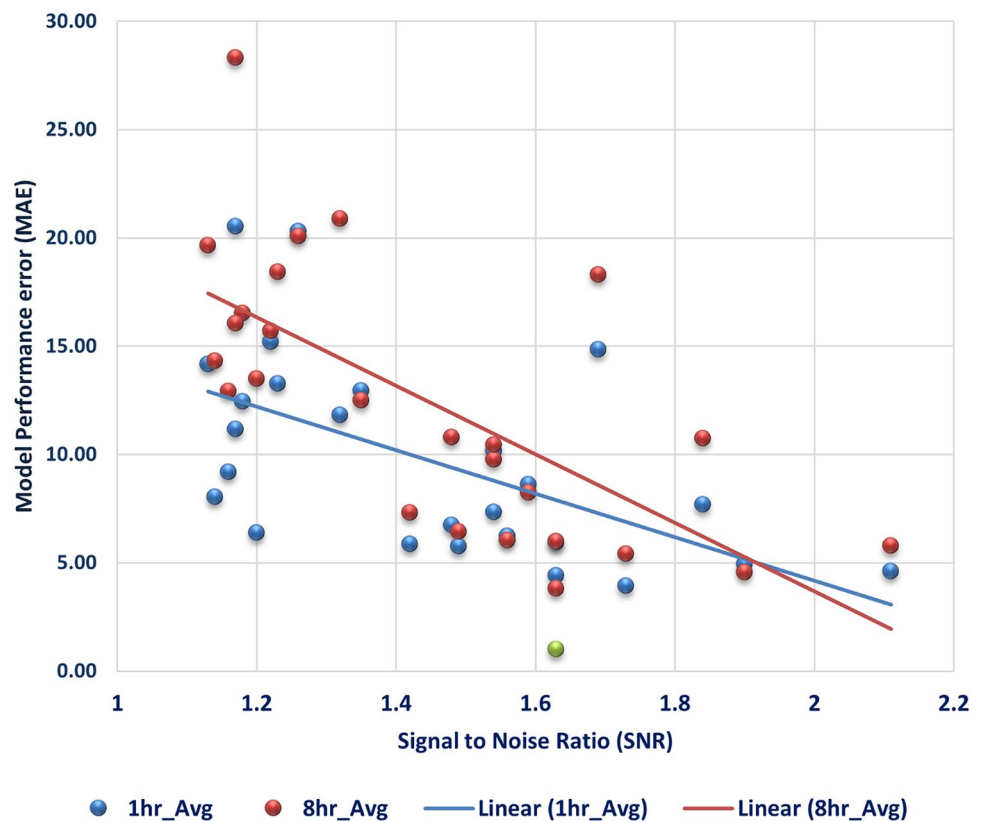
4.4 Comparison with Other Studies

The comparative analysis of the results reveals a significant improvement in the model errors in terms of RMSE when compared with the multiple output in a sequence. [48] applied multi-output auto encoder model for forecasting $PM_{2.5}$ and PM_{10} concentrations at Beijing city and obtained the best RMSE value of 39, although the applied model has used multiple inputs such as meteorological variables in addition to the time series data of the pollutant concentrations. [49] have applied an ANN model to achieve an error of 0.0191 in terms of MSE with a correlation coefficient of 0.7301. However, the model prediction horizon is of single-step only, and the model viability for multistep ahead prediction horizon was not examined. Similarly, [46] evaluated a simple feed-forward artificial neural network model for Kolkata region in eastern India using multivariate input parameters to predict single-step $PM_{2.5}$ concentrations during the COVID-induced lockdown period and reported the RMSE value of 3.74 and MAE value of 1.14. Similarly, [50] has tested 8 different models including Stacked LSTM, LSTM-autoencoder, BiLSTM and Conv2DLSTM models on different air

Table 5 Signal-to-noise ratio

Sl. no.	Agroclimatic zone	Station name	SNR
1	Eastern Himalayan Region	Shillong	1.633768849
		Guwahati	1.17804288
		Siliguri	1.203006861
2	Lower Gangetic Plain Region	Howrah	1.1400944798
3	Middle Gangetic Plain Region	Gaya	1.216926641
		Patna	1.170136375
		Varanasi	1.228970146
4	Upper Gangetic Plains Region	Agra	1.160642418
		Kanpur	1.13443877
		Lucknow	1.167957241
5	Trans-Ganga Plains Region	Amritsar	1.479871748
		Delhi	1.320374418
6	Eastern Plateau and Hills	Jamshedpur	1.83514448
		Talcher	1.259120932
7	Central Plateau and Hills	Mandideep	1.415810484
		Nagpur	1.492675234
8	Western Plateau and Hills	Aurangabad	1.734633454
		Mumbai	1.539312566
		Solapur	1.629221215
9	Southern Plateau and Hills	Bengaluru	1.556344153
		Hyderabad	2.114678224
10	Eastern Coastal Plains and Hills	Chennai	1.346752821
		Vijayawada	1.899935939
11	Western Coastal Plains and Ghats	Thiruvananthapuram	1.588071357
12	Gujarat Plains and Hills	Ahmedabad	1.542437308
13	Western Dry Region	Jodhpur	1.688448862

Fig. 8 Scatter plot of MAE vs SNR



pollutants in Kolkata and observed RMSE and MAE values more than $10 \mu\text{g}/\text{m}^3$. Similarly, [51] has achieved an MAE value of ~ 15 in case of $\text{PM}_{2.5}$ forecasting in Delhi. Furthermore, [48] reported RMSE values of 31, 56 and 68 for 3-h, 5-h and 9-h ahead prediction using ANN model for Talcher station in India. Using LSTM and BiLSTM, they have reported RMSE values of 26, 41, 80 and 42 and 155 and 168, respectively in comparison to the RMSE values of 29.04 and 40.41 for 1-h and 8-h ahead prediction horizon. In the present study, the proposed model has achieved RMSE values ranging from 7.09 to 53.81 across different data centres spread over 13 different agroclimatic zones in India. Out of the total 26 locations, 18 locations have RMSE values less than 30 in India. The results indicate the robustness of the model to be applicable to different locations in India without alterations.

5 Conclusion

Air pollution data mostly contained seasonal trends, multiple periodicities and stochastic components. To address multiple complexities, present in the air pollution data, a hybrid deep learning model was formulated by integrating Convolutional LSTM, 3D Convolutional Neural Network and Bidirectional LSTM network and examined its forecasting efficiency across India on a univariate $\text{PM}_{2.5}$ time series data. There is universality in the $\text{PM}_{2.5}$ data series across India as all of these data rejected the null hypothesis of normal distribution. They largely follow either Gamma or Beta distribution apart from Patna in MGP and Shillong in EH region that follow log normal distribution. The results obtained for 8-h ahead sequential prediction reveal significant variations across the region with minimum (7.09) and maximum (53.81) RMSE values obtained at Shillong in EH and Patna in MGP, respectively. Similar results (minimum: -5.41 and maximum: -34.09) were also found in terms of MAE values at Shillong and Patna respectively. In terms of MAPE, minimum and maximum values were observed to be 18.6 and 52.7% at Hyderabad and Chennai, respectively. The robustness of model performance was evident from the little variations observed in the model error estimation for 1-h ahead and 8-h sequential forecasts. The results (MAE) were further analysed against SNR and found strong association between level of error and SNR values. As SNR decreases, model performance decreases (MAE values increases). The variations in the SNR may be attributed to anthropogenic activities in the region. The results reveal weak performance in and around IGP in comparison to the rest of India. The model has the potential to be utilized for policy and planning for pollution control. It could be a useful tool for forewarning about lurking air pollution events.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10666-023-09902-4>.

Acknowledgements The authors would like to acknowledge the Tezpur University for providing the necessary infrastructure for carrying out this study. The database maintained and provided by Central Pollution Control Board (CPCB) is also duly acknowledged. We also acknowledge the following state level pollution control boards (PCB) of India—Uttar Pradesh PCB, Gujrat PCB, Punjab PCB, Maharashtra PCB, Bihar State PCB, Assam PCB, West Bengal PCB, Telengana State PCB, Jharkhand State PCB, Rajasthan State PCB, Madhya Pradesh PCB, Meghalaya PCB, Odisha State PCB, Kerala PCB and Andhra Pradesh PCB. We are pleased to acknowledge the Open-Source technologies such as Linux, Python, TensorFlow, Keras, etc., which are used for postprocessing model outputs.

Author Contribution Mr. Pranjol Goswami. The present work is part of the Ph.D. research work for his thesis. He has developed the Deep Learning Network and carried out major part of data analysis and preparation of first draft of the paper. Mr. Manoj Prakash: He has contributed to model testing and refinement and further analysed the network design and efficiency. Dr. Rakesh Kumar Ranjan: He has contributed in GIS related work in data presentation and statistical analysis. Dr. Amit Prakash: the lead investigator of the present work; conceptualized the research work and provided the overall supervision. All the authors have contributed to the data analysis and manuscript preparation.

Availability of Data and Materials The data used in this study were procured from Central Pollution Control Board (CPCB), Government of India (<http://www.cpcb.nic.in/>). These are the public repositories maintained by the government of India.

Declarations

Ethical Approval Not applicable.

Competing Interests The authors declare no competing interests.

References

- Li, X., Zhang, X., Zhang, Z., Han, L., Gong, D., Li, J., Wang, T., Wang, Y., Gao, S., Duan, H., & Kong, F. (2019). (D. J. Schroeder (1999). *Astronomical optics* (2nd ed.). Academic Press. p. 278. ISBN 978-0-12-629810-9., p.278). Air pollution exposure and immunological and systemic inflammatory alterations among schoolchildren in China. *Science of The Total Environment*, 657, 1304–1310. <https://doi.org/10.1016/j.scitotenv.2018.12.153>
- Chen, Z., Cui, L., Cui, X., Li, X., Yu, K., Yue, K., Dai, Z., Zhou, J., Jia, G., & Zhang, J. (2019). The association between high ambient air pollution exposure and respiratory health of young children: a cross sectional study in Jinan, China. *Science of the Total Environment*, 656, 740–749. <https://doi.org/10.1016/j.scitotenv.2018.11.368>
- Organization, W. H. (n.d.). Ambient air pollution: a global assessment of exposure and burden of disease. World Health Organization. <https://apps.who.int/iris/handle/10665/250141>
- Coats C. J., Jr. (1996). High-performance algorithms in the Sparse Matrix Operator Kernel Emissions (SMOKE) modeling system. In *Proceedings of Ninth AMS Joint Conference on Applications of Air Pollution Meteorology with A&WMA*. American Meteor Society, GA (pp. 584-588). <https://www.osti.gov/biblio/422986>
- Olatinwo, R. O., Prabha, T., Paz, J. O., Riley, D. G., & Hoogenboom, G. (2010). The weather research and forecasting (WRF) model:

- Application in prediction of TSWV-vectors populations. *Journal of Applied Entomology*, 135(1–2), 81–90. <https://doi.org/10.1111/j.1439-0418.2010.01539.x>
6. Vautard, R., Builtjes, P. J. H., Thunis, P., Cuvelier, C., Bedogni, M., Bessagnet, B., Honore, C., Moussiopoulos, N., Pirovano, G., Schaap, M., Stern, R., Tarrason, L., & Wind, P. (2007). Evaluation and intercomparison of Ozone and PM10 simulations by several chemistry transport models over four European cities within the CityDelta project. *Atmospheric Environment*, 41, 173–188. <https://doi.org/10.1016/j.atmosenv.2006.07.039>
 7. Stern, R., Builtjes, P. J. H., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., Memmesheimer, M., Feldmann, H., Renner, E., Wolke, R., Kerschbaumer, A., Liu, B. C., Binaykia, A., Chang, P. C., Tiwari, M. K., Tsao, C. C., Srivastava, N., Mansimov, E., Salakhutdinov, R., ... Bui, T. (2017). A model inter-comparison study focussing on episodes with elevated PM10 concentrations. *Atmospheric Environment*, 42(19), 4567–4588. <https://doi.org/10.1016/j.neucom.2018.06.049>
 8. Saide, P. E., Carmichael, G. R., Spak, S. N., Gallardo, L., Osses, A. E., Mena-Carrasco, M. A., & Pagowski, M. (2011). Forecasting urban PM10 and PM2.5 pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model. *Atmospheric Environment*, 45(16), 2769–2780. <https://doi.org/10.1016/j.atmosenv.2011.02.001>
 9. Goyal, P., Chan, A. T., & Jaiswal, N. (2006). Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmospheric Environment*, 40(11), 2068–2077. <https://doi.org/10.1016/j.atmosenv.2005.11.041>
 10. Antanasijević, D. Z., Pocajt, V. V., Povrenović, D. S., Ristić, M. Đ., & Perić-Grujić, A. A. (2013). PM10 emission forecasting using artificial neural networks and genetic algorithm input variable optimization. *Science of the Total Environment*, 443, 511–519. <https://doi.org/10.1016/j.scitotenv.2012.10.110>
 11. Mishra, D., & Goyal, P. (2016). Neuro-fuzzy approach to forecast NO2 pollutants addressed to air quality dispersion model over Delhi. *India. Aerosol and Air Quality Research*, 16(1), 166–174. <https://doi.org/10.4209/aaqr.2015.04.0249>
 12. Paschalidou, A., Karakitsios, S., Kleanthous, S., & Kassomenos, P. (2011). Forecasting hourly PM10 concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environmental Science and Pollution Research International*, 18, 316–327. <https://doi.org/10.1007/s11356-010-0375-2>
 13. Kolehmainen, M., Martikainen, H., & Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, 35, 815–825. [https://doi.org/10.1016/S1352-2310\(00\)00385-X](https://doi.org/10.1016/S1352-2310(00)00385-X)
 14. Kang, Z., Qu, Z., Kim, M. H., Kim, Y. S., Lim, J., Kim, J. T., Sung, S. W., & Yoo, C. (2017). Data-driven prediction model of indoor air quality in an underground space. *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCI/A)*, 27(6), 1675–1680. <https://doi.org/10.1007/s11814-010-0313-5>
 15. Feng, Y., Zhang, W., Sun, D., & Zhang, L. (2011). Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmospheric Environment*, 45(11), 1979–1985. <https://doi.org/10.1016/j.atmosenv.2011.01.022>
 16. Prakash, A., Kumar, U., Kumar, K., & Jain, V. (2011). A wavelet-based neural network model to predict ambient air pollutants' concentration. *Environmental Modeling & Assessment*, 16, 503–517. <https://doi.org/10.1007/s10666-011-9270-6>
 17. Díaz-Robles, L. A., Ortega, J. C., Fu, J. S., Reed, G. D., Chow, J. C., Watson, J. G., & Moncada-Herrera, J. A. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco. *Chile. Atmospheric Environment*, 42(35), 8331–8340. <https://doi.org/10.1016/j.atmosenv.2008.07.020>
 18. Chen, Y., Shi, R., Shu, S., & Gao, W. (2013). Ensemble and enhanced PM10 concentration forecast model based on stepwise regression and wavelet analysis. *Atmospheric Environment*, 74, 346–359. <https://doi.org/10.1016/j.atmosenv.2013.04.002>
 19. Alimissis, A., Philippopoulos, K., Tzanis, C. G., & Deligiorgi, D. (2018). Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmospheric Environment*, 191, 205–213. <https://doi.org/10.1016/j.atmosenv.2018.07.058>
 20. Yang, Z., & Wang, J. (2017). A new air quality monitoring and early warning system: air quality assessment and air pollutant concentration prediction. *Environmental Research*, 158, 105–117. <https://doi.org/10.1016/j.envres.2017.06.002>
 21. Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, 231(December), 997–1004. <https://doi.org/10.1016/j.envpol.2017.08.114>
 22. Reddy, V., Yedavalli, P., Mohanty, S., & Nakhat, U. (2017). Deep air: forecasting air pollution in Beijing, China. https://www.ischool.berkeley.edu/sites/default/files/sproject_attachments/deep-airforecasting_final.pdf
 23. Kök, İ., Şimşek, M. U., & Özdemir, S. (2017). A deep learning model for air quality prediction in smart cities. *IEEE International Conference on Big Data (Big Data)*, 2017, 1983–1990. <https://doi.org/10.1109/BigData.2017.8258144>
 24. Liu, B., Yan, S., Li, J., Qu, G., Li, Y., Lang, J., & Gu, R. (2019). A sequence-to-sequence air quality predictor based on the n-step recurrent prediction. *IEEE Access*, 7, 43331–43345. <https://doi.org/10.1109/ACCESS.2019.2908081>
 25. Soh, P. W., Chang, J. W., & Huang, J. W. (2018). Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *IEEE Access*, 6, 38186–38199. <https://doi.org/10.1109/ACCESS.2018.2849820>
 26. Qi, Y., Li, Q., Karimian, H., Liu, D., Gong, Y., Liu, L., Yang, M., Bourdev, L., Soh, P., Chang, J., Huang, J., Stojov, V., Koteli, N., & Lameski, P. (2019). A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. *Science of The Total Environment*, 664(2014), 1–10. <https://doi.org/10.1016/j.scitotenv.2019.01.333>
 27. Fan, J., Li, Q., Hou, J., Feng, X., Karimian, H., & Lin, S. (2013). A spatiotemporal prediction framework for air pollution based on deep RNN. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4(4W2), 15–22. <https://doi.org/10.5194/isprs-annals-IV-4-W2-15-2017>
 28. Zhang, C., Yan, J., Li, C., Rui, X., Liu, L., & Bie, R. (2016). On estimating air pollution from photos using convolutional neural network. *MM 2016 - Proceedings of the 2016 ACM Multimedia Conference*, 297–301. <https://doi.org/10.1145/2964284.2967230>
 29. Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 23(22), 22408–22417. <https://doi.org/10.1007/s11356-016-7812-9>
 30. Wang, J., & Song, G. (2018). A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, 314, 198–206. <https://doi.org/10.1016/j.neucom.2018.06.049>
 31. Bui, T., Le, V.-D., & Cha, S.-K. (2018). A deep learning approach for forecasting air pollution in South Korea using LSTM. <http://arxiv.org/abs/1804.07891>
 32. Zhao, X., Zhang, R., Wu, J. L., & Chang, P. C. (2018). A deep recurrent neural network for air quality classification. *Journal of Information Hiding and Multimedia Signal Processing*, 9(2), 346–354.
 33. Lee, S., & Shin, J. (2019). Hybrid model of convolutional LSTM and CNN to predict particulate matter. *International Journal of*

- Information and Electronics Engineering*, 9(1), 34–38. <https://doi.org/10.18178/ijee.2019.9.1.701>
34. Pak, U., Ma, J., Ryu, U., Ryom, K., Juhyok, U., Pak, K., & Pak, C. (2020). Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: a case study of Beijing, China. *Science of The Total Environment*, 699, 133561. <https://doi.org/10.1016/j.scitotenv.2019.07.367>.
 35. Ma, J., Cheng, J. C. P. P., Lin, C., Tan, Y., & Zhang, J. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, 214(July), 116885.
 36. Bhatla, R., Sarkar, D., Verma, S., Sinha, P., Ghosh, S., & Mall, R. K. (2020). Regional climate model performance and application of bias corrections in simulating summer monsoon maximum temperature for agro-climatic zones in India. *Theoretical and Applied Climatology*, 142(3), 1595–1612. <https://doi.org/10.1007/s00704-020-03393-z>
 37. Gnauck, A. (2004). Interpolation and approximation of water quality time series and process identification. *Analytical and Bio-analytical Chemistry*, 380(3), 484–492. <https://doi.org/10.1007/s00216-004-2799-3>
 38. Cho, K., Merriënboer, van, Gülçehre, Ç. B., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *BT - Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*, October 25–29, 2014, Doha, Qatar, A meeting of SIGDAT, (pp. 1724–1734). <https://doi.org/10.3115/v1/d14-1179>
 39. Wu, Z., Rincon, D., Luo, J., & Christofides, P. D. (2021). Machine learning modeling and predictive control of nonlinear processes using noisy data. *AIChE Journal*, 67(4), e17164. <https://doi.org/10.1002/aic.17164>
 40. Zhang, B., Zhang, H., Zhao, G., & Lian, J. (2020). Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environmental Modelling & Software*, 124, 104600. <https://doi.org/10.1016/j.envsoft.2019.104600>
 41. Essien, A., & Giannetti, C. (2020). A deep learning model for smart manufacturing using convolutional LSTM neural network autoencoders. *IEEE Transactions on Industrial Informatics*, 16(9), 6069–6078. <https://doi.org/10.1109/TII.2020.2967556>
 42. Rolling analysis of time series. In: Zivot, E., Wang, J. (Eds.), *Modeling Financial Time Series with S-PLUS®*. Springer New York, New York, NY, pp. 313–360. https://doi.org/10.1007/978-0-387-32348-0_9
 43. Ramachandran, P., Zoph, B., & Le, Q. V. (2018). Searching for activation functions. 6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings. <https://arxiv.org/pdf/1710.05941.pdf>
 44. Kingma, D. P., & Ba, J. L. (2015). Adam: a method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.
 45. Chelani, A., & Gautam, S. (2021). Lockdown during COVID-19 pandemic: a case study from Indian cities shows insignificant effects on persistent property of urban air quality. *Geoscience Frontiers*, 101284. <https://doi.org/10.1016/j.gsf.2021.101284>
 46. Bera, B., Bhattacharjee, S., Shit, P. K., Sengupta, N., & Saha, S. (2021). Significant impacts of COVID-19 lockdown on urban air pollution in Kolkata (India) and amelioration of environmental health. *Environment, Development and Sustainability*, 23(5), 6913–6940. <https://doi.org/10.1007/s10668-020-00898-5>
 47. Schroeder, D. J. (1999). *Astronomical optics* (2nd ed.). Academic Press. p. 278. ISBN 978-0-12-629810-9, p.278.
 48. Samal, K. K. R., Panda, A. K., Babu, K. S., & Das, S. K. (2021). Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites. *Urban Climate*, 39, 100943. <https://doi.org/10.1016/j.uclim.2021.100943>
 49. Masood, A., & Ahmad, K. (2020). A model for particulate matter (PM2.5) prediction for Delhi based on machine learning approaches. *Procedia Computer Science*, 167, 2101–2110. <https://doi.org/10.1016/j.procs.2020.03.258>
 50. Middy, A. I., & Roy, S. (2022). Pollutant specific optimal deep learning and statistical model building for air quality forecasting. *Environmental Pollution*, 301, 118972. <https://doi.org/10.1016/j.envpol.2022.118972>
 51. Kumar, S., Mishra, S., & Singh, S. K. (2020). A machine learning-based model to estimate PM2.5 concentration levels in Delhi's atmosphere. *Heliyon*, 6(11), e05618. <https://doi.org/10.1016/j.heliyon.2020.e05618>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.