Proceedings of the 9ICEG
9th International Congress on Environmental Geotechnics
25-28 June, 2023 | Chania, Greece
© Authors: All rights reserved, 2023
https://doi.org/10.53243/ICEG2023-422

A Study on the Estimation of Surface Ozone Pollution in the Indian Megacity, Delhi at pre-, during- and post-COVID Years using Statistical and Machine Learning Models

Subhojit Mandal[1], and Mainak Thakur[2]

[1]*Indian Institute of Information Technology, Sricity, India, email: subhojit.m@iiits.in*

[2]*Indian Institute of Information Technology, Sricity, India, email: mainak.t@iiits.in*

ABSTRACT

The characteristics of a trace gas: surface ozone (O_3) concentrations for the Indian city of Delhi are presented from January 2018 to June 2022 in this study. The comparative analysis reveals the ozone concentration of Delhi follows a decreasing trendline during 2021 and 2022 compared to previous years. Locally, low-income areas of Sonia Vihar, densely populated areas like Nehru Nagar, Sirifort, Sri Aurobindo Marg, and green areas like Dr. Karni Singh Shooting Range are identified as high ozone concentration hotspots (average 40-48 $\mu g/m^3$). The O_3 distribution all over Delhi in 2020 and 2021 is concentrated between 30-35 $\mu g/m^3$ and 25-30 $\mu g/m^3$, respectively. Delhi faces high ozone pollution during summer (March-June). While the lowest O_3 concentration can be observed during January month, the month of May faces the highest O_3 pollution events in Delhi for each of the years except 2019, when the highest monthly mean ozone concentration was observed during June. Due to significant traffic emissions, high temperatures, and favorable weather conditions, O_3 concentrations in the megacity are on the higher side during the pre-noon and noon while they are on the lower side during the early morning (3-5 AM). In this study, ARIMA and Long Short Term Memory (LSTM) models are further developed and employed to forecast daily ozone concentration for the next year of 2023.

Keywords: Ozone pollution, LSTM modeling, Time series modeling, pollutant forecasting

1 INTRODUCTION

The degradation of urban air quality in India over the last decade has posed a significant barrier to the long-term development of the Indian ecosystem. Severe weather condition is most prevalent in metropolitan areas. Increasing early death of its citizens due to diseases related to respiratory problems, economic loss due to climate change-related problems, changes in local atmospheric conditions, and other disorders have plagued Delhi, India's capital city, in recent years. These afflictions are directly related to Delhi's rising air pollution. Surface O_3 in the lower troposphere is one of the air pollutants that contribute significantly to the atmosphere's ability to oxidize and has a great potential to alter the local atmosphere (Sharma et al. 2016 [14]). Mills et al. [5] revealed that long-term ozone levels above 40 parts per billion (ppb) could cause crop and ecological losses. Notably, according to Levi et al. 2001 [8], human mortality changes by 0.3-0.9% per 10 μg/m^3 of 24-hr average O_3 concentrations. The formation of O_3 takes place in the presence of sunlight and volatile organic compounds (VOCs), NOx, and Carbon monoxide (Zhang et al. 2004 [13]). Chen et al. 2021 [12] found that a 50% reduction in NOx emissions results in a 10-50% increase in surface ozone, whereas a 50% reduction in VOC emissions results in a ~60% reduction in O_3 concentrations. Thus, being a secondary pollutant, the emissions of VOCs, NOx, and CO control the budget of atmospheric O_3. The exchange process between stratospheric and tropospheric O_3 is also prevalent under certain meteorological conditions. Additionally, the tropospheric O_3 distribution is controlled by air chemistry, turbulent diffusion, and deposition on the earth's surface, and it varies significantly with pollutant transport and photochemistry (Seinfield 1986 [6]).

The emission of Volatile organic compounds, Carbon Monoxide (CO), Chloromethane (CH_3Cl), Hydrogen (H_2), and Nitrogen oxides (NO_x) from broad-scale biomass burning, traffic congestions, and stubble burning causes the indirect formation of surface O_3. Notably, Indian farmers from Punjab and Haryana burn their crop residuals during the post-monsoon season. The intricate combination of

precursor emissions, chemistry, depositional processes, and meteorological constraints on dispersion determines the spatiotemporal patterns of O_3 concentrations.

In this regard, a proper examination of the tropospheric ozone's patterns, temporal distribution, and change are highly helpful for comprehending the urban environment's sustainability perspective and associated chemical processes. From January 2018 to December 2022, the O_3 measurements from 40 Central Pollution Control Board (CPCB, 2014 [17]) ground monitoring stations dispersed around Delhi city are examined in this study. Furthermore, we consider analyzing the temporal properties of the spatially and 1-Day temporally averaged time-series O_3 concentration data to discover the effect of the Nationwide COVID-19 lockdown and un-lockdown on the O_3 concentration of Delhi. Moreover, the spatial distribution of the O_3 concentration has been analyzed, and the sites of high O_3 concentrations have been identified.

Time series forecasting is a challenging task for the modelers to understand the future behavior of the time series data based on historical patterns of the observed data. The observed time series of the spatially averaged O_3 concentration for Delhi can serve as the starting point for investigating the underlying causes of the complex characteristics of the concerned time series. Moreover, which sudden occurrences had an impact on the seasonal or cyclical changes in the concentration of the concerned pollutant in the atmosphere is a highly demanding question. In this connection, both statistical and machine learning models have been employed by researchers for short-term or multi-timestamp time series forecasting. This paper explored the utility of Auto-Regressive Integrated Moving Average (ARIMA) (Newbold. 1983 [15]) as a statistical modeling technique and Long-Short-Term Memory (LSTM) as a Recurrent Neural Network (RNN) model (developed by [1]) in forecasting the O_3 concentrations. Many researchers found that LSTM models can preserve long-term dependency in the model forecasts. Zhang et al. 2021 [2], Mbatha [3], and Cheng et al. 2021 [4] used the LSTM model to forecast the O_3 concentration time series for future timestamps. In this work, the O_3 concentration values for the year 2023 are forecasted using the historical data from 2018-2022. The average O_3 concentration for the year 2023 is predicted to be 32 and 29.6 ($\mu g/m^3$) by ARIMA and LSTM models, respectively.

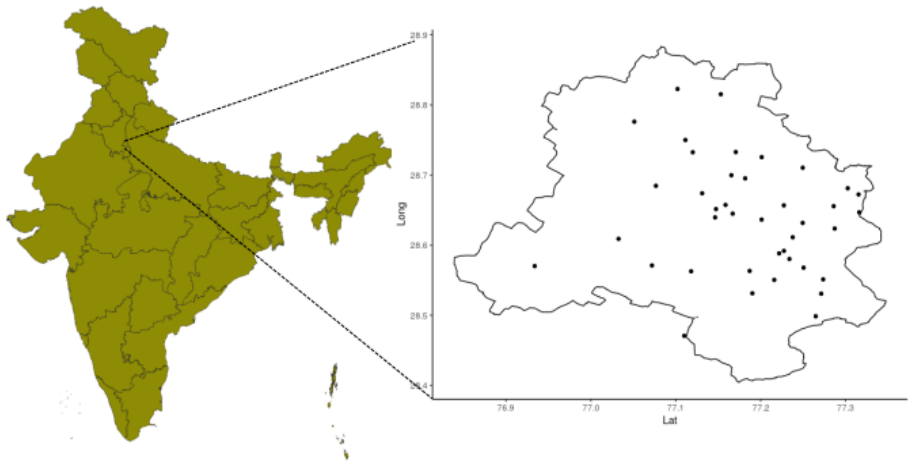


Figure 1. *Monitoring stations in Delhi*

2.1 Methodology and Study Area:

The Indian National Capital city of Delhi covers an area of 1484 square kilometres within the central northern part of the Indo-Gangetic plain bordering the state of Punjab and Haryana. It is one of the most densely populated cities in the country. The Central Pollution Control Board installed 40 ground monitoring stations to capture ambient atmospheric pollutant concentrations. They continue to provide pollutant observations in a span of 15 minutes. The locations of these monitoring stations are shown in Fig. 1. This study considers the O_3 concentration data for the duration of January 2018 to December 2022. The O_3 concentration data over the 40 monitoring stations were spatially averaged. In 2020, the government of India called for a strict nationwide lockdown (Lockdown-1) to tackle COVID-19 spread all over India starting from 25th March. Following that, Lockdown-2 (15th April 2020- 3rd May 2020), Lockdown-3 (4th May 2020- 17th May 2020), and Lockdown-4 (19th May 2020- 31st May 2020) were implemented all over the country.

We analyze the year-wise statistical summary and related exploratory data analysis of the O_3 concentration time series data. Following this, the post-COVID lockdown years' O_3 concentration data was investigated. Furthermore, we checked the stationarity conditions of the year-wise O_3 concentration data using the Augmented Dicky Fuller (ADF) test. The nature of the temporal autocorrelation function (ACF) and Partial autocorrelation function (PACF) was explored to check the properties of the O_3 time series data. Afterward, the ARIMA model was developed and used to forecast the O_3 concentration for the next year (2023). On top of that, a weekly ARIMA model was established for weekly averaged O_3 concentration data to explore the ability of ARIMA model to capture the periodicity present in the time-series data. Subsequently, a Long Short-Term Memory (LSTM) network was utilized for the next seven-day O_3 concentration forecasting in a multi-step fashion using daily historical O_3 patterns. Further, the LSTM model is used to forecast the next year's O_3 concentration. Notably, the utility of both the models ARIMA and LSTM were checked using the train-test validation method. In this regard, 80% of the total data was used for training, and the remaining was kept for testing.

2.2 Data Analysis

Table 1. *Statistical summary of the spatially and daily averaged O_3 concentration data*

	2018	**2019**	**2020**	**2021**	**2022**
Mean ($\mu g/m^3$)	36.14	33.23	33.45	29.71	30.52
Std ($\mu g/m^3$)	11.69	12.26	11.83	8.89	8.85
Min ($\mu g/m^3$)	13.88	10.65	11.6	13.6	10.56
25% ($\mu g/m^3$)	27.71	24.37	24.9	23.2	24.17
50% ($\mu g/m^3$)	34.14	29.27	31.6	28.37	29.07
75% ($\mu g/m^3$)	43.77	41.72	40.28	34.62	34.71
Max ($\mu g/m^3$)	66.4	71.46	76.34	60.74	57.3

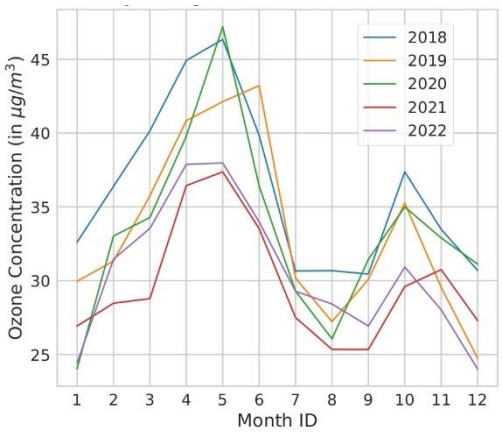


Figure 2. *Monthly averaged ozone concentration (Year-wise separated)*

In the years 2018, 2019, 2020, 2021, and 2022, the daily average O_3 concentration was 36.14, 33.23, 33.45, 29.7, and 30.52 μg/m^3, respectively. The year-wise statistical summary of the data is shown in Table 1. The Std indicates the standard deviation of the data; 25%, 50%, and 75% represent the 25 percentile, 50 percentile, and 75 percentile of the data, respectively. Each year two major high O_3 concentration peaks can be observed for Delhi (Fig. 2). The one peak is observed during the summer and another during post-monsoon season. The 14-day moving average data of the spatially averaged O_3 concentration time-series for Delhi is shown in Fig. 3. A decrement in the O_3 concentration peak heights can be observed after the lockdown and unlock phases. From the third week of March to the last week of

June, during the year 2018, 2019 the weekly mean O_3 concentration surpasses the threshold of 40 µg/m^3 (the summer peak).

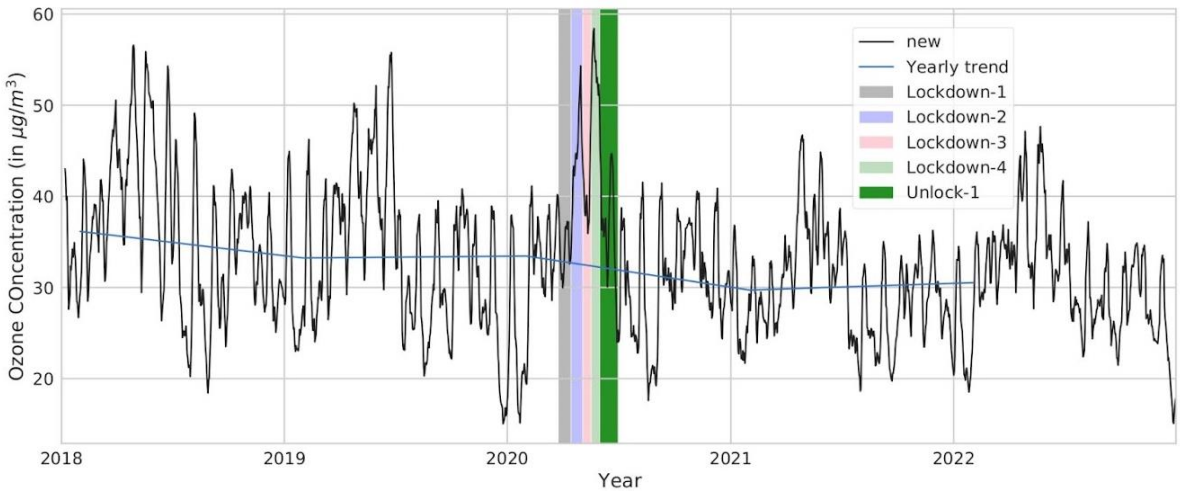


Figure 3. *14-day rolling average of Ozone concentration during 2018-2022*

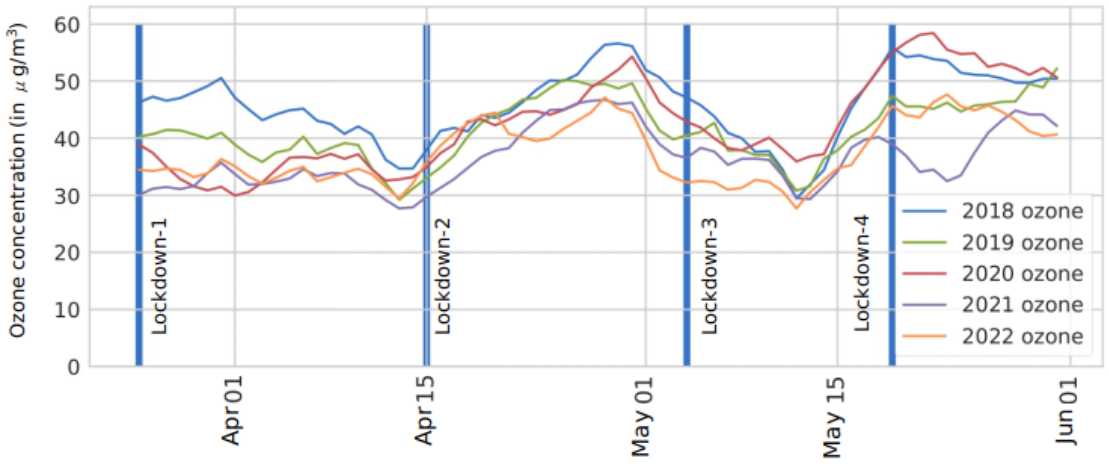


Figure 4*. 7-day rolling mean of ozone concentration during lockdown period of 2020 and comparison with other years (year-wise separated)*

Notably, Mills et al. [5] revealed 40 µg/m^3 as a potentially dangerous threshold for human sustainability. During 2020 and 2021, the weekly mean crossed 40 µg/m^3 around the first week of April and the last week of April, respectively. Again, during the month of October 2018 and 2019, the weekly mean crossed 40 µg/m^3 three times, whereas, on the 2nd week of October 2020, the weekly mean crossed 40 µg/m^3. In contrast, for the year 2021 and 2022, the weekly mean has not crossed the threshold of 40 µg/m^3 for any of the weeks. The mean O_3 concentration data across Delhi throughout the 21-day lockdown-1 period, from March 25th to April 14th, 2020, was 33.71 µg/m^3, representing a 21% and 7% decrease over the 2018 and 2019 O_3 concentration, respectively. Following that, the same period in 2021 and 2022 reported O_3 concentrations of 32.31 and 33.64 µg/m^3 , respectively. Fig. 4 depicts the comparative year-wise 7-day rolling mean of the O_3 concentration during the total lockdown period of 25th March to 31st May 2020 and corresponding dates for the years 2018-2022. During the Lockdown-2 time period, the O_3 concentration did not change much during all the years. The average O_3 concentration during the time period of lockdown-1 to Lockdown-4 (25th March to 31st May) were 46.01, 41.76, 42.9, 36.82 and 37.78 µg/m^3 for the years 2018, 2019, 2020, 2021 and 2022 respectively. During the total time period of lockdown-1 to lockdown-4, the average O_3 concentration decreases in the succeeding years of COVID-19 (2020). The summertime (April-June) average O_3 concentrations have decreased by 12% and 13%, respectively, compared to 2020. These summer-time average O_3 concentration values for 2021 and 2022 have been noted as 35 and 36 µg/m^3, respectively. Overall, summer-time average O_3 concentration for the years 2018-2022; follows a decreasing trendline. The 2021 O_3 data, followed by that of 2022, have been relatively lower than the prior years' O_3 concentration

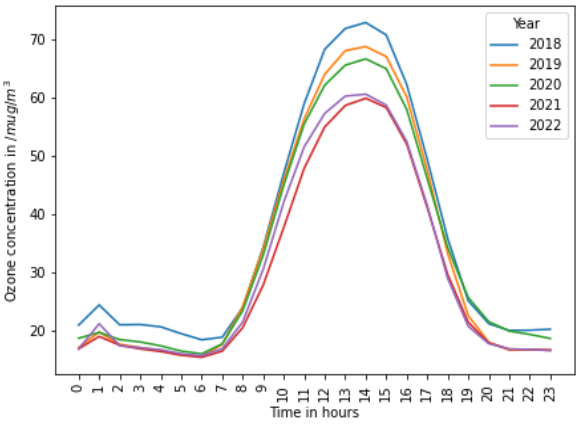


Figure 5. *Year-wise diurnal variation of O_3 concentration*

data. This may be due to a generalized after-effect of the lockdown. The diurnal variation of the O_3 concentration between the year 2018 to 2022 is shown in Fig. 5. It can be observed that between noon 12 o clock and afternoon 4 PM Delhi faces high O_3 concentrations. Following the lockdown year of 2020, O_3 concentration has reduced between Noon 1 and 3 PM for the years 2021 and 2022. Usually, VOC/NO_x gases exhausted in the morning hour get time to react in the presence of sunlight and pile up the O_3 production at noon. As a result, the decrease in the O_3 concentration for 2021 and 2022 can be linked to either the lower temperatures or lowered exhaust VOC/NOx gases. The average O_3 concentrations between 12-3 PM during 2018, 2019, 2020, 2021, and 2022 were 72.85, 69.18, 65.58, 63.44, and 56.75 µg/m³ respectively. This confirms the decreasing nature of the O_3 concentrations during noon-time throughout Delhi. The year-wise comparative boxplot has been shown in Fig. 6. It can be seen that the median value (specified by red color) is higher for 2018, among the considered dataset timespan. The highest monthly average O_3 concentration for the year 2019 was observed for the month of June. There is a dip in the median O_3 concentration during the year 2021 with respect to all other dataset timespans. Also, it can be seen that the variation in the O_3 concentration is less in the after-lockdown years of 2021 and 2022 compared to other years. This can also be confirmed by the standard deviation values in Table 1. The year-wise monthly-averaged O_3 concentration data is shown in Fig. 2. In the year 2021, a major decrease in O_3 concentration has been observed for all the months. The statistical properties of the year-wise daily O_3 concentration time series data have been investigated in this study. The Augmented Dicky Fuller (ADF) test (Cheung 1995 [16]) has been conducted for each year-wise daily O_3 concentration component (2018, 2019, 2020, 2021, and 2022) to check the non-stationarity of the individual year-wise time-series data. The null hypothesis (H_0) is considered as the data is non-stationary. If the ADF statistic p-value ≥ 0.05 (for 5% critical value), the null hypothesis is accepted, and the data is non-stationary. On the contrary, if the ADF statistic p-value is less than the critical value threshold, the null hypothesis is rejected. The ADF test results for the whole dataset as well as year-wise time series data have been shown in Table 2. By observing the p-value for each of the time series data, it can be observed that the null hypothesis is rejected at 1%, 5%, and 10% significance level for all the time series, and each of these time series are stationary.

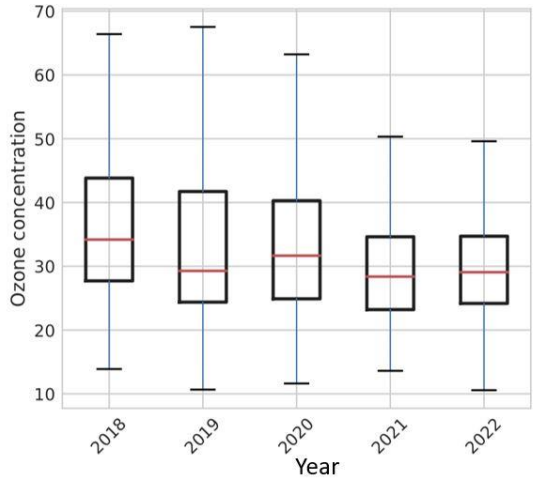


Figure 6. *Year-wise boxplot of O_3 concentration*

Table 2. Stationarity test statistics of the ozone time series data

		Whole dataset	2018	2019	2020	2021	2022
ADF statistic		-10.74	-8.32	-4.64	-4.92	-4.58	-3.037
	1%	-3.43	-3.34	-3.34	-3.45	-3.45	-3.45

Critical values of ADF (Confidence interval)	5%	-2.86	-2.87	-2.87	-2.87	-2.87	-2.87
	10%	-2.57	-2.571	-2.571	-2.57	-2.571	-2.57
p-value		0.00	0.026	.00010	0.00003	0.00013	0.031
H_0 (Null Hypothesis)		Reject	Reject	Reject	Reject	Reject	Reject
Time series		Stationary	Stationary	Stationary	Stationary	Stationary	Stationary

Notably, the National Ambient Air Quality Standards, India (NAAQS) suggested hourly monitored O_3 demarcation limit as 180 µg/m^3. The hourly O_3 concentration for 40 monitoring stations exceeded 180 µg/m^3 for 1416, 1424, 1180, 956, and 1218 times (counting at individual monitoring stations) during 2018, 2019, 2020, 2021, and 2022 respectively. It can be observed among all the years, 2021 has the least occurrences of O_3 concentration exceeding the demarcation limit of 180 µg/m^3. The most exceeding events of O_3 concentrations beyond 180 µg/m^3 were observed at Dr. Karni Singh shooting range, Jawaharlal Nehru Stadium, Nehru Nagar, and Sri Aurobindo Marg monitoring stations during

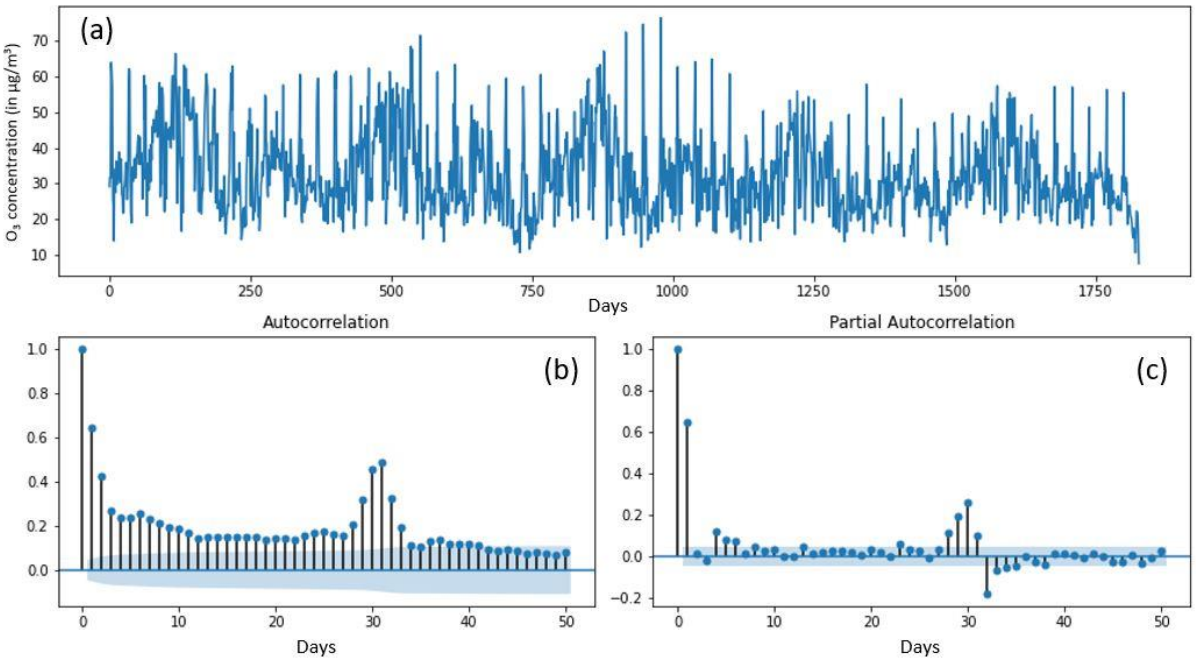


Figure 7. *(a) Total ozone concentration time series data, (b) Autocorrelation (ACF), (c) Partial Autocorrelation function (PACF)*

summer each year. Notably, these monitoring stations have a high standard deviation of the O_3 concentrations (34-52.9 µg/m^3). Notably, during the month of April, 2022 the CPCB monitoring stations recorded higher counts of O_3 concentration exceeding events crossing the threshold 180 µg/m^3. Temperature records from the Delhi Pollution Control Board (DPCC) show that during the last week of April 2022, the extreme O_3 concentration events (surpassing the demarcation limit of 180 µg/m^3) are linked with the corresponding rising temperatures within a range of 27 to 43 degrees Celsius, which is higher than average temperature. This provides a clue that the formation of these high O_3 concentrations is correlated with high temperatures and the presence of sunny days (summer). Notably, East Arjun Nagar, Dr. Karni Singh shooting range, Jawaharlal Nehru Stadium, and densely populated Nehru Nagar station observed an increasing yearly mean O_3 concentration. Apart from these stations, Sonia Vihar station was identified as the site with high average O_3 concentration. In the year 2018 and 2019, the IHBAS Dilshad garden monitoring station observed the highest O_3 concentration. In 2020, Nehru Nagar station observed the highest mean O_3 concentration. The highest mean O_3 concentration was observed for the years 2021 and 2022 in East Arjun Nagar station. As is widely known, the increasing rate of

precursor gas emissions builds O_3 concentrations at a larger scale in the presence of sunshine and favorable climatic conditions. In a study using persistence analysis, Varotsos et al. (2003) [10] and Chelani (2009) [11] attributed the high O_3 concentrations to solar radiation along with the change in weather conditions. These factors may be the reasons for higher O_3 build-up in those monitoring stations with higher O_3 concentrations.

2.3 Ozone concentration time series forecasting

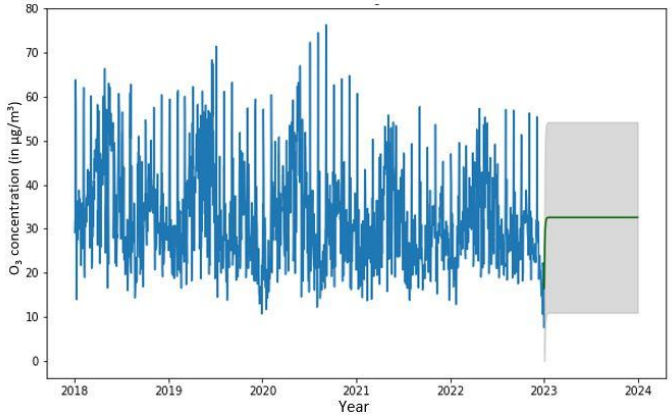


Figure 8. *ARIMA forecasts for the year 2023*

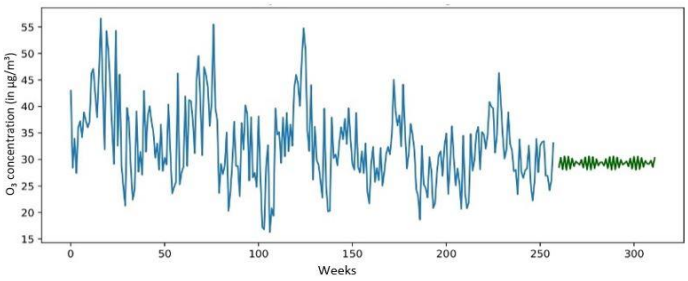


Figure 9. *Weekly ARIMA model forecasts for weekly averaged O_3 concentration data*

Time series forecasting is considered to be an important data exploration study where the objective is to predict the random variable value for near future time stamps by exploring the historical pattern from the observed data. In this work, the Auto-Regressive Integrated Moving Average (ARIMA) model ARIMA(p,d,q) is applied for the prediction of the O_3 concentration for the year 2023. The ARIMA model has three components: the Autoregressive (AR) model and the Integrated and Moving average model. In ARIMA(p,d,q), p is the AR parameter value; d is the degree of differencing, and q is the MA parameter value. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) are used to find the time series' autoregressive behavior. The actual time series, ACF, and PACF for the daily-averaged O_3 data are shown in Fig. 7. After 30-35 days of lag, the O_3 concentration autocorrelation goes towards 0 within a 95% confidence of interval. The ADF test confirms the stationarity of the time series. The best model is identified as the SARIMAX(1,0,0) model fitted with an AIC score of 12988. Notably, the Seasonal ARIMA has been considered for ARIMA model fitting as seasonality is present in the time series. The ARIMA model is employed for forecasting for the year 2023. The forecasting plot is shown in Fig. 8. The ARIMA model provides an almost similar value as forecasts, which depicts the inability of the ARIMA model to capture autoregressive behavior efficiently. Following that, further investigations were carried out on weekly averaged O_3 concentration data. The weekly ARIMA model was fed with the weekly averaged ozone concentration data. Fig. 9 displays the ARIMA weekly ozone estimates until 2023. The observed ARIMA weekly forecast mean for the 52 weeks of 2023 was 29.32 µg/m^3. We found that the weekly ARIMA model may, to a certain extent, reflect the periodicity of the data.

Similarly, the selected time series data is prepared for the input of the LSTM model also. Notably, the LSTM model is implemented in a multi-step forecasting setting. Each sample of the dataset has 30 historical timestamps, and seven future forecasting timestamps, following the autocorrelation of the time series (as in Fig. 7 (b)). The LSTM model takes the last 30 days' daily O_3 data as input and forecasts the next seven days' daily O_3 concentration. The dataset contains a total of 1821 such samples. Further, the dataset is separated into training and testing datasets. 80% data were used for developing the model, and the 20% remaining data were kept for testing. Afterward, the dataset is standardized using ($\hat{x} = \frac{x-\mu}{\sigma}$), where μ is the training sample mean, and σ is the standard deviation of the samples. The prepared samples of shape [batch size, historical data] were sent in a batch-wise fashion to the LSTM model. The modeling equation can be written as $(x_{t+1}, .. x_{t+7}) = f(x_{t-n}, x_{t-n+1}, .. x_t)$; where x_{t+i} represents the O_3 time series value at i^{th} timestamp. The LSTM works as a recurrent neural network that uses LSTM cells in a recurrent manner to learn the inherent properties of the time series. The LSTM cells consist of an input gate, a forget gate, an output gate, and the cell state. For each timestamp, the

Table 3. LSTM model parameters

Layers	Layer dimension	Parameters
LSTM 1	64	16896
LSTM 2	20	6800
Fully Connected	7	147

Table 4. ARIMA and LSTM model accuracy on 20% test data

Model	MAE ($\mu g/m^3$)	RMSE ($\mu g/m^3$)	MAPE (in %)	Corr
ARIMA(2,0,1)	8.81	11.67	24	0.05
LSTM	6.09	7.26	20	0.51

information is controlled via these three gates, and essential patterns are learned by the weight vectors corresponding to these three gates. The LSTM model architecture is shown in Table 3. The network contains two layers of LSTM cells and a fully connected multi-layer perceptron layer. The first layer of LSTM has 64 hidden units, and the second layer of LSTM contains 20 hidden units. The last layer of the fully connected MLP layer contains 147 trainable parameters with a layer output dimension of 7. The model accuracy metrics for both the ARIMA and LSTM model is presented in Table.4. The accuracy metrics are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Average Percentage Error (MAPE), Correlation (Corr). The LSTM model performs better and provides better correlation values with the actual data. ARIMA model couldn't capture the periodicity of the data. For getting the forecasts for the year 2023, a sliding approach has been utilized. The dataset's last sample with the last 30 days' data was utilized to predict the O_3 levels for the following seven days. The previous historical data was then added to this forecasted data, which was then used to forecast the O_3 concentration for the following next seven days. In a continuous fashion, this approach was utilized to generate the O_3 forecasts from the LSTM model. The O_3 concentration forecasts for the year 2023 using the LSTM model are shown in Fig. 10. Since 2021, a decreasing trend of O_3 concentrations has been observed. The LSTM model forecasts a further decrement in O_3 concentrations during 2023. During the summer of 2023, the model O_3 forecasts are lesser than that of the year 2022. The mean O_3 concentration for the year 2023 using the model ARIMA, weekly ARIMA model, and LSTM model are forecasted as 32.1, 29.3, and 29.6 µg/m^3 respectively.

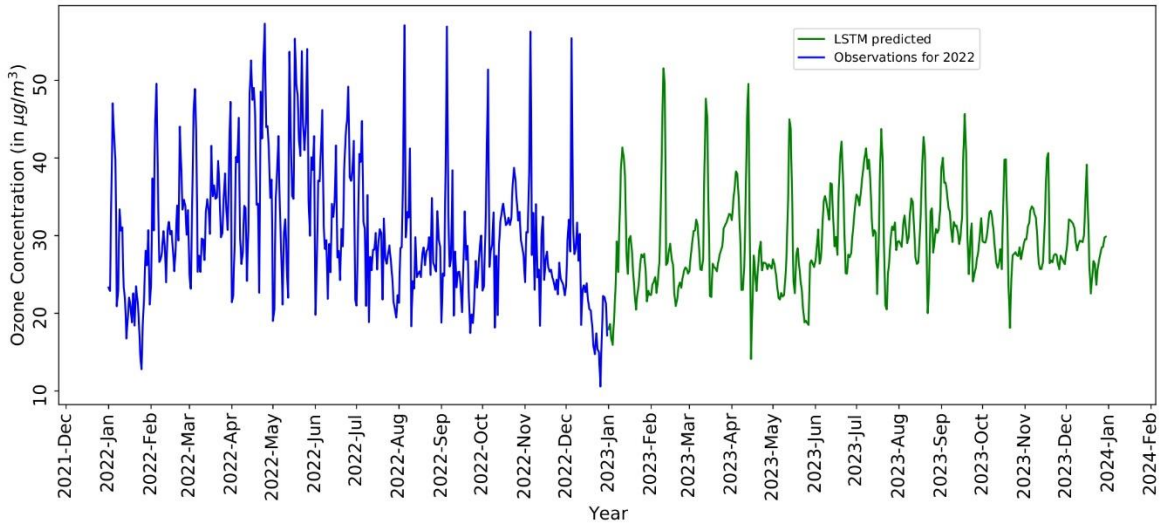


Figure 10. *LSTM forecasts for 2023*

3 CONCLUSION:

The spatially and daily averaged O_3 concentration time series data for Delhi, Indian capital city, from December 2018 to December 2022 are analyzed in this study. A detailed investigation of the characteristics of the O_3 concentration time series data was conducted. Following this, a statistical model ARIMA and a machine learning model LSTM model were utilized to provide a daily O_3 concentration forecast for Delhi during the year 2023. The study's major findings are as follows: (1) Decrease in the daily-averaged O_3 concentration in the atmosphere of Delhi was noticed after the lockdown during 2020 to 2021. Notably, the morning and evening traffic congestion levels of Delhi dipped by 19% and 17% with respect to 2020 for the year 2021 (tomtom.com [17]). In addition, as the years pass by, an overall decreased and lower concentrations pattern is observed in the diurnal variations of the O_3 concentrations each year starting from 2018. Generally, traffic and industrial emissions are the sources of the O_3 precursors, which can be a prevalent factor for the decrement. (2) In the summer, Delhi faces high O_3 concentration, (3) During October-November, another episode of O_3 concentration spikes are observed. Since 2020, these ozone concentration spikes are following a decreasing trend. (4) The ARIMA and LSTM models were fitted on the 80% data and tested on the 20% data with MAE values 8.81 and 6.09 $\mu g/m^3$ respectively. The average forecasts by ARIMA model and LSTM model for 2023 Delhi ozone concentrations are observed as 32.1 and 29.6 $\mu g/m^3$, respectively. Future research could look into the spatiotemporal variation of ozone concentrations throughout Delhi.

4 ACKNOWLEDGEMENTS

We would like to thank the Central Pollution Control Board, India for pollutant concentration data.

REFERENCES:

[1] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[2] Zhang, X., Zhang, Y., Lu, X., Bai, L., Chen, L., Tao, J., ... & Zhu, L. (2021). Estimation of lower-stratosphere-to-troposphere ozone profile using long short-term memory (LSTM). Remote Sensing, 13(7), 1374.

[3] Mbatha, N., & Bencherif, H. (2020). Time series analysis and forecasting using a novel hybrid LSTM data-driven model based on empirical wavelet transform applied to total column of ozone at buenos aires, argentina (1966–2017). Atmosphere, 11(5), 457.

[4] Cheng, Y., Zhu, Q., Peng, Y., Huang, X. F., & He, L. Y. (2021). Multiple strategies for a novel hybrid forecasting algorithm of ozone based on data-driven models. Journal of Cleaner Production, 326, 129451.

[5] Mills, G., Buse, A., Gimeno, B., Bermejo, V., Holland, M., Emberson, L., & Pleijel, H. (2007). A synthesis of AOT40-based response functions and critical levels of ozone for agricultural and horticultural crops. Atmospheric Environment, 41(12), 2630-2643.

[6] Seinfeld, J. H. (1986). ES&T books: atmospheric chemistry and physics of air pollution. Environmental science & technology, 20(9), 863-863.

[7] https://www.tomtom.com/traffic-index/new-delhi-traffic/

[8] Levy, J. I., Carrothers, T. J., Tuomisto, J. T., Hammitt, J. K., & Evans, J. S. (2001). Assessing the public health benefits of reduced ozone concentrations. Environmental health perspectives, 109(12), 1215-1226.

[9] Chelani, A. B. (2013). Study of extreme CO, NO2 and O3 concentrations at a traffic site in Delhi: Statistical persistence analysis and source identification. Aerosol and Air Quality Research, 13(1), 377-384.

[10] Varotsos, C., Efstathiou, M., & Kondratyev, K. Y. (2003). Long-term Variation in Surface Ozone and its Precursors in Athens, Greece- A Forecasting Tool. Environmental science and pollution research international, 10(1), 19-23.

[11] Chelani, A. B. (2009). Statistical persistence analysis of hourly ground level ozone concentrations in Delhi. Atmospheric Research, 92(2), 244-250.

[12] Chen, Y., Beig, G., Archer-Nicholls, S., Drysdale, W., Acton, W. J. F., Lowe, D., ... & Wild, O. (2021). Avoiding high ozone pollution in Delhi, India. Faraday Discussions, 226, 502-514.

[13] Zhang, R., Lei, W., Tie, X., & Hess, P. (2004). Industrial emissions cause extreme urban ozone diurnal variability. Proceedings of the National Academy of Sciences, 101(17), 6346-6350.

[14] Sharma, A., Sharma, S. K., & Mandal, T. K. (2016). Influence of ozone precursors and particulate matter on the variation of surface ozone at an urban site of Delhi, India. Sustainable Environment Research, 26(2), 76-83.

[15] Newbold, P. (1983). ARIMA model building and the time series analysis approach to forecasting. Journal of forecasting, 2(1), 23-35.

[16] Cheung, Y. W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. Journal of Business & Economic Statistics, 13(3), 277-280.

[17] CPCB. (Central Pollution Control Board) (2014). http://cpcb.nic.in/air.php. accessed in 15 June 2022.