STATISTICAL ANALYSIS OF OZONE POLLUTION IN DELHI: BEFORE AND AFTER LOCKDOWN

Mimansha Agrawal[1*], Tanisha Agrawal** and Shubham Gupta*
*Barclays Technology Centre India PVT Ltd, Kharadi, Pune, Maharashtra, India
**Bayes Business School, City University of London, Bunhill Row, London EC1Y 8TZ, United Kingdom (UK)

ABSTRACT
Ozone (O3) is technically known as, a Greenhouse Gas, has its own importance or harmful depending on where it is found in the earth's atmosphere. The present work mainly focuses on the study of concentration variation levels in O3 over a 5–year period (2015–2020), using the Multiple Imputation by Chained Equations (MICE) and Time Series analysis on multivariate data. Machine learning models based on SARIMA is built to for analysis of Ozone pollution especially in Delhi. Data were obtained from National Pollution Board of India which was decomposed into seasonal, reside and trend components. Through the time–series analysis of O3 in Delhi, the results showed values above average during the cold seasons. The study analysis the effects of O3 on the health during the 5-years and prediction of future Ozone pollution level. The performance evaluation of the prediction model is done by calculating mean square error (MSE), root mean square error (RMSE), and mean absolute error, may help control the degraded ozone quality. The SARIMA model for O3 gave 20.41 RMSE values for Delhi for the year 2021 and tested through one-way ANOVA hypothesis and got a positive result.

KEYWORDS: Ozone, Atmospheric pollution, SARIMA, ANOVA, Time-Series, Delhi.

MSC: 65-C60, 62-H15, 62-M10, 62-F03

RESUMEN
El ozono (O_3) es conocido técnicamente como un Gas de Efecto Invernadero, tiene su propia importancia o es dañino dependiendo de dónde se encuentre en la atmósfera terrestre. El presente trabajo se centra principalmente en el estudio de los niveles de variación de la concentración en O_3 durante un período de 5 años (2015-2020), utilizando la Imputación Múltiple por Ecuaciones Encadenadas (MICE) y el análisis de Series Temporales sobre datos multivariados. Los modelos de aprendizaje automático basados en SARIMA están diseñados para el análisis de la contaminación por ozono, especialmente en Delhi. Los datos se obtuvieron de la Junta Nacional de Contaminación de la India, que se descompuso en componentes estacionales, de residencia y de tendencia. A través del análisis de series de tiempo de en Delhi, los resultados mostraron valores por encima del promedio durante las estaciones frías. El estudio analiza los efectos del O_3 en la salud durante los 5 años y la predicción del futuro nivel de contaminación por ozono. La evaluación del rendimiento del modelo de predicción se realiza calculando el error cuadrático medio (MSE), el error cuadrático medio (RMSE) y el error absoluto medio, lo que puede ayudar a controlar la calidad degradada del ozono. El modelo SARIMA para O3 arrojó valores RMSE de 20,41 para Delhi para el año 2021 y se probó a través de la hipótesis ANOVA unidireccional y obtuvo un resultado positivo.

PALABRAS CLAVE: Ozono, Contaminación atmosférica, SARIMA, ANOVA, Time-Series, Delhi.

1. INTRODUCTION

The WHO states: " Many epidemiological studies show the chronic study of ozone on human health. Asthma episodes, decreased lung function, lung cancer, and total mortality are the main research findings. Very little evidence shows an independent long-term effect of O3 on lung cancer or even death." According to https://www.epa.gov, Ozone is a gas that is made up of three oxygen atoms. The Earth's upper as well as lower level of the atmosphere consists of Ozone, but at ground level, ozone is regarded as a significant source of harmful pollutants for humans and the environment. As per the website, https://www.eia.gov, Technically, Ozone is a greenhouse gas, but ozone can be either useful or even death-causing depending upon where it is found in the earth's atmosphere. On inhaling ozone, we may either damage our lungs, may have chest pain; Cough, shortness of breath,

[1] : mimanshaagrawal1998@gmail.com

etc. which may also worsen chronic respiratory illnesses such as asthma and jeopardize the body's ability to fight off respiratory infections. Overall, air pollution levels have risen by almost 5.5 percent since 2016, and Carnegie Mellon researchers have identified a number of reasons for this, including increased natural gas consumption and more people driving Jiang et al. (2016). The paper focuses on the difference in O3 concentrations before and after the lockdown period in Delhi. Sicard et al. (2020) The study observes that even in the lockdown, the level of O3 was higher, as the concentration of ozone increases in warmer climates. The year 2020 was recorded to be the hottest year with heat-related accidents like forest fires etc. Ozone depletion is on the rise, escalating, and has been a problem for many reasons, according to its new analysis. Ozone is not only a summer problem, the scientist detects ozone levels exceeding even in winter, making winter air extremely dangerous. An extraordinary level of ozone is visible even at night as ozone is produced through other gases by chemical reactions in the presence of sunlight. The top 10 ozone hotspots in Delhi include stations from the south of Delhi and smaller towns in the region. This situation requires the development of a clean air-conditioning system to increase ozone depletion strategies, to take stronger measures on cars, industry, and burn waste.

2. DATA DESCRIPTION

Real-time data is collected from the official portal of Pollution Control Board for Air Quality Management online and obtained information from around 81 metrological stations. The reason why Delhi is considered in this study as, it is known as the hotspot city for Ozone pollution in India. The data was mainly focused on the 5 years which included the years before and after lockdown. Also, further prediction is made for future. The Fig.1 (a) shows the study area of Delhi and Fig.1 (b) shows the date starting from 2015 till 2020 and the concentration of Ozone for the corresponding dates in Delhi.

	O3
2015-01-01	25.868710
2015-02-01	30.829643
2015-03-01	28.738387
2015-04-01	58.172000
2015-05-01	69.061613
...	...
2020-03-01	37.859355
2020-04-01	46.185667
2020-05-01	56.768065
2020-06-01	39.616000
2020-07-01	29.730000

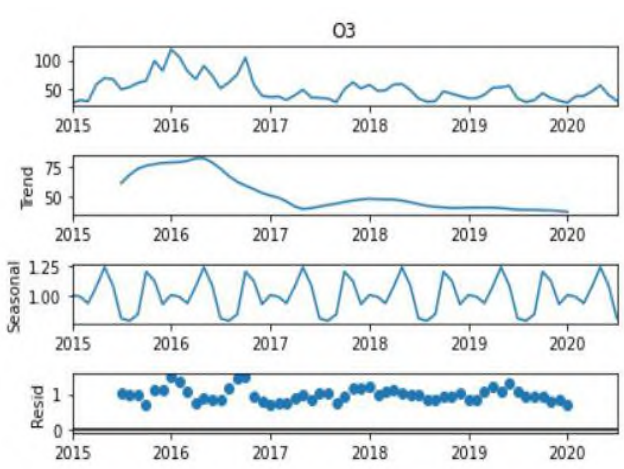
67 rows × 1 columns

Figure 1: (a)Boundary of the Study Area, Delhi (b) Dates and the concentration of Ozone in Delhi

3. METHODOLOGY

A series of O_3 concentrations per day are analyzed using the SARIMA modelling method. SARIMA represents a standard category of time series models with seasonal variations which includes several time series strategies such as variations, automatic models, and dynamic scale. The SARIMA simulation method follows certain steps. First, the SARIMA model checks, that the time series is stationary or non-stationary Krishnan et al. (2004). If it is stationary, then its fine, but if its non-stationary then it is firstly converted to stationary series using the appropriate degree of variance which gives the value of d. Then, the interim values of p and q are obtained through the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the time series Krishnan et al. (2004). After determining the p, d, and q, the coefficients of the default and moving average values are measured using the indirect square method. This is followed by a test of the model, and, if required this model may be transformed according to the values of p and q to achieve the required level of model validity. Fig.2 shows the seasonal decomposition of the

average yearly Ozone. The trend plot shows a decline after 2016 and till 2020 it shows the lowest values. This sudden decline after 2016 is because the Odd-Even rule by the Delhi government, which reduced the vehicular pollution and in 2020 may be due to the COVID-19 lockdown, hence there is a up-down trend for the data. In the seasonal plot, there is definite seasonality in the data, causing concentration of the ozone to fluctuate in a similar trend. This makes sense that the concentration of the ozone depends upon its seasonal production cycle. Because of this reason, we include a seasonal (S) piece in our model.



SARIMAX Results

Dep. Variable:	O3	No. Observations:	41
Model:	SARIMAX(0, 1, 0)	Log Likelihood	-170.422
Date:	Mon, 22 Nov 2021	AIC	342.845
Time:	19:49:06	BIC	344.534
Sample:	01-01-2015	HQIC	343.456
	- 05-01-2018		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
sigma2	293.9248	61.166	4.805	0.000	174.042	413.808

Ljung-Box (L1) (Q):	0.60	Jarque-Bera (JB):	0.20
Prob(Q):	0.44	Prob(JB):	0.90
Heteroskedasticity (H):	0.30	Skew:	-0.07
Prob(H) (two-sided):	0.04	Kurtosis:	3.32

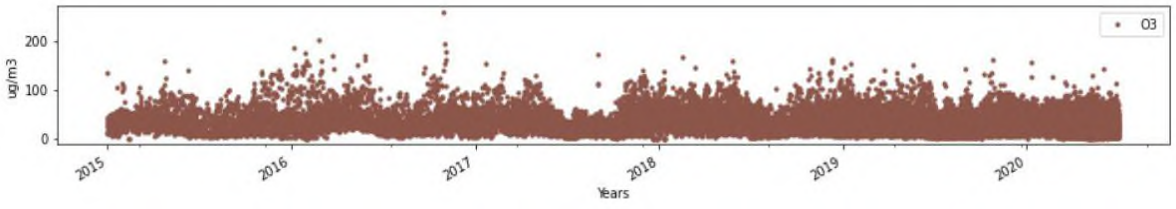
Figure 2: (a) Seasonal Decomposition of Ozone (b) Summary of SARIMA model used

4. RESULTS AND DISCUSSION

As per the website, https://www.cseindia.org, The Centre of Science and Environment analysis claims that ozone is not only a problem of warmer days but, also a topic of discussion in the colder months. Observations tells that mostly the annual increase in the number of transit days is due to the unprecedented increase in ozone during the rainy and colder months. It was also observed that, the winter months of the year 2020-21 shows an astonishing increase of 84% in the days with many localities surpassing the normal standard. According to Zhang (2019), Some Meteorologist confirms that, during colder months of the year 2020-21, as compared to the previous years, there is around 32% rise in the days when more than 15 areas excel the normal approved standards. On the other hand, the survey examines that the summer data is relatively stable since years in 2021. This fact was despite the fact that the summers of 2021 was mild and no serious heat waves were reported. In addition, the monsoon of the year 2020 saw some increase of the standard level as compared to 2019. Considering the days without extremes, there were only 36 days of rainfall in 2020, while 2019 had 45 days and 2018 had 60 days. Ozone during the winter solstice is still relatively an unknown phenomenon. Xu et al. (2018) Data analysis of ozone and other harmful pollutant levels in the year 2020 smoke season shows that ozone is one of the main pollutants which is on the peak throughout the season. Ma et al. (2015) There was a slight decrease in the plot of ozone pollution when the sun's rays were too low and increased $PM_{2.5}$ and vice versa.

Meteorological reasons such as Sunlight etc., contributes a lot to the formation of ozone in summer as well as winter months. Sunlight is very vital as it carry out the process of photochemical reaction between various gases to form ozone as an end product. Ozone data which are obtained from 2 channels, RK Puram and Siri Fort, - are related with sun’s radiation, wind, temperature, NO_2 to determine geographical differences. The quantum of ozone produced at 600-700 W/mt2 of solar radiation in the summer was similar to the ozone produced at 100-300 W/mt2 of solar radiation in colder months. However, not much difference is seen with varying temperature on the concentration. On

the other hand, NO_2 and ozone, has negative correlation during the summer months is just opposite in the winter months. At RK Puram, the continuous average of eight hours remained above the average noted for some 31 hours from 5-6 May, 2021. A homogeneous pattern was recognized at Siri Fort. As per https://www.downtoearth.org.in, The impact of the epidemic and ozone depletion continued and the duration of ozone depletion also increased: Although particle and NOx levels dropped drastically due to the lockdown in the Delhi, high levels of ozone were noted in several areas. Ananda Vihar showed a very impressive data, with less than 5 days for exceeding the levels than the normal standard, as compared to the past summers of 2018 and 2019, but during the summer of pandemic i.e. 2020 it was noted 49 days of increase. Ma et al (2016) Also, it was seen that number of hours of increased levels showed a rise during the lockdown. For example, the stations that monitor the 8-hour ozone level in the 2019 summer, showed a regular rise of 5.3 hours each day. Due to this, the level has exceeded 20% during the lockdown of 2020 summer. As per Turner (2016), At night, ozone was recorded higher during pandemic curfews. Ozone usually get dispersed after the dusk as no sunlight and NO_x reacts with ozone. But high ozone concentration has been observed during the night. But this information is a controversy as, evening shifts and traffic gets reduced, but still ozone levels are rising.



Ali et al. (2012) researched that Ozone is a severe problem in the NCR with Ghaziabad exceeding the normal standard concentration of ozone pollution. Faridabad Sector 16A and The New Industrial Town recorded 152 days and just three extremes in 2020 respectively. In Gurugram, Vikas Sadan which was the highly polluted area and recorded 139 days exceedance, while neighboring area Trigram noted as low as 35 days. Noida has shown a limited number of days with higher concentration. Top 10 areas of Delhi including southern Delhi and small towns of NCR including Bulandshahr in Uttar Pradesh and Bhiwani Haryana were observed in the top 20 ozone-depleted cities and towns Bernstein et al. (2009). Areas south of Delhi top the list as they rank 4 times out of 10, Including stations from Ghaziabad, Gurugram and Faridabad, the Dr K S Shooting Range became the most polluted area in the NCR. The AQI report of Central Pollution Control Board (CPCB), looks at the concentration level from 8 AM-4 PM ozone to calculate the concentration and compare the 8h measure. But this fails to include hours with high altitude i.e. in the evening and at night. As per the observations, Ozone level rises usually at dusk in Delhi and the average high hour can be 4-5 PM or 5-6 PM in summer. In view of the serious health risks associated with ozone, we must incorporate these National Clean Air Programme for Delhi-NCR action plan. A firm stand on high emitters of VOCs that combine automotive and industrial and NO_x is required to reduce the Ozone pollution level. Recent Delhi resource

Figure 3: Analysis of Ozone pollution in various Years

research has shown that the transport sector is the largest emitter of NO_x and VOCs (producing ozone) in Delhi apart from industry and other sources. SAFAR claims that, transport is responsible for 62.5% NO_x, 90% VOC, 24% NO_x industries and 9% VOCs. According to ARAI-TERI list, transport is responsible for 81.4% NO_x, 80% VOC. Now, it is very worrying that ozone levels have been found above the 100 µg/m3 in winter and are highly sensitive to sunlight Chen et al. (2019).

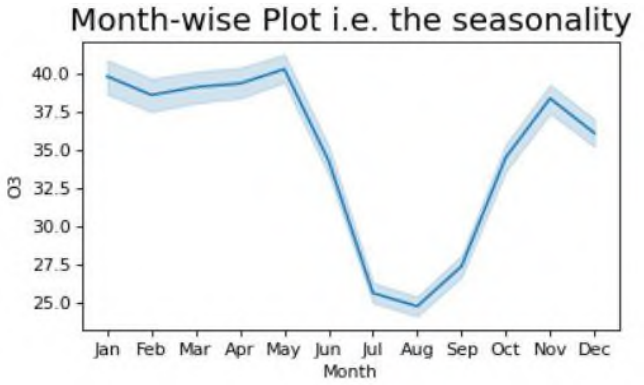


Figure 4: Concentration Levels of Ozone measured in Delhi.

After a long break from COVID commercial as well as industrial activities have emerged in the days of Diwali months in October and November, which led to an increase in car pollution, which as a result had a major impact on Delhi pollution Jacko et al. (2009). The reports reveal that due to increase in commercial activity, both car pollution and dust pollution are increasing at an ever-increasing rate. Fig.3 shows the Analysis of Ozone pollution in various years in last 5 years from 2015-2020. The concentration almost remains same in all 5 years as the ozone was regularly produced through photochemical reaction. CPCB claims that Air Pollution data Of Delhi shows that pollution levels have dropped in some years. By 2020, it has decreased due to lockdown (some restrictions) etc. To neutralize air pollution in Delhi, CM Arvind Kejriwal announced a 10-point "winter action plan" focused to control the dust pollution, using Pusa, which is a bio-decomposer installed smoke panels and waste inspection incinerator and vehicle extraction. Delhi, however, is exacerbated by a combination of tropical thermal smoke plants and brick schools in the capital's area, pollution from the dense transport network, burning of races or biomass of farmers in neighboring provinces, and a lack of clean air that creates air pollution over the city. Ghude et al. (2014) In the year 2020, the share of burns on Delhi pollution increased to 42% on 5 November. In 2019, the burning of crop residues accounted for 44% of Delhi's $PM_{2.5}$ on November 1st.

Fig.4 shows the monthly data which depicts a significant reduction in ozone pollution in the month of August. This reason is likely due to the rainy season, which have less sunlight for the ozone production. Fig.5 shows the various plots of the Ozone concentration in Delhi with Training, testing and predicted model. The model was taken in a ratio of 80:20 of the whole dataset The training data includes data from 2015-2018 and the testing data includes data of the year 2019. The year 2020 was kept as an outlier as it has lockdown months also included.

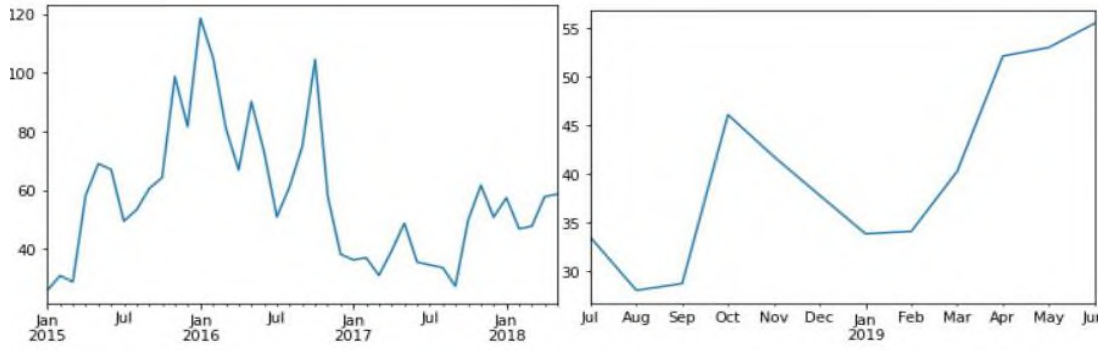
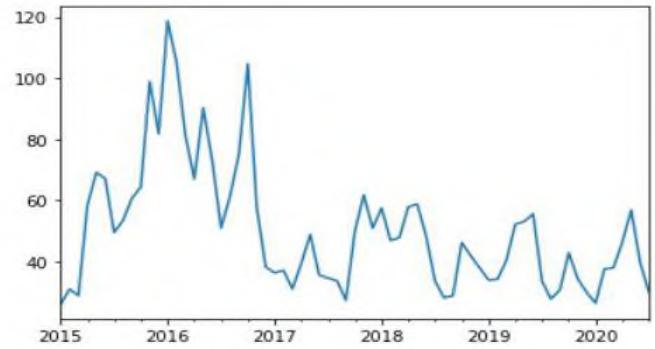
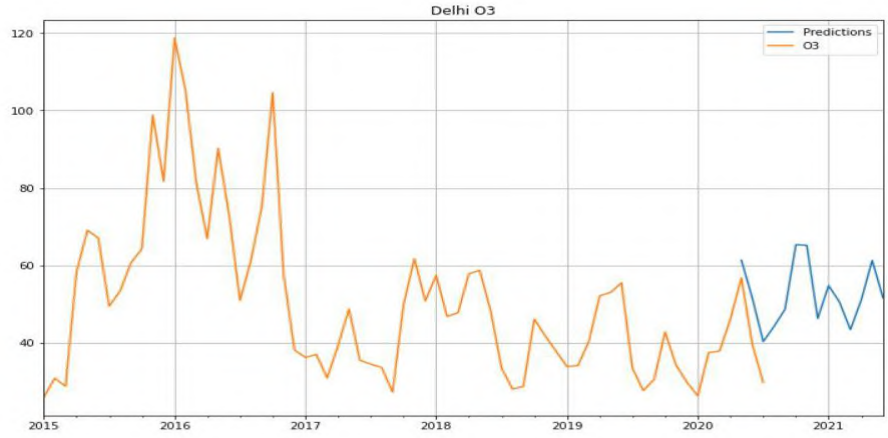


Figure 5: (a) Train Plot of Ozone concentration In Delhi (b) Test Plot of Ozone Concentration in Delhi



(c) Ozone plot of Delhi for Train and Test plot combined



(d) Predicted Plot of Ozone in Delhi for the year 2021.

5. MODEL VALIDATION

To investigate the goodness of the model equity, it is also necessary to validate model predictions. SARIMA models usually offer the best forecasts in one day because future predictor errors are unknown. Therefore, in this study, model validation was performed by obtaining one-time predictions before the last 30 observations of the time series, which was not part of the validity of the original model. After every prediction, the prediction error (predicted value of the predicted value) was compounded and the perceived value became part of the data for estimating the model for the next forecast. The performance measurement models for air quality models are, Root Mean Square Error (RMSE), Bias and Mean Square Error (MSE). The calculated values of these measurements are presented in Table 1. The values indicate that the performance of the model is very satisfactory.

Measurement	Values
Root Mean Squared Error (RMSE)	20.414402404788888
Mean AQI	40.40149276753712
Bias	-18.296572
Mean Squared Error (MSE)	416.7478255446503
Mean Absolute Error (MAE)	18.296571748591916

Table 1: Measurements of the Proposed Model

6. TEST OF HYPOTHESIS

The average O_3 of Delhi is greater or less than other cities let's suppose Mumbai.
Null Hypothesis (H_0): Delhi have less average O_3 than other cities due to high population.
Alternate Hypothesis (H_1): Delhi have greater average O_3 than other cities due to high population.
Checked on 5% significance and the p-value came: 2.4347660095335485e-139. We reject the null hypothesis *(H_0)*. Therefore, the null hypothesis *(H_0)* has rejected which means, "Delhi have greater average O3 than other cities due to high population". The null hypothesis is a wrong assumption, and the alternate assumption *(H_1)* is true, and which is also proved graphically. Hence, we confirm the assumption at 5% of significance.

7. CONCLUSION AND DISCUSSION

The SARIMA order model (0, 1, 0, 12) is equivalent to a series of high O3 focus time series between year 2015-2020. Jenkins et al. (2002) The coefficients of the model and patterns of the ACF and PACF show the dominance of the Autoregressive Process of order 0 in the model. Examination of the first 20 residual combination combinations with the remaining ACF structure shows that the model balance is good. Model predictions are found to be reasonably close (13% of the Total Average Mistake) in the targeted quantities of O3 daily face concentrations. The calculated values of the model performance measurements (RMSE 20.41, Bias -18.29 and MSE 416.74) also confirm the strength of the forecast model. Agrawal et al. (2021) Research shows and confirms that the SARIMA modelling system can be used effectively to obtain short-term air quality forecasts. Khemani et al. (1995) Particularly in cases where data limitations limit the use of prescribed modelling methods, the SARIMA modelling approach may serve as an effective means of obtaining O3 predictions. In some cases, the overall air quality forecast can be very accurate and useful when SARIMA modelling results are applied in conjunction with the results of determining climate-based models and other relevant variables. Comparative research of SARIMA comparisons with other methods like Classification, Regression models and neural network models in the case of air quality modelling may be highly desirable. Also, ANOVA test is applied on the assumed hypothesis at 5% significant and the null hypothesis was rejected.

REVISED: JANUARY, 2023.
RECEIVED: SEPTEMBER, 2023-

REFERENCES

[1]. AGRAWAL, M., AGRAWAL, M., ASLAN, A., DÖNMEZ, I., GÜNEŞ, A. (2021): Monitoring Air Pollution Impacts of COVID-19 in India. **Euras Journal of Health**, 10.17932/EJOH.2020.022/ejoh_v02i2001. https://ejoh.euras-edu.org/2021-december-volume-2-issue-2/

[2]. ALI, K., INAMDAR, S. R., BEIG, G., GHUDE, S., & PESHIN, S. (2012): Surface ozone scenario at Pune and Delhi during the decade of 1990s. **Journal of Earth System Science**, 121, 373–83.

[3]. BERNSTEIN, J. A., ALEXIS, N., BARNES, C., BERNSTEIN, I. L., BERNSTEIN, J. A., NEL, A., WILLIAMS, P. B. (2004): Health effects of air pollution. **Journal of Allergy and Clinical Immunology**, 114, 1116–23.

[4]. CENTRE FOR SCIENCE AND ENVIRONMENT. (2023): **CSE rings the alarm bells on ozone pollution in Delhi-NCR. CSE India.** https://www.cseindia.org/cse-rings-the-alarm-bells-on-ozone-pollution-in-delhi-ncr-10912.

[5]. CHEN, Z., ZHUANG, Y., XIE, X., CHEN, D., CHENG, N., YANG, L., & RUIYUAN, L. (2019): Understanding long-term variations of meteorological influences on ground ozone concentrations in Beijing during 2006-2016. **Environmental Pollution**, 245, 29–37.

[6]. DOWN TO EARTH. (2019): **Ozone pollution in Delhi worse this summer than last: CSE. Down To Earth.** https://www.downtoearth.org.in/news/pollution/ozone-pollution-in-delhi-worse-this-summer-than-last-cse-65171.

[7]. GHUDE, S. D., JENA, C., CHATE, D. M., BEIG, G., PFISTER, G. G., KUMAR, R., & RAMANATHAN, V. (2014): Reductions in India's crop yield due to ozone. **Geophysical Research Letters**, 41, 5685–91.

[8]. JACKO, R., & LA BRECHE, T. (2009): Air pollution and noise control. **Environmental Engineering**.

[9]. JIANG, X. Q., MEI, X. D., & FENG, D. (2016): Air pollution and chronic airway diseases: What should people know and do? **Journal of Thoracic Disease**, 8(1), E31-E40.

[10]. JENKIN, M. E., & CLEMITSHAW, K. C. (2002): Chapter 11 Ozone and other secondary photochemical pollutants: Chemical processes governing their formation in the planetary boundary layer. **Development in Environmental Science**, 1, 285–338.
[11]. KRISHAN KUMAR, A., YADAV, A. K., SINGH, M. P., HASSAN, H., & JAIN, V. K. (2004): Forecasting Daily Maximum Surface Ozone Concentrations in Brunei Darussalam—An ARIMA Modeling Approach. **Journal of the Air & Waste Management Association**, 54, 809-814.
[12]. MA, Z., HU, X., SAYER, A. M., LEVY, R., ZHANG, Q., XUE, Y., ET AL. (2015): Satellite-based spatiotemporal trends in PM2.5 concentrations: China, 2004–2013. **Environmental Health Perspectives**, 124, 184–92.
[13]. MA, Z., XU, J., QUAN, W., ZHANG, Z., LIN, W., & XU, X. (2016): Significant increase of surface ozone at a rural site, north of eastern China. **Atmospheric Chemistry and Physics**, 16, 3969–77. doi:10.5194/acp-16-3969-2016.
[14]. SICARD, P., AGATHOKLEOUS, E., XU, X., PAOLETTI, E., DIÉGUEZ, J., & CALATAYUD, V. (2020): Amplified ozone pollution in cities during the COVID-19 lockdown. **Science of The Total Environment**, 735, 139542.
[15]. TURNER, M. C., JERRETT, M., POPE III, C. A., KREWSKI, D., GAPSTUR, S. M., DIVER, W. R., ET AL. (2016): Long-term ozone exposure and mortality in a large prospective study. **American Journal of Respiratory and Critical Care Medicine**, 193, 1134–42.
[16]. **U.S. Energy Information Administration**. (2021): How much carbon dioxide is produced from burning gasoline and diesel fuel? https://www.eia.gov/tools/faqs/faq.php?id=84&t=11
[17]. **U.S. Environmental Protection Agency**. (2021). Ozone generators are sold as air cleaners. https://www.epa.gov/indoor-air-quality-iaq/ozone-generators-are-sold-air-cleaners.
[18]. WEI, Y.; ZHANG, J., WEI, Y., & FANG, Z. (2019): Ozone Pollution: A Major Health Hazard Worldwide. **Frontiers in Immunology**, 10, 2518. https://www.frontiersin.org/articles/10.3389/fimmu.2019.02518/full.
[19]. WORLD HEALTH ORGANIZATION. (2021): **Air pollution.** https://www.who.int/southeastasia/health-topics/air-pollution.
[20]. XU, W., XU, X., LIN, M., LIN, W., TARASICK, D., TANG, J., ET AL. (2018): Long-term trends of surface ozone and its influencing factors at the Mt Waliguan GAW station, China-Part 2: the roles of anthropogenic emissions and climate variability. **Atmospheric Chemistry and Physics**, 18, 773–98.