Forecasting Air Quality in Amritsar

***Ankita Nandy, **Rashmi Kujur**
*Independent Researcher
Hooghly, West Bengal, India
** Assistant Professor, Information Technology,
B.C.S Government PG College,
Dhamtari, Chattisgarh, India

DOI:10.37648/ijrst.v13i03.010

[1]*Received: 19 August 2023; Accepted: 17 September 2023; Published: 17 September 2023*

ABSTRACT

Tourism in Amritsar happens to be an important contributor towards the economy of Punjab. The holy sites of Amritsar attract several pilgrims all around the year, several of them being in the autumn of their lives. However, the alarming levels of air pollution in the area housing the Golden temple, measured by the Air Quality Index (AQI), make it lethal to the visitors. The surfaces of several heritage monuments have witnessed corrosion. Suspended particulate matters of diameter 2.5 micrometres, or less, referred to as PM2.5 are the primary contributors to this deteriorating air quality. Employing the daily AQI data recorded for past three years, a statistical time forecasting model is built. Reliable forecasts can aid authorities in monitoring and improving the air pollution in this region, and safeguarding the health of both visitors and residents in the holy city of Amritsar.

Keywords: *Air Quality Index; PM2.5; Forecasting; Amritsar; Tourism*

INTRODUCTION

Amritsar, a city in Punjab, India, is well known for its splendid Gurdwaras, ancient temples, monuments reminding us of the struggle for Independence, and undoubtedly, the mouth-watering dishes, and attracts numerous tourists and pilgrims all year round. The Golden temple is of great religious and spiritual significance to the Sikh community. As per popular belief, Luv and Kush, the sons of Sri Ram and his wife Sita, had captured and bound Hanuman in the courtyard of the Durgiana Temple. The massacre of thousands of innocents at the Jallianwala Bagh was instrumental at propelling the struggle for Independence to the finish line. Soaked in spirituality, culture and history, Amritsar has an important place on the bucket list of any traveler. The Punjab government has invested heavily towards the conveniences and comforts of the visitors, with elaborate walkways, seating arrangements, several modes of transportation such as horse driven carriages as well as rickshaws and autorickshaws. There are numerous shops lined up selling traditional handicrafts, costumes, snacks et cetera around the tourist attractions, giving this space a pleasant ambience. The religious significance of the city draws aged pilgrims and young explorers alike.

The Air Quality Index (AQI) evaluates the presence and concentrations of various pollutants in the air. This index was developed by the United States Environmental Protection Agency (US-EPA). In India these values are published by the Central Pollution Control Board (CPCB). The values are divided into size categories as follows, a value less than 50 is considered Good, 51-100 Moderate. AQI of 101-150 is unhealthy for sensitive people. At this level, those with existing respiratory issues can experience discomfort in breathing. On extended exposure to air with a high AQI, even healthy individuals might report irritation in the eyes and discomfort in breathing even. 151-200 is unhealthy, 201-300 very unhealthy, and any value beyond is hazardous. AQI is calculated based on several pollutants such as oxides of nitrogen and sulfur, however this work focusses on Particulate Matter with diameter less than 2.5 micrometers, referred to as PM2.5. Their tiny size facilitates their easy passage through the nasal pathways to the lungs, and even the bloodstream, causing serious damage to the pulmonary system [10]. High concentration of PM2.5 in the air can prove to be lethal. Common sources of this pollutant are unfiltered smoke from vehicles and industries, and burning of firewood.

[1] How to cite the article: Nandy A. Kujur R. (September 2023); Forecasting Air Quality in Amritsar; *International Journal of Research in Science and Technology*, Vol 13, Issue 3, 98-102, DOI: http://doi.org/10.37648/ijrst.v13i03.010

The walled area housing the Golden Temple regularly records high values of AQI. In 2015 itself, environmentalists had highlighted the corrosion of the surface of heritage monuments and reduction in their sheen, such that of the Golden Temple. The walled area being a busy tourist place experiences high volume of traffic. Negligible green cover and throngs of aged pilgrims makes such high concentrations of PM2.5 even extremely fatal. Foresight is undeniably valuable in the planning and management of any tourist locale. Estimates of the footfall and economic impacts of such influx aids in the judicious use of available resources. On similar lines, the impact of anthropogenic activities on the environment ensures the visitors and the locals managing these tourists can enjoy not just comfortable but healthy lifestyles. This document presents a statistical time series forecasting model for the PM2.5 levels in the walled area housing the Golden Temple in Amritsar.

LITERATURE REVIEW

World over, the derogatory impact of air pollution has been felt not just on human health, but on the flora and fauna as well. These pollutants have consequently led to the climatic changes and global warming we are witnessing on a widespread level. The literature is rich with numerous models predicting the levels of these pollutants in the atmosphere, as in [8]. Chemical Transport Models employ satellite imagery and the numerical air quality predication model WRF-Chem, an example being presented in [6], to dynamically assess air quality and issue alert warnings in the event of sever deterioration. Other research works, such as [1][9], use statistical models to estimate the AQI. Neural networks, which are designed to mimic the firing of neurons in the human brain, are famed to deliver reliable forecasts even for highly non-linear time series, and have been demonstrated in [3][5][7]. However, when limited data is available, statistical models outdo neural networks in performance [9]. In [2][4], the forecasting problem is converted into a classification task by categorizing the AQI values in Good, Moderate et cetera and predicting the category.

MATERIALS AND METHODS

A. Data

Measuring the characteristics of a time-varying process at fixed intervals of time aids in the study of its dynamic behavior. To capture the variations in PM2.5 levels in the atmosphere, the past three years' records for the Golden temple region have been collected from [www.aqicn.org/here]. The very few days with missing values have been imputed with the values located on the previous day.

The COVID pandemic in 2020 and the subsequent lockdowns led to a massive drop in the inflow of tourists in the otherwise busy Amritsar. Businesses remained closed and traffic was minimal. There, as observed in Figure 1, except for the winter months where people lit bonfires to counter the chill, PM2.5 levels remained low. In 2021 and 2022, with the social distancing and vaccinations limiting the COVID infections, Amritsar returned to its former vibrancy. Form Figure 2 and 3, PM2.5 levels are observed to spike in winters, possibly because of the festive celebrations and the smoke from bonfires which stay suspended in the dense fog. In July, the rains potentially deter tourists and consequently the traffic around this region drops, further reducing the PM2.5 values.

B. Methods

Stationarity implies time invariance of statistical properties like mean and variance. When data is stationary, autoregressive (AR), moving average (MA), or their combination (ARMA) models have been used. AR indicates the next values have some dependency on previous value(s). MA indicates the series is a weighted combination of lagged white noise components. Following is the mathematical representation of the ARMA model.

$$(1-\sum_{i=1}^{p}\alpha_i L^i)X_t = (1-\theta_i L^i)\varepsilon_t$$

However, real time series are mostly non stationary. Therefore, as per the Box Jenkins methodology, ARIMA models are built which incorporate the differencing of the time series to make it stationary.

$$(1-\sum_{i=1}^{p}\alpha_i L^i)(1-L)^d X_t = (1-\sum_{j=1}^{P}\theta_i L^i)\varepsilon_t$$

The COVID pandemic in 2020 made it an unusual year. Thus, PM2.5 values for 2020 will be excluded from the modelling task.

An ARIMA model will be developed using the "auto.arima" function in the forecast package in R. In contrast to the Box Jenkins method of determining the parameters of an ARIMA model based on the autocorrelation function and the partial autocorrelation function, "auto.arima" iteratively fits several ARIMA models with different combination of the parameters. The model with the best fit, determined as per the Akaike Information Criteria (AIC) is output at the end.

PM2.5 levels starting from the 1st of January, 2021 till the 12th of August, 2023 were split into training and testing. Data will the 31st of July, 2023 will be used for training the ARIMA model while the remaining 12 data points will be used for testing. The fit will be evaluated based on three error measures, Root Mean Squared Error, Mean Absolute Error, and the Mean Absolute Percentage Error, as shown in Table 1.

RESULTS

ARIMA (5,1,0) was reported to be the best model by the "auto.arima" function. In simple words, the next day PM2.5 level could be estimated using the levels observed today and in the past five days. As seen in Table 2, the level is consistently higher than 100. Following is the model equation.

$$Y(t+1) = Y(t) - 0.26\,Y(t-1) - 0.301\,Y(t-2) - 0.194\,Y(t-3) - 0.193\,Y(t-4) - 0.05\,Y(t-5) + \varepsilon_t$$

The fit reports an RMSE of 13.79, MAE of 11.33 and MAPE of 7 percent.

FUTURE WORK

The inclusion of meteorological factors, such as the air temperature, humidity, wind speed et cetera can be incorporated in the ARIMA model, for better forecasts. Principles from the numerical weather prediction models can be added to estimate these levels for a longer horizon. The impact of agricultural practices in surrounding regions, such as stubble burning (which coincides with harvest festivals), can also be studied to better understand the causes of consistently high PM2.5.

CONCLUSION

Amritsar is a prime tourist magnet in Punjab. Rising levels of pollutants in the air is detrimental to not just the visitors and locals, but to the contribution made by tourism towards the state economy as well. The analysis of the PM2.5 levels around the Golden temple shows that these levels are affected by the volume of tourists and the seasons. For most of the year it stays in the "unhealthy for sensitive people" category and progresses towards the "unhealthy" category during festivals, ARIMA (5,1,0) model forecasts the level above 100 for the 12 days in August. The use of face masks in outdoor settings and air purifiers indoors is recommended for visitors, and especially for the shopkeepers who spend most of the year in this region. Switch to non-polluting electric vehicles will be another progressive move. Bonfires in this region should be discouraged. For long term improvement in air quality, the green cover of this region should be increased.

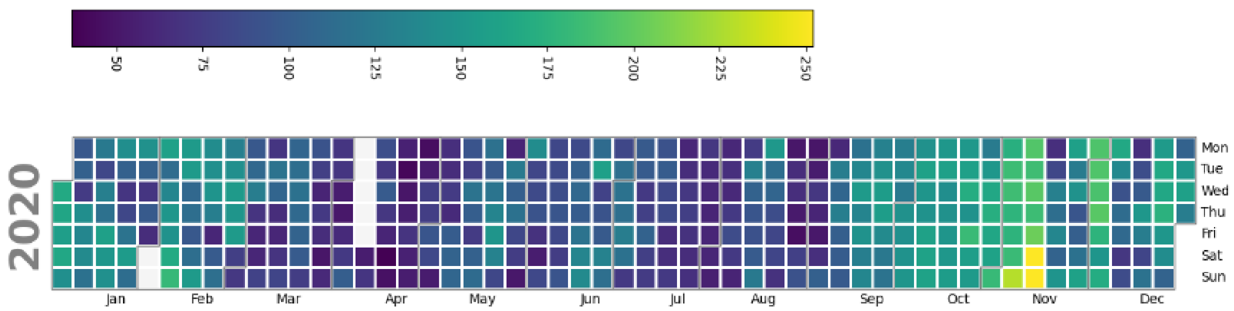

Figure 1. Calendar chart for 2020 PM2.5 values

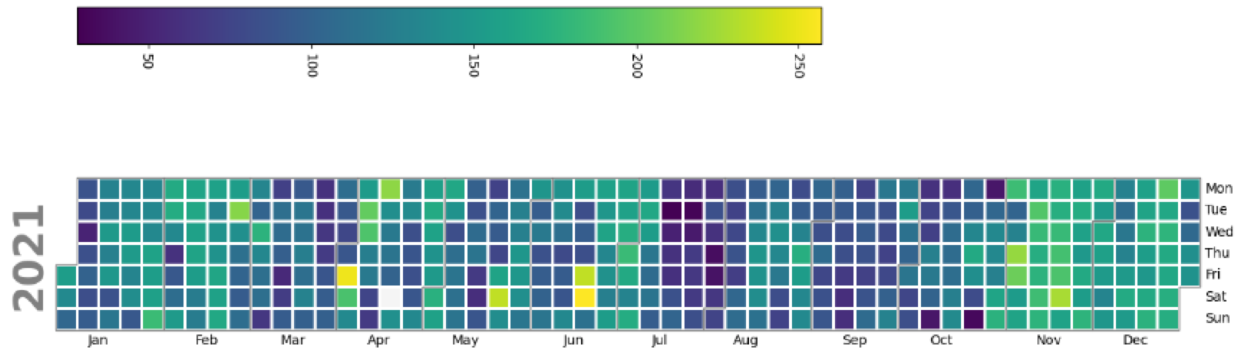

Figure 2. Calendar chart for 2021 PM2.5 values

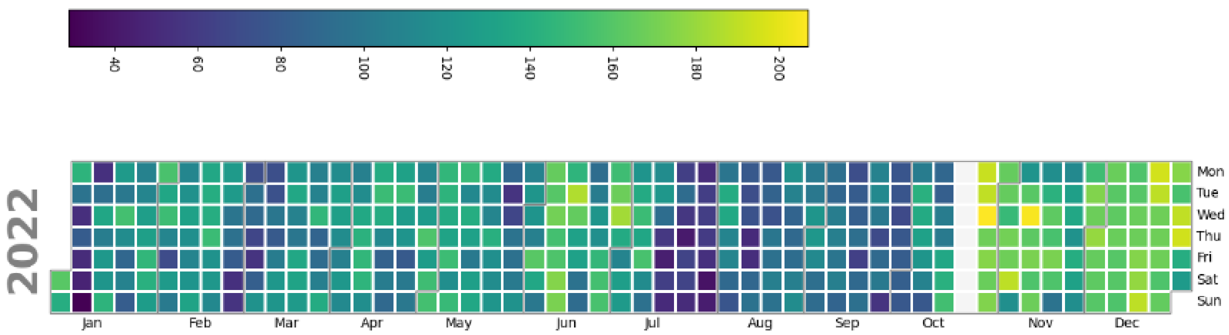

Figure 3. Calendar chart for 2022 PM2.5 values

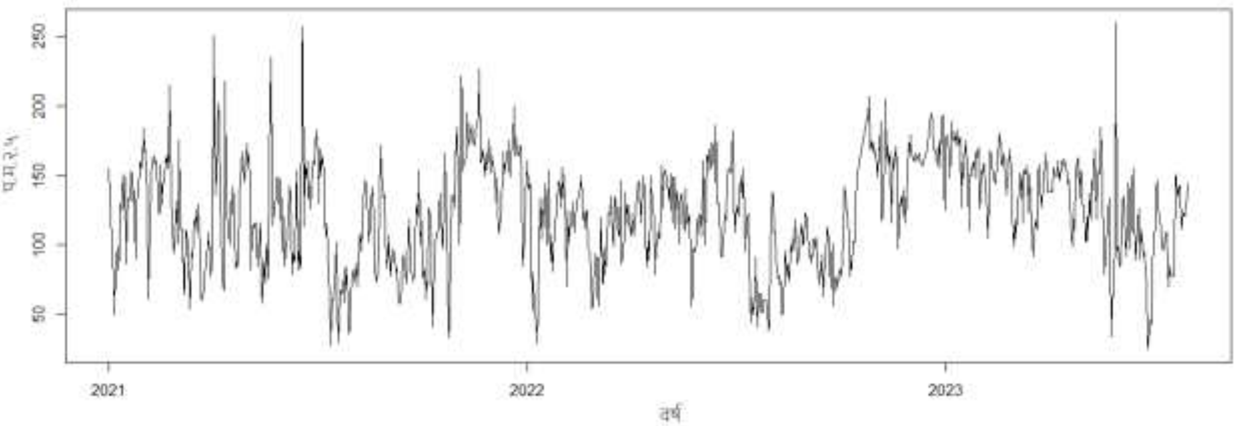

Figure 4. Line plot of PM2.5 values in Amritsar 2021 onwards

Table 1 Error Measures

Error Measure	Formula
Mean Absolute Error (MAE)	$\text{MAE} = \frac{1}{n}\sum_{1}^{n} \lvert y_i - \hat{y}_i \rvert$
Mean Absolute Percentage Error (MAPE)	$\text{MAPE} = \frac{1}{n}\sum_{1}^{n} \frac{\lvert y_i - \hat{y}_i \rvert}{y_i} \times 100$

Error Measure	Formula
Root Mean Squared Error (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{1}^{n}(y_i - \hat{y}_i)^2}$

Table 2. Comparison of actual and estimated PM2.5 values for 12 days in August 2023

ARIMA (5,1,0)	*1/8*	*2/8*	*3/8*	*4/8*	*5/8*	*6/8*	*7/8*	*8/8*	*9/8*	*10/8*	*11/8*	*12/8*
Actual	141	146	135	151	130	131	140	162	151	149	125	149
Estimated	138	134	134	135	136	136	136	136	135	136	136	136

REFERENCES

1. Atoui, A., Slim, K., Andaloussi, S. A., Moilleron, R., & Khraibani, Z. (2022). Time Series Analysis and Forecasting of the Air Quality Index of Atmospheric Air Pollutants in Zahleh, Lebanon. *Atmospheric and Climate Sciences*, *12*(4), 728-749.
2. Chen, C. W., & Chiu, L. M. (2021). Ordinal time series forecasting of the air quality index. *Entropy*, *23*(9), 1167.
3. Handhayani, T. (2023). An integrated analysis of air pollution and meteorological conditions in Jakarta. *Scientific Reports*, *13*(1), 5798.
4. Haq, M. A. (2022). SMOTEDNN: A novel model for air pollution forecasting and AQI classification. *Computers, Materials & Continua*, *71*(1).
5. He, H., & Luo, F. (2020, February). Study of LSTM air quality index prediction based on forecasting timeliness. In *IOP Conference Series: Earth and Environmental Science* (Vol. 446, No. 3, p. 032113). IOP Publishing.
6. Jena, C., Ghude, S. D., Kumar, R., Debnath, S., Govardhan, G., Soni, V. K., ... & Rajeevan, M. (2021). Performance of high resolution (400 m) PM2. 5 forecast over Delhi. *Scientific reports*, *11*(1), 4104.
7. Jin, N., Zeng, Y., Yan, K., & Ji, Z. (2021). Multivariate air quality forecasting with nested long short term memory neural network. *IEEE Transactions on Industrial Informatics*, *17*(12), 8514-8522.
8. Liu, H., Yan, G., Duan, Z., & Chen, C. (2021). Intelligent modeling strategies for forecasting air quality time series: A review. *Applied Soft Computing*, *102*, 106957.
9. Nath, P., Saha, P., Middya, A. I., & Roy, S. (2021). Long-term time-series pollution forecast using statistical and deep learning methods. *Neural Computing and Applications*, 1-20.
10. Xing, Y. F., Xu, Y. H., Shi, M. H., & Lian, Y. X. (2016). The impact of PM2. 5 on the human respiratory system. *Journal of thoracic disease*, *8*(1), E69.