RESEARCH PAPER

ANALYZING AND FORECASTING AMBIENT AIR QUALITY OF CHENNAI CITY IN INDIA

Imran Nadeem[1]*, **Ashiq M. Ilyas[1], P.S. Sheik Uduman[1]**

[1]Department of Mathematics, B.S Abdur Rahman Crescent Institute of Science and Technology, GST Road Vandalur 600048, Chennai, India
*Corresponding author: malikimrannadeem218@gmail.com

ABSTRACT. Regardless of the existing governmental and public preventive actions for surveillance and controlling the air quality in several regions of the Chennai city in India, the air quality does not meet the desired standard. In this regard, this study employs an ARMA/ARIMA modelling approach for forecasting Respirable Suspended Particulate Matter (RSPM), Sulphur dioxide (SO_2) and Nitrogen dioxide (NO_2) concentration for three most polluted sites in Chennai city. A total of nine univariate linear stochastic models have been developed, three for each of the stations which includes one for each of the specific pollutants in order to forecasts the concentration of each pollutant. The evaluation of the model statistics R^2 values and index of agreement values evince that a significant level of real-time forecasting of the pollutants can be achieved by employing the developed ARMA/ARIMA models. Moreover, the comparisons of actual air pollutant concentration have been carried out with the permissible limit as prescribed by the National ambient air quality standards (NAAQS) of India for assessing the level of pollution of all three locations.

CITATION: Imran Nadeem, Ashiq M. Ilyas, P.S. Sheik Uduman (2020). Analyzing and forecasting ambient air quality of Chennai city in India. Geography, Environment, Sustainability.
https://DOI-10.24057/2071-9388-2019-97

Conflict of interests: The authors reported no potential conflict of interest.

INTRODUCTION

In recent years, the massive decline in air standard is predominately attributed to a swift increase in industrialisation and density of vehicles that increase the air pollution in the environment. Reliable forecasts for the concentration of pollutants in the atmosphere are required with time and space for managing the air standard up to non-hazardous level and to formulate the air pollution control policy. Most of the air polluted countries have launched an active surveillance system to reduce major air pollutants in highly polluted areas of their dominion.

The air quality prediction for assessing air pollution can be established either by analytical or statistical models. Analytical models are usually more appropriate to make long-term forecasting and planning decisions (Juda 1989; Zannetti 1989). But, such models do not produce satisfactory results for air pollutant series characterised by rapid dynamics (Cats and Holtslag 1980; Jakeman et al. 1988; Raimondi et al. 1997). In addition, the analytical methods are unable to bring a quantitative assessment of the environmental pollution if the data of extra input factors such as temperature, wind, traffic features for evaluating the emission rate is not available (Petersen 1980; Benson 1989). In the cases when the data of extra input factors is unavailable, stochastic modelling provides an alternative approach to deal with the time series of air pollutants.

The forecasts of air quality can be attained either by air standard report from running monitoring sites after analysing the general pattern or by the air pollution predictor models. In stochastic models, ARMA/ARIMA approach is most suitable for linear time series assessment and forecasting (Box & Jenkins 1976). For regulatory bodies, forecasting is essential in order to apply counter techniques for maintaining the pollutant level in check.

The ARMA/ARIMA model, also known as the Box-Jenkins model, is widely acknowledged as one of the most efficient statistical methods for forecasting from time-series data (Adebiyi et al. 2014). ARIMA models are comparatively more robust and competent than other complex fundamental models with respect to short-term forecasting (Meyler et al. 1998). The ARIMA modelling is recognized for remarkable forecasting precision and for the suitable presentation of different kinds of time series in an effective manner for optimal model formulation (Khandelwal et al. 2015).

The ARMA/ARIMA models are employed to attain the best fit model from the past values of a time series. The use of ARMA/ARIMA forecasting model is not only restricted to air pollutant time series, but it is widely used in many other fields for forecasting. The maximum of ozone aggregation was predicted by Slini et al., (2002) using the past nine years of air standard data. Kumar et al., (2004) study applied the ARMA approach to get maximum daily ozone forecasts

at Brunei Darussalam. Duenas et al., (2005) use a stochastic model to forecasts ground-level ozone aggregation in urban and rural regions. Liu, (2009) forecasted day by day aggregation level using Box–Jenkins time series models and multivariate analysis. Numerous univariate ARMA/ARIMA models were developed by Sharma et al., (2009) for assessing and predicting a monthly maximum of the 24-hours average time series data for sulphur dioxide, nitrogen oxide and suspended particulate matter aggregation in an urban region of Delhi city. Kumar and Jain, (2010) developed univariate ARIMA models for predicting the daily mean of ambient air pollutant such as ozone, carbon monoxide, nitric oxide and nitrogen dioxide aggregation at an urban traffic location. ARIMA modelling was applied by Jian et al., (2012) to forecast submicron particle aggregation. Naveen & Anu, (2017) forecasts by employing ARIMA and SARIMA approach on the ambient air quality data of Thiruvananthapuram District of Kerala and arrive at a result that ARIMA model gave better forecasting in comparison to SARIMA model.

There is no evidence of forecasting air pollutants of metropolitan cities of India such as Chennai, Mumbai, Calcutta and Hyderabad. Therefore this study attempt to fill the gap using Chennai city as its case study.

With this motivation, our study best-fits ARMA/ARIMA models for forecasting pollutants level of three sites in Chennai city and then graphically represent the trend of air pollutants accompanied with comparison to NAAQS for January 2004 to December 2018. The rest of the paper is compiled as follow. In Section 2, we discuss the background of air pollution in India, particularly in Chennai. In Section 3, we discuss the collection of data and illustration of all the three sites. In Section 4, the general strategy regarding the formation of best-fit models is provided. Section 5 demonstrates the performance of the best-fitted models. Section 6 discusses the results obtained from the best-fitted models and their implications. Section 7 provides a conclusion reflecting on this research.

BACKGROUND OF AIR POLLUTION IN INDIA

The reports of past studies analyse that particulate matter is increasing at a rapid rate among all the air pollutants in India. The census of 2011 indicates that, out of the 640 districts in India, annual $PM_{2.5}$ concentration exceeded in 27% districts in 1998, 45% districts exceeded it in 2010 and 63% districts exceed it in 2016 compared to the annual standard value of 40µg/m³ (Guttikunda et al. 2019). Further, Venkataraman et al., (2018) affirms that 99.5% of these districts cross the limit of WHO guideline of 10µg/m³ (annual average) for $PM_{2.5}$ concentration in 2016, and about 50% of the population lives in an area where the annual average concentration of $PM_{2.5}$ is exceeded than 40µg/m³ of admissible limit as per NAAQS of India. The report of the World Health Organization (WHO 2014) stated that, of the top 20 most polluted cities in the world, 14 are in India. While the pollution level is not uniform in different cities all over India. The report further indicates that north India is worst polluted than south India, as none among these 14 cities is from south India. The north Indian cities like Uttar Pradesh, Delhi, Jharkhand, and Punjab are mostly polluted with quite higher amount of PM_{10} concentration as compared to other pollutants (Pant et.al. 2019). Most of the Indian cities have the only SO_2 pollutant in compliance with NAAQS (Guttikunda et.al. 2014). Delhi, the capital of India, is the worst-ranked city in term of air pollution in India (Kaushik et.al. 2019). The tremendous increase in the number of motorised vehicles is the major cause of pollution in Indian

cities (Dhyani et.al. 2017). Unlike the northern cities, the air pollution in a southern city like Chennai is not much higher than NAAQS. Sivaramasundaram and Muthusubramanian (2010) indicate that particulate matter (PM) level is the major air pollutant in Chennai and is more than the NAAQS at those urban sites where vehicular movement is highest. Guttikunda et al., (2015) study examined that the vehicle exhaust contributes about (34%), industries (21%), power plants (12%), road dust (9%), brick kilns (7%), domestic wastages (4%), and open waste burning (3%) to PM_{10} pollution in Chennai. The diesel exhausts contribute about 50% to PM_{10} and gasoline about 15% to PM_{10} level in Chennai (Srimuruganandam and Nagendra 2012).

DATA COLLECTION AND STUDY SITES

Chennai, the capital of Indian state, Tamil Nadu, is situated at 13.0827° N, 80.2707° E. Chennai city has Tropical savanna climate with dry summers and winters (Koppen climate classification) and is situated close to the southern coastal part of India. May and June are the hottest with a daily mean temperature of 38°C and December and January are coldest with a daily mean temperature of 21°C. The average annual precipitation falls down between October and December. The air quality in most of the regions of Chennai is decaying from the past decade. Anna Nagar (Fig. 1), a major residential area, lies in the north-western part of Chennai. It has good road and railway networks comparing to other parts of Chennai. It is situated about a distance of 10km from Chennai beach. Theagaraya Nagar is a very prosperous commercial and residential neighbourhood district of Chennai. It is one of the major business districts in Chennai. It is about 9 km away from the famous Marina beach. Kilpauk is a commercial (traffic intersection) area located in Chennai. It is about a distance of 8 km from the famous Marina beach and about 18km from Chennai airport. It has a good road and railway connectivity with other parts of Chennai City.

As of September 2018, three key air pollutants, Sulphur dioxide (SO_2) , Nitrogen dioxide (NO_2) and Respirable Suspended Particulate Matter (RSPM/PM_{10}) have been identified for continuous monitoring at above-mentioned stations like all the other stations in India. All other pollutants are also monitored, but not at all stations across the country. The pollutant is monitored for 24-hour (4-hourly sampling for gaseous pollutants and 8-hourly sampling for particulate matter) manually twice in a week to obtain 104 observations in a year (https://cpcb.nic.in/monitoring-network-3/) under the National Air Monitoring Programme (NAMP). In Chennai, 8 ambient air quality monitoring stations are running, and data is sampled manually once a day to cover two stations per day on all working days (http://tnenvis.nic.in/Database/TN-ENVIS_793.aspx). The data collected at these stations is descriptive rather than absolute. This approach is applied for forecasting when the long-term data records are available. An elementary requirement to use ARMA/ARIMA approach on the time series data is the continuity of data. Our study makes possible to employ directly ARMA/ARIMA approach for forecasting due to the absence of missing values in each of the time-series data for all the three sites. The data for each of the three sites Anna Nagar, Theagaraya Nagar and Kilpauk has been acquired from January 2004 to December 2018 from the Central Pollution Control Board (CPCB) of India, (www.cpcb.nic.in; accessed in April-2018). After collecting the data, the data have been split into two parts, the training and test data sets. The data set from January 2004 to December 2016 act as «training data set» and from
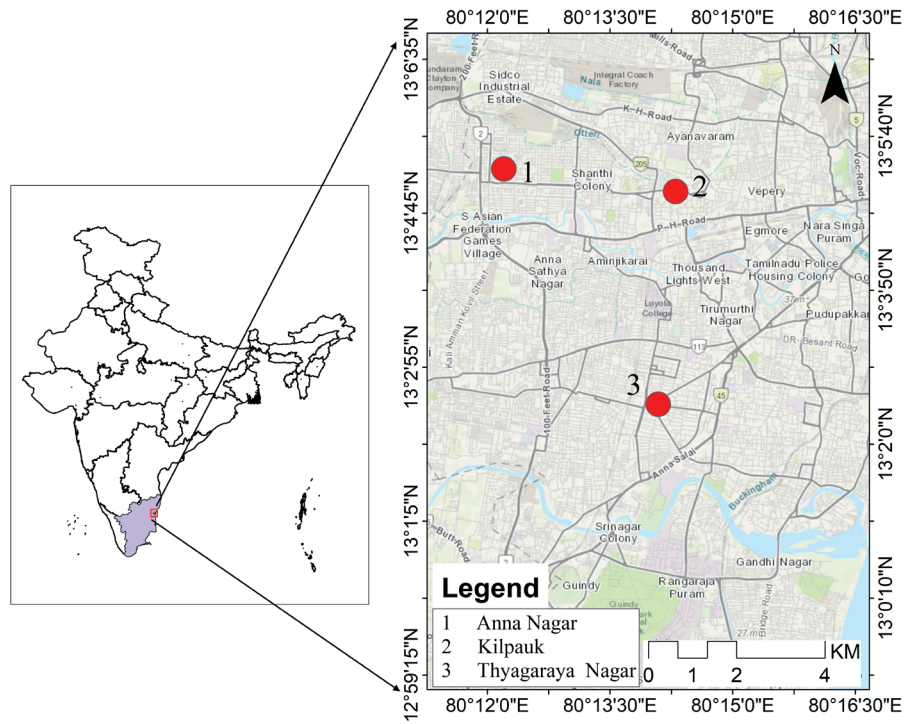
Fig. 1. Depicting the location of considered ambient air quality monitoring stations in Chennai on map of India

January 2017 to December 2018 acts as «test data set». The training data set is used to obtain the best fit model, while the test data set serves as an unobserved data set for comparing with the efficiency of the forecasting obtained from the best fit model of the training data set.

METHODOLOGY

The present study adopts a univariate linear stochastic ARMA/ARIMA modelling approach for forecasting the monthly average concentrations of each of the ambient air pollutants (RSPM, SO_2 and NO_2) for each of the three most polluted stations of the Chennai city. The basis of this study is to apply the ARMA/ARIMA approach for forecasting the air pollutants in an efficient manner. This approach is an integrated framework consisting of several interrelated steps to be applied until the best-fitted model is attained for forecasting as shown in figure 2.

Formulation of ARMA/ARIMA modelling

The ARIMA modelling brings out the predictable trend, variation and correlation from the observed data until a series of white noise is attained in ACF of residuals to indicate the best-fit model. The process is followed by disintegrating the time series into three constituents, autoregressive (AR), the integration (d; difference) and the moving average (MA) operators.

In practice, the formulation of the most suitable ARMA/ARIMA model is not convenient and requires four phases to be applied to the time-series data.

Model identification: In the preliminary phase, the time series is investigated for stationarity. If the series is non- stationary then after the conversion of series into stationary, the tentative values of non-seasonal AR and MA function are evaluated on the basis of plotting of Autocorrelation function (ACF) and Partial autocorrelation function (PACF).

Parameter estimation: After determining the tentative values of (p, d, q) parameters for AR, differential operator and MA, the linear coefficients of the models are evaluated using maximum likelihood or minimum least-squares method. AIC is defined by (Brockwell and Davis 2002) as

$$AIC = 2(v \, / \, m) - 2ln(L) \, / \, m \qquad (1)$$

BIC is given by (Schwarz1978) as:

$$BIC = vln(m) \, / \, m - 2ln(L) \, / \, m \qquad (2)$$

If the model is univariate, linear in parameters and the residuals are normally distributed, then the AICc is given by (Burnham and Anderson 2004) as:

$$AICc = AIC + \frac{2v^2 + 2v}{m - v - 1} \qquad (3)$$

where L denotes the likelihood function in Eqn. (1), (2) and v, m denote the number of variables and number of observations respectively in the Eqn. (1), (2) and (3). The idea of Portmanteau goodness-of-fit test is applied on tentative models for choosing such a model among numbers of tentative models that have the least values of AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and AICc (another version of Akaike information criterion). These statistical information criteria are estimators used to determine the tentative models for a given set of data. All of these three Information Criteria has its own benefits and drawbacks. Thus in our studies, we make use of all these criteria instead of relying upon any one of these to choose tentative models.

Validation: The proficiency of the selected ARIMA (p, d, q) model is determined by employing the certain statistical test after the best-fit model achieves the white noise (residuals do not have autocorrelation). Following this, two types of diagnostic tests are generally applied to determine the statistical competency of the selected best-model. The first test analyses the correlation of the residuals series by plotting the ACF of residual. Further, if the plot of the ACF of residuals is not correlated then the residuals are white noise. The second is the Chi-square statistics test that depends on the residual autocorrelations of first 25 lags in our studies, the Portmanteau goodness-of-fit test (Box et al. 1994). On testing both of these two criteria, if either of these two criteria is not fulfilled on a tentative model obtain from (step 2) then the re-estimation of the model parameters is needed to apply on other tentative models to test for validation until the model satisfying both conditions is achieved.
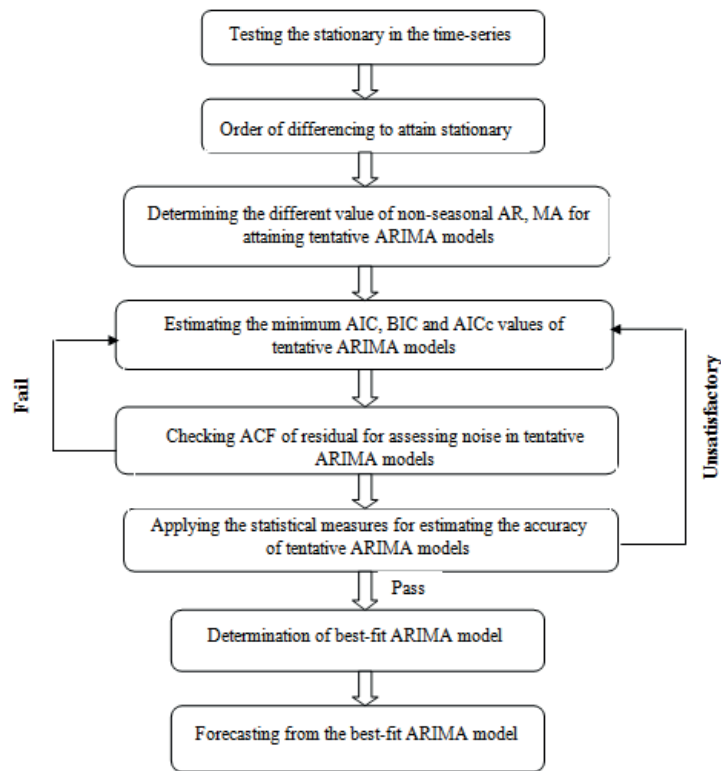
15

Fig. 2. Flowchart depicting the outline of the general methodology

Forecasting: After obtaining the best-fitted model from the trained data with good accuracy, the same best-fitted model is applied to the test data set of that series for attaining the forecasting. In this manner, a total of nine models have been developed, one for each specific pollutant for each of the three stations.

The ARMA/ARIMA modelling approach is applied only to a stationary time series. If a series is non-stationary, «logarithmic», «square root» or «power transformations» are applied for stabilizing the variance in the time series (Mills 1991). ARIMA approach suggested regular and seasonal differencing transformation for removing the non-stationary created by trend and seasonality. The difference operator "∇" and "∇" given by $\nabla x_t = x_t - x_{t-1}$ and $\nabla_s x_t = x_t - x_{t-s}$ are operated on the time series y_t.

The preliminary phase is to initiate the process of building the model after the time-series have been converted to stationarity by applying the differencing of suitable order. The pattern of ACF and PACF plots suggest the suitable tentative models by defining the behaviour, trends, stationary and order of AR and MA operator in the time series data (Tabachnik and Fidell 2005; Pankratz 1983). The ACF identifies the extent of linear dependence between the observations of the time series apart by lag[1]. The PACF plot helps in identifying the numbers of AR terms required for the model. A common representation of a AR(p) is

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \varepsilon_t \qquad (4)$$

where P determine the number of terms in the past required for forecasting the present value with random error term ε_t at the time $\beta_0, \beta_1, \beta_2... \beta_p$ having as coefficients.

A moving average MA(q) model having an order q, is one in which y_t depends only on the random error term following a white noise process (ε_t having zero mean and constant variance to lag q). A moving average MA (q) model with $\varphi_0, \varphi_1, \varphi_2...\varphi_q$ being the coefficients is generally represented by

$$y_t = \varphi_0 + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + ... + \varphi_q \varepsilon_{t-q} \qquad (5)$$

A combination of both AR(p) and MA(q) is referred to ARMA(p,q), depending upon *p* of its own past values and *q* of past values of the white noise distribution expressed by (Shumway and Stoffer 2006) for some constant α.

$$y_t = a + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \varphi_1 \varepsilon_{t-1} +$$
$$+ \varphi_2 \varepsilon_{t-2} + ... + \varphi_q \varepsilon_{t-q} + \varepsilon_t \qquad (6)$$

But if differencing is employed to time series to make it stationary then the resulted model is referred as ARIMA(p,d,q) where d indicates the order of differencing (Shumway and Stoffer DS 2006; Brockwell and Davis 2002).

Best fitted model for different monitoring sites

The selection of the final best fit model is based upon the given criteria: A stationary model having the least value of AIC, BIC and AICc are chosen among the tentative stationary models. And in addition to this, if ACF of residuals of the chosen model has white noise then it is selected as a best-fit model only when the numerical error on applying statistical tests is least compared to other tentative white noised models. Otherwise, rests of models from tentative stationary models have to be checked until one of them is best-fitted.

In each model, the Ljung-Box test is applied on the first 25 lag all at once for determining the p-value (significance level for comparison) and Q*-statistics. The value of p greater than 0.05 from the Ljung-Box for all the nine best-fit models suggests the acceptance of the null hypothesis, that the given models are the best fit. Similarly, the Q*-statistics is performed for testing the competency of the different tentative models for each of the time series. The Ljung- Box test Q*(m) statistic is characterized by asymptotic chi-square distribution with degrees of freedom. The null hypothesis Q*(m) is satisfied if the obtained model has a non-serial correlation and is computed as:

$$Q*(m) = n(n+2) \sum_{k=1}^{m} \frac{p_k}{n-1}$$

where n is the number of values in the data set, m is the maximum number of lags used in the test and P_k is the autocorrelation of data at lag k. The low values of Q*-

statistics suggest the validity of a model. The Q*m value is less than 20 for all the nine models, suggesting the adequacy of all the best-fit models.

Auto-correlation function of the residuals for best fitted ARIMA (p, d, q) models

The portmanteau test of Ljung-Box test is mostly applied to check the level of goodness of fitness of a time series model. If significant autocorrelation is not present in the residuals from the model, then the model is claimed to fulfil the test. Each of the series of RSPM, SO_2 and NO_2

for all the stations used the 156 observations (training data set) for the formulation of the model. All the sub-figures in Figure 3 for ACF of the residuals show that the individual residual autocorrelations are very small and are lying within

$$\pm \frac{2}{\sqrt{m}} = \pm 0.16$$

for $m=156$ significance bounds. It signifies ACF of the residuals of all the nine best-fit models that are represented in Figure 3 shows white noise character (as all of the values lie inside the confidence interval).
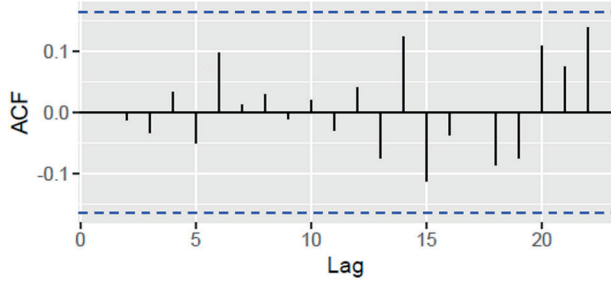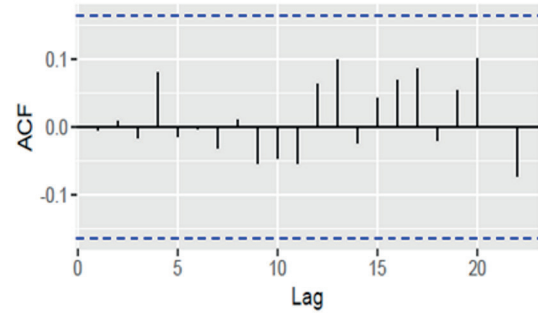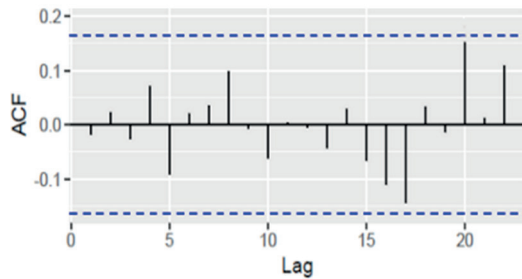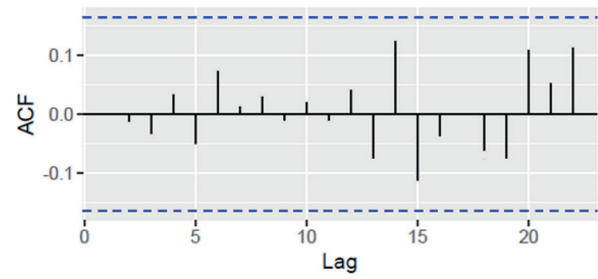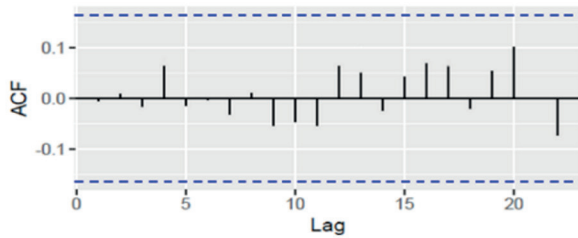

Fig. 3a

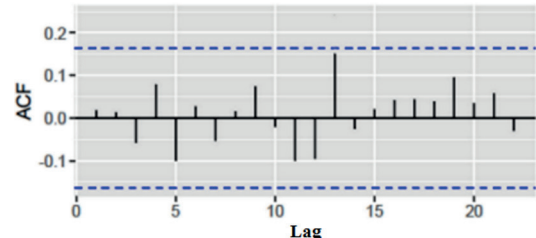

Fig. 3b

Fig. 3c

Fig. 3d
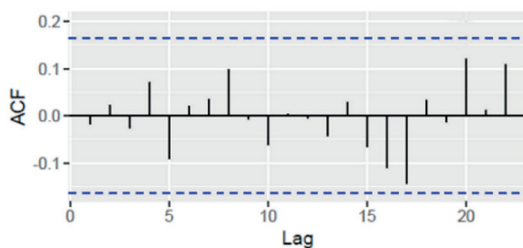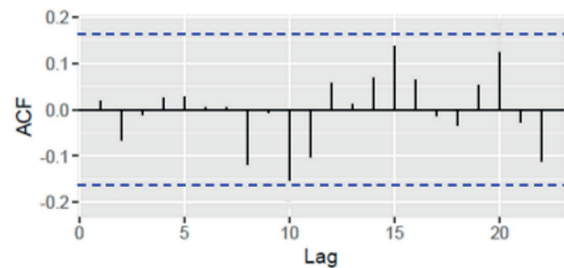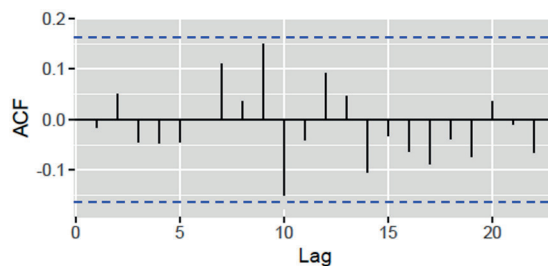

Fig. 3e

Fig. 3f

Fig. 3g

Fig. 3h

Fig. 3i

Fig. 3. ACF of the Residuals for the best-fitted ARMA/ARIMA model

Figures 3a-3c depicts the ACF of the Residuals for best fitted ARMA/ARIMA model of RSPM, and respectively for Anna Nagar. The ACF of the Residuals for best fitted ARMA/ARIMA model of RSPM, and for Theagaraya Nagar are shown in Figure 3d, 3e, 3f respectively. Figures 3g-3i presents the ACF of the Residuals for best fitted ARMA/ARIMA model of RSPM, SO_2 and NO_2 respectively for Kilpauk.

Assessment of the Performance of the best-fitted Models

The forecasting efficiency and reliability of best-fitted models are judged by employing the statistical techniques on the «test data set». Different techniques are employed to estimate the degree of accuracy and reliability of the time series forecasting models. The most commonly used method for estimating the degree of forecasting accuracy is to visualize by plotting the measured and forecasted values. The visualization of the plotting directly helps to analyse the extent up to which the performance of the model is convincing. However, this method of evaluation lacks objectivity. In order to make the numerical error analysis free from the subjectivity, the current studies use two quite effective statistical measures for assessing the forecasting efficiency of the developed models. The root mean square error (RMSE) and it's constituent, systematic $RMSE_S$ and unsystematic $RMSE_u$ proposed by Willmott (1981) and Willmott et al., (1985). The RMSE is given as, states the division of the total error into systematic and unsystematic where

$$RMSE_S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - O_i)^2}$$

and

$$RMSE_u = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}$$

The O_i and y_i represent the observed and predicted values respectively. Also $\hat{Y}_i = mO_i + b$, where and are the slope and intercept of the least square regression respectively on the observed parameters. Willmott (1981) and Willmott et al., (1985) also proposed the index of the agreement (d) as:

$$d = 1 - \frac{\sum_{i=1}^{n}(Y_i - O_i)^2}{\sum_{i=1}^{n}(|Y_i - \overline{O}| + |O_i - \overline{O}|)^2} \quad 0 \le d \le 1 \quad (7)$$

where $\overline{O}$ represent the mean of the observed values.

The index d evaluates the limit to which signs and the magnitude of the observed values about the $\overline{O}$ are associated to the forecasted deviations about $\overline{O}$ and determines the variation not only in O and Y but also in the proportionality of O and Y (Rao et al. 1985). The value of the index d ranges from 0.0 to 1.0. The former value implies no agreement while the latter defines the best agreement. The d can be regarded as standardized (in terms of the difference in the forecastings and observations about the observed mean) estimate of the mean square error. The index d was suggested by Willmott (1981) as a substitute to R (coefficient of correlation) an R^2 (coefficient of determination). The index d is a dimensionless and bounded technique with values near to one implies a strong agreement. However, Willmott and wicks (1980) analyses that the high or statistical significant values of R and R^2 may be inaccurate, as they frequently are not associated with the size of variation between O_i and y_i.

RESULTS AND DISCUSSION

The results shown in Table 1 summaries the statistical analysis of the training data set and test data set of all the nine different models. One model for each of the pollutant $RSPM/PM_{10}$, SO_2 and NO_2 at the three stations are developed in our studies. In the evaluation of model statistics, the values of R^2,d, RMSE lie between the ranges of 0.89 to 0.94, 0.87 to 0.91 and 0.12 to 0.43 respectively for forecasting all the nine models. The range of d value suggests that there exist a good level of agreement between the training data set and test data set for all the nine models. Moreover, the range of R^2 value signifies that at least 89% of the forecasting evaluated in all the models is free from the errors. The low percentage of errors in the models suggests that the forecasting is quite convincing.

The forecasting results acquired in the present study (Table 1) are compared with the results of Sharma et al., (2009) study who applied an identical approach for developing ARMA/ARIMA models and for determining the statistical efficiency to forecasts the ambient air quality of Delhi City. Our studies obtain at least 89% of forecasting free from error for all of the models in comparison to at least 86.93% forecasting accuracy of (Sharma et al. 2009) for all the models.

The Central Pollution Control Board (CPCB) of India has installed hundreds of ambient air quality surveillance centres across the country and made it mandatory to monitor at least three pollutants(RPSM, SO_2 and NO_2) at each of the sites to keep the air pollutants level in control. It has specified certain permissible limits of the different pollutants given by NAAQS in µg/m³ (Table 2) on the basis of annual pollutant concentration for comparing with the concentration of the actual pollutant of a site for assessing pollution level.

The annual average concentration of pollutants from January 2004 to December 2018 for five locations has been compared

Table 1. Model evaluation statistics for best-fitted models

Stations	Pollutant	ARIMA model	AIC	BIC	AICc	d		R^2		RMSE	
						Fit	Forecast	Fit	Forecast	Fit	Forecast
Anna Nagar	RSPM	3,1,2	502.5	504.3	510.8	0.88	0.90	0.89	0.91	0.22	0.28
	SO_2	3,0,0	712.3	713.4	721.3	0.88	0.87	0.87	0.89	0.19	0.23
	NO_2	1,0,1	603.3	604.2	616.2	0.89	0.89	0.90	0.92	0.38	0.34
Theagaya Nagar	RSPM	1,2,2	721.1	722.6	732.8	0.90	0.90	0.92	0.92	0.44	0.37
	SO_2	1,1,2	593.2	594.4	602.3	0.89	0.88	0.90	0.90	0.33	0.29
	NO_2	1,0,2	857.3	858.9	863.6	0.90	0.91	0.91	0.91	0.39	0.43
Kilpauk	RSPM	2,1,1	934.9	935.3	946	0.90	0.90	0.92	0.93	0.40	0.33
	SO_2	2,1,3	576.4	578.3	587	0.91	0.89	0.93	0.94	0.16	0.12
	NO_2	1,1,1	673	674.3	682.4	0.90	0.88	0.94	0.94	0.31	0.27

Table 2. The main characteristics of soils in the different landscape types (*Eh, pH, ρ*) and the depth of pitting corrosion in the metal of gas pipelines of the inter field collector

Types of Location	Annual Permissible limits in		
	RPSM/PM$_{10}$	SO$_2$	NO$_2$
Industrial areas	60	50	40
Residential and other areas	60	50	40
Ecologically sensitive areas	60	20	30
Anna Nagar	102	9	21
Theagaraya Nagar	105	11	23
Kilpauk	104	11	22
Adyar	51	8	14
Nungambakkam	92	11	20

with NAAQS (Table 2). Adyar and Anna Nagar are residential areas while Theagaraya Nagar, Kilpauk, and Nungambakkam are commercial (traffic intersection) areas of Chennai city. The mean annual concentration of *RSPM/PM$_{10}$* is at least 1.5 times higher than the permissible limit of 60*μg/m³* at all the five considered stations except Adyar from January 2004 to December 2018 (Table 2). But, unlike *RSPM/PM$_{10}$,* the mean annual concentration level of both *SO$_2$* and *NO$_2$* at all the five given stations over the period from 2004 to 2018 was well within the admissible limits of 50*μg/m³* and 40*μg/m³* respectively as specified by NAAQS (Table 2). The study of Rajamanickam & Nagan (2018) reports that *RSPM/PM$_{10}$* is not only the main contributor to the pollution, but also exceed the limit as prescribed by NAAQS in all the regions of Chennai. While *SO$_2$* and *NO$_2$* is well within the limit of NAAQS at all the regions of Chennai. The mean annual ambient air quality of three stations considered in our studies, from 2004 to 2018 has been compared with the mean annual NAAQS in order to assess the pollution level at each site.

The figures 4, 5 and 6 clearly shows that the mean annual concentration range of   for Anna Nagar, Theagaraya Nagar and Kilpauk lies within 71*μg/m³* to 135*μg/m³* 79*μg/m³* to 128*μg/m³* and 73*μg/m³* to 160*μg/m³* respectively over the period of 2004 to

2018. These ranges show that the mean annual concentration of was too much higher for all three stations over the given period of time than the permissible limits of 60*μg/m³* as prescribed by NAAQS.

The figures 4, 5 and 6 depict that the mean annual concentration of *SO$_2$* in Anna Nagar, Theagaraya Nagar and Kilpauk ranges from 6*μg/m³* to 14*μg/m³*, 7*μg/m³* to 19*μg/m³* and 7*μg/m³* to 19*μg/m³* respectively. Further, the above-mentioned figures signify that the mean annual concentration of *NO$_2$* in Anna Nagar, Theagaraya Nagar and Kilpauk lies in the range from 15*μg/m³* to 37*μg/m³*, 17*μg/m³* to 83*μg/m³* and 16*μg/m³* to 30*μg/m³* respectively. These ranges suggest that the mean annual concentration level of *SO$_2$* and *NO$_2$* of three mentioned stations over the period of 2004 to 2018 was well within the permissible limits of 50*μg/m³* and 40*μg/m³* respectively as specified by NAAQS.

The report of National Ambient Air Quality Monitoring of India,2014-15 states that the low concentration of *SO$_2$* and *NO$_2$* in Anna Nagar, Theagaraya Nagar and Kilpauk areas of the Chennai city over the years has been influenced by local atmospheric circulation that regularly rushes from the sea into these areas. While the main reason behind the quite high level of in these
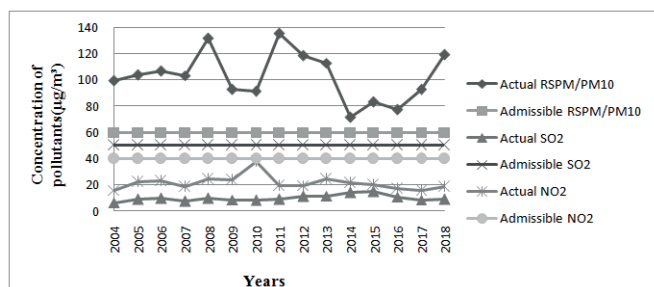

Fig. 4. Temporal variability of the mean annual concentration of the RSPM, *SO$_2$* and *NO$_2$* pollutants (*μg/m³*) with an admissible limit of Anna Nagar from 2004 to 2018

Fig. 5. Temporal variability of the mean annual concentration of the RSPM,*SO$_2$* and *NO$_2$* pollutants (*μg/m³*) with an admissible limit of Theagaraya Nagar from 2004 to 2018

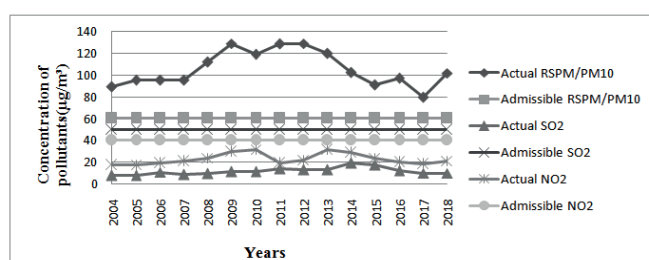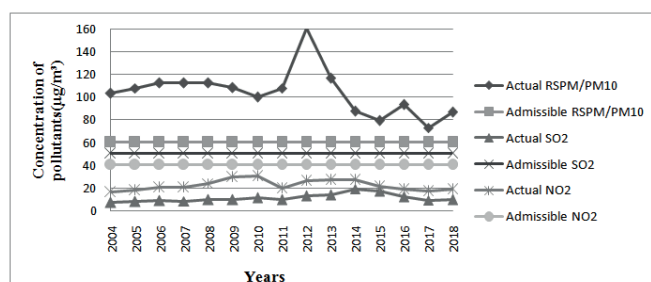**Fig. 6. Temporal variability of the mean annual concentration of the RSPM, *SO$_2$* and *NO$_2$* pollutants (*μg/m³*) with an admissible limit of Kilpauk from 2004 to 2018**

areas has been constant emission from a growing number of vehicles and dust from traffic. The number of motorized vehicles rises to 24-fold since 2005, and private vehicles now constitute 55% of daily all-person trips (Basic Road Statistics of India, Urban Infrastructure: Twelfth Five Year Plan). Every day, at least 700 new vehicles go to the Chennai streets triggering the level of pollution (Sharma et al. 2019).

The fairly higher concentration of RSPM and under controlled concentrations of SO_2 and NO_2 of mentioned stations in Chennai City in comparison with the permissible limits of these pollutants as defined by NAAQS suggests that the forecasting of air pollutants is requisite for monitoring and checking the further air pollution.

CONCLUSIONS

The present study has introduced an application of ARMA/ARIMA modelling approach that yields the convincing results for forecasting the ambient air quality of Chennai city in India. A total of nine models are selected from the number of tentative models, one for each

pollutant at each of the three sites after attaining the white noise in ACF of residual and fulfilling certain other criteria. These models are quite beneficial for forecasting air quality as the forecasting assessed from all nine different models is almost free from errors. The level of accuracy suggests that the present forecasting approach is quite convincing but still more efforts are needed to improve the efficiency of the forecasting. The study shows that SO_2 and NO_2 is under NAAQS, but RSPM/PM_{10} is quite higher than NAAQS at all the three stations. The tremendous increase in numbers of vehicles is the major source of the excessive level of RSPM/PM_{10} in Chennai. The actual pollutants level on comparing with a permissible limit of National ambient air quality standards of India and forecasting accuracy of ARMA/ARIMA best-fitted models of three sites provide an inclusive approach for framing a suitable policy to handle the degrading level of air standard in Chennai City.

This study can be extended by analysing the impact of air pollutants on atmospheric properties and human health in India (Chubarova et al. 2019) ◼

REFERENCES

Adebiyi A.A., Adewumi A.O. and Ayo C.K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. Journal of Applied Mathematics, 2(1), 1-7, DOI: 10.1155/2014/614342.

Box GEP and Jenkins G.M. (1976). Time series analysis, forecasting and control, revised ed. Holden-Day, San Francisco.

Box GEP. Jenkins GM. and Reinsel, GC. (1994). Time series analysis: Forecasting and control (3rd ed.). Englewood Cliffs, New Jersey: Prentice Hall.

Benson P. E. & Pinkerman, K.O. (1984). CALINE4, a dispersion model for predicting air pollution concentration near roadways. State of California, Department of Transportation, Division of Engineering Services, Office of Transportation Laboratory.

Brockwell J.B. and Davis R.A. (2002). Introduction to time series and forecasting. New York: Springer.

Burnham K.P. and Anderson D.R. (2004). Multimodel inference: understanding AIC and BIC in Model Selection. Sociological Methods & Research, 33, 261-304, DOI: 10.1177/0049124104268644.

Cats G.J. and Holtslag A.A.M. (1980). Prediction of air pollution frequency distribution, The lognormal distribution. Atmospheric Environment, 14, 255-258.

Chubarova N.E., Androsova E.E., Kirsanov A.A., Vogel B., Vogel H., Popovicheva O.B. and Rivin G.S. (2019). Aerosol And Its Radiative Effects During The Aeroradcity 2018 Moscow Experiment. Geography, Environment, Sustainability, 12(4), 114-131, DOI: 10.24057/2071-9388-2019-72.

Duenas C., Fernandez M.C., Canete S., Carretero J. and Liger E. (2005). Stochastic model to forecast ground-level ozone concentration at urban and rural areas. Chemosphere. 61(10), 1379-1389.

Dhyani R., Sharma N. and Maity A.K. (2017). Prediction of PM2.5 along urban highway corridor under mixed traffic conditions using CALINE4 model. Journal of Environmental Management, 198, 24-32.

Guttikunda S.K., Goel R., Mohan D., Tiwari G. and Gadepalli R. (2015). Particulate and gaseous emissions in two coastal cities – Chennai and Vishakhapatnam, India. Air Quality, Atmosphere & Health, 8(6), 559-572.

Guttikunda S.K., Nishadh K.A. and Jawahar P. (2019). Air pollution knowledge assessments (APnA) for 20 Indian cities. Urban Climate, 27, 124-141.

Jakeman A.J., Simpson R.W. and Taylor J.A. (1988). Modelling distributions of air pollutant concentrations — III. The hybrid deterministic-statistical distribution approach. Atmospheric Environment, 22, 163-174.

Jian L., Zhao Y., Zhang M.B. and Bertolatti D. (2012). An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. Science of the Total Environment, 426, 336-345.

Juda K. (1989). Air pollution modelling. In P. N. Cheremisinoff (Ed.), Encyclopedia of environmental control technology, air pollution control, USA: Gulf Publishing Company 2, 83-134.

Kaushik G., Chel A., Patil S. and Chaturvedi S. (2019). Status of Particulate Matter Pollution in India: A Review. Handbook of Environmental Materials Management, 167-193.

Khandelwal I., Adhikari R. and Verma G. (2015). Time series forecasting using hybrid Arima and ANN models based on dwt decomposition. Procedia Computer Science, 48, 173-179.

Koppen climate classification | climatology. Encyclopedia Britannica. Archived from the original on 2020-02-19. Retrieved 2020-02-19.

Kumar K., Yadav A.K., Singh M.P., Hassan H. and Jain VK. (2004). Forecasting daily maximum surface ozone concentrations in Brunei Darussalam – An ARIMA modelling approach. Journal of Air Waste management Association, 84, 809-814.

Kumar U. and Jain V.K. (2010). ARIMA forecasting of ambient air pollutants (O_3, NO, NO_2 and CO). Stochastic Environmental Research and Risk Assessment, 24, 751-760.

Liu P.W.G. (2009). Simulation of the daily average PM_{10} concentrations at Ta-Liao with Box–Jenkins time series models and multivariate analysis. Atmospheric Environment, 43, 2104-2113.

Meyler A., Kenny G. and Quinn T. (1998). Forecasting Irish inflation using ARIMA models, 1-48.

Mills T.C. (1991). Time series techniques for economists. Cambridge: Cambridge University Press.

National Ambient Air Quality Monitoring NAAQMS/……/2014-2015, Retrieved 19 November, 2018. [online] Available at: www.indiaenvironmentportal.org.in/files/file/NAAQ Status_ Trend_Report_2012.pdf. [Accessed 20 May 2020].

Naveen V. and Anu N. (2017). Time Series Analysis to Forecast Air Quality Indices in Thiruvananthapuram District, Kerala, India. International Journal of Engineering Research and Application, 7(6), 66-84.

Pankratz A. (1983). Forecasting with Univariate Box–Jenkins models: concepts and cases. Wiley, New York, DOI: 10.1002/9780470316566.

Pant P., Lal R.M., Guttikunda S.K., Russell A.G., Nagpure A.S., Ramaswami A. and Peltier R.E. (2019). Monitoring particulate matter in India: Recent trends and future outlook. Air Quality, Atmosphere & Health, 12(1), 45-58.

Petersen W.B. (1980). User's guide for HIWAY-2: A highway air pollution model. US Environmental Protection Agency.

Raimondi P.M., Rando F., Vitale M.C. and Calcara A.M.V. (1997). Short-time fuzzy DAP predictor for air pollution due to vehicular traffic. WIT Transactions on Ecology and the Environment, 19.

Rajamanickam R. and Nagan S. (2018). Assessment of air quality index for cities and major towns in Tamil Nadu, India. Journal of Civil and Environmental Engineering, 8(2).

Rao S.T., Sistla G., Petersen W.B., Irwin J.S. and Turner D.B. (1985). Evaluation of the performance of RAM with the regional air pollution study database. Atmospheric Environment, 19, 229-245.

Road Statistics of India. [online] Available at: www.indiaenvironmentportal.org.in/files/file/basic%20road%20statistics%20of%20india.pdf [Accessed 20 June 2019].

Schwarz G. (1978). Estimating the dimension of a model. The Annals of Statistics 6(2), 461-464.

Sivaramasundaram K. and Muthusubramanian P. (2010). A preliminary assessment of and TSP concentrations in Tuticorin India. Air Quality, Atmosphere and Health, 3(2), 95-102.

Slini Th., Karatzas K., Moussiopoulos N. (2002). Statistical analysis of environmental data as the basis of forecasting: an air quality application. Science of the Total Environment, 288(3), 227-237.

Sharma P., Chandra A. and Kaushik S.C. (2009). Forecasts using Box–Jenkins models for the ambient air quality data of Delhi City. Environmental Monitering and Assessment, 157(1-4), 105-112.

Sharma R. Kumar R. Sharma DK. Priyadarshini I. Pham BT. Bui DT. and Rai S. (2019). Inferring air pollution from air quality index by different geographical areas: case study in India. Air Quality, Atmosphere & Health, 1-11.

Shumway R.H., Stoffer D.S. (2006). Time series analysis and its applications – with R examples. Springer Science, Business Media, LLC.

Srimuruganandam B. and Nagendra SMS. (2011). Characteristics of particulate matter and heterogeneous traffic in the urban area of India. Atmospheric Environment, 45(18), 3091-3102.

Srimuruganandam B. and Nagendra S.M.S. (2012a). Source characterization of and mass using a chemical mass balance model at urban roadside. Science of the Total Environment, 433, 8-19.

Tabachnik B.G. and Fidell L.S. (2005). Using multivariate statistics, 5th edition. Pearson Int. Edition, Boston.

Urban Infrastructure: Twelfth Five Year Plan (2012–2017). [online] Available at: www.planningcommission.gov.in/plans/planrel/12thplan/pdf/12fyp_vol2.pdf [Accessed 27 August 2017].

Venkataraman C., Brauer M., Tibrewal K., Sadavarte P., Ma Q., Cohen A., Chaliyakunnel S., Frostad J., Klimont Z.,Martin R.V., Millet D.B., Phillip S., Walker K. and Wang S. (2018). Source influence on emission pathways and ambient PM2.5 pollution over India (2015–2050). Atmospheric Chemistry and Physics Discussions, 18, 8017-8039.

Willmott C.J. and Wicks D.E. (1980). An empirical method for the spatial interpolation of monthly precipitation within California. Physical Geography, 1, 59-73.

Willmott C.J. (1981). On the validation of models. Physical geography, 2(2), 184-194.

Willmott C.J., Ackleson S.G., Davis R.E., Feddema J.J., Klink K.M. and Legates D.R. (1985). Statistics for the evaluation and comparison of models. Journal of Geophysical Research, 90, 8995-9005.

World Health Organization. (2014). Seven million premature deaths annually linked to air pollution, World Health Organization, 25 March 2014, viewed on 15 Jan 2016. [online] Available at: www.who.int/mediacentre/news/releases/2014/air-pollution/en/ [Accessed 20 May 2020].

Zannetti P. (1989). Simulating short-term, short-range air quality dispersion phenomena. In Encyclopedia of environmental control technology, Gulf Publishing Company Houston.