ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2017; 3(8): 443-447
www.allresearchjournal.com
Received: 04-06-2017
Accepted: 05-07-2017

Tripti Singh Rajput
Department of Civil Engineering (Infrastructure), Birla Institute of Technology & Science, Pilani Rajasthan, India

Nandini Sharma
Department of Civil Engineering (Infrastructure), Birla Institute of Technology & Science, Pilani Rajasthan, India

Correspondence
Tripti Singh Rajput
Department of Civil Engineering (Infrastructure), Birla Institute of Technology & Science, Pilani Rajasthan, India

Multivariate regression analysis of air quality index for Hyderabad city: Forecasting model with hourly frequency

Tripti Singh Rajput and Nandini Sharma

Abstract
The present study gives a description on air quality index (AQI) for the major city of India i.e. Hyderabad. Major parameter considered for AQI computation are NO, NO_2, NO_x, SO_2, Ambient temperature, relative humidity, bar pressure, solar radiation, wind speed, wind direction, benzene, toluene, xylene, $PM_{2.5}$ and rack temperature. An approach to assess and represent air quality status through an Air Quality Index (AQI), in Hyderabad city is done by representing the variation of AQI with a multivariate regression model. The utility of this study lies for major metropolitan cities, where different types of activities, viz. industrial, commercial and residential are in progress, on a short and a long term basis this model can be useful for better forecasting of air quality parameters. To make the index more informative, air quality status is classified into five different categories, viz. Clean, Moderate, Poor, Bad and Dangerous. Long term air quality indices are then calculated using hourly basis data for Hyderabad city which is a metropolitan city of India and is fast developing in terms of different climatologically defined features. The included application of this formal analysis includes accounting for atmospheric processes, ambient measurements, emissions characterization, air quality modeling of emissions to ambient concentrations, and characterization of human and ecological responses to ambient pollutant exposure. There is a need for new management strategy that would expand the current practice of accountability that relates emission reductions and attainment of air quality derived from air quality criteria and standards. Conceptually, achievement of accountability would establish goals optimizing risk reduction associated with pollution management.

Keywords: AQI, activity zones, pollutant standards, multivariate regression model, quality parameter forecasting

1. Introduction

An air quality index is one of the important tools available for analyzing and representing air quality status uniformly. The Air Quality Index (AQI) can be used as a measure to assess the relative change in the concentrations of groups of pollutants in two situations. The two situations may represent either two time periods or two regions. The relative change may also be with respect to the concentrations of pollutants and respective stipulated standards. The index should be based on stipulated guidelines and take into account each and every aspect of air pollution. The determination of the overall status of air pollution, however, may result in raising unnecessary alarms if a less polluted area is declared as highly polluted or vice versa. Standard air quality index development should take into consideration the aspects, like consideration of major pollutants, scope for incorporation of additional features, uniformity, inbuilt provision for consideration of standards and episodic conditions, acceptable tolerances with respect to precision, consistency in prediction, spatial and temporal variations and forecasting capabilities. There have been attempts to develop air quality indices, which present a clear and true picture of air quality. Researchers developed an index based on sulfur dioxide (SO2) concentrations and coefficients of haze. Babcock developed an index known as Pindex based on Particulate Matter, oxides of sulfur (SOx) and nitrogen (NOx), carbon monoxide (CO) and hydrocarbons. Swamee and Tyagi presented air quality index based on aggregation of air pollution sub-indices. Specific pollutants, industrial emissions and inter-urban air quality, these three major aspects of air quality were taken into consideration in most of the studies for air quality quantification.

A healthy city is a city designed with the consideration of an environmental place inhabited by people who are dedicated to the minimization of waste output and pollution, among other things. The gaseous blanket above the earth surface provides the basic elements for survival of all lives on the Earth, such as nitrogen, oxygen, and carbon dioxide, along with various gaseous materials such as inert gases helium, neon, and argon, etc. However, due to various natural (e.g., forest fires, volcano eruptions, wind) and anthropogenic (e.g., industrial activity, transportation, deforestation) activities, many components such as SPM, (pollen grains, aerosols, bacteria, viruses, and various tiny particles suspended in the atmosphere), NO2 (nitrogen oxides), SO2 (sulfur oxides), carbon monoxide (CO; due to incomplete combustion of fuels) get released to the atmosphere. In due course of time, if no prevention measure is taken it would reach a harmful level. Air pollution-related problems have resulted in an increased public awareness of the air quality in both developing and developed countries. 1 The high concentration of the pollutants can be life threatening, causing breathing difficulties or other troubles such as headaches and dizziness. The maximum permissible limit per India's Central Pollution Control Board (CPCB) is given by certain quality standards. Air quality index (AQI) is an important factor for the general public to easily understand how bad or good the air quality is for their health and to assist in data interpretation for decision-making processes related to pollution mitigation measures and air quality management. Basically, the AQI is defined as an index or rating scale for reporting daily combined effect of ambient air pollutants recorded in the monitoring sites.2 Also, it helps to track changes over time and to view and examine the balance development of the city to compare different cases. The AQI has been computed through a method used by U.S. Environmental Protection Agency (USEPA 1999) for different criteria pollutants: RSPM, NO_2, SO_2 etc.

Fine particles, (particulate matter [PM] with aerodynamic diameter to 2.5 m in diameter, or $PM_{2.5}$ have been linked to various health effects including asthma and cardiac arrest. The U.S. Environmental Protection Agency (EPA) promulgated the annual and daily $PM_{2.5}$ National Ambient Air Quality Standard (NAAQS) in 1997 and again revised it in 2006. The revised $PM_{2.5}$ NAAQS consists of a 24-hr daily standard of 35 g/m^3 and an annual standard of 15 g/m^3. Historically, most daily air quality forecasts for a region have been based on a combination of statistical methods and human judgment using point measurements at observational sites. In contrast, the application of a three-dimensional air quality model can provide spatially and temporally resolved forecasts, but its use in real-time forecasts in the past was often limited by computational speed. Because of increases in computational speeds and the maturity of air quality models, the past decade has witnessed the emergence of model-based forecast guidance as an increasingly relevant component of air quality forecasting for ozone and $PM_{2.5}$ concentrations. The increased availability of forecast guidance based on photochemical models had stimulated interest in the assessment of the performance of these models when applied in a forecasting mode. This study reports the application and evaluation of the performance of the Community Multi scale Air Quality (CMAQ) model in predicting daily $PM_{2.5}$ air quality over the eastern United States over a 2 year period.

A lot of detailed studies have been done in developed countries for most of the part of the country. This has led these countries to control and regulate air quality standard with greater accuracy. For developing countries like India, there is a need for comprehensive models and assessment for these high air quality pollutant issues. This study aims to develop an accurate framework for one of the most significantly polluted city of India by using multi correlation approach. Multivariate regression model is one of those models that considers non-comparable parameter with different dimensionality and let us comment on accuracy of the model by evaluation of R squared value of percent fit.

2. Study Area & Data collection

Data for various air quality parameters are collected from authorized department (state and central), Central pollution control board (CPCB). The board has recently. All laboratories are equipped with high precision laboratory equipment for analysis of various environmental samples. Data processing in the present study is conducted through analysis protocols with an consideration for zero error policy.

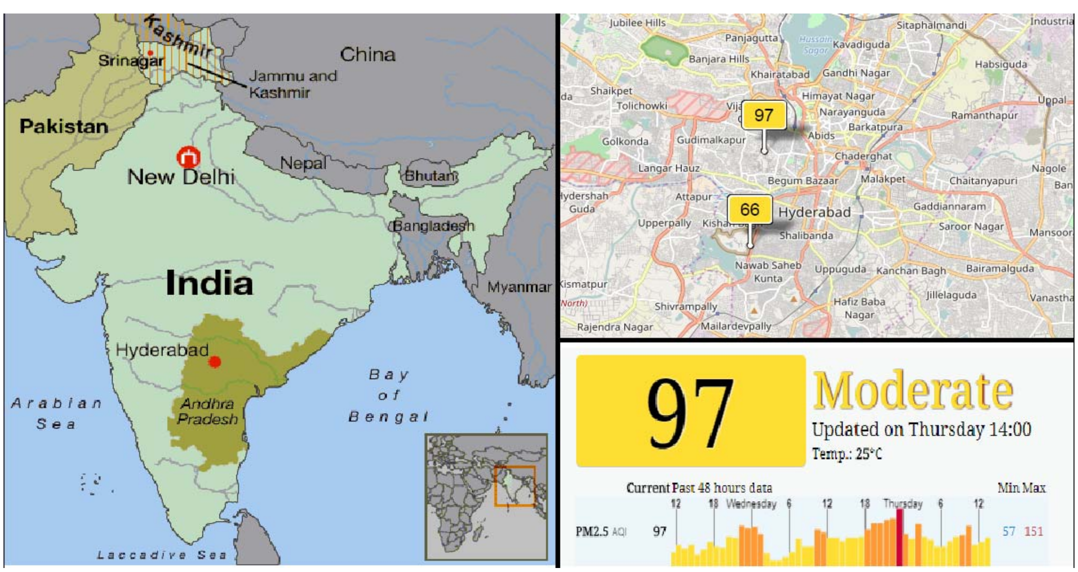

Fig 1: Hyderabad City as Study Area.

Hyderabad is the capital of the southern Indian state of Telangana and de jure capital of Andhra Pradesh. Occupying 650 square kilometers (250 sq mi) along the banks of the Musi River, it has a population of about 6.7 million and a metropolitan population of about 7.75 million, making it the fourth most populous city and sixth most populous urban agglomeration in India. At an average altitude of 542 metres (1,778 ft), much of Hyderabad is situated on hilly terrain around artificial lakes, including Hussain Sagar—predating the city's founding—north of the city centre.

Hyderabad has a tropical wet and dry climate bordering on a hot semi-arid climate. The annual mean temperature is 26.6 °C (79.9 °F); monthly mean temperatures are 21–33 °C (70–91 °F). Summers (March–June) are hot and humid, with average highs in the mid-to-high 30s Celsius; maximum temperatures often exceed 40 °C (104 °F) between April and June. The coolest temperatures occur in December and January, when the lowest temperature occasionally dips to 10 °C (50 °F). May is the hottest month, when daily temperatures range from 26 to 39 °C (79–102 °F); December, the coldest, has temperatures varying from 14.5 to 28 °C (57–82 °F).

Heavy rain from the south-west summer monsoon falls between June and September, supplying Hyderabad with most of its mean annual rainfall. Since records began in November 1891, the heaviest rainfall recorded in a 24-hour period was 241.5 mm (10 in) on 24 August 2000. The highest temperature ever recorded was 45.5 °C (114 °F) on 2 June 1966, and the lowest was 6.1 °C (43 °F) on 8 January 1946. The city receives 2,731 hours of sunshine per year; maximum daily sunlight exposure occurs in February.

As of January 15, 2016 air quality index (AQI) measured by the many agencies working in the study area. While SAFAR (System of Air quality, Weather Forecasting and Research), a Government of India-recognized institute, put Hyderabad's AQI at 80 - which is 'good/satisfactory' by its standards - a well-known international portal, aqicn.org, declared on its website that the city's air was clearly "unhealthy". The AQI figure, according to them was 158, which implied that "everyone may begin to experience health effects, members of sensitive groups may experience more serious health effects". Worse, Plume Labs - a French company best known for real-time monitoring of air quality worldwide - concluded that Hyderabad, on that particular day, was very highly polluted according to the AQI that read 105. "This means that the air reaches a very high level of pollution.

3. Methodology

3.1 Index computation

The estimation of AQI is through USEPA methods using observed monthly concentration of air pollutant. Maximum concentration and minimum concentration month wise and year wise, along with existing concentrations were considered for the index computation. Equation 1 is followed for computation of AQI based on inclusive assessment of air quality at a given site.4 Subindices for all monitoring stations are computed using equation 1, then an overall index for each station is computed and, finally, the highest overall maximum index becomes the final index for the city. The final index is therefore the highest value of the subindices for each component (pollutant). The overall AQI is now determined on the basis of the AQI for pollutant "p" and highest among them is stated as the overall AQI for that particular year or month. The breakpoint and sub-index values for individual pollutant are set for certain standards.

$$I_p = \{(I_h - I_{low}) \times (C_p - Bp_{low}) / (Bp_h - Bp_{low})\} + I_{low}$$

The sub-indices and breakpoint concentrations in the formula are made according to Indian National Ambient Air Quality Standard. Good (0–100): Air quality is acceptable; however for some pollutants there may be a moderate health concern for a very small number of people. Moderate (101–200): members of sensitive groups may experience health effects like as a runny nose, sore throat, cough, and allergies. Unhealthy for sensitive or Poor (201–300): members of sensitive groups suffering from respiratory diseases like asthma, bronchial diseases, emphysema, pneumonia, and cardiovascular diseases may experience more serious health effects. Very Poor (301–400): Triggers health changes, everyone may experience more serious health effects with physiological stress on individuals with cardiovascular and respiratory disease; possibly increased mortality. Hazardous or Severe (401–500): Triggers health warning of emergency conditions indicating severe ill effects and death.

3.2 Multivariate Regression Modeling

In the study a statistical modeling called regression analysis is used. In this process the statistical process for estimating the relationships among variables is analyzed. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or 'predictors'). More specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function which can be described by a probability distribution. A related but distinct approach is necessary condition analysis (NCA), which estimates the maximum (rather than average) value of the dependent variable for a given value of the independent variable (ceiling line rather than central line) in order to identify what value of the independent variable is necessary but not sufficient for a given value of the dependent variable. Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables. However this can lead to illusions or false relationships, so caution is

advisable; for example, correlation does not imply causation.
The following formulations are used for the calculations of the coefficient sand residues of the regression model:

$$Y_i = \alpha + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \varepsilon_i \quad \text{Eq.1}$$
$$Y_i = E(Y|X_i) + \epsilon_i \quad \text{Eq. 2}$$
$$\hat{\sigma} = \sqrt{\textstyle\sum_i r_i^2/(n-p-1)} \quad \text{Eq. 3}$$

4. Results and Discussions
A total of seven independent parameters are used for finding correlation with six dependent parameters taken separately one at a time. Ozone forecasting model shows acceptable results with R squared values of 83.03% which is well above standard acceptance limit of 75%. The constant for the model is positive and quite low, which is a good sign for forecasting for less time ahead in future. Models for NO, NO_2, AQI and SO_2 does not qualify the percent fit R squared threshold for their model to be accepted. The first three have close to 75% values for R squared value, and later parameter shows close to inverse correlation with the considered parameters. However the model for $PM_{2.5}$ shows significant correlation value and is taken as most accurate result for this study. With the R squared value of 83.89% and a positive constant value of slightly more than that of ozone, the model shows best fit results for the available data. It can be used to short term forecasting.

Table 1: Results for Regression Modelling

Parameters	Regression Equation	R^2
Ozone	Ozone = 20.56 + 0.6283 Ozone - 0.0570 Nitrogen dioxide(NO_2)- 0.1261 Oxides of Nitrogen(NOx) - 0.512 Ambient Temperature(AT) + 0.08901 Solar Radiation(SR) + 3.022 Wind Speed(WS) + 0.1428 PM $_{2.5}$(PM2.5)	83.03%
NO	NO = 11.49 + 0.0895 Ozone - 0.4143 Nitric Oxide(NO)+ 0.5062 Nitrogen dioxide(NO_2) + 0.5910 Oxides of Nitrogen(NO_x) - 0.04210 Solar Radiation(SR) - 0.05836 PM $_{2.5}$($PM_{2.5}$)	73.80%
NO_2	NO_2 = 11.49 + 0.0895 Ozone - 0.4143 Nitric Oxide(NO) + 0.5062 Nitrogen dioxide(NO_2) + 0.5910 Oxides of Nitrogen(NO_x) - 0.04210 Solar Radiation(SR) - 0.05836 PM 2.5($PM_{2.5}$)	73.80%
SO_2	SO_2 = 5.126 + 0.6004 Sulfur Dioxide(SO2)	35.54%
$PM_{2.5}$	PM2.5 = 28.41 + 0.0408 Ozone + 0.1446 Nitrogen dioxide(NO_2) + 0.2323 Oxides of Nitrogen(NO_x) - 0.5784 Ambient Temperature(AT) - 0.01694 Solar Radiation(SR) - 3.343 Wind Speed(WS) + 0.6606 PM 2.5($PM_{2.5}$)	83.89%
AQI	AQI = -7.57 + 2.2923 AQI - 2.0249 Nitric Oxide(NO) + 4.651 Nitrogen dioxide(NO_2) - 1.5366 $PM_{2.5}$	68.35%

The forecasted values for both the standard models of Ozone and Pm2.5 are ploted for error quantification:

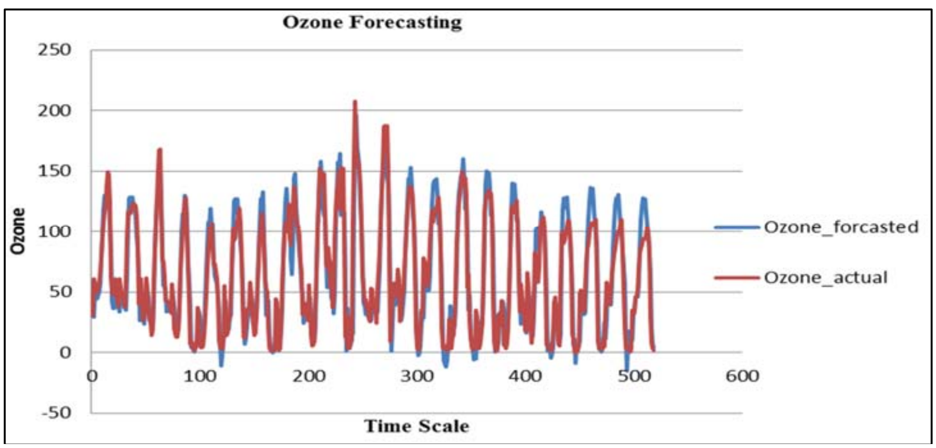

Fig 2: Ozone forecast model error

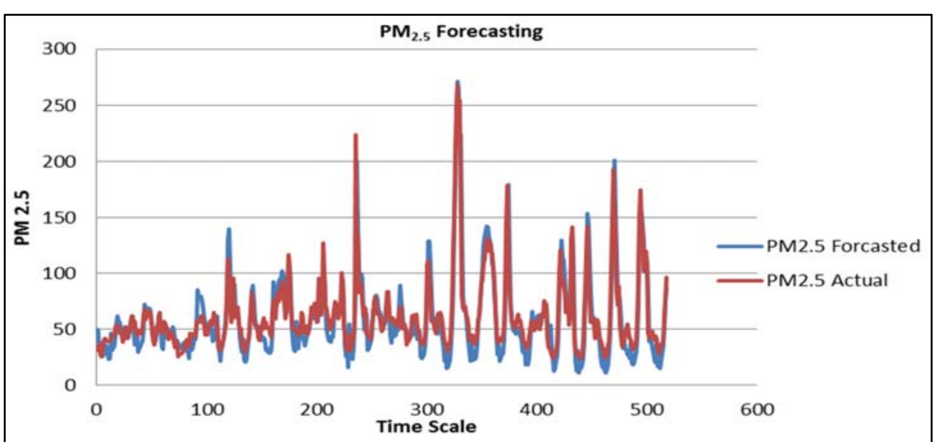

Fig 3: PM $_{2.5}$ forecast model error

For a short term forecasting the error found are in acceptable limits and these two models can be used as a base structure models for forecasting similar parameters with greater accuracy by using larger data sample. There is a recommended need for apply principle component analysis on these data sets to check for any co linearity and hence reducing the error.

5. References

1. Swamee PK, Tyagi A. 'Formation of an air pollution index, J of Air and Waste Manag. Assoc. 1999; 49:88-91.
2. Green MH. An air pollution index based on sulfur dioxide and smoke shade, J. of Air Pollut. Contr. Assoc. 1966; 11(12):703-706.
3. Babcock LR. A combined pollution index for measurement of total air pollution, Atmos. Environ. 1970; 20(10):653-659.
4. Inhaber H. A set of suggested air quality indices for Canada, Atmos. Environ. 1975; 9:353-364.
5. Ott WR, Thom. A critical review of air pollution index systems in the United States and Canada, Atmos. Environ. 1976; 126(5):460-470.
6. Environmental Protection Agency, Guidelines for reporting of daily air quality-Air quality index (AOI) (EPA – 454=R-99-010, 1999.
7. NEERI Report, Air Quality Status: India (National Environmental Engineering Research Institute, Nagpur, India, 1991-1992.
8. NEERI Report, Ambient Air Quality Status in Ten Major Cities: India (National Environmental Engineering Research Institute, Nagpur, India, 1993-94, 1995-96).
9. Thomas WA, Babcock LR, Schults WB. Oak Ridge air quality index, ORNLNSF-EP-8, Sept. Oak Ridge National Laboratory, Oak Ridge, TN, 1971.
10. Shah JJ, Nagpal T. Urban air quality management strategy in Asia, Greater Mumbai Report World Bank Technical. 1997, 381.
11. Klemm RJ, Mason RM, Heilig CM, Neas LM. Dockery DW. Is Daily Mortality Associated Specifically with Fine Particles? Data Reconstruction and Replication of Analyses; J Air & Waste Manage. Assoc. 2000; 50:1215-1222.
12. Pope III, CA, Dockery DW. Health Effects of Fine Particulate Air Pollution: Lines that Connect; J Air & Waste Manage. Assoc. 2006; 56:709-742.
13. McHenry JN, Seaman N, Coats C, Stauffer D, Lario-Gibbs A, Vukovich J, *et al*. The NCSC-PSU Numerical Air Quality Prediction Project: Initial Evaluation, Status and Prospects. Presented at the 11th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA; A&WMA: Pittsburgh, PA, 2000; pp 403-408.
14. McHenry JN, Ryan WF, Seaman NL, Coats CJ, Pudykiewicz J, Arunachalam S, *et al*. A Real-Time Eulerian Photochemical Model Forecast System—Overview and Initial Ozone Forecast Performance in the Northeast U.S. Corridor; Bull. Am. Meteorol. Soc. 2004; 85:525-548.
15. Otte TL, Pouliot G, Pleim JE, Young JO, Schere KL, Wong DC, *et al*. Linking the Eta Model with the Community Multiscale Air Quality (CMAQ) Modeling System to Build a National Air Quality Forecasting System; Weather Forecast. 2005; 20:367-384.
16. McKeen S, Wilczak J, Grell G, Djalalova I, Peckham S, Hsie EY, *et al*. Assessment of an Ensemble of Seven Real-Time Ozone Forecasts over Eastern North America during the Summer of J Geophys. Res. 2004, 2005, 110; doi: 10.1029/2005JD005858.