Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 21 | No. 3 | pp. 1269-1277 | 2022

Original Research Paper

https://doi.org/10.46488/NEPT.2022.v21i03.033

Open Access Journal

Comprehensive Modeling of Seasonal Variation of Surface Ozone Over Southern Tropical City, Bengaluru, India

G. Dhanya*†, T. S. Pranesha*, Kamsali Nagaraja, D. M. Chate*** and G. Beig******
*Department of Physics, BMS College of Engineering, Bengaluru-560019, India
**Department of Physics, Bangalore University, Bengaluru-560056, India
***Centre for Development of Advanced Computing, Pune-411008, India
****National Institute of Advanced Studies, Bengaluru-560012, India
†Corresponding author: G. Dhanya; gopidhanya85@gmail.com

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 05-09-2021
Revised: 10-11-2021
Accepted: 21-11-2021

Key Words:
Ozone
Tropospheric pollutant
Seasonal variation
Meteorological parameters
Statistical variations

ABSTRACT

Surface ozone (O_3) is an important pollutant. In this study we investigated the effects of precursor gases on the difference in ozone concentration utilizing various statistical methods like Multiple Linear Regression (MLR), Principal Component Regression (PCR), Artificial Neural Network (ANN), and Principal Component and Artificial Neural Network (PC-ANN) in conjunction with meteorological parameters for forecasting. The pollutants ozone (O_3), carbon monoxide (CO), nitric oxide (NO), nitrogen dioxide (NO_2), oxides of nitrogen (NOx), and the meteorological parameters temperature (temp), relative humidity (RH), solar radiation (SR), wind speed (WS) and wind direction (WD) observed during 2019 are taken as inputs for MLR, PCR, ANN, and PCANN. The mathematical models obtained from the numerical analysis showed that O_3 concentration was significantly affected by the CO, NO, NO_2, NO_X, temp, RH, SR, WS, and WD factors. PCR model's regression coefficient was lower than the MLR model, but the same for ANN and PCANN models was much better in all the seasons than the linear models such as MLR and PCR. The efficiency of all methods is inspected using several performance metrics.

INTRODUCTION

Ozone (O_3) is among the main air pollutants often related to global environmental degradation. O_3 is a secondary contaminant arising from complex chemical effects in the atmosphere (Seinfeld & Pandis 2006). High concentrations of O_3 are closely connected to meteorological backgrounds. They typically arise during hot days, when principal pollutants, oxides of nitrogen (NOx) and Non-Methane Hydrocarbons (NMHC) react photochemically, followed by intense solar radiation and higher temperatures (San José et al. 2005). Ozone destruction happens through various mechanisms, including the chemical reactions happening in the atmosphere and surface deposition. The photochemical loss of Ozone occurs by the photolysis of Ozone and its consequent reaction. The reaction with the water molecule produces OH and HO_2 during midday through the generation of an O^1D molecule. Other ways in which Ozone is eliminated is by either dry or wet precipitation (Varotsos et al. 2000). Ozone has detrimental effects on both humans as well as plants. The analysis of Ozone variations is complicated due to the various precursors, and photochemical and meteorological processes in the air (Abdul-Wahab et al. 2002). Toh et al. (2013) have shown that meteorological parameters significantly influence the effectiveness of the photochemical system linked to the creation and depletion of Ozone. Therefore, it is useful to establish a relation between these primary pollutants and meteorological variables that could be used to determine Ozone concentration (Elkamel et al. 2001).

Several models have been used to satisfy both these criteria: the influence of primary pollutants and meteorological parameters on Ozone and predicting the pollutant levels. Accordingly, these models are sorted into two groupings deterministic and the other as statistical. Deterministic models involve high resources and precision data (Azzi et al. 1995). Therefore, relying on statistical models is far more practical and economical than the deterministic models. Statistical models are simplistic, and they can be implemented in the real data that is available. The statistical models can classify output parameter relationships with input parameters deprived of cause and effect methodology.

Several statistical methods are proposed for analyzing and forecasting concentrations of Ozone. Graphical analysis, Multiple Linear Regression (MLR), Principal Component Analysis (PCA), and Artificial Neural Networks (ANN) have been used to explain the variability of extensive data on air pollution (Abdul-Wahab et al. 2005, Sousa et al. 2007, Özbay et al. 2011). A commonly used technique for achieving a normalized input-output model for a given data set is the MLR process. It is used in biology, medicine, and environmental studies (Smith & Wachob 2006, Özbay et al. 2011).

The present study addresses the effect of precursor gases on the difference in O_3 concentration using MLR, PCR, ANN, and PCANN laterally with meteorological parameters. The primary objective is to link Ozone concentration to meteorological parameters and concentrations of prime pollutants, including the precursors.

MATERIALS AND METHODS

Study Area

The measurement site is located on the campus of B.M.S. College of Engineering, Basavangudi, a residential location in Bengaluru, India. Bengaluru (12.9410^o N, 77.5655^{o}E, 910 m above MSL) one of India's fastest-growing cities, is known as the 'Silicon Valley of India' due to information technology-based industries.

Surface O_3 concentration is measured continuously using a Serinus 10 O_3 analyzer based on the absorption of ultraviolet radiation with a specific wavelength, 254 nm. An Ecotech analyzer provides O_3 estimates in the range of 0-250 ppbv with a 0.5 ppb detection limit. The NO, NO_2, and NO_X concentrations are measured by a Serinus 40 NO_X analyzer based on the chemiluminescence technique. The lower detectable value is 0.40 ppbv with a processing time of 40 s for the NO_X analyzer. Wind speed, wind direction, ambient temperature, relative humidity, and rainfall are recorded using an Automatic Weather Station (AWS) which is also mounted in this location.

Statistical Techniques

MLR, PCR, ANN, and PCANN were used to build four models. Statistical packages for Social Sciences 22.0 (SPSS 22.0) were used for MLR and PCR, while MATLAB R2019a was used for ANN and PCANN research.

Multiple linear regression: One of the most used methodologies for defining the linear relationship among the dependent variable and many other independent variables is Multiple Linear Regression (MLR). MLR uses a least square technique to suit the model, thus reducing the number of square differences amid experimental and predicted concentrations of O_3.

The General MLR equation is given as follows:

$$Y = b_o + \sum_{i=1}^{p} b_i x_i + \varepsilon$$

Where Y would be the dependent variable, b_o is a constant, b_i is independent variable coefficients, x_i is independent variables, and ε is regression model error. Despite its effectiveness in several applications, MLR can experience tremendous difficulty once the independent variables are associated (Mcadams et al. 2000). Multicollinearity, or strong association in a regression equation between the independent variables, may make it hard to classify the utmost significant contributors to a physical method.

Principal component regression (PCR): PCR is Principal Component Analysis (PCA) and MLR collective method. Principal Components (PCs) produced by PCA are being utilized in MLR by employing an input variable, which reduces the multicollinearity and simplifies the model. The selected PCs with extreme PCA loadings assured that most of the original parameters were used in these models, making them suitable for usage in MLR as independent factors (Gvozdić et al. 2011).

Principal component analysis (PCA): PCA is a multivariate data technology that is applied in this analysis to minimize parameters and to include the most significant important parameter in variations of O_3. PCA transforms these into a limited number of independent variables named principal components. PCs were extracted in such a way whereby its first principal component (PC1) represented the most considerable amount of overall variability in the collected data, while the remaining components (PC2, PC3) reported for the rest of the variations (Kovač-Andrić et al. 2009).

PCs are overall, expressed in terms

$$PC_i = l_{1i}X_1 + l_{2i}X_2 + \cdots + l_{mi}X_m$$

PC_i is the ith PC, and l_{mi} is the observed loading X_m.

In PCA, Bartlett's sphericity method is adopted to check whether variables are correlated with each other or not. Kaiser-Meyer-Olkin (KMO) verifies PCA's applicability to the set of data, and the value of KMO's > 0.5 suggests PCA's data suitability. The varimax rotation has been implemented, simplifying the model by making smaller and larger loads and ensuring that each component has a maximum association with just one printable factor and is minimally associated with some variables (Dominick et al. 2012).

Artificial neural network (ANN): ANN applications in atmospheric/environmental sciences started in the late 1990s and were proven effective in model forecasting (Luna et al. 2014). The neuron is an essential building block for any ANN architecture. This network has three layers, namely input mode, hidden mode, and output mode. The input neurons encrypt information from an outside world, all the neurons in the previous layer receive signals from the hidden neurons, while the output neurons transmit back useful data to the outside world. The output value is obtained by applying a mapping function such as Sigmoid, Tangential, and linear hyperbolic. Maier & Dandy (2000) have explained in detail the application of ANN in environmental modeling.

Principal component and artificial neural network (PC-ANN): PCANN is a combination of PCA and ANN. PCs created from PCA are used instead of original variables as input variables. The accuracy of individual forecasts can be enhanced by combining predictions from various models (Zhang 2003). Even though such models are of rather complex architecture, they could be more useful in predicting the levels of Ozone.

Performance Indicators

Various performance indicators were evaluated towards assessing the errors and accuracies of the models developed.

Normalized Absolute Error (NAE)

NAE summates the expected and calculated value difference separated by a summation of the observed values (Elbayoumi et al. 2014).

$$NAE = \frac{\sum_{i=1}^{n} |P_i - O_i|}{\sum_{i=1}^{n} O_i}$$

Root Mean Square Error (RMSE)

RMSE is the standard method of calculating a model's error when predicting quantitative data. It is formally defined as

$$RMSE = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(P_i - O_i)^2}$$

Index of Agreement (IA)

IA is a standardized model prediction error degree indicator, which ranges from 0 to 1. The formula is given by

$$IA = 1 - \left[\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2}\right]$$

IA equals 0 means that the expected and observed values do not match, while IA equals 1 means that the observed and predicted values are in perfect harmony.

Mean Biased Error (MBE)

Mean biased error means that the prediction degree is either above or below the values. MBE level > 0 would be an indicator of over-prediction, while the level < 0 is an indicator of under-prediction.

$$MBE = \bar{P} - \bar{O}$$

where n indicates sample size, P_i has Predicted Ozone concentration; O_i is a concentration of detected Ozone; $\bar{P}$ is expected concentration of Ozone; $\bar{O}$ is average of the concentration of Ozone.

RESULTS AND DISCUSSION

For the present study dataset of the year 2019 has been used for training and then these data were segregated into distinct seasons for model validation. The entire set of data was checked for normalization before being utilized in different models. In this study, four models for different seasons, winter, summer, monsoon, and post-monsoon, were used to estimate Ozone concentration using precursors and meteorological parameters.

MLR

Regression-based methods were adopted to predict the impacts of identified variables on O_3 differences (Abdul-Wahab et al. 2005). Stepwise multiple linear regression modeling was used in this study to determine the predictive equation designed for the concentration of Ozone as a result of the different parameters that are being measured. CO, NO, NO_2, NO_X, Temperature (temp), Relative Humidity (RH), Solar Radiation (SR), Wind Speed (WS), and Wind Direction (WD) are the independent predictors while Ozone is the dependent or response variable that has been used in this study. The method automatically chooses the most important parameters and excludes those that are of the least significance.

As Variance Inflation Factor (VIF) fell under 5, multicollinearity problems were excluded. Autocorrelation has been assessed using the Durbin Watson test and the autocorrelation values are 1.78, 1.71, 0.92, and 1.13 for winter, summer, monsoon, and post-monsoon for the current data. The model description is presented in Table 1, which provides many correlations R, R^2, adjusted R^2, and best fit model solution. MLR model developed was evaluated by the coefficient of determination (R^2) used to measure the capability of the designated parameters to elucidate the O_3 concentration variations (Abdul-Wahab et al. 2005). The findings showed how MLR showed success with R^2 at 0.933, 0.90, 0.95, and 0.90 during winter, summer, monsoon, and post-monsoon. For winter, the table's variables describe more than 93%

of the variance in Ozone concentration. NO_2 appears to be the primary source of Ozone during summer. Among these meteorological parameters, temperature and solar radiation make significant contributions to high Ozone concentrations. The variation in NO_2 and temperature explains the change in Ozone up to 95% for monsoon. However, post-monsoon, it accounts for 90% depending on the variable, as shown in the table.

For the MLR model using precursor gases and meteorological parameters, R^2 calculated was found to range from 0.88 to 0.91. MLR model shows that the independent variables listed in the table explain more than 85 % of the variance in Ozone concentration during the different seasons. Fig. 1(a-d) shows a strong positive correlation between measured and modeled Ozone levels with a coefficient of 0.92, 0.89, 0.95, and 0.89 for winter summer, monsoon, and post-monsoon seasons. Fig. 2(a-d) shows the diurnal variation of observed and modeled Ozone for winter, summer, monsoon, and post-monsoon seasons. It indicates a strong agreement of diurnal variability between observed and predicted values for the peak hours. It showed an overestimation for winter, summer, and monsoon during the morning and night hours while there was an underestimation during morning hours for the post-monsoon season.

PCR

The main aim of PCR is to acquire a limited quantity of components that would elucidate much of the maximum variation in the predictor variables. The adequacy of data collected for PCA is evaluated by Bartlett's studies and Kaiser-Meyer-Olkin (KMO). Before using PCA for extraction, 10 linear variables, O_3, CO, NO, NO_2, NO_X, temp, RH, SR, WS, and WD, were chosen. After PCA extractions, three variables were selected as the PCs built on eigenvalue greater than unity. PCs obtained in Table 2 obey varimax rotation as this maximizes the correlation of the parameters to only one factor and minimizes correlations between the parameters and the other factors (Dominick et al. 2012). The cumulative variances during winter, summer, monsoon, and post-monsoon were 96%, 95%, 93%, and 95%, respectively.

Table 1: Correlation of O_3 concentration with models during the different seasons for MLR.

Season	Method	R	R^2	Adjusted R^2	Equation of Model
Winter	MLR	0.963	0.926	0.915	O_3=((2.433*Temp) -(1.029*WD)+(0.051*SR)+(241.964))
Summer	MLR	0.950	0.903	0.889	O_3=((-1.884*NO_2) +(4.543*Temp)+(0.029*SR)-(48.529))
Monsoon	MLR	0.977	0.954	0.950	O_3=((-1.367*NO_2) +(1.906*Temp) -(8.838))
Post Monsoon	MLR	0.950	0.902	0.888	O_3=((-0.096*WD)-(1.34*CO)+(0.009*SR)+(34.508))

Table 2: Rotated Principal Components loadings during the different seasons.

	Winter			Summer			Monsoon			Post-Monsoon		
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
NOx	0.991			0.974				0.992			0.971	
NO2	0.965			0.892				0.903			0.942	
CO	0.963			0.979				0.984			0.984	
NO	0.962			0.926				0.925			0.927	
Temperature		0.977			0.964		0.964			0.932		
RH		-0.983			-0.971		-0.957			-0.91		
WS			0.939		0.835		0.975			0.938		
WD	0.881					-0.802	-0.936					0.867
SR		0.621	0.594		0.589	0.701	0.897		0.403	0.835		
Ozone	-0.703		0.542		0.819		0.918			0.834		
Eigen value	5.263	2.635	1.744	4.019	3.903	1.619	5.535	3.841		4.405	4.103	1.034
% of Variance	52.626	26.35	17.437	40.188	39.032	16.191	55.346	38.405		44.054	41.034	10.341
Cumulative %	52.626	78.975	96.413	40.188	79.22	95.411	55.346	93.751		44.054	85.088	95.429

During winter, the first factor represented 52.6% of the total change and itemized six variables (NO, NO_2, NOx, CO, WS, WD) as the crucial variables. While the second factor signified 26.3% of the total change and contained loadings of temperature and solar radiation, PCA then classified Ozone, SR, and WS onto the third factor, which characterized almost 17.4% of the total change. In summer, we observe on the first PC, NO, NO_2, NO_X, and CO have significant loadings with 52.6% of the variation of independent parameters. The second PC describes 39% of the discrepancy, and it is profoundly loaded on Ozone, temp, and WS. The third PC was heavily ladened on wind direction and solar radiation which explains 16.2% changes.

The first PC stated 55.3% of the whole data variability for the monsoon results. They are heavily encumbered onto Ozone, temp, RH, WS, WD, and SR. The subsequent PC, which reported about 38.4% of the total changes, showed a distinction of only pollutants, namely CO, NO, NO_2 and NO_X. The rest of the parameters were signified by the remaining main components, which successfully accounted for less of the total variation. The post-monsoon's first parameter included most of the carefully chosen meteorological variables with a collected variance of 44%. The second factor was constituted by many of the primary sources of pollutants with a difference of 41%, although WD exclusively characterized the third factor with approximately 10%. Thus, principal components obtained from PCA are implemented in the MLR model to find the best fit for variations in Ozone.

Table 3 shows the result of the PCR model. The analysis indicates that PCR during winter and post-monsoon showed optimum performance with R^2 values equivalent to 0.69 and 0.76, correspondingly. Therefore, more than 70% of the O_3 differences throughout winter and post-monsoon were described by the selected parameters. A significantly higher R^2 (0.90 and 0.93) was obtained during summer and monsoon, which indicated that these selected variables explained higher possibilities of O_3 variations. During these periods, R^2 values for the PCR analysis were crucially lower than R^2 values for the MLR analysis. The use of PCs as MLR inputs could not increase the models' efficiency to clarify the disparities in O_3 concentrations through the different seasons. A scatter plot of modeled vs observed shows that the correlation between observed and modeled is moderately related for winter and post-monsoon, and shows a good correlation between the two for summer and monsoon (Fig. 1(a-d). Fig. 2 (a-d) shows the model offers strong consistency during peak Ozone hours in the summer, winter, and monsoon seasons, overestimating Ozone rates in the early morning and night hours. While for the post-monsoon season it over and underestimates the entire diurnal variation.

ANN

Models of artificial neural networks (ANN) can detect complex non-linear underlying interactions amid responses and forecasters and could be accomplished using many successful training methods. In this analysis, ANN was developed using MATLAB 2019a and used a feedforward backpropagation method. It has a layer of input, hidden, and output. There are 9 input parameters and one output parameter. The 9 input data that have been in the investigation are CO, NO, NO_2, NO_X, temp, RH, SR, WS, and WD. The entire data set is divided into 70% of training, 15% validation data set, and 15% the test data set. The number of neurons and the appropriate activation function are two of the major issues encountered when building the architecture of the hidden layer. Yang et al. (2005), have explained the number of neurons in the hidden layer can be evaluated using the following formula

$$n_h = 2n_i + 1$$

n_h represents the number of neurons present in the hidden layer and n_i represents the number of neurons present in the input layer. The ideal number of neurons in the hidden layer was 19 ($n_i = 9$). The model with 19 neurons was explored, but the results were not promising. For better output, the model was optimized by using 5,10,15,19 neurons in the hidden layer while 10 neurons in the hidden layer only showed better results and so this is incorporated in the study.

For training, validation, and testing, datasets correlation coefficients were 0.99, 0.96, and 0.99, respectively for winter. The cumulative regression is 0.99. While for summer and

Table 3: Correlation of O_3 concentration with models during the different seasons for PCR.

	Method	R	R2	Adjusted R2	Equation of Model
Winter	PCR	0.810	0.670	0.630	O3=((-0.282*PC2)+(0.057*PC3)+(49.478))
Summer	PCR	0.950	0.902	0.887	O3=((-8.13*PC1)+(2.877*PC2)+(0.036*PC3)-(41.728))
Monsoon	PCR	0.966	0.934	0.928	O3=(1.434*PC1)-((-0.372*PC2)-11.019))
Post-monsoon	PCR	0.871	0.758	0.735	O3=((0.529*PC1)-(0.078*PC2)+(4.445))

monsoon it is 0.98, 0.99, and 0.99 for training, validation, and testing. The overall regression was 0.98. For post-monsoon, the training, validation, and testing were found to be 0.99, and the overall regression was 0.99. The complete regression using the ANN model showed closure to unity which is stronger than the other models. Fig. 1(a-d) shows the scatter plot of observed and predicted Ozone levels for different seasons. We observe that the correlation between observed and measured Ozone levels during winter, summer, monsoon, and post-monsoon is 0.99, 0.97, 0.97, and 0.99. ANN model is better when compared to the MLR and PCR models. Figure 2a-d displays the diurnal variation of Ozone observed and modeled by ANN. There was good agreement between the observed and modeled Ozone levels during peak hours for winter, monsoon, and post-monsoon seasons. While for summer season there is an overestimation during the night-time.

PC-ANN

PC-ANN model was used, which was devoid of multicollinearity to make the ANN model efficient and straightforward. Model construction was initiated by applying PCA analysis to the data and then applying ANN on the generated three most explaining PCs. As a result, the input layer has 3 neurons, while the number of neurons in the hidden layer is 7 ($2n_i + 1$). The model's performance was tested again by 5, 7, and 9 neurons in the hidden layer. The results displayed best with 7 neurons in the hidden layer. The dataset is divided into training (70%), validation (15%), and testing datasets (15%). The correlation coefficients of PC-ANN-based modeled datasets of training, validation, testing, and overall were 0.98, 0.99, 0.93, and 0.97 for the winter season. For summer and monsoon, the training, validation, testing, and overall were 0.99, 0.96, 0.99, and 0.99, respectively. During the

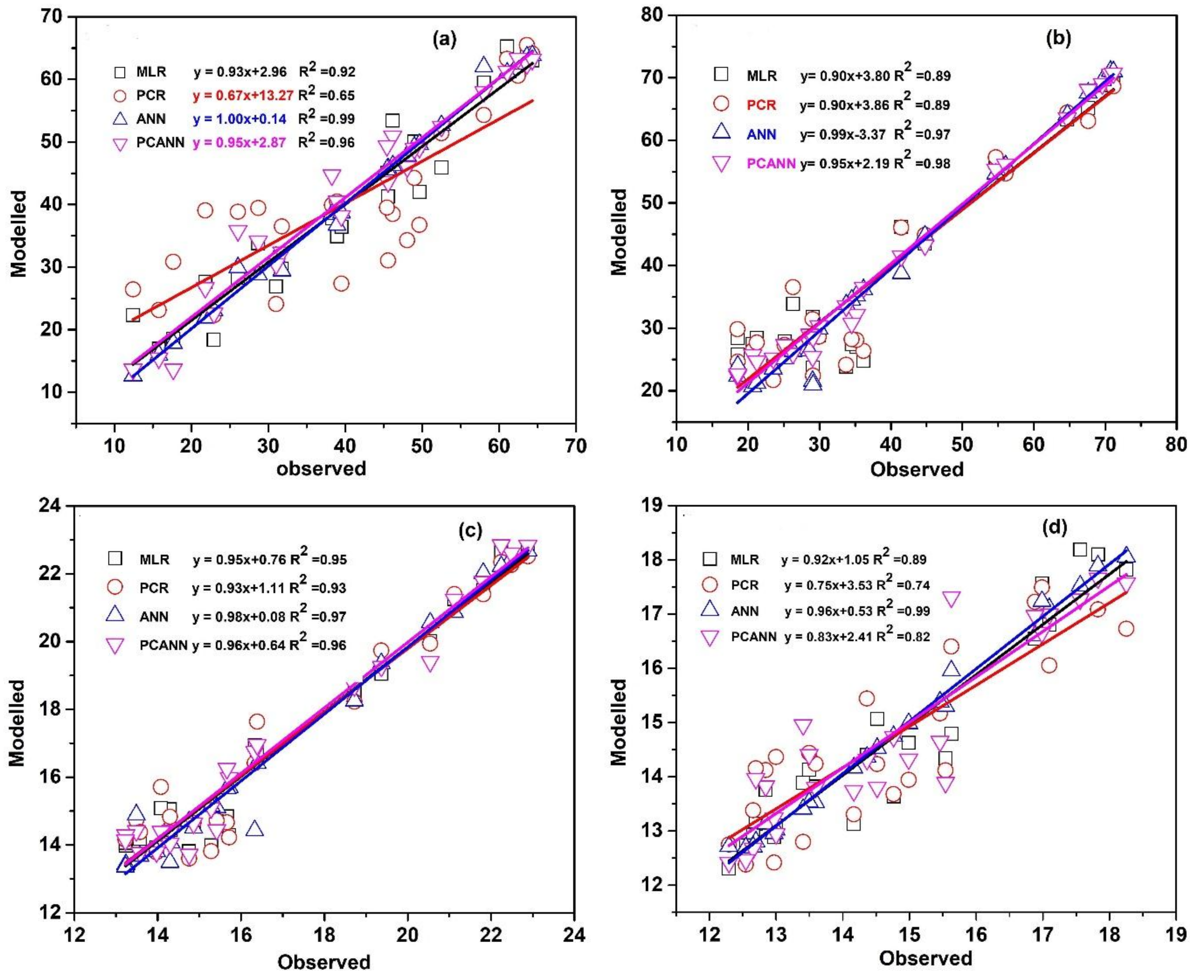

Fig. 1: Modelled vs Observed correlation graph for a winter, b summer, c monsoon, d post-monsoon seasons.

post-monsoon season, it was 0.89, 0.99, 0.94, and 0.90 for training, validation, testing, and overall results.

A scatter plot shows a strong correlation between modeled and observed Ozone levels for all the seasons and is shown in Fig. 1 (a-d). Fig. 2 (a-d) exhibits the diurnal variation of modeled and observed Ozone for PC-ANN analysis. There is a good agreement between the observed and the modeled values for peak hours' observations and overestimation was observed for all seasons during the early morning and late evening hours.

Performance Indicators

The performance of each of these models was measured using different error terms such as normalized absolute error, root mean square error, index of agreement, prediction accuracy, and mean biased error. The value of NAE, RMSE, IA, PA, and MBE are shown in Table 4 for different seasons.

For the most accurate model, the NAE value must be near zero (Sharma et al. 2016). In the present data, the NAE is close to zero, indicating the accuracy of the models. The lower RMSE value shows that the model works well, though the higher RMSE value does not imply that the model is entirely incorrect as peak data take a high RMSE impact (Vlachogianni et al. 2011).

RMSE remained highest for MLR and PCR model and the smallest for the ANN and PC-ANN model. If IA in the model is nearer to one, it means that the forecasted values are adjacent to the observed values. It was nearest to 1, suggesting that the ANN-based method represents this model's best alignment with the observed set of data. The MBE

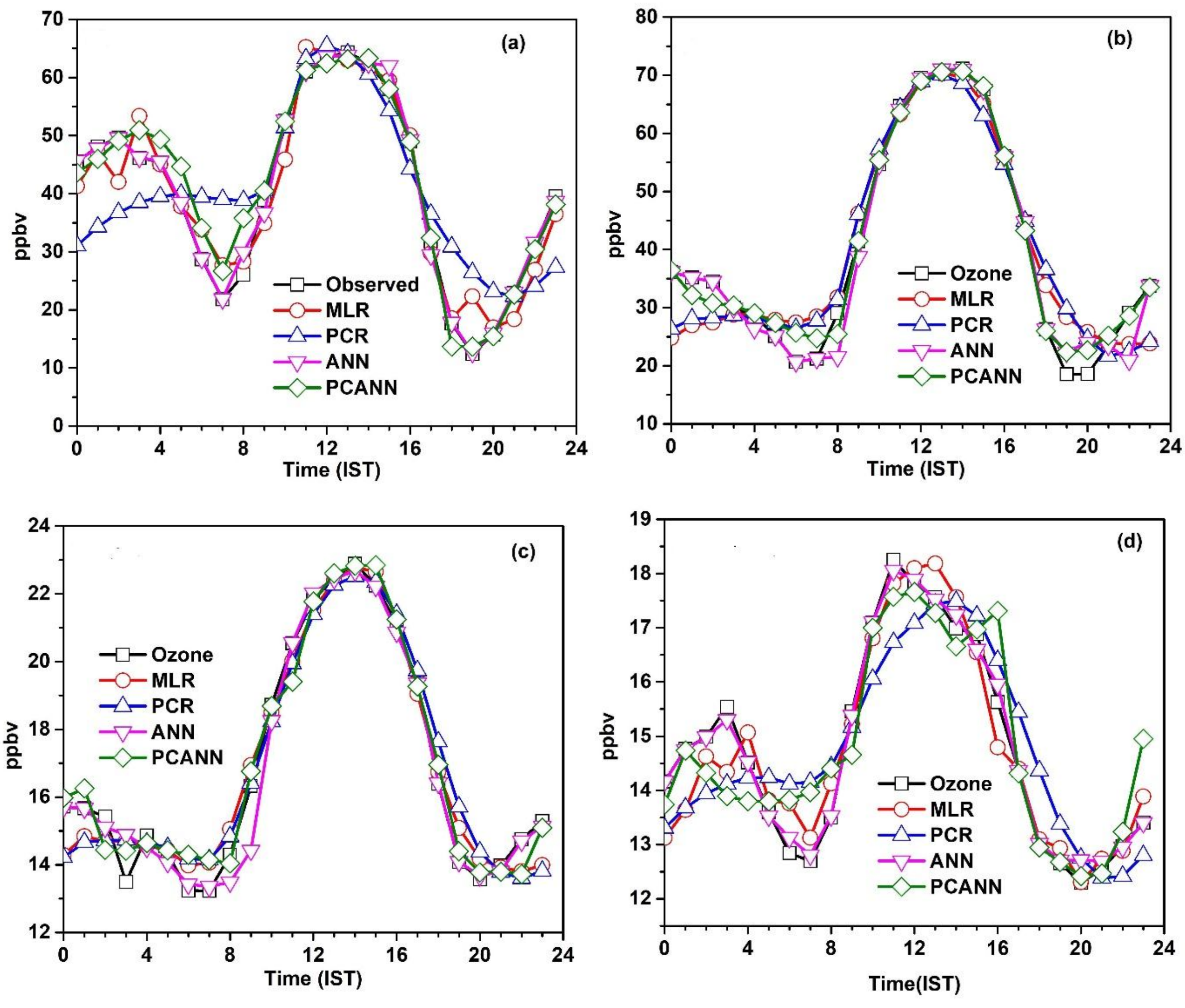

Fig. 2: Diurnal variation of experimental, MLR, PCR, ANN, and PCANN exhibited ozone level a winter b summer c monsoon d post-monsoon.

Table 4: Summary of Performance Indicators.

Winter PI	NAE	RMSE	IA	PA	MBE
MLR	0.115	0.697	0.905	0.77	-0.016
PCR	0.26	1.508	0.587	0.676	0.048
ANN	0.02	0.143	0.986	1.398	0.17
PC-ANN	0.074	0.462	0.883	1.527	1.09
Summer- PI	**NAE**	**RMSE**	**IA**	**PA**	**MBE**
MLR	0.152	0.891	0.924	1.667	0.012
PCR	0.154	0.91	0.919	1.062	-0.043
ANN	0.05	0.271	0.951	1.062	-0.492
PC-ANN	0.062	0.331	0.989	1.113	0.321
Monsoon- PI	**NAE**	**RMSE**	**IA**	**PA**	**MBE**
MLR	0.038	0.121	0.915	1.235	-0.017
PCR	0.045	0.145	0.925	1.461	0
ANN	0.019	0.062	0.976	1.795	-0.124
PC-ANN	0.028	0.091	0.919	0.109	0.082
Post-monsoon- PI	**NAE**	**RMSE**	**IA**	**PA**	**MBE**
MLR	0.032	0.098	0.971	1.102	-0.026
PCR	0.056	0.168	0.917	1.423	-0.008
ANN	0.008	0.023	0.997	0.946	0.032
PC-ANN	0.038	0.114	0.903	0.945	0.036

results show over and under prediction for the observed and predicted value. All these performance indicators suggest that ANN and PCANN models were the best compared to MLR and PCR-based models for predicting Ozone levels for all the seasons. Ozone concentrations predicted by these models vary. These variations are attributed to the complex set of reactions involving Ozone's precursor other than the ones used in the present study.

CONCLUSION

Surface Ozone simulation was carried out with four statistical models. The first concept is built on MLR and the coefficients of regression for this model are 0.926, 0.903, 0.954, and 0.902 during winter, summer, monsoon, and post-monsoon. The second concept is the PCR model developed that uses factor analysis of principal components as the input parameter in multiple linear regression. The third model used was ANN which showed regression analysis results of training, validation, test, and overall unity. The fourth model is PCANN, and their regression coefficient is also close to unity. The reliability of all models is tested using various performance indicators. The governing equations developed from the regression techniques showed that O_3 concentration was significantly affected by parameters such as CO, NO, NO_2, NO_X, temp, RH, SR, WS, and WD. The PCR model's regression coefficient was less than the MLR model, but the same for ANN and PCANN models was much better in all the seasons than the linear models such as MLR and PCR.

The present work is preliminary research on Bengaluru air pollution that might be used in various data analytical approaches to develop a forecasting model. An effective forecasting model of ambient air pollution is essential for demonstrating it to be a valuable tool for public health protection to build rigorous pollution control technologies and procedures.

ACKNOWLEDGEMENTS

The Ministry of Earth Sciences (MoES) supports the Indian Institute of Tropical Meteorology (IITM), Pune. The authors sincerely express their thanks to Professor R. S. Nanjundiah, Director, IITM and MoES for the MAPAN air quality station at BMS College of Engineering, Bengaluru for the mutual benefit of academic research. The contents and conclusions in this research paper are of the authors and do not necessarily reflect the organizations' views. The authors from BMS College of Engineering sincerely express their thanks to the Principal, Professor B. V. Ravishankar.

REFERENCES

Abdul-Wahab, S.A. and Al-Alawi, S.M. 2002. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. Environ. Model Software, 17(3): 219-228. doi: 10.1016/S1364-8152(01)00077-9.

Abdul-Wahab, S.A., Bakheit, C.S. and Al-Alawi, S.M. 2005. Principal component and multiple regression analysis in modeling of ground-level ozone and factors affecting its concentrations. Environ. Model. Software, 20(10): 1263-1271. doi: 10.1016/j.envsoft.2004.09.001.

Azzi, M., Johnson, G.M., Hyde, R. and Young, M. 1995. Prediction of NO_2 and O_3 concentrations for NO, plumes photochemically reacting in urban air. Math. Comput. Model, 21: 39-48.

Dominick, D., Juahir, H., Latif, M.T., Zain, S.M. and Aris, A.Z. 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmos. Environ., 60: 172-181. doi: 10.1016/j.atmosenv.2012.06.021.

Elbayoumi, M., Ramli, N.A., Md Yusof, N.F.F., Yahaya, A.S.B., Al Madhoun, W. and Ul-Saufie, A.Z. 2014. Multivariate methods for indoor PM10 and PM2.5 modeling in naturally ventilated school buildings. Atmos. Environ., 94: 11-21. doi: 10.1016/j.atmosenv.2014.05.007.

Elkamel, A., Abdul-Wahab, S., Bouhamra, W. and Alper, E. 2001. Measurement and prediction of ozone levels around a heavily industrialized area: A neural network approach. Adv. Environ. Res., 5(1): 47-59. doi: 10.1016/S1093-0191(00)00042-3.

Gvozdić, V., Kovač-Andrić, E. and Brana, J. 2011. Influence of meteorological factors NO2, SO2, CO, and PM10 on the concentration of O3 in the urban atmosphere of Eastern Croatia. Environ. Model. Assess., 16(5): 491-501. doi: 10.1007/s10666-011-9256-4.

Kovač-Andrić, E., Brana, J. and Gvozdić, V. 2009. Impact of meteorological factors on ozone concentrations modeled by time series analysis and multivariate statistical methods. Eco. Inf., 4(2): 117-122. doi: 10.1016/j.ecoinf.2009.01.002.

Luna, A.S., Paredes, M.L.L., De Oliveira, G.C.G. and Corrêa, S.M. 2014. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. Atmos. Environ., 98: 98-104. doi: 10.1016/j.atmosenv.2014.08.060.

Maier, H.R., and Dandy, G.C. 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and applications. Environ. Model. Software., 15(1): 101-124. doi: 10.1016/S1364-8152(99)00007-9.

McAdams, H.T., Crawford, R.W. and Hadder, G.R. 2000. A vector approach to regression analysis and its application to heavy-duty diesel emissions. SAE Tech. Papers, 1: 19-61. http://dx.doi.org/10.4271/2000-01-1961

Özbay, B., Keskin, G.A., Doğruparmak, Ş.Ç. and Ayberk, S. 2011. Multivariate methods for ground-level ozone modeling. Atmos. Res., 102(1–2): 57-65. doi: 10.1016/j.atmosres.2011.06.005.

San José, R., Stohl, A., Karatzas, K., Bohler, T., James, P. and Pérez, J.L. 2005. A modeling study of an extraordinary nighttime ozone episode over the Madrid domain. Environ. Model. Software., 20(5): 587-593. doi: 10.1016/j.envsoft.2004.03.009.

Seinfeld, J.H. and Pandis, S.N. 2006. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. 2nd edition. J. Wiley, Hoboken, N.J.

Smith, C.M. and Wachob, D.G. 2006. Trends associated with residential development in riparian breeding bird habitat along the Snake River in Jackson Hole, WY, USA: Implications for conservation planning. Bio. Cons., 128(4): 431-446. doi: 10.1016/j.biocon.2005.10.008.

Sharma, S., Sharma, P., Khare, M. and Kwatra, S. 2016. The statistical behavior of ozone in an urban environment. Sus. Environ. Res., 26(3): 142-148. doi: 10.1016/j.serj.2016.04.006.

Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M. and Pereira, M.C. 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environ. Model. Software., 22(1): 97-103. doi: 10.1016/j.envsoft.2005.12.002.

Toh, Y.Y., Lim, S.F. and Von Glasow, R. 2013. The influence of meteorological factors and biomass burning on surface ozone concentrations at Tanah Rata, Malaysia. Atmos. Environ., 70: 435-446. doi: 10.1016/j.atmosenv.2013.01.018.

Varotsos, C.A., Kondratyev, K.Y. and Cracknell, A.P. 2000. New evidence for ozone depletion over Athens, Greece. Int. J. Rem. Sen., 21(15): 2951-2955. doi: 10.1080/01431160050121366.

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S. and Kukkonen J. 2011. Evaluation of a multiple regression model for the forecasting of the concentrations of NOx and PM10 in Athens and Helsinki. Sci. Total Environ., 409(8): 1559-1571. doi: 10.1016/j.scitotenv.2010.12.040.

Yang, J., Rivard, H. and Zmeureanu, R. 2005. Online building energy prediction using adaptive artificial neural networks. Energy Build., 37(12): 1250-1259. doi: 10.1016/j.enbuild.2005.02.005.

Zhang, G.P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50: 159-175. doi: 10.1016/S0925-2312(01)00702-0.