MAUSAM, **73**, 3 (July 2022), 617-626

MAUSAM

DOI : https://doi.org/10.54302/mausam.v73i3.5933
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 502.2 : 213.52 : 551.509.3

Regression analysis of ventilation coefficient at a semi-arid IGP region using forward selection technique

PRIYANKA SINGH[#,*], KIRTI SONI[@,*], ANJALI S. NAIR[#,*] and MAHAVIR SINGH[#,*]

[#]*CSIR-National Physical Laboratory, Dr. K. S. Krishnan Marg, New Delhi – 110 012, India*

[@]*CSIR-Advanced Materials and Processes Research Institute, Bhopal – 462 026, India*

**Academy of Scientific and Innovative Research (AcSIR), Ghaziabad – 201 002, India*

(*Received 2 March 2021, Accepted 14 March 2022*)

e mail : soniks@ampri.res.in

सार – एक सुदूर संवेदी तकनीक वाले ध्वनिक साउंडर, यानी सोडार (ध्वनि और उसकी पहुँच का पता लगाने वाला) का उपयोग करके मौसम संबंधी प्राचलों और वायु प्रदूषक सांद्रता के साथ वायु संचार गुणांक (वीसी) के रैखिक और एकाधिक प्रतिगमन विश्लेषण के लिए एक अग्रचयन (एफएस), तकनीक का प्रयोग किया गया है; यह रीयल-टाइम पर एबीएल (वायुमंडलीय सीमा परत) ऊंचाई की निरंतर जानकारी देता है और उससे लगे मौसम विज्ञान संवेदक मौसम संबंधी प्रतिमानों जैसे पवन की गति/दिशा, सापेक्षिक आर्द्रता और तापमान की जानकारी देता है। दैनिक परिवर्तनशीलता एबीएल की एक प्रमुख विशेषता है, जो सतह और मुक्त वातावरण के मध्य गर्मी, आवेग, नमी और रासायनिक घटकों के आदान-प्रदान में महत्वपूर्ण भूमिका निभाती है। वायुसंचार गुणांक एक वायुमंडलीय स्थिति है जो वायु गुणवत्ता और प्रदूषण क्षमता को इंगित करता है। विभिन्न वायु प्रदूषकों की सांद्रता के साथ वीसी के सहसंबंध पर चर्चा की जाती है और संवहन और पवन की गति के संवहन गुणांक पर प्रभाव का भी विश्लेषण किया जाता है। वायुसंचार गुणांक के दैनिक, मासिक और ऋतुनिष्ठ बदलाव वायु गुणवत्ता प्रबंधन के लिए दिन-प्रतिदिन की मौसम की घटनाओं के बारे में जानकारी देते हैं। परिणाम से पता चलता है कि नौ चर (WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, ओजोन, PM_{10}) पवन की गति (WS), सापेक्षिक आर्द्रता (RH), तापमान (तापमान), और कार्बन मोनोआक्साइड (CO) वायुसंचार-गुणांक को अत्यधिक प्रभावित करने वाले चर हैं इन चार सर्वश्रेष्ठ चरों को नौ चरों की तुलना में बेहतर प्रदर्शन प्राप्त करने के लिए किसी भी पूर्वानुमान और एफएस द्वारा चुने गए मॉडल में इनपुट प्रतिमान के रूप में उपयोग किया जा सकता है।

ABSTRACT. In this study, a forward selection (FS), technique has been employed for linear and multiple regression analysis of Ventilation coefficient (VC) with meteorological parameters and air pollutants concentrations; using a Remote sensing technique acoustic sounder, *i.e.*, SODAR (Sound Detection and ranging); it gives real-time continuous ABL (Atmospheric Boundary Layer) height data and attached meteorological sensors give meteorological parameters data such as wind speed/direction, relative humidity and temperature. The diurnal variability is a dominant feature of the ABL, which plays an important role in the exchanges of heat, momentum, moisture and chemical constituents between the surface and free atmosphere. The ventilation coefficient is an atmospheric condition that indicates the air quality and pollution potential. The correlation of VC with different air pollutants concentrations are discussed and the effect of convection and wind speed on the ventilation coefficient is also analysed. Diurnal, monthly and seasonal variation of ventilation coefficient gives knowledge about the day to day weather phenomena for air quality management. The result shows that among nine variables (WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, Ozone, PM_{10}), wind speed (WS), relative humidity (RH), temperature (Temp) and CO are the highly influencing variable to the ventilation coefficient and these four best variables selected by FS can be used as an input parameter in any prediction model to get better performance in comparison to the nine variables.

Key words – Remote sensing, SODAR, Ventilation coefficient, Atmospheric boundary layer height, Semi-arid.

1. Introduction

Air pollution is one of the most serious problems of urban developing city like Delhi, so it is attracting the most attention all over the world. India is developing country with increasing urbanization hence increasing air pollution in urban capital city Delhi due to the rapid development of industry, road and other constructions, which is affecting the urban environment, climate and health. Delhi is within the top ten cities with the utmost polluted air quality (Arora, 2021). In the previous two decenniums, Delhi has been growing expeditiously, accordingly, demands for conveyance, waste generation, energy consumption, construction activities and domestic cooking and industrial activity grew considerably. From last two decades, Delhi's air pollution is a coalescence of factors, including power plants, industries, burning of coal and biomass and convey (direct conveyance exhaust and indirect road dust) that contribute to air quality reduction (Goyal *et al*., 2006; Soni *et al*., 2010, 2011; Roychoudhury *et al*., 2012; Rizwan *et al*., 2013). For further orchestrating to ameliorate the quality of air, it is consequential to ken the vicissitude in air quality over Delhi after the implementation of sundry schemes of switching over to CNG (Compressed Natural Gas) and ameliorations in fuel quality and conveyance technology, etc. India is spending 150 billion $ per year to overcome air pollution. Vehicular pollution alone contributes about 72% of the total air pollution load in Delhi as estimated using emission factor and activity-based approach recommended by IPCC (Sindhwani & Goyal, 2014). So, the air pollution dispersion study is important, for that the atmospheric boundary layer height is an important parameter for the vertical mixing of air pollution and is highly influenced by boundary layer height and wind speed.

In current years, the run of the mill urban barometrical condition has declined continually, uncovering an intriguing complex synthetic and physical trademark. It is additionally detailed that urban air contamination scenes follow certain pre-decided examples that are related with the nearby meteorological conditions and discharges of essential toxins (Assimakopoulos and Helmis, 2003). India is the greatest creating nation with urban area seriously expanding air contamination because of the quick improvement of urbanization, essential developments and monetary turn of events, which applies an antagonistic effect on the urban condition, wellbeing and environmental change. Until this point, the air contamination in India has pulled in much consideration around the globe. The ventilation coefficient (VC) assumes a significant job in the weakening and expulsion of all poisons including gases that can be considered as one of the elements for the portrayal of contamination potential over an area of concern (Chan *et al*., 2012). The VC tells us about the rate at which air inside the blended layer (blending stature) is shipped away from a district of intrigue and is a record of air contamination or air quality over that area (Holzworth, 1967). Various examinations have been performed for the varieties in the VC over different locales in India just as in different places of the world (Goyal *et al*., 2006; Genc *et al*., 2010; Iyer & Raj 2013; Mahalaxmi *et al*., 2011).

The nocturnal study of ventilation coefficient was done with help of LIDAR in Pune (Devara *et al*., 2002**)** they used the LIDAR derived mixing depth and pilot balloon wind data, in that they found relatively high VC in pre-monsoon and low during winter and monsoon month. Ventilation coefficient value high in pre-monsoon maybe due to cloud scavenging of aerosol and precipitation, low VC values during winter says that there is poor air quality in winter. A more detailed analysis using Radiosonde revealed that in Delhi, the ventilation coefficient decreased at the rate of 49 and 32 m^2/s/year in the months of December and February, respectively during the 30-year period (1971-2000) (Iyer and Raj, 2013). On the other hand, eight years (2006-2014) study reported increase in trend, *i.e*., 70 m^2 /s/year using SODAR over the same location but for the different period (Saha *et al*., 2019).

SODAR (Sound Detection and ranging) is a remote sensing technique used to give real time ABL height and meteorological sensors provide wind speed, wind direction, temperature and humidity for various atmospheric studies. ABL is the lowermost part of the troposphere which plays an important role for pollution dispersion. Ventilation can be known of a particular area with the help of wind speed and boundary layer height. This study emphasizes the two most important meteorological parameter Atmospheric boundary layer height and the average wind speed which is given by a weather sensor and SODAR instrument. The ABL height controls the vertical extent, concentration and transformation of atmospheric pollution to some extent and has importance on the study of trace gases and ozone in the lower troposphere (Krishnan & Kunhikrishnan, 2004; Mahalaxmi *et al*., 2011; Sujata *et al*., 2016). In this study on the temporal variations of ventilation coefficient (VC), a parameter that depends on the ABL height and mean wind within the mixed layer and its possible implications on the dispersion of air pollutants using SODAR and wind observations over capital city station, Delhi, India discussed.

2. Data and methodology

All the measurements were taken at the CSIR-National Physical Laboratory, Delhi. The capital city of India, Delhi was selected as the study area for this work. Delhi is situated about 1100 km from the nearest coast on the North Arabian Sea. The area around Delhi lies on Yamuna flood plains and the Delhi ridge. Delhi features dry winter and bordering a hot semi-arid climate. The warm season lasts from March to May with an average daily high temperature above 39 °C (102 °F). The cold season lasts from November to February with an average daily high temperature below 20 °C (68 °F). In early March, the wind direction changes from north-westerly to south-westerly. The monsoon arrives at the end of June, along with an increase in humidity. The month of June, July, August and September are under Monsoon Season. The brief, mild winter starts in November and peaks in January and heavy fog often occurs. The whole data set has been classified into four seasons on the basis of existing meteorological conditions, as per IMD (India Meteorological Department) namely Winter (January-February), Pre-monsoon (March-May), Monsoon (June-September) and Post-monsoon (October-December).

The urban city with undulated terrain which is high in population. There were increase in number of vehicles and so, increment in traffic for almost 24 hrs on roads in Delhi. In nearby areas, there are a lot of small industrial plants developed in this 21st century, which release more pollution to the city, these economic developments contribute to maximum aerosol ($PM_{2.5}$ and PM_{10}) level in the city, which leads to Delhi as most polluted city in India.

For the measurement of pollution condition in the city, CSIR-NPL developed the instrument SODAR which continuously provides data of ABL height in real space and time in form of SODAR ecogram. The SODAR ecogram not only used for Atmospheric pollution study but also for other meteorological phenomena study, like fog, rain, wind shear and turbulence. In this work the total data taken from March 2019 to February 2020 and during Monsoon, the ABL height data is limited due to rainfall.

The weather sensor is placed near the SODAR antenna, above 10 meters from the ground. The weather sensor gives average wind speed and other different parameters of weather but in this study, the average wind speed data is used for estimation of the Ventilation coefficient.

2.1. *Ventilation coefficient*

The Ventilation coefficient is a product of Atmospheric boundary layer height and average wind speed through the boundary layer. The ventilation coefficient reflects the ability of the atmosphere to dilute and disperse the pollutants in the mixing layer. VC can be calculated by (Saha *et al*., 2019).

$$\mathrm{VC} = H \times V \tag{1}$$

where, H is the ABL height and V is the average wind speed within that height.

The unit of ventilation coefficient will be meter squared per second (m^2/s).

The values of ABL height and Wind speed were taken at the height of 10 m from the ground, so the ventilation coefficient value is also 10 m above from the ground surface.

2.2. *Forward Selection technique*

The Forward Selection (FS) method has been used successfully by many researchers to develop various accurate predictive models. FS is a linear model. First, a reliable variant is selected and a combination of other descriptions is performed. After that, it is organized from the most related variables to the less closely related variables. The best variety included is considered the first entry. After that the best second variable is selected. It is evident that there is a greater change than the R^2 value known as the coefficient of integration. The coefficient of correlation is the degree of correlation between any two variables '*x*' and '*y*'. The value of R is between -1 and 1. If the value of R is 1 it is said to correspond perfectly. At a value of -1, they are said to be completely opposite and at 0, they are said to have no connection at all. The R formula is shown below:

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\left(\sum x^2\right) - \left(\sum x\right)^2\right]}\sqrt{\left[n\left(\sum y^2\right) - \left(\sum y\right)^2\right]}} \tag{2}$$

Correlation can be correctly defined by simple linear regression but in the case of multiple linear regressions it becomes tedious to define one distinct relationship. Therefore, we use positive interaction (R^2) to better define relationships. It is important as it provides a measure of the variability of one variation rather than the predictable value of the other variables. As it is a square of the value of R and the range is between 0 and 1. The R^2 can be used to describe both linear regression and multiple linear regression.

The FS method step is ongoing for the N-1 periods which helps us to assess the effect caused by each variation of the output model. Finally, we get the number N of the subsets. The best subset of R^2 is considered to be the subset of modeling. After finding the R^2 that is most compatible with the group of variables there is no significant change in the value of the R^2 when new variables are added to it (Noori *et al*., 2011).

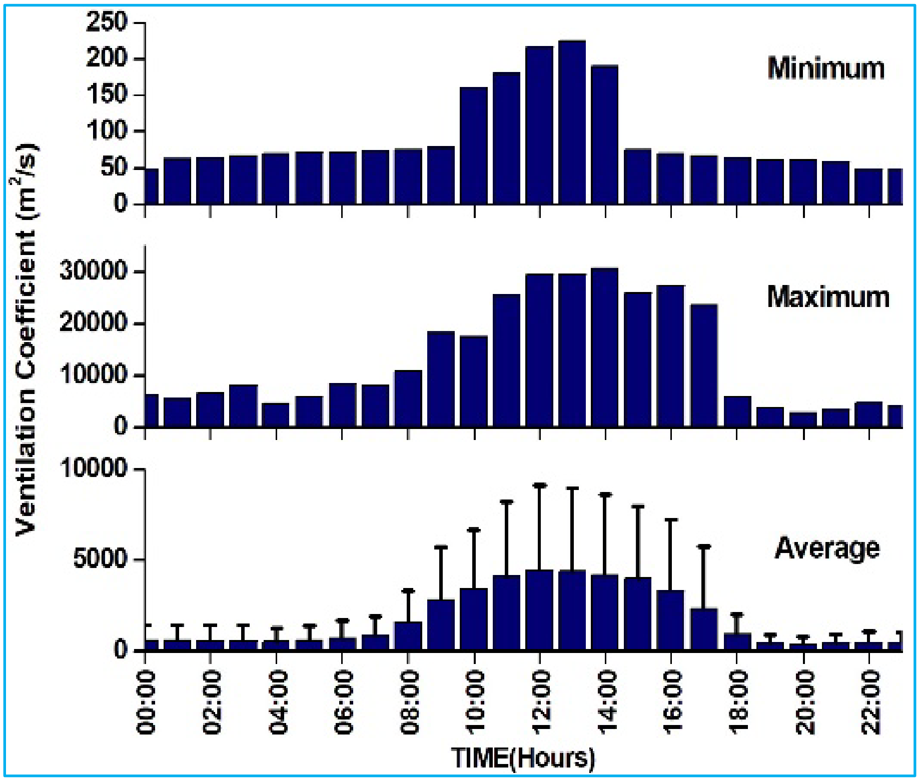

Fig. 1. Temporal average variation of VC over Delhi for the period from March 2019 to February 2020

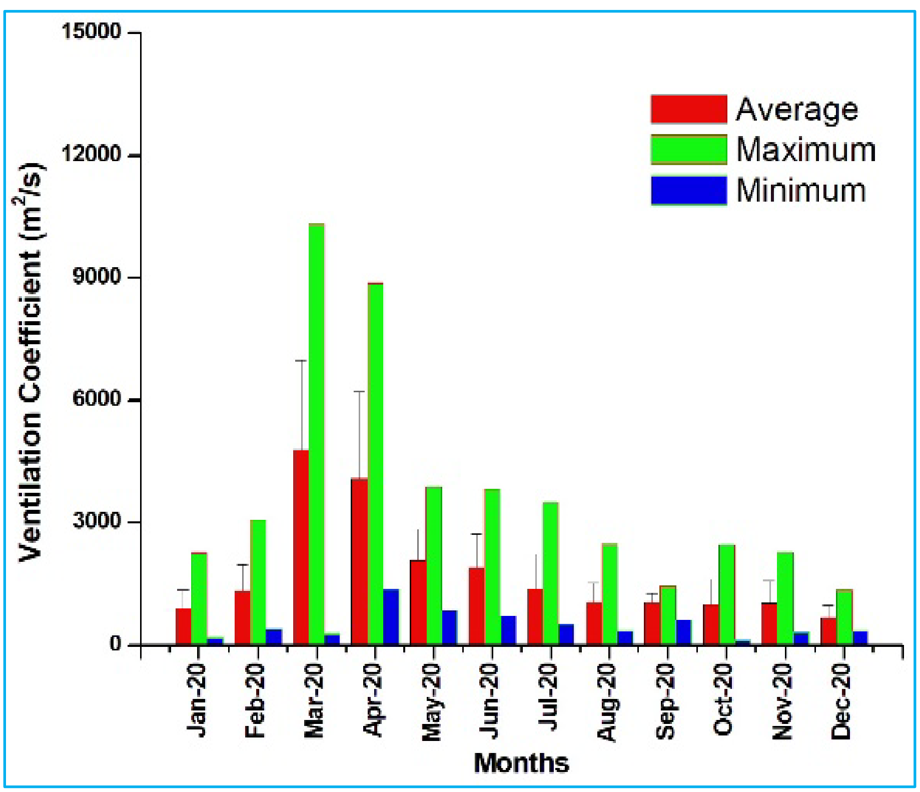

Fig. 2. Monthly average variation of VC over Delhi for the period from March 2019 to February 2020

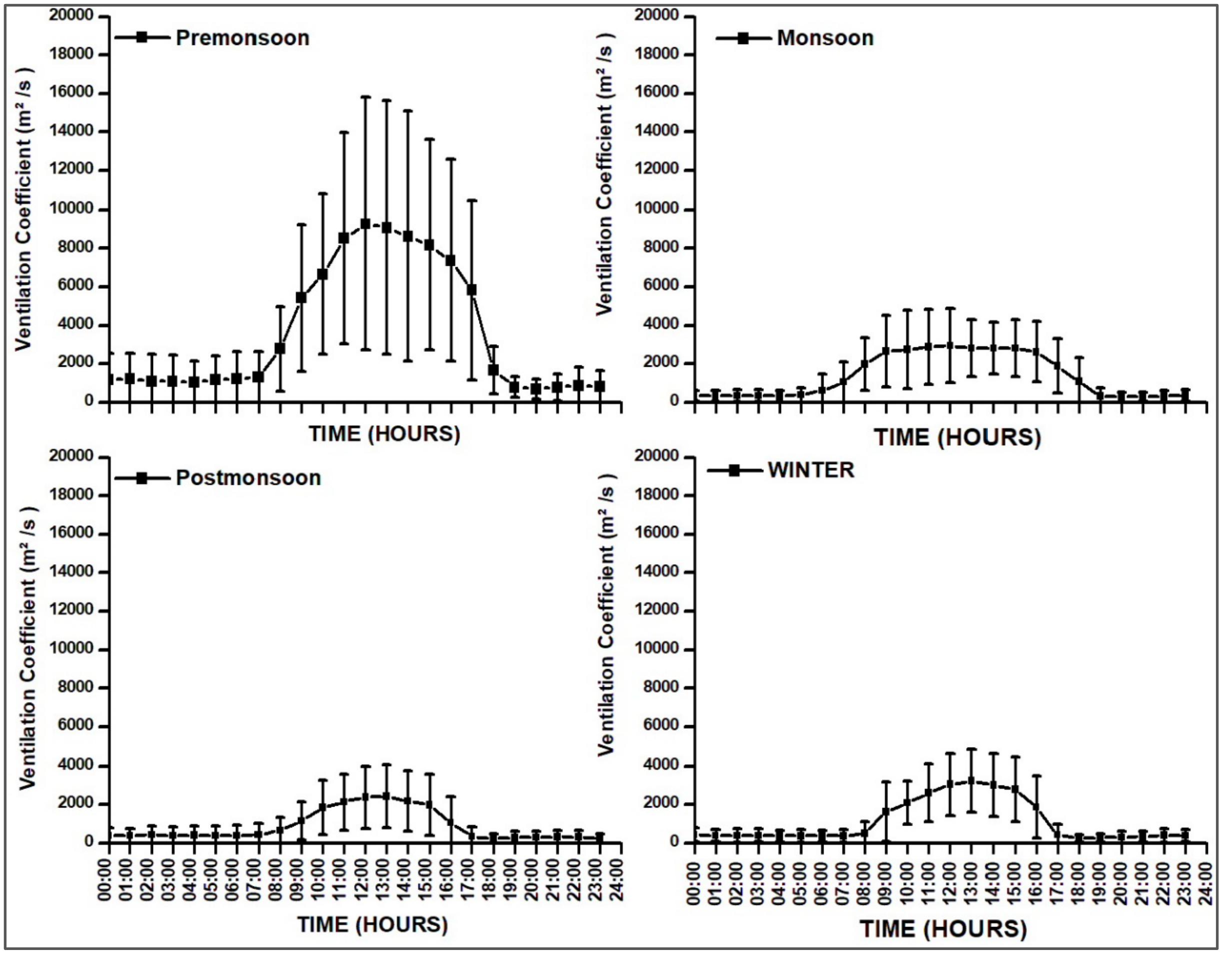

Fig. 3. Temporal variations of ventilation coefficient during different seasons (from March 2019 to February 2020)

3. Results and discussion

3.1. *Temporal and Seasonal variation of VC*

On the basis of Figs. 1 to 4 it has been observed that there is an existence of significance variation of Ventilation coefficient in hourly average, daily average, monthly average and seasonal from March 2019 to February 2020. Fig. 1. Fig. 2 and Fig. 3 represent the temporal, monthly and seasonal variations of ventilation coefficient respectively from March 2019 to February 2020.

It is observed from Fig. 1 and Fig. 2 that VC varies monthly in different manner while diurnally it varies as at 0700 hrs to 1800 hrs (*i.e.*, in day hours) it is high and from 1900 hrs to 0600 hrs (*i.e.*, in night) it is low, it indicates ventilation is maximum value during day time (30000m²/s) and minimum value during night (upto 250 m²/s). While if we observe it seasonally (Figs. 3 to 5,

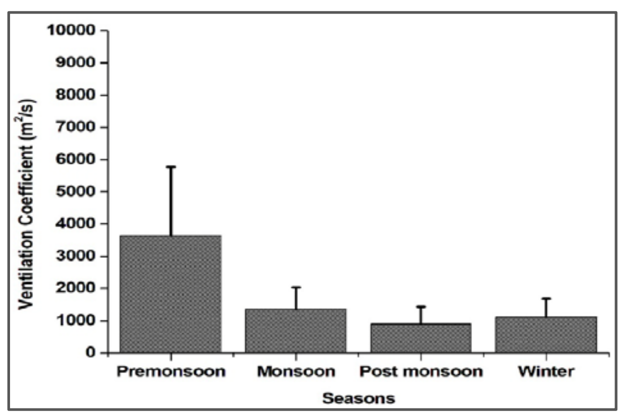

Fig. 4. Variation of Ventilation Coefficient for one year during different seasons (March 2019-Febryary 2020)

Temporal, daily and monthly average during different seasons) then it is maximum in pre-monsoon (March, April, May) and Monsoon (June, July, August,

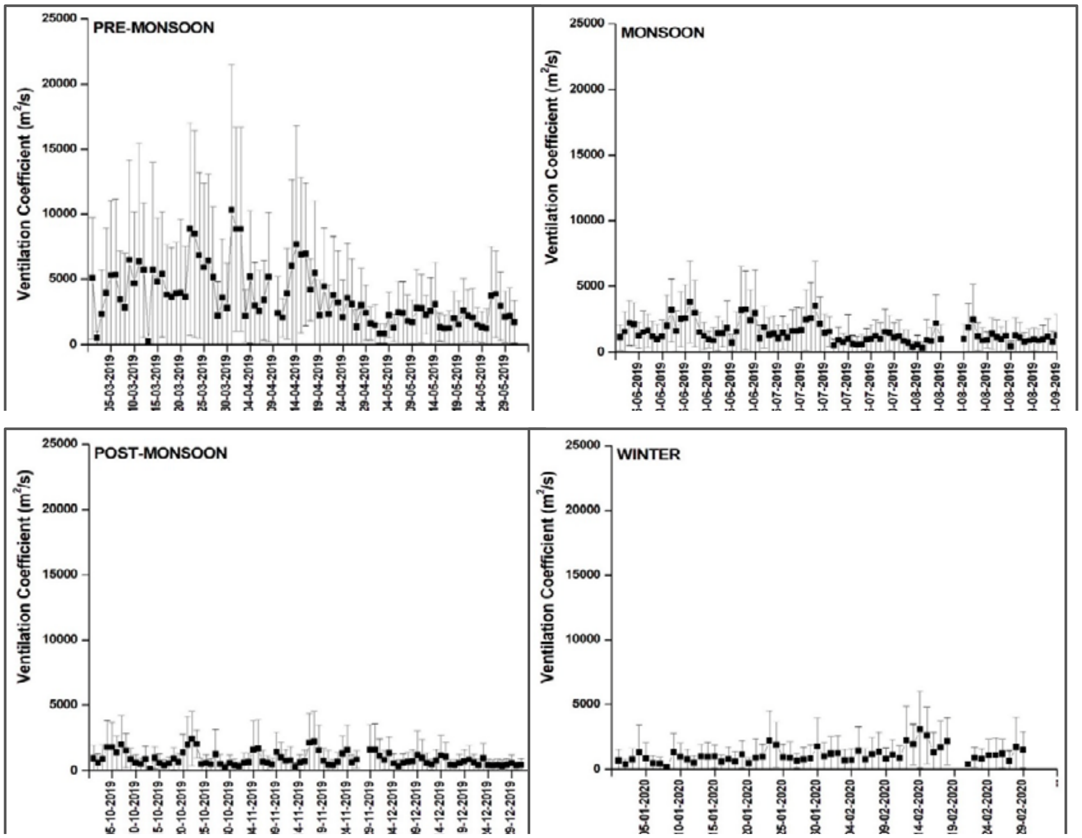

Fig. 5. Daily average variation of ventilation coefficient for one year during different seasons (March 2019-February 2020)

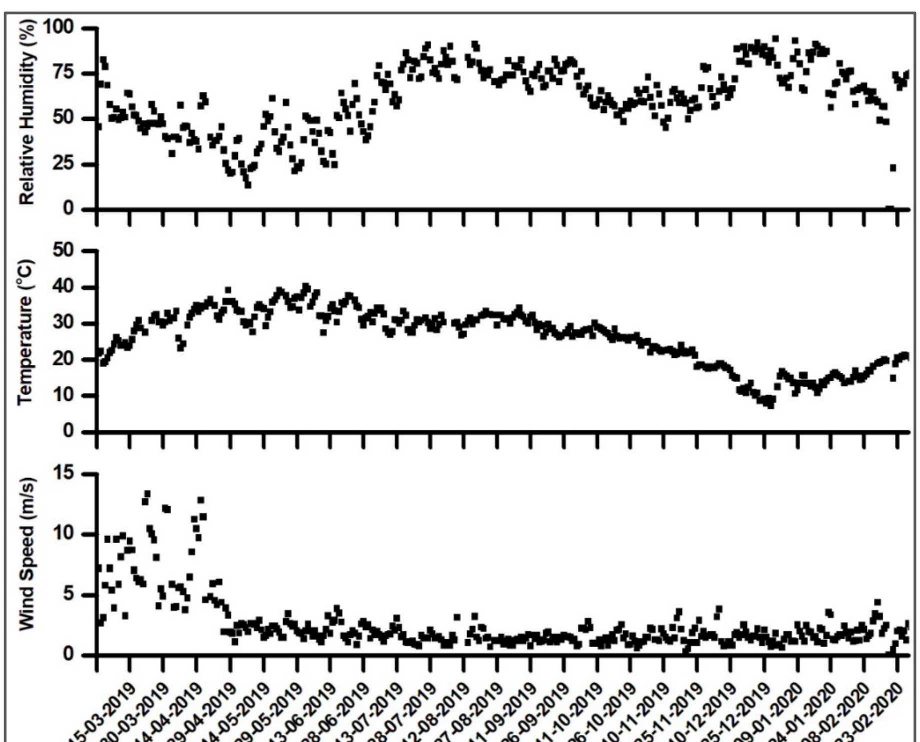

Fig. 6. Variation of meteorological parameter from March 2019 to February 2020

September) and the lowest VC is observed in winter (January, February) and Post-monsoon (October, November, December) season. The average values of the ventilation coefficient indicated the highest value in Pre-monsoon (3667±2140 m²/s) followed by in the Monsoon season (1358±680 m²/s), in the winter season (1112±574 m²/s) and lowest in the Post-monsoon season (907±522 m²/s). The ventilation coefficient value during afternoon period in pre-monsoon is higher than 6000 m²/s in comparison to other seasons, which shows very poor atmospheric pollution conditions during remaining seasons (Iyer and Raj, 2013). This shows that the ventilation coefficient mostly depends on temperature, humidity and wind speed as the wind speed and temperature were maximum and humidity was minimum during Pre-monsoon and monsoon season (Fig. 6). When compared with other Indian cities, the VC values over Delhi are slightly low. It can be compared with hilly industrial sites like Visakhapatnam bowl area which had VC in Summer was 9781 m² s⁻¹ and 13,924 m² s⁻¹ in winter (Krishnan *et al*., 2004) and over Manali where it was in the range 7900 m² s⁻¹ (Summer) to 2226 m² s⁻¹ (winter) (Manju *et al*., 2002), however, our values are comparable with the VC obtained over the Delhi area where the study done for different timeframe, these values are lower than the values obtained in present study, this shows that over these times the ventilation condition get improved in Delhi. The result of this study is comparable to Changsha city, that the value of mixing layer height during daytime was higher than that in night time (18 hrs IST to 6 hrs IST) (Chan *et al*., 2012). Chan *et al*. (2012) reported values for the daytime mixing layer height presented large variations during the year. The Mixing height was deepest of about 925-1070 m from May through October and the shallowest of about 600-650 m in December and January with rapid changes in transition months. The maximum values of ventilation coefficient were 5,122 m²/s in summer followed by 3,711 m²/s in autumn and 3,694 m²/s in spring, respectively. Winter records had the smallest value of 2,537 m²/s which is slightly higher than Delhi for winter season. (Chan *et al*., 2012)

In order to observe the variation of pollution dispersion during the year, during pre-monsoon months (Mar-May) there is highest value of ventilation coefficient (30000 m²/s) and lowest value of VC was 278 m²/s shown in the Fig. 5, which indicates that the pollution dispersion is high during the Pre-monsoon and monsoon season due to better convection, low humidity and comparatively high wind speed. During October, November, December and January months the ventilation coefficient value (Fig. 2.) is 1007±615 m²/s, 1028±557 m²/s, 697±286 m²/s and 916±441 m²/s, respectively, which is very low and showing poor air quality in these months and December

TABLE 1

Individual correlation of dependent variable Ventilation Coefficient (VC) with other independent variables

Correlation of ventilation coefficient with	R^2
WS	0.8628
RH	0.1886
Temp	0.0878
CO	0.0875
$PM_{2.5}$	0.0867
SO_2	0.0693
NO_2	0.0377
Ozone	0.0262
PM_{10}	0.0035

month has lowest VC value of 697±286 m²/s (Average value), showing worst air quality in this month in 2019. During pre-monsoon season VC value was 47812 m²/s (Average) in March month, which indicates better air quality due to low humidity and high wind speed, in April (4082±2128 m²/s), May (2070±763 m²/s) and June (1905±825 m²/s) also there are higher VC values, which indicates a little improvement in VC but still poor air quality in Pre-monsoon season comparative to other season.

Based on the variation of Atmospheric boundary layer height and wind speed during day and night, the ventilation coefficient varied differently in different seasons but in Post-Monsoon and Winter season it is almost similar (Fig. 3). It is also reported by Ashrafi *et al*. (2009) that the VC stands maximum in spring at 22329.17 m²/s and minimum fall at 7901.4 m² /s in urban area Tehran. In a similar study over an urban city of Hyderabad India, monthly mean VC highest value (19,735 m²/s) was observed during the month of May and the lowest (13,331 m²/s) observed in the January month (Sujatha *et al*., 2016).

3.2. *Forward selection technique*

It can also be called an input selection method. There arenine candidates for input in this study (WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, Ozone, PM_{10}). The Forward Selection Method is used as a direct input selection process that determines the most appropriate set of these eight variables. It can also be said that it is a straightforward model developed using a very good subset incorporated into the input. Initially the interaction between each input candidate is detected and the result is calculated. Table 1 shows the effect of each combination

TABLE 2

Result for forward selection procedure

Input Subset	R^2
WS	0.863
WS, RH	0.885
WS, RH, Temp	0.881
WS, RH, Temp, CO	**0.885***
WS, RH, Temp, CO, $PM_{2.5}$	0.885
WS, RH, Temp, CO, $PM_{2.5}$, SO_2	0.886
WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2	0.888
WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, Ozone	0.888
WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, Ozone, PM_{10}	0.889

*After this value, variation of R^2 are negligible

of dependent variables (Ventilation Coefficient) with other variables. In the very first step, the high flexibility and adjustment wind speed (WS) and $R^2 = 0.86$, were carefully selected as the first and most important inputs. This shows that the WS have a profound effect on the VC. After that all the remaining variables candidates (RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, Ozone, PM_{10}) with the addition of increments are added to each model in a step-by-step format. Model quality is measured by one of the parameters called the coefficient of determination or also called R-squared (R^2). Second, it is the Relative Humidity (RH) that plays the most important role, according to our dataset, *i.e.*, coefficient of determination for RH is 0.885. After that we have temperature at number three and the pattern goes like CO, $PM_{2.5}$, SO_2, NO_2, Ozone and finally it becomes PM_{10}. It is the SO_2, NO_2, Ozone and PM_{10} that rarely affects the VC as its value in the R-squared is less than 0.1.

Forward Selection method is used as a linear input selection technique which determines the best suited subset from many variables. Step by step method is given below:

Step 1: Calculate the coefficient of determination or also called as R-squared (R^2) between the Dependent (VC) and Independent variables (WS, RH, Temp, CO, $PM_{2.5}$, SO_2, NO_2, Ozone, PM_{10})

⇩

Step 2: Arrange the R-squared between VC and Independent variables (WS, RH, Temp, CO, PM2.5, SO2, NO2, Ozone, PM10) in decreasing order (given in Table 1)

⇩

Step 3: The variable with the maximum correlation that is WS with $R^2 = 0.8628$, is carefully chosen as the first and the utmost vital input.

⇩

Step 4: Then the entire remaining variable according to the decreasing correlation is added to the model individually in step by step format (given in Table 2).

⇩

Step 5: After obtaining the individual values of R- squared for each single parameter, this process is repeated many times until that adding a new variable to input does not significantly advances the model output.

This step is continued until the new variable is so selected that the change in the value of correlation coefficient is over 5%.

⇩

Step 6: Lastly, input variables (WS, RH, Temp, CO) with utmost vital effect on output are selected and rest variables are removed.

After obtaining the individual values of the R square per parameter, this process is repeated several times until the addition of new input variables does not significantly improve the model effect. It is evident that the value of R^2 is increasing as the new group is made with new variables such as the number of WS is 0.863 but as soon as the RH is combined with the value it also increases to 0.885. This step continues until a new variant is selected so that after this value, variation of R^2 are negligible. Finally, the input variable with the most important output effect is selected and the remaining variable is removed. The outcome of the Transfer Selection process is shown in Table 2 where four variables are selected according to their importance as input variables. It is the wind speed (WS), relative humidity (RH), temperature (Temp) and CO some of which are removed. After these variables it is not uncommon for there to be an increase in the amount of coefficient of determination so the variance is almost negligible. Therefore, the input corresponding to this value is carefully selected.

3.3. *Correlation between VC and air pollutants concentrations*

Fig. 7 represents the comparisons between the variation in VC and pollutant's concentration on daily basis. The ventilation coefficient values are better during February, March, April, May and June, while lower values observed in July, August, September, October, November and December. The VC values variation on comparison with CO concentration is showing that on increasing the VC value the CO concentration decreases and on decreasing the value of VC, the CO concentration get increased during November, December and January. While the other pollutant like NO_2 and SO_2 is showing opposite correlation with VC variation as it is increasing on increase of VC values. During rainy season the SO_2 and NO_2 concentration is minimum due to precipitation and high during summer and winter (Fig. 7).

In order to see the correlation between VC and different pollutants the scatter plotting done, these plot shows not good positive correlations with SO_2 (0.27), NO_2 (0.19) and Ozone (0.14). But the VC value and the CO, $PM_{2.5}$ concentration lacks any good correlation (-0.29, -0.28). Ozone (O_3) concentration also depends on Ventilation and it shows 0-50 µg/m³ at lower VC value of <4000m²/s. All pollutants like SO_2, NO_2 shows almost same correlation with VC values but the particulate matter (PM_{10}) shows quite different result as it shows poor correlation of -0.05, at ≤3000m²/s VC value the PM_{10} concentration is 0-600 µg/m³, the $PM_{2.5}$ values also shows high concentration of 0-300 µg/m³ at lower VC value (≤3000m²/s) (Fig. 7). So, we can say that VC values must be greater than 4500m²/s for lower concentration of

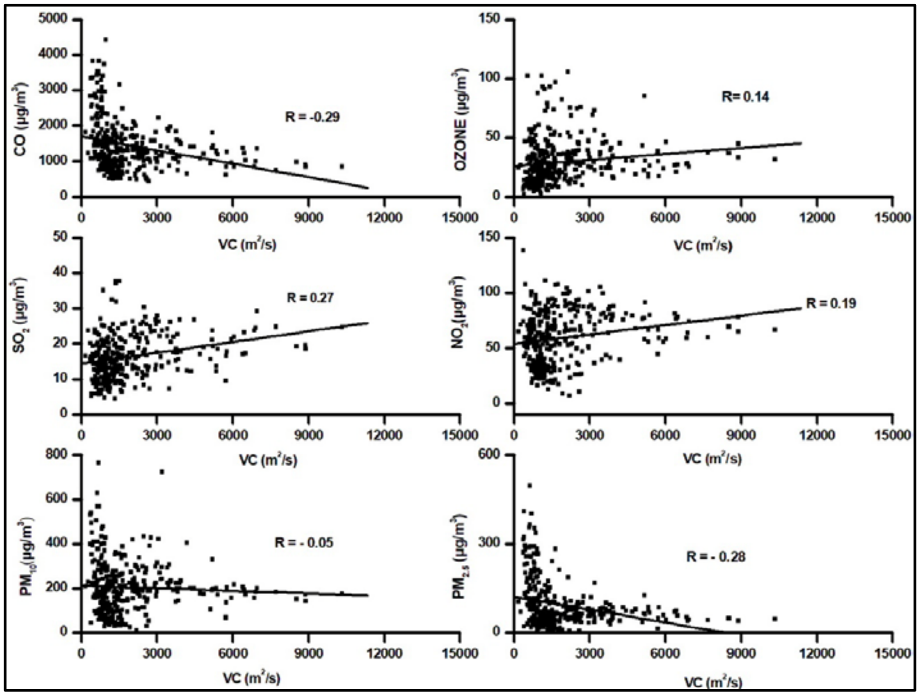

Fig. 7. Scatter plot between VC and air pollutants

Pollutants in ABL and more dispersion of pollutants are essential for better air quality.

4. Conclusions

The present study explores the significance of Ventilation coefficient in dispersion of air pollutants in the atmosphere. Forward Selection (FS) scheme is used for selecting the highly effecting input variables to the ventilation coefficient. After analysis four variables, *i.e.*, wind speed (WS), relative humidity (RH), temperature (Temp) and CO are found as a highly effecting parameters for the VC. These four best variables selected by FS can be used as an input parameter in any prediction model to get better performance. As the number of features increases, the performance of the prediction model decreases. It's mean that by using FS many irrelevant variables that can cause bias can be removed. Additionally, from the results it is also observed that the ventilation coefficient is variable during all day and vary seasonally as well which indicates that at afternoon the convection is high which helps the pollution to dispersed and night time is prone to lock the pollution within the inversion boundary layer. This suggest that specially day time is suitable for industrial, vehicular emission etc. in the atmosphere when the ventilation is highest while night time is not suitable for emission in atmosphere for better air quality condition. The high ventilation coefficient values of 30800m²/s, in March month shows better air quality in Delhi from previous year as compared from previous research work done by Iyer and Raj (2013) on ventilation coefficient study for Delhi with help of radiosonde data. In previous study the VC values comes very low as compared to this study by Iyer and Raj (2013), it shows that the air quality condition gets improved in this year mostly in spring and summer.

Acknowledgement

This research was carried out at the CSIR-NPL. The authors are grateful to the Director of the CSIR National Physical Laboratory (NPL).

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Arora, Sumit, 2021,"IQAir Air Quality Index: Delhi, Kolkata, Mumbai among world's top 10 polluted cities", Published On *Current Affairs* November 16th, 2021 (https://timesofindia.indiatimes.com/india/delhi-kolkata-mumbai-among-worlds-top-10-polluted-cities/articleshow/87685824.cms).

Ashrafi, K., Shafipour, M. and Kamalan, H., 2009, "Estimating temporal and seasonal variation of ventilation coefficients", Int. J. Environ. Res., **3**, 4, 637-644.

Assimakopoulos, V. D. and Helmis, C., 2003, "Sodar mixing height estimates and air pollution characteristics over a Mediterranean big city", *Environmental technology*, **24**, 1191-9.

Chan, Lu., Deng, Q. H., Liu, W. W., Huang, B. L. and Shi, L. Z., 2012, "Characteristics of ventilation coefficient and its impact on urban air pollution", *Journal of Central South University*, **19**, 3, 615-622.

Devara, P. C. S., Maheskumar, R. S., Raj, P. E., Pandithurai, G. and Dani, K. K., 2002, "Recent trends in aerosol climatology and air pollution as inferred from multi-year lidar observations over a tropical urban station", *International Journal of Climatology : A Journal of the Royal Meteorological Society*, **22**, 4, 435-449.

Genc, D. D., Yesilyurt, C. and Tuncel, G., 2010, "Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere", *Environmental monitoring and assessment*, **166**, 1-4, 11-27.

Goyal, P., Anand, S. and Gera, B. S., 2006, "Assimilative capacity and pollutant dispersion studies for Gangtok city", *Atmospheric Environment*, **40**, 9, 1671-1682.

Holzworth, G. C., 1967, "Mixing depths, wind speeds and pollution potential for selected locations in the USA", *Journal of Applied Meteorology*, **6**, 1034-1044.

Iyer, U. S. and Raj, P. E., 2013, "Ventilation coefficient trends in the recent decades over four major Indian metropolitan cities", *Journal of earth system science*, **122**, 2, 537-549.

Krishna, T. R., Reddy, M. K., Reddy, R. C. and Singh, R. N., 2004, "Assimilative capacity and dispersion of pollutants due to industrial sources in Visakhapatnam bowl area", *Atmospheric Environment*, **38**, 39, 6775-6787.

Krishnan, P. and Kunhikrishnan P. K., 2004, "Temporal variations of ventilation coefficient at a tropical Indian station using UHF wind profiler", *Current Science-Bangalore*, **86**, 3, 447-450.

Mahalaxmi, D. V., Badarinath, K. V. S. and Naidu, C. V., 2011, "Influence of boundary layer dynamics on pollutant concentrations over urban region - A study using ground based measurements", *Indian Journal of Radio & Space Physics*, **40**, 147-152.

Manju, N., Balakrishnan, R. and Mani, N., 2002, "Assimilative capacity and pollutant dispersion studies for the industrial zone of Manali", *Atmospheric Environment*, **36**, 21, 3461-3471.

Noori, R., Karbassi, A. R., Moghaddamnia A., Han D., Zokaei-Ashtiani M.H., Farokhnia A. and Ghafari Gousheh M., 2011, "Assessment of input variables determination on the SVM model performance using PCA, Gamma test and forward selection techniques for monthly stream flow prediction", *Journal of Hydrology*, **401**, 177-189. ISSN 0022-1694. https://doi.org/10.1016/j.jhydrol.2011.02.021.

Rizwan, S., Nongkynrih, B. and Gupta, S. K., 2013, "Air pollution in Delhi : Its Magnitude and Effects on Health", *Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine*, **38**, 1, 4-8. https://doi.org/10.4103/0970-0218.106617

Roychoudhury, S., Mondal, N. K., Mukherjee, S., Dutta, A., Siddique, S. and Ray, M. R., 2012, "Activation of protein kinase B (PKB/Akt) and risk of lung cancer among rural women in India who cook with biomass fuel", *Toxicology and applied pharmacology*, **259**, 1, 45-53.

Saha, D., Soni, K., Mohanan, M. N. and Singh, M., 2019, "Long-term trend of ventilation coefficient over Delhi and its potential impacts on air quality", *Remote Sensing Applications: Society and Environment*, **15**, 100234.

Sindhwani, R. and Goyal, P., 2014, "Assessment of traffic-generated gaseous and particulate matter emissions and trends over Delhi (2000-2010)", *Atmospheric Pollution Research*, **5**, 3, 438-446.

Soni, K., Singh, S., Bano, T., Tanwar, R. S. and Nath, S., 2011, "Wavelength dependence of the aerosol Angstrom exponent and its implications over Delhi, India", *Aerosol Science and Technology*, **45**, 12, 1488-1498.

Soni, K., Singh, S., Bano, T., Tanwar, R. S., Nath, S. and Arya, B. C., 2010, "Variations in single scattering albedo and Angstrom absorption exponent during different seasons at Delhi, India", *Atmospheric environment*, **44**, 35, 4355-4363.

Sujatha, P., Mahalakshmi, D. V., Ramiz, A., Rao, P. V. N. and Naidu, C. V., 2016, "Ventilation coefficient and boundary layer height impact on urban air quality", *Cogent Environmental Science*, **2**, 1, 1125284.