IDENTIFICATION AND COUNTERMEASURES OF SPATIAL CONCENTRATION FACTORS FOR PARTICULATE MATTER 2.5

Ibrahim*

Department of Engineering Management, U.E.T, Taxila, mibrahimbjr@gmail.com
Corresponding author: Email ID: mibrahimbjr@gmail.com

Abstract-As we know PM2.5 has adverse effects on human health and decrease the progress of humanity. For this purpose, air quality monitoring departments try to find the composition of the PM2.5 and their source of origin and identify the high polluted areas. Geographic information systems (GIS) provide useful tools, providing the spatial concentration of a region and by these spatial concentrations, an equation is made to predict the values of PM2.5 concentration. The research's basic purpose is to identify the PM2.5 concentration and create an Equation which is composed of main factors of meteorology, which are collected using an inverse distance raster analysis, results to carry out analysis at temporal resolution and finer spatial, and to identify the yearly and monthly PM2.5 concentrations in Ohio Country. Factors of meteorology and concentration of PM2.5 are studied and projected at a large scale in the Ohio state, by using measurements from 55 meteorological stations in the study region Ohio for annual and for three months (July, August, September). A GIS spatial interpolation procedure is used for them. An Ordinary least squares method was used for each of these three months' data, and in this way, models were created. The research results show the model performance was at this accuracy level (98, 99, and 97). To validate the results of these models, the predicted value of PM2.5 data was compared with the value of PM2.5 data observed, and the accuracy level for that is at (97, 98, and 96 percent, respectively). The show results of the models have a very strong fit of the Ordinary least Squares model to the data observed, which confirms, that the results of this work will examine and forecast PM2.5 accurately.

Keywords- GIS, IDW, OLS, PM 2.5

1 Introduction

Because of the hazardous impacts on human health of PM2.5 pollution, it is critical to identify the changing levels of PM2.5 and to take future preventative steps to overcome the concentration level up to the standard level. $PM_{2\cdot 5}$ is now the leading environmental contributor to the global burden of disease. PM2.5 rising from being the fifth leading contributor among environmental risk factors in 1990, in part driven by declines in household air pollution and unsafe water and sanitation [1]. Cities' PM2.5 concentrations can be improved at various sectors (industrial, transportation, and residential, for example) and geographic levels. The US Environmental Protection Agency categorized and identified a comprehensive list of harmful air pollutants. These pollutants may go beyond the high risks in various metropolitan areas and harm people's health[2]. Nowadays, there is prevalent concern about that, the increasing number of vehicles on the road and their impact on public health. One of the health studies showed that a variation in cardiovascular mortality rates is associated with PM2.5, concerning the geographical distance from the central monitoring sites. Spatiotemporal variation has, therefore, become a matter of concern for environmental scientists, health researchers, public health officials, and the public[3]. It is thus important to spot regions where PM2.5 levels exceed the standard level to make a realistic PM2.5 prediction method [4]. Given the severity of the air pollution problems in regional or national, measurement of PM2.5 has recently attracted more consideration to identify spatial concentration.

A useful indicator to describe air quality is Air Quality Index (AQI). Which provides information on the recent air quality of an area and its health consequences [2]. Government air quality departments used these values to express the degree of air pollution in the atmosphere and their public relation, and its value can change according to the air emissions variation. To point out the harmfulness of air contamination, the United States Environmental Protection Agency (www.epa.gov/aqi) developed an AQI indicator. There are five foremost pollutants of air involved: CO2, PM2.5, O3, NO2, and SO2 [5]. AQI is one of the best ways to continuously assess and explain air quality of an area. AQI aims to show the severity of ambient air pollution levels in an easy way to the public by showing the concentration of each pollutant on a common scale where health effects, occur at a value that is common to all pollutants. Such transformations presented air pollution in an easy accessible way with little knowledge of how the public uses the data [6]. The air quality index is reported as the combination of diverse pollutants and this combination is represented by a single value [7].

Geographic information systems (GIS) have the ability of strong data processing and modeling. It has analysis ways to support decision-makers and academics to enhance interpretation and explain data more precisely, hence advancing helpful and more significant data[8]. Where analysts and academics use the GIS results to conclude supportive arguments on harsh conditions [9]. In recent years different methods and data are presented for quantifying air pollution. In environmental research, Geographic Information Systems (GIS) and Remote sensing (RS) techniques have been widely used, because of their easy use and accessibility of temporal environmental satellite-driven data[10].

Various land base regression methods based are made to look at spatial patterns of PM2.5 concentration in a specific region [11]. The two fundamental types of linear regression methods are non-parametric (such as Passing Bablok's Linear Regression) and parametric (such as Ordinary or Least-Square Linear Regression and Deming's Linear Regression) [12]. The accessibility of extensive tools for fitting linear regression models, fitting the linear mean function to geo statistical data is simple and easy [13]. Accurate and precise estimation is necessary for appropriate exposure assessments of PM2.5 concentration. Land use regression (LUR) models have generally been used to prevent misclassification and biased risk evaluation in estimation [14].

In this paper, we construct an alternative way for measuring PM2.5 based on data obtained from the GIS toolbox of geo statistical way analysis. Following that, the proposed method used the GIS-based OLS strategy to evaluate three individual equations of PM2.5 estimation. The estimations are made for three months from July to September based on the data obtained from Ohio State. Using Ordinary linear regressions between the dependent variable (PM2.5) and the other independent Metrological factors to form the associations; for this reason, the predictable values will be used to represent PM2.5 levels, and the predictor variables will correctly clarify the geographical variations of PM2.5 concentration in a specific region.

2 Measuring Procedures

IDW

In Arc GIS, the main method used for spatial patterns is inverse distance weighting (IDW). Because inverse distance weighting (IDW) is a precise interpolator, data is apprehensively taken into account in this method. The interpolation method's purpose is to use measurements z(si), i= 1,2,3,..., n, placed at different point locations, for the reason to provide a possibility of the point value of the experienced possessions in the position s0 when observations of the point do not exist. The weights assigned to different elements are determined by the element's distance from the estimated location. Weights are commonly obtained using the inverse squared distance (represented by the exponent -2), and the interpolation approximation function can be specified as:

$$\hat{z}\,(s_0)\frac{\sum_{i=1}^{n} z\,(s_i)d_{i0}^{-2}}{\sum_{i=1}^{n} d_{i0}^{-2}}$$

Here di0 represented the path between s_0 and si

OLS model

Ordinary linear squares are the most often used method for fitting a preliminary linear mean function to geo-statistical[13]. The formula for OLS regression with k independent parameters is:

$$z_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon$$

Here i represent locations, and x1i, x2i, x3i…, xki represents independent variables at location i, β0, β1, β2, β3 …,βk represents variables that will be predicted, and ε represents the term of error.

3 Research Methodology

The essential capabilities like data management and mapping, Arc GIS includes logical approaches that have long been a basis for Arc GIS. The spatial data tools of Arc GIS are a collection of approaches for characterizing and modeling geographic data. They also examine the geographical pattern of a specific region, processing, trends, distributions, and linkages. The study's procedure was implemented in three steps for the prediction of PM2.5: inverse distance weighted (IDW) raster analysis interpolation method, regression models formation, and confirmation of the regression models. This study's analysis is based on two types of data: an Ohio climate record acquired from NASA satellite data using the RET Screen software, and yearly data of PM2.5, AQI data downloaded from the source https://www.epa.gov/outdoor-air-quality-data/download-daily-data. Arc GIS version 10.3 was used for Ordinary linear squares model development and interpolation procedure, as well as making final maps outputs, which were built in the legends and scale bar in the respective prediction maps.

4 Results

To calculate values of PM2.5 for any site in Ohio State, Inverse Distance Weighting (IDW) method was performed by utilizing the measured parameter data of the 55 stations in Ohio State. The resulting Inverse Distance Weighting (IDW) map includes AQI, PM2.5, TEMP, HUMIDITY, PRESSURE, and WIND SPEED values are grouped as classes with Different colors ranging from lowest value to highest value. The predicted values are limited to the interpolation range of the values. Because Inverse Distance Weighting (IDW) is an average distance weighted, this average distance weighted cannot be more than the upper value or less than the lower value of the given values. As a result, ridges or valleys are not constructed if these limitations are not sampled previously. The Inverse Distance Weighting (IDW) was used to distinguish the values of the cell using a set of sample points of a site. In comparison to tools of the same function, the Inverse Distance Weighting (IDW) is easy to use and execute in the program because there is no need for pre-modeling or subjective assumptions like other tools. Figure1, 2, and 7 shows the IDW resulting maps of the Ohio state for 2018 and from July to September 2018 respectively.

Annual IDW analysis of PM2.5

To show the spread of PM2.5 in Ohio State for yearly Inverse Distance Weighting (IDW) procedure has been used and the role of the higher and lower value map by colors range (Figure 1). The IDW results have shown the uppermost values of PM2.5 at the Western Side stations and Southern stations of State Ohio with a value of 8.89ug/m3 and 11ug/m3 respectively. The lower PM2.5 values at southern sites are 7ug/m3 and 5ug/m3. The Northern sides of Ohio have an unhealthy value of 8ug/m3.

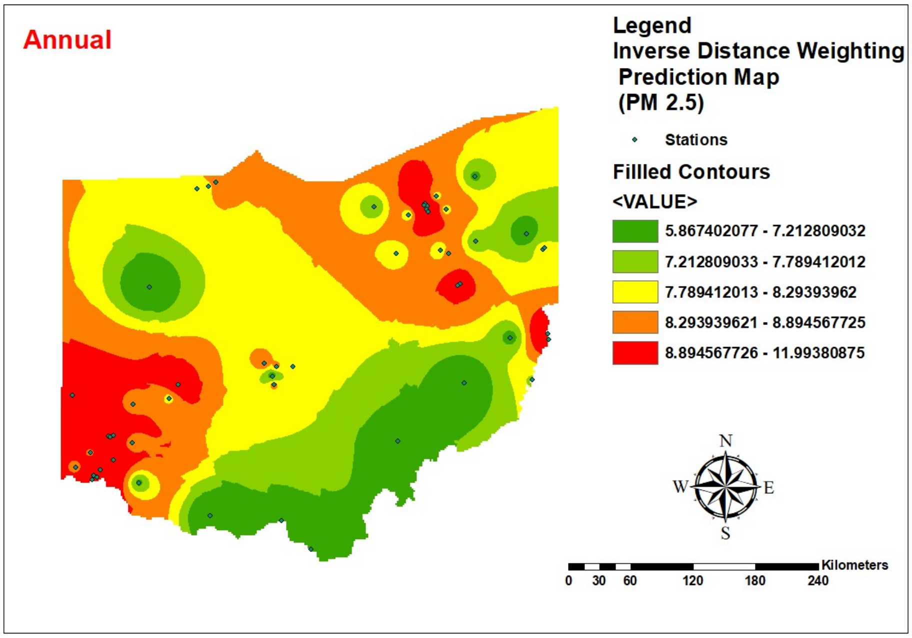

*Fig. **1** PM 2.5 map of Ohio State.*

Annual IDW analysis of AQI

Air Quality index increases as the concentration of PM 2.5 increases in the respective regions (Figure 2). The IDW results of AQI have shown the higher values of AQI at the Western Side stations and Southern stations of State Ohio with a value of 35.48ug/m3 and 46.30ug/m3 respectively. The lower AQI values at southern sites are 24.4ug/m3 and 29.4ug/m3. The Northern sides of Ohio have an unhealthy value of 33.33ug/m3.

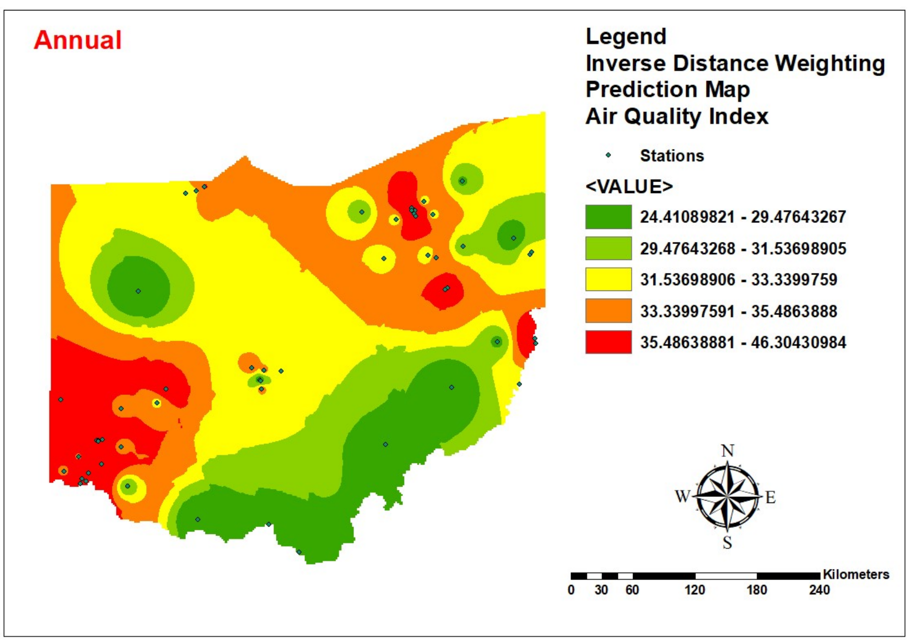

*Fig. **2** AQI map of Ohio State.*

Correlation between PM2.5 and Wind Speed

The highest concentration regions have Wind speed that influences the parallel and perpendicular transport of air pollutants. As wind speed rises, the concentration of PM2.5 decreases noticeably. When wind speed increases from 0-5 km/h to 5-10 km/h, the mean PM2.5 concentration decreases from 55 g/m3 to 33 g/m3. This is because strong wind speeds support the increase of PM2.5. Low wind speed, on the other hand, hinders PM2.5 diffusion and causes air pollutants to build up at the surface. Next, wind direction has a considerable impact on PM2.5 levels in different regions. PM2.5 levels varied depending on the direction of the wind blowing. In this way, the west wind carried the maximum PM2.5 throughout the spring. In the summer, the South East wind, North West wind, and West wind conveyed additional PM2.5 than other directions of the wind. In this way, in fall, the West wind, and South East wind bring more PM2.5 than any other wind direction. In the winter, the South West wind, East South wind, North Wind, and West wind transport more PM2.5 concentration than the other winds. In every season, the west wind brings more PM2.5 than the other winds. Thus, one reason may be air pollutants in Ohio State come mostly from Indiana State.

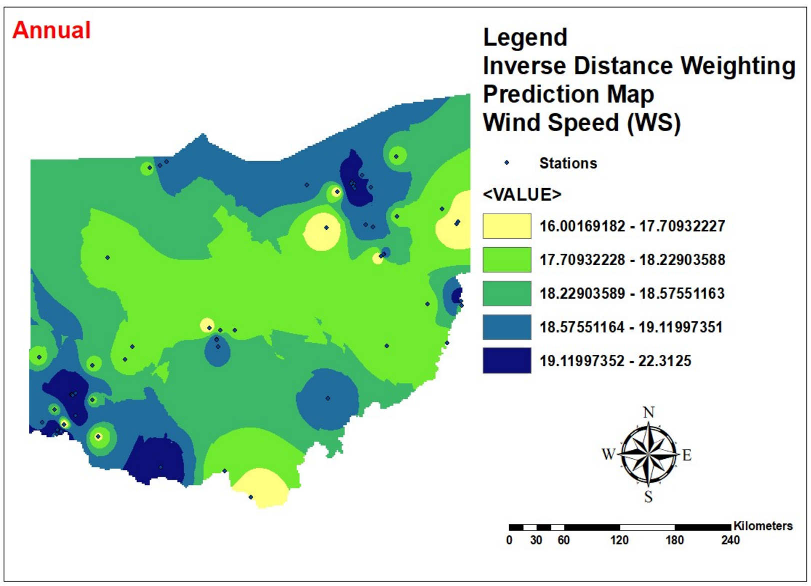

*Fig. **3** Wind speed map of Ohio State.*

Correlation between PM2.5 and Temperature

The Temperature values in 2018 are divided into five ranges which are: 10-15, 15-20, 20-25, 25-30, and 30-35. Due to the subtropical monsoon climate in Ohio State temperature variation is smaller than in other high latitude areas. PM2.5 concentrations mean values decrease with the increase of temperature up to some level. PM2.5 mean value is highest (11 μg/m3) in the temperature range of 10.1-10.5°C. When the temperature value is equal to 10.5°C or greater than 10.5°C, the air convection at the lower surface is strong enough to reduce the PM2.5 concentration up to 10μg/m3 by upward transportation. Some months of Ohio State show a positive correlation of PM2.5 with temperature. One of the reasons for the positive correlation is high temperature contribute to the formation of particulate matter due to the Photochemical process between the particles.

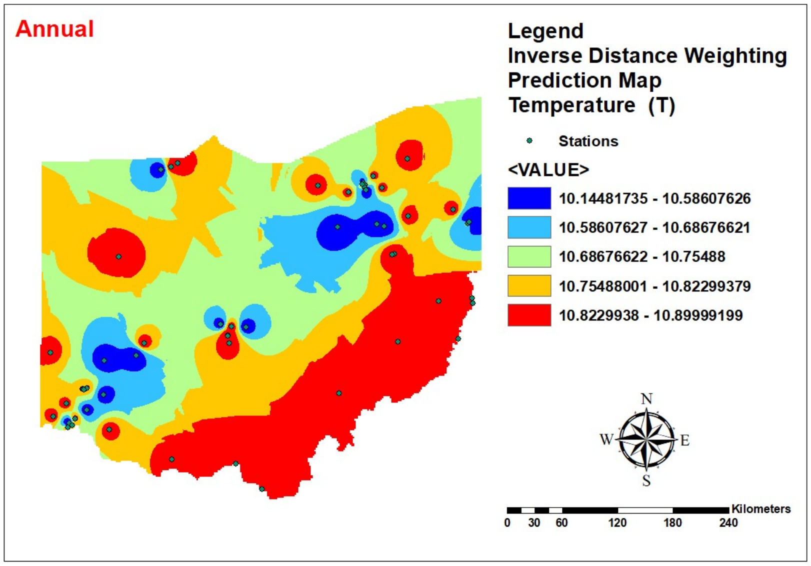

*Fig. **4** Temperature map of Ohio State.*

Correlation between PM2.5 and Precipitation

The rainfall in 2018 was divided into six ranges: 80-90, 90-100,100-110,110-120,120-130, >150. PM2.5 values were calculated for each interval of RH values. About 60% of days in 2018 have rainfall less than 115 mm. The frequency in 110-115 is 15%. The rainfall >120 mm has a low frequency of 2%. PM2.5 concentrations mean values reduce with rainfall for different RH intervals. Especially when the rainfall changes from<110 to 115, PM2.5 concentration has reduced from 8 µg/m3 to 7 µg/m3. Therefore, precipitation can improve air quality by wet deposition, and PM2.5 concentration can be reduced significantly.

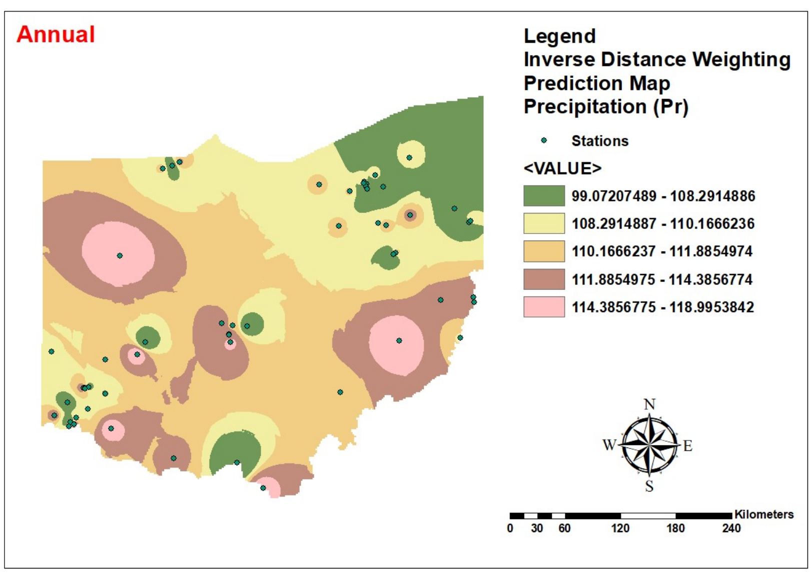

*Fig. **5** Precipitation map of Ohio State.*

Correlation between PM2.5 and Relative Humidity

The relative humidity in 2018 ranging from 80% - 90% was classified into 3 intervals: 60-70, 70-80, 80-90.Ohio State has a humid climate and has a strong negative correlation between PM2.5 concentration and RH. The frequency of RH between 80% and 90% exceeds 60% in Ohio State. PM2.5 concentration means the value reduces in each interval with the increase of RH value. Some months show positive correlations in Ohio State but the correlation coefficient is very low. A negative correlation exists between PM2.5 and RH in summer in Ohio State because the humidity value is higher than 80%. When the humidity value is low in other months the concentration of PM2.5 increases due to hygroscopic growth present in the air. As a result of the heavy weight of dry deposition, the particles fall and the concentration of PM2.5 reduces.

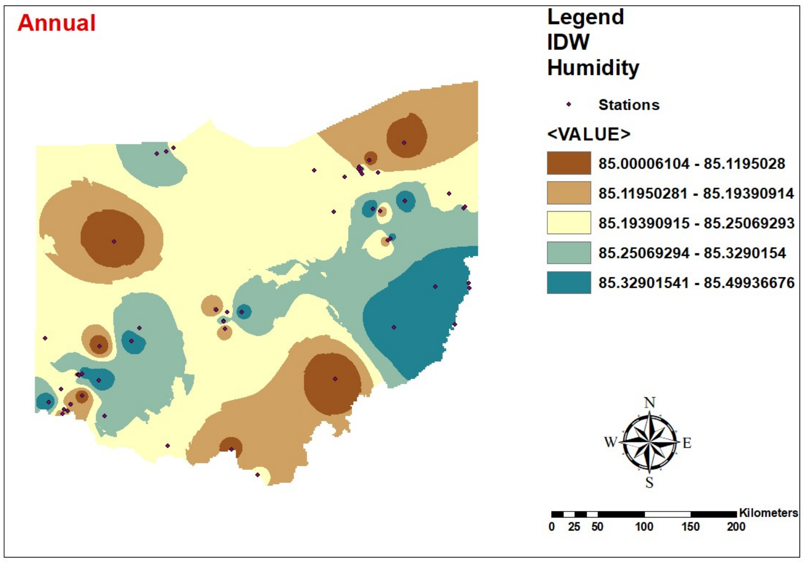

*Fig. **6** Humidity map of Ohio State*

Three months IDW analyses (July-September)

The highest concentration region in August and September indicates Wind speed influences the horizontal and vertical transport of Air pollutants. Humidity and Precipitation also have a positive correlation with Air quality in these months, temperature has a negative correlation with PM2.5 in these months.

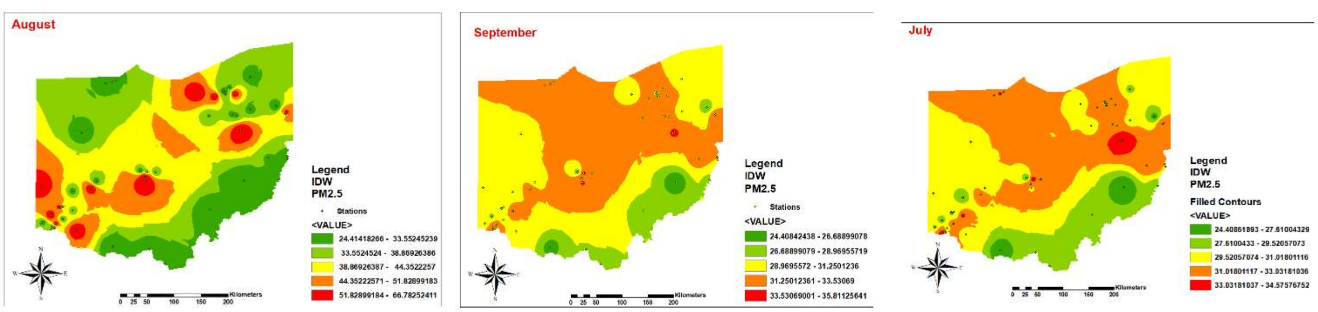

*Fig. **7** PM2.5 maps of July to September 2018*

Ordinary linear squares models of PM2.5

By using Ordinary linear squares regression to develop equations for PM2.5 based on the use of independent variables with a typical accurateness of 97%. OLS results in summary for July, August, and September 2018 are shown in Tables 1, 2, and 3 respectively.

4.8.1 OLS results in July 2018

Table 1 OLS results summary for July

Variable	Coefficient	Std Error	t_ Statistic	Probability	Std Coefficient
Intercept	−8.9877	29.7821	−0.2976	0.8231	0
Precipitation	3.1334	0.1972	9.9832	4.2372E − 14	0.8824
Temperature	0.0828	0.4972	0.2011	0.9402	0.0207
Wind speed	2.9822	0.6973	3.9521	1.9721E − 06	0.0697
Humidity	0.2011	0.1894	0.7923	0.3922	0.0683

4.8.2 OLS results in August 2018

Table 2 OLS results summary for August

Variable	Coefficient	Std Error	t_ Statistic	Probability	Std Coefficient
Intercept	−396.8748	198.8743	−1.9824	0.0452	0
Precipitation	−0.6321	0.8734	−0.6021	0.6012	−0.3032
Temperature	−2.1021	0.5228	−1.7341	0.0424	−0.3141
Wind speed	22.4264	1.6731	6.7824	1.397E − 09	0.6972
Humidity	6.6374	1.9372	1.7866	0.0200	0.8774

4.8.3 OLS results in September 2018

Table 3 OLS results summary for September

Variable	Coefficient	Std Error	t_ Statistic	Probability	Std Coefficient
Intercept	−498.3970	206.5856	−2.5156	0.02175	0
Precipitation	9.6224	3.2465	3.9762	4.9747E − 06	0.9872
Temperature	−6.5832	1.9982	−2.1556	0.0357	−0.7797
Wind speed	16.1574	4.3215	3.2155	0.0029	0.3917
Humidity	8.2432	1.9858	3.9826	2.626E − 06	2.0022

When ordinary linear squares investigation was completed based on data, in this way three analytical models have been set to study Particulate Matter (PM2.5). The association and modeling among dependent and independent variables according to the circumstances mentioned have been experienced.

Equations

The obtained three modeling equations from regression analysis have been made for the measurement of Particulate Matter (PM2.5) from July to September represented by three Equations (1, 2, and 3) respectively.

$$\text{PM2.5 July} = -8.9877 + 2.9822\ Ws + 3.1334P \tag{1}$$

$$\text{PM2.5 August} = -396.8748 - 2.1021T + 6.6374\ H + 22.4264\ Ws \tag{2}$$

$$\text{PM2.5 September} = -498.3970 - 6.5832\ T + 8.2432\ H + 16.1574\ Ws + 9.6224P \tag{3}$$

Where, Particulate Matter value is represented by PM2.5, Relative Humidity is represented by H (%), Temperature is represented by T (°C), Wind speed is represented by Ws (m/sec), and Precipitation is represented by P (mm). The Parameters used as independent variables are; T, H, P, and Ws, with PM2.5 as a dependent variable. Based on both of the results of IDW and NASA satellite data, which are combined through RET screen software and OLS PM2.5 prediction equations were made to approximate the value of PM2.5 at all the points recorded in the region. It's depending upon the multiple regressions formula.

In the first case to analyze the relationships between independent and dependent variables OLS regression method was used. Precipitation and wind speed are included in the first equation used in the modeling. In July temperature has no impact on PM2.5. Linear relationships between variables are limited by the correlation. Most of the variables have a high level of co-linearity with each other but the co-linearity of independent variables temperature and humidity, in this case, is low. The Temperature probability value as shown in Table 1 is 0.97 which is high than the 0.05 value, for this reason, it has not been included in the first equation. The studies consider a good correlation which is more than 0.70. But in this case, the goal is to find minimum relations between the variables to find the impacts of these variables. Therefore, based on the use of these two variables, a regression model was made represented in Equation (1).

In the second case, the model represented by Equation (2) included three independent variables: wind speed, humidity, and temperature. In this case, precipitation does not correlate so we excluded it from the equation and model construction. In the third case, the model that we have represented by Equation (3) included temperature, humidity, wind speed, and precipitation. In this case, all independent variables correlate.

These three equations find the values of the predicted variable for three months. It does not produce the predicted variable absolute values but represents the closed or near approximated to the real observed value. In these three equations, the coefficients' physical significance shows us a good optimization level, which represented the near value of the real observed value can be determined accurately. The three estimated mathematical equations have accuracy corresponding to R^2 and the probability value obtained from the regression analysis. Forward computations were used to achieve the extra high R^2 adjusted in estimated model equations. The correlation shown by the three equations, wind speed has the highest share in the increased concentration of PM2.5 and next the precipitation. When the speed of the wind is strong it increases dust storms and resulting in increased particle sizes.

OLS model justification

In the below Figure, by using 40 points, the justification has been experienced. This justification uses the polynomial linear polynomial fitting to calculate PM2.5 values by the OLS model and compare them to the observed values of PM2.5.

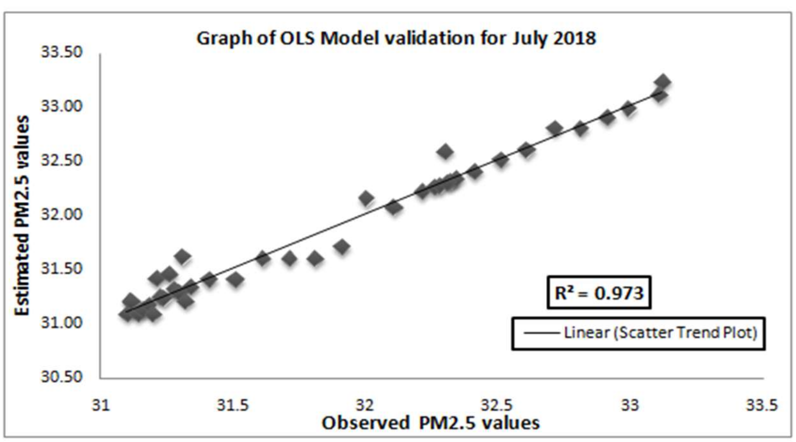

Fig.8 Graph of OLS Model validation for July 2018

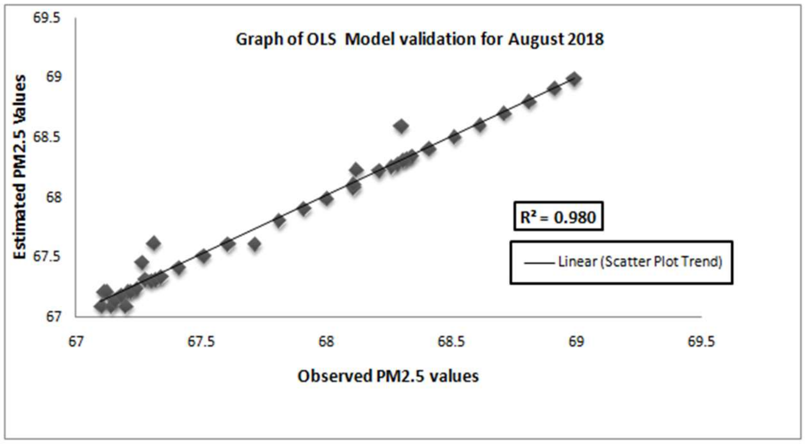

Fig.9 Graph of OLS Model validation for August 2018

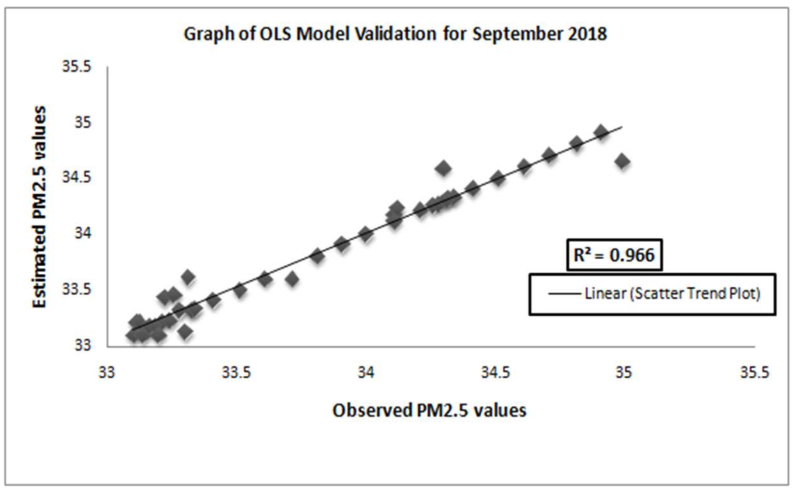

Fig 10 Graph of OLS Model validation for September 2018

The justification applied for the months (July, August, and September). The results with R^2 up to 0.97 in July and R^2 up to 0.98 in August, and 0.96 in September showed that the total records are within the boundary of confidence. Here the observed PM2.5 values refer to the values entered in the PM2.5 model and the estimated PM2.5 represents the values that are calculated from the result regression. The predicted method by putting the point's objective is to accuracy attainment and to test the concentration of PM2.5 at the position point.

Practical Relevance

This study will contribute to air quality monitoring Agencies in the following ways. The IDW and OLS proposed in this study are important because it can support efficient future spatial PM2.5 monitoring. Next, various other hazardous pollutants which increase the air quality index can be estimated by adopting these proposed methods. Engineers will be able to develop Air quality monitoring systems and mapping systems based on including meteorological factors. This study will also help in regulatory or policy assessments to reduce the impacts of PM2.5 by estimating the concentration of pollutants.

5 Conclusion

1 The Inverse Distance Weighting (IDW) method is valuable for measuring the consequence of particular meteorological factors on PM2.5 concentrations.
2 The Ordinary least squares (OLS) method is valuable for investigating the general influence on PM2.5 concentrations by meteorological factors.
3 PM2.5 concentrations are interrelated to meteorological factors: wind speed, temperature, wind direction, precipitation, relative humidity, and air pressure in several ways, including spreading, expansion, chemical creation, photolysis, and deposition.
4 To enhance the perception of PM2.5 meteorological relations, further field tests (e.g., smog chambers and sounding) should be conducted to confirm model results. Second, the accuracy of CTM methods should be increased. Third, quantitative data should be extracted for forecasting and regulating PM2.5 concentrations.
5 To mitigate PM2.5 pollution in a specific area, wind corridors and artificial precipitation can be used according to the geographical circumstances.

Acknowledgment

The authors would like to thank every person/department who helped thorough out the research work, particularly the EM department, and IE department. The careful review and constructive suggestions by the anonymous reviewers are gratefully acknowledged.

Reference

[1] V. A. Southerland *et al.*, "Global urban temporal trends in fine particulate matter (PM2· 5) and attributable health burdens: estimates from global datasets," vol. 6, no. 2, pp. e139-e146, 2022.
[2] C.-M. J. E. M. Liu and Software, "Effect of PM2. 5 on AQI in Taiwan," vol. 17, no. 1, pp. 29-37, 2002.
[3] M. Roy, C. Brokamp, and S. J. A. Balachandran, "Clustering and Regression-Based Analysis of PM2. 5 Sensitivity to Meteorology in Cincinnati, Ohio," vol. 13, no. 4, p. 545, 2022.
[4] A. Gattrell and M. Loytonen, *GIS and Health*. Houghton Mifflin Harcourt, 1998.
[5] P. Poursafa, M. Mansourian, M.-E. Motlagh, G. Ardalan, and R. J. E. r. Kelishadi, "Is air quality index associated with cardiometabolic risk factors in adolescents? The CASPIAN-III Study," vol. 134, pp. 105-109, 2014.
[6] P. L. Fung *et al.*, "Improving the current air quality index with new particulate indicators using a robust statistical approach," vol. 844, p. 157099, 2022.
[7] A. P. Dadhich, R. Goyal, P. N. J. T. E. J. o. R. S. Dadhich, and S. Science, "Assessment of spatio-temporal variations in air quality of Jaipur city, Rajasthan, India," vol. 21, no. 2, pp. 173-181, 2018.
[8] Y. Murayama and R. B. J. T. G. L. Thapa, "Spatial analysis and modeling in geographical transformation process," vol. 100, 2011.
[9] T. L. Nyerges and P. Jankowski, *Regional and urban GIS: a decision support approach*. Guilford Press, 2009.
[10] D. Xu *et al.*, "PM2. 5 Exposure and Health Risk Assessment Using Remote Sensing Data and GIS," vol. 19, no. 10, p. 6154, 2022.
[11] Y. Tian, X. Yao, L. J. C. Chen, Environment, and U. Systems, "Analysis of spatial and seasonal distributions of air pollutants by incorporating urban morphological characteristics," vol. 75, pp. 35-48, 2019.
[12] P. Twomey and M. J. I. j. o. c. p. Kroll, "How to use linear regression and correlation in quantitative method comparison studies," vol. 62, no. 4, pp. 529-538, 2008.
[13] A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, *Handbook of spatial statistics*. CRC press, 2010.
[14] T. Thongthammachart, S. Araki, H. Shimadera, T. Matsuo, A. J. E. M. Kondo, and Software, "Incorporating Light Gradient Boosting Machine to land use regression model for estimating NO2 and PM2. 5 levels in Kansai region, Japan," vol. 155, p. 105447, 2022.