Statistical Analysis of Meteorological Factors and Carbon Dioxide as an Air Pollutant in Al-Hilla City, Iraq
Ala’a Hamed Emran Al-Husseini*
College of Engineering, Babylon University, Babil, Iraq.

ABSTRACT
The objective of this study is to show the connection among an air pollution parameter and meteorological factors in Al-Hilla town. Data consists of daily averages of three meteorological factors (1. wind speed WS (m/s), 2. air temperature T (°C) and 3. relative humidity RH (%)) also, an air pollution parameter (carbon dioxide CO₂) for the period 2015-2016. In order to describe the connection between the 4 variables, the multivariate statistical analysis is used. Multiple linear regression equations are used to express the connection between meteorological parameters which represent the independent variables and concentration of air pollutant which is the dependent variable, by SPSS and Excel software. The result is: decreasing in wind speed and air temperature be accompanied by moderate decreased in CO₂ values. Increasing relative humidity be accompanied by moderate increased in CO₂ values. Linear regression analysis indicates a moderate level of connection between carbon dioxide CO₂ concentrations and the meteorological parameters in Al-Hilla city.

*Corresponding author.
E-mail addresses: a_alhusseini16@yahoo.com

130
Introduction

In urban areas the meteorological factors, the geographical and topographical are periodic, so air pollution has various properties. A mixture of the different meteorological elements are effect on air pollution concentrations. Therefore, air pollutants concentrations and the meteorological information may be correlated them by statistical models [1].

The relationship between dependent and independent variables can be classically descriptive by linear regression, which is a powerful and overall analysis method. Regression analysis consists of estimating the parameters of model, then predicting the response variable [2].

The national and regional areas need to make studies on air pollution. There are many factors effect on air pollution such as distributions of pollution sources, geographical locations and climatological circumstances [3].

Meteorological conditions influence the levels of air pollutants. Serious pollution episodes do not come from local sources but they are related to a given weather conditions, during which pollutants can be accumulated. It has been found that nocturnal temperature inversions with light winds or calm weather, namely an anticyclonic circulation pattern, are a favorable condition for pollutants accumulation [4].

Recently, the issue of air pollution has become one of the most important environmental problems due to overcrowding, traffic and industrial growth [5].

The dispersion is the most important factor affecting on air pollution because of the mitigation of concentrations of air pollutants from the ambient ambience. There is a need for studies that describing the relation between meteorological factors and air pollution. [6]

The affecting of meteorological factors (precipitation, air temperature) on concentrations of particulate matter is well explained by and other meteorological factors. [7].

Pollution exposure depends on the contamination components and the source as well as meteorological elements. Traffic-related pollutants such as nitrogen dioxide, ozone and particulate matter with size less than 10 μm have been linked to asthma exacerbation and allergic responses [8].

Many factors affecting on air pollution in cities such as population, industrialization, traffic load and meteorological parameters [9].

The Climate Change Intergovernmental Panel presented that the universe becomes warmed in the last years, and the increasing in temperature at the middle of 20th century is related with the increasing in anthropogenic CO₂ concentration [10].

Materials and Methods

1. Description of Studied Area

Al-Hilla, city is located in the central of Iraq on Shutt Al- Hilla (branch of the Euphrates River), south of Baghdad at about 100 km and it's adjacent to the ancient city of Babylon. It's located around 32° 29' N and 44° 26 E and about 35 m above the sea level. The annual precipitation in Al-Hilla city is about 0.37 inches and humidity of 23%. It is dominated by desert climate characterized by low rainfall and high temperature in summer, which reach 50°C, there is a warm atmosphere in winter and residential population was 970,000 in 2015. The area of Al-Hilla city is 5,119 km².

2. Records and Methodology

Records of this study involves of daily averages (each 60 minutes) of 3 meteorological elements (1. wind speed WS (m/s), 2. air temperature T (°C) and 3. relative humidity RH (%)) and the concentration of the air pollutant carbon dioxide (CO₂) form 2015-2016. All records have been done by Abu Khastawi monitoring station, which is situated in the western north of Al-Hilla town (shown in Fig.1). This station was managed by the Alleged Environmental Protection in the province of Babylon as a branch of the Iraqi Ministry of Environment.

![Figure 1: Map with air observing station in Al-Hilla urban.](image)

Regression Analysis

To describe the linear relation between more than one variable, regression analysis is used. Regression is primarily used for causal inference
and prediction [11]. Regression problem is when study variable depends on more than one independent or explanatory variables, called as multiple linear regression model. This model specifies the simple linear regression in two ways. Let \( y \) denotes the study (dependent) variable that is linearly related to \( k \) explanatory (independent) variables \( x_1, x_2, \ldots, x_k \) through the parameters \( \beta_1, \beta_2, \ldots, \beta_k \) the multiple linear regression model is[12]:

\[
y = a + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon
\]

(1)

\( a \) is the regression constant and \( (\beta_1, \beta_2, \ldots, \beta_k) \) are the coefficients of regression associated with \( (x_1, x_2, \ldots, x_k) \) respectively and \( \varepsilon \) is the random error component reproducing the difference between the observed and fitted linear relationship.

To show the best regression fitting model, coefficient of determination \( R^2 \) is calculated as below [5]:

\[
R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{\text{pre},i} - y_{\text{obs},i})^2}{\sum_{i=1}^{N}(y_{\text{obs},i} - \bar{y})^2} 
\]

(2)

\( y_{\text{pre},i} \): predicted value from application of the regression equation, \( \bar{y} \): \( y \) mean value and \( y_{\text{obs},i} \): \( y \) observed value. If \( R^2 \) value is equal to one, this certain that all the predicted data are on the fitted regression line and vise inverse for \( R^2 \) equals to 0. \( R^2 \) decreases with a decrease in the variation of the independent variables [5].

This study using a stepwise regression method, it's a combination of the forward and backward selection techniques, it was very popular. Stepwise regression needs two significance levels, one for removing variables and the other for adding variables [13]. The correlation between independent variables and dependent variable is expressed in equation 1 can express in stepwise model as follows [5]:

\[
y = f(x_1), \quad y = f(x_2), \ldots, \quad y = f(x_1, x_2), \ldots, \quad y = f(x_1, x_2, \ldots, x_k),
\]

(3)

Additionally to \( R^2 \) (coefficient of determination), MBE (mean bias error) and RMSE (root mean square error) are used for examining the goodness of fitting, which are determined as follows [5]:

\[
MBE = \frac{1}{N} \sum_{i=1}^{N}(y_{\text{pre},i} - y_{\text{obs},i})
\]

(4)

\[
RMSE = \left[ \frac{\sum_{i=1}^{N}(y_{\text{pre},i} - y_{\text{obs},i})^2}{N} \right]^{1/2}
\]

(5)

Where the number of observations is \( N \). In multiple linear regression analysis, \( CO_2 \) concentrations data are represented the dependent variable and the climatic factors: (1. wind speed, 2. air temperature and 3. relative humidity) are the independent variables. Analysis was performed by using the SPSS and Excel programs.

Results and Discussion

1. Statistics of \( CO_2 \) concentration

The daily and monthly averages of \( CO_2 \) values in 2015 are graphed in Fig.2 to show the daily and monthly variations in concentrations. However, the maximum values of \( CO_2 \), were observed during January, February and December the temperature in these months has the minimum values in the study periods in Al-Hilla city. Table 1 shows the statistics of daily concentrations of \( CO_2 \) and daily information of meteorological elements for the period 2015-2016.
Table 1: Statistics of CO$_2$ concentration and meteorological parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO$_2$ concentration, (ppm)</td>
<td>369.633</td>
<td>27.325</td>
<td>176</td>
</tr>
<tr>
<td>WS (m/s)</td>
<td>5.471</td>
<td>1.980</td>
<td>176</td>
</tr>
<tr>
<td>T (°C)</td>
<td>26.796</td>
<td>10.115</td>
<td>176</td>
</tr>
<tr>
<td>RH (%)</td>
<td>32.338</td>
<td>15.642</td>
<td>176</td>
</tr>
</tbody>
</table>

The measurements of meteorological parameters are done by predictable meteorological instruments at Abu Khastawi monitoring station. Fig. 3 shows the daily averages of meteorological parameters values in the year 2015-2016. It can be seen that the values of wind speed are stable with time and have no trend, temperature reaches to the highest values in June, July, August and September months. Finally, as it is clear that, the daily average relative humidity is increased in the colder months January, February, November and December.

The relation between daily variations in wind speed, temperature and relative humidity with CO$_2$ concentration, shows in Fig. 4. It can be seen from these figures that CO$_2$ concentration are moderate decreased with increasing of both wind speed and air temperature. This situation shows that pollutants dilute by dispersion when wind speed is high. With increasing of relative humidity, CO$_2$ concentration is increased also, this adverse related is also indicated by many researchers such as Akpinar and et al [5].
2. Models establishment

The analysis of the air quality data and meteorological data sets are continued by applying multiple linear regression (MLR) with the additional aim of developing effective operational forecasting models for CO₂ concentrations. In Table 2 the correlations R between daily averages meteorological parameters and daily averages CO₂ concentration are shown. CO₂ is correlated with meteorological factors as a moderate to weak level but it is significant ρ for temperature and relative humidity (ρ < 0.05) as seen in Table 2. The regression equations between CO₂ concentration and meteorological factors with correlation coefficients for each step of stepwise regression method in Table 3, the last step gives the best equation which has the higher value of R.

Table 2: The correlation coefficients between various meteorological parameters and CO₂

<table>
<thead>
<tr>
<th>Statistics</th>
<th>WS (m/s)</th>
<th>T (°C)</th>
<th>RH (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation</td>
<td>0.019</td>
<td>-0.613</td>
<td>0.329</td>
</tr>
<tr>
<td>Sig.(2-tailed)</td>
<td>0.399</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>N</td>
<td>176</td>
<td>176</td>
<td>176</td>
</tr>
</tbody>
</table>

Table 3: The regression equations and the coefficients between CO₂ concentration and different meteorological parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>414.039-1.657*T</td>
<td>0.613</td>
</tr>
<tr>
<td>2</td>
<td>468.740-2.695<em>T-.832</em>RH</td>
<td>0.675</td>
</tr>
<tr>
<td>3</td>
<td>459.246-2.736<em>T-0.808</em>RH+1.8*WS</td>
<td>0.687</td>
</tr>
</tbody>
</table>

Figure 4: CO₂ concentrations versus: (a) wind speed, (b) air temperature and (c) relative humidity in 2015-2016.

Figure 5 shows histograms of the residuals of CO₂ model, the residual analysis indicated that the residuals are normally distributed. Fig.6 indicates the normality of residuals (all points are near to the straight fitted line). Fig.7 displays the plot of fitted values with residuals for CO₂ model indicating that the uncorrelated residuals are because the residuals are contained in a horizontal band and clearly that variance are constant this also indicates normality.

Figure 5: Regression standardized residuals histogram of CO₂.

Figure 6: Regression standardized residuals normal P-P plot of CO₂.
3. Models verification

To determine a goodness-of-fit of the models the predicted daily CO$_2$ concentrations for the model resulting for Al-Hilla city during the year 2015 was plotted in Fig. 8 against the observed values. The regression lines showing 95% confidence interval are drawn also, which is including almost the points. For regression model line A and B are the upper and the lower 95% confidence limits. The accuracy of the forecasted model is 95%, $R^2$ is 0.472, $MBE$ is 2.525 and $RMSE$ is 33.497.

$$\text{Forecasted CO}_2 = 459.246 - 2.736*T - 0.808*RH + 1.8*W$$ (6)

Figure 7: Scatterplot of regression standardized residuals of CO$_2$.

Figure 8: Measured and predicted concentration CO$_2$ value according to equation 6 for 2015-2016 year.

Conclusions

The connection between meteorological variables and the concentration of the air pollutant CO$_2$ was studied at Al-Hilla city during 2015-2016, by using the statistical method of multiple regression analysis. According to the results, meteorological variables effects on air pollutant CO$_2$ were moderate investigated. As a general result, with decreasing wind speed and temperature CO$_2$ values was moderate decreased. High wind speed is accompanied with best ventilation conditions and, consequently, with low concentration levels of primary pollutants and vice versa. With increasing relative humidity it was moderate increased. Finally, the outcome of this research is; there is a moderate correlated among CO$_2$ (air pollutant concentration) and the climatological parameters in Al-Hilla urban. Statistical models were developed in order to predict the CO$_2$ concentration with regard to meteorological parameters. The coefficient of correlation for statistical model of CO$_2$ including meteorological parameters is 0.687. This equation can be employed for many purposes, such as, when data are not detected when the instruments cannot be used or intended for early warnings system for public health as well as for local authorities to formulate strategies in improving the air quality at Al-Hilla city.

References


8. NCSS Statistical Software NCSS.com, "Stepwise Regression", Chapter 311. © NCSS, LLC. All Rights Reserved.


