❏   1131

Towards an accurate Ground-Level Ozone Prediction

Eiman Tamah Al-Shammari
College of Computing Sciences and Engineering, Kuwait University, Kuwait

Article Info	ABSTRACT
	This paper motivation is to find the most accurate technique to predict the ground level ozone at Al Jahra station, Kuwait. The data on the meteorological variables (air temperature, relative humidity, solar radiation, direction and speed of wind) and concentration of seven pollutants of environment (SO_2, NO_2, NO, CO_2, CO, NMHC, and CH_4) were applied to forecast the ozone concentration in atmosphere. In this report, three methods (PLS regression, support vector machine (SVM), and multiple least-square regression) were used to predict ground-level ozone. We used Fifteen parameters to evaluate the performance of methods. Multiple least-square regression, partial least square regression (PLS regression), and SVM using linear and radial kernels were the best performers with MAE (mean absolute error) of 9.17x 10-03, 9.72 x 10-03, 9.64 x 10-03, and 9.12 x 10-03, respectively. SVM with polynomial kernel had MAE of 5.46 x 10-02. These results show that these methods could be used to predict ground-level ozone concentrations at Al Jahra station in Kuwait.

Corresponding Author:

Eiman Tamah Al-Shammari,
Departement of Information Science,
Kuwait University,
Safat 13060, Kuwait.
Email: eiman.tamah@gmail.com

1. INTRODUCTION

This paper intends to find the most accurate technique in forecasting the ground level ozone. Kuwait has a number of civic areas where the pollution of air has dramatically grown up due to the industrialization and technical development, and over population [1]. All such factors have been contaminating the air and causing enhanced air pollution in the areas with high number of population. These populated areas of Kuwait might experience severe health related issues in near future due to the pollutants in their nearer proximity. At the initial stage, the ambient of ozone layer has been severely damaged by the pollution and this has become a matter of concern due to the increasing pollutants that are being originated from developed and industrialized nations. Air pollution can have more drastic affects for populations living in areas closer to the industrial estates that are causing the pollution in the air. This might be dangerous for the health of all living things especially chronic respiratory infection, aggravate asthma, lung inflammation, damaged lung defense mechanisms and decreased immunity in human beings [2].

Disturbance to the balanced atmosphere has become the major cause of damage to the ozonosphere. This damage might lead to severe foliar injuries, reduction in biomass production and agricultural yield. A considerable shift in the competitive benefits from different species of plants in varied populations has also been seen [3]. The ozone layer's damage is consistently increasing which is dangerous for the health of humans and the environment [4]. According to scientists, the ozone layer is formed by O_3 that results the depletion from the reaction of ultraviolet rays and chemical interaction of oxides of nitrogen NOx along with organic species [5].

All such basic pollutants have generally originated from industrial and other factors such as urban development. Lack of protection from ozone might create health issues because of its high reaction, to

chemicals, rubber materials and fabrics. It also affects severely to some crops. Therefore, the ozone's concentration in the lower atmosphere has been greatly consid erable due to its reaction against oxidizing photochemical fog. As a vital index substance of photochemical smog, ozone has been considered as the major pollutant affecting the quality of air and environment [6], [7].

Two main chemical precursors of ozone along with other photochemical oxidants have been recognized as Hydrocarbons (HC) and Nitrogen Oxides (NOx) [6]. Petrochemical processes and fuel or oil burning along with transportation have been the major causes of atmospheric HCs and NOx. Most of such pollutants are measured by emission rates that are counted with the help of activities going on in urban and industrial areas. The association of such basic pollutants with meteorological elements might help reach to a determination of ozone levels. A model that determines the chemical processes and atmospheric movements might be an appropriate strategy. The chemistry of organic species might also be difficult to be accurately collected. The complicated ozone's formation has been uncontrollable. The stratospheric ozone's layer that protects the earth from ultraviolet rays causing harms for both humans as well as for crops. But the lower level ozone concentration has been the main concern in terms of causing health related issues and other material and vegetation effects [8]. Ozone inhalation can initiate numerous health problems like respiratory tract irritation, eyesight problem, coughing, chest tightness, wheezing. Children who usually go outside during the daytime in summer are likely to be at risk when the concentration of ozone remains higher in the air. Further, this ozone is also become a reason of loss in the agricultural production.

In the recent past, the air pollution in the urban areas of the country is heavily increased [9]. This is the due to the overpopulation, technical growth, and rapid industrialization in the country. The levels of ozone pollution as observed in the residential district of Salmiya were increasing the ambient quality standards of air during certain times of year. Therefore, there is an immense need of accurate forecasting of the surface ozone; as with the forecasting it would be assisting in the successful implementation of the warning strategies for public especially during the episodic days in country.

There are three wide areas in terms of meteorology that need to be focused for ozone's concentration through statistical methods. Every area of approach is considerably unique from others: first one is regression based method, the second is extreme value method, and third one is Space-Time approach. Ozone's variability is decreasing that is yet to be understoo d and that has also been under consideration very commonly with the help of meteorological adjustment. The change in the climate, or a change in the policy, ultimately creates a change in the process. These changes might be considerably smaller and hard to be identified, which needs efforts of separation of it from weather and climate [10]. Regression based method and extreme value method, both concentrate on forecasting, estimation and revealing the fundamental mechanisms.

Similarly, the studies carried out for the analysis of ozone level were focused on the comparison of ozone levels with the standard limits internationally. This comparison includes the study of seasonal trends of ozone levels, understating of behavior diurnally in ozone, assessment of effects on health by ozone pollution [11], [12].

Few studies on developing a robust system for a public warning system of forecasting that can be utilized, most of the forecasting systems were developed for the prediction of concentration in the ambient ozone in Kuwait with the use of precursor concentrations and meteorological data [11], [12].

In [1], predicting the levels of ozone from meteorological conditions and precursor concentrations at (SIA) Shuaiba Industrial Area of Kuwait during the daylight hours was achieved by using step wise multiple regression modeling.

The application of artificial neural networks and the principal component regression was done for the prediction of ozone of concentration in the lower atmosphere of Kuwait. The prediction was done using five variables of meteorology (air temperature, relative humidity, solar radiation, wind direction, and wind speed) and the data from seven concentrations of environmental pollution (SO_2, NO_2, NO, CO_2, CO, NMHC, and CH_4).

As linear regression is a well known method ,a number of studies such as [1], [13-16] have discussed multiple linear regression method to associate the ozone's measurement with contemporary meteorological measurements. These models have presented with lack of autocorrelation and cross-correlation and this is why these models do not fulfill the basic requirement of merit, proven scientifically. Time series regression has been another complex factor that associates the still relationship along with a correlation structure such as a simple AR (1), for the residuals [17]. This is considerable only when the fits are diagnosed by appropriate methods. As this technique is robust however, these methods might be inappropriate for obtaining the interactions and nonlinear response of ozone's concentration.

In the paper [18], the comparison of performances of various forecasting systems is done across different locations in Kuwait. Fuzzy modeling and time series are the tools which are used for the analysis. The two forecasting models of analysis depicted a significant improvement in comparison to the currently

used model of forecasting air pollution i.e. pure persistence forecast. Large proportion of ozone variation is described by the daily maximum temperature.

According to [19] the statistical linear models have been experienced as complex for gathering the multifaceted association in ozone and meteorological variables. For gathering the data and developing a parametric nonlinear model, around forty five monitoring stations were established in the Chicago region and they provided considerable data during 1981 to 1991. The data was gathered at the AIRS database. The authors presented a model of the daily median across different sites, with a maximum of one hour of ozone's average values, with the help of nonlinear least squares. During the exploratory graphical evidences and through nonparametric modeling, several relationships were observed such as, contemporary still ozone, relative humidity, upper earth surface temperature, and seven hundred HPA surface wind speeds, that present a trend of parametric forms. The Fourier series is applied in the modeling of seasonal waves.

This helps in calculating the standard errors in the coefficients and authors confirms the occurrence of serial autocorrelation in the model residuals. The authors accordingly apply suitable adjustments by applying the Galant's methods [20] applied the linear method of stepwise multiple regression so that the best fit equation could be formed that relates to the ozone's maximum concentration during the daylight period in the air and meteorological conditions with a twenty four hour air's upwind parcel trajectory. There were four variables involved in the equation, such as maximum upwind ozone on earlier day, maximum temperature upwind of last day, and the average upwind speed in both the upper and the lower layer. The rate of emissions for upwind along with hydrocarbons and nitrogen oxides that form the lower layer of ozone, were also examined and found to be lacking in improvement of multiple correlation coefficients.

A number of evaluations of variables used in meteorological adjustments are aimed at being stepwise but lacking in gathering all appropriate subsets. These subsets were computationally very difficult due to huge number of variables. Stepwise method was found to be missing a global phenomenon of model selection, therefore, it has been causing the problem when a variable is eliminated earlier, might have vital interactions with the others. They are later dropped from the model after being masked. [21] Applied a different strategy for linear models for the determination of health related issues with regards to specific matter and air pollution. It was an approach that set down earlier probabilities of containing the numerous variables and then calculates the uncertainty of associated model with regards to posterior probabilities for a vast number of models.

It is difficult to predict the levels of ozone by using theoretical method (for example detailed atmospheric diffusion model). For the development of forecasting system, the empirical analysis is needed. The well evaluated forecasting model of ozone would be the factor which can raise the chances of a successful control strategy. In addition to this, the daily forecasting of maximum ozone concentrations would be helpful in reducing and avoiding the damages and injuries related with ozone. The conducted research is important in this manner because it gives comparison between different statistical methods for the prediction of ground level ozone at Al Jahra station Kuwait.

Because of its major significance to atmospheric chemistry, ozone has been widely studies for several years both theoretically as well as experimentally [21]. Despite being just a triatomic, learning ozone kinetically, spectroscopically as well as dynamically has been highly difficult for the theory. There is considerable divergence between anticipated as well as detected low temperature rates of $O+O_2$ isotope exchange [23]. At low temperature, experimental rates are three to five times greater than predictions and reveal a negative temperature on the basis of the fact that has been evidenced tough reproducing theoretically. In the troposphere, the level of ozone is of immense significance due to its negative impact on vegetation, materials and human health.

Through complicated photochemical reactions in the sunlight, ground-level ozone is mainly produced from its precursor of NOx as well as volatile organic compounds (VOC), according to [24]. By physical and chemical processes as well as by the meteorological conditions, accumulation of ozone at ground level is affected. Over numerous temporal and spatial scales, atmospheric pollution extents reveal complicated inconsistency with harmful impacts on the environment [25].

The existing condition of improvement in the measurement studies as well as modeling of ozone precursors, transport processes and photochemical behavior has been currently evaluated [26]. Although ozone chemistry has been widely examined in several chamber experiments as well as in the photochemical modeling analysis [27], there are still difficulties in perfectly forecasting ambient ozone levels and its spatial distribution, behavior and related patterns. One has to develop comprehension of not just ozone itself for tracking and predicting ozone, however also the situations that integrate to its formation. It is essential to implement the models describing and helping to know the complicated associations between several variables and ozone levels causing or hindering ozone production.

To predict ozone variations time to time, photochemical models are sometimes used and assist to develop lucrative sources of minimizing ambient ozone to control, particularly the emissions of NOx and

VOC from different sources [28]. Ozone formation differs on the basis of hours, days and seasons, due to the complicated series of reactions are handled by sunlight and temperature. For predicting ozone levels, a survey of the significance of meteorology in the surface ozone levels is portrayed as well as linear regression ways were employed [29]. In the United States as well as other industrialized countries, ground-level ozone pollution is said to be severe health issues in several cities.

For several ozone-sensitive individuals, specifically people who suffer from respiratory diseases like asthma, high summertime levels may cause distress [30]. In most of the cities, it is clear that the public forecasts or announcements of potential unhealthy ozone air quality for future may be of great advantage to those at the risk of respiratory discomfort [31]. Moreover, "ozone action" processes was intended in several cities to control episodic emission due to its health effect. These actions rely on forecasting of ground level ozone.

2. THE PROPOSED METHOD
2.1. Datasets
The datasets used in this report were discussed in the public warning systems for forecasting ambient ozone pollution in Kuwait. Data from 2006 to 2008 were used for training the methods, and data from 2009 and 2010 were used for testing the methods. Datasets were pre-processed in Microsoft Excel using information from [32] before analyses.

2.2. Methods
2.2.1. Multiple Least-squares Regression
Multiple least-squares regression (MLR) models the relationships between dependent variables(Y) and independent variables (X) as shown in Equation (1).

$$Y = XB + E \qquad (1)$$

where B is the matrix of regression coefficients and E contains the residuals.

2.2.2. PLS Regression
PLS regression predicts a relationship between a set of predictor variables, X, and a set of dependent variables, Y. In the first stage of PLS regression, w and q are weight vectors derived from X and Y respectively, where the corresponding scores $t = Xw$ and $u = Yq$. Next, calculating least squares regression between u and t, the inner relationship r is determind. At last, rank-one reductions of X and Y are performed such that $Xj-1 = Xj-tjpjT$ and $Yj-1 = Yj-rjqjT$, where $pj = XjTtj/(tjTtj)$, and j is the latent variable being calculated. Thes stages are repetitive until the desired number of latent variables (K) has been extracted
The general model is given by Equation (2) [33].

$$Y = \sum_{j=1}^{K} (X_j \mathbf{w}_j) \mathbf{q}_j^{T} + \mathbf{E} \qquad (2)$$

2.2.3. SVM
SVMs are used widly due to their use of kernel functions to represent data, The differentkernel functions of SVM are Polynomial, Linear, Sigmoid, and Radial Basis Function (RBF) [34]. In this paper linear, radial, and polynomial kernels were used for training and predictions.

2.3. Statistical Analysis
R3.1.2 was used to perform the statistical analyses, with the following packages (pls, and e1071). The performances of the methods were assessed using mean absolute error.

2.4. Experimental Setting
2.4.1. Multiple Least-square Regression
MLR was applied to the normalized data using Equation (1). The regression coefficients and the p-values of the training of the MLR method are represented in Table 1. All variables were significant at $p < 0.001$. Variables with positive regression coefficients were SO2, NO2, WG, SOLAR, TEMP-AMB, and CO2. On the other hand variables with negative regression coefficients were NOX, PM10, CO, CH4, NCH4, WS, TEMP-IND, and Wind-deg. The variable NO had the largest positive effect on the Ozone with the

regression coefficient of 0.5219, and the variable NOX had the largest negative effect on Ozone with the regression coefficient of -0.5369 (Table 1).

Table 1. Regression coefficients and p-values of the predictor variables from training dataset (2006 to 2008)

	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	0.0122023	0.0008534	14.299	<2e-16
SO2	0.0452040	0.0028250	16.001	<2e-16
NO	0.5218786	0.0390372	13.369	<2e-16
NOX	-0.5368621	0.0398481	-13.473	<2e-16
NO2	0.0582727	0.0078268	7.445	1.01e-13
PM10	-0.0058920	0.0008625	-6.831	8.67e-12
CO	-0.0039184	0.0010157	-3.858	0.000115
CH4	-0.0130075	0.0013198	-9.856	<2e-16
NCH4	-0.0077966	0.0014514	-5.372	7.89e-08
WS	-0.0295952	0.0048163	-6.145	8.17e-10
WG	0.0384623	0.0049445	7.779	7.70e-15
SOLAR	0.0123449	0.0003574	34.540	<2e-16
TEMP-IND	-0.0088203	0.0005373	-16.415	<2e-16
TEMP-AMB	0.0202759	0.0003742	54.181	<2e-16
CO2	0.0187256	0.0029063	6.443	1.20e-10
Wind-deg	-0.0020436	0.0003279	-6.233	4.67e-10

2.4.2. PLS Regression

With PLS regression, the question is the number of PLS regression components to consider so that we don't select noise in training the model. In this report, root mean square error of prediction was used to select the number of PLS regression components (Figure 1). Five PLS regression components were selected and used for training the method. The model created was then used for predicting the ground-level ozone concentrations using test dataset (2009 and 2010). The model mean absolute error was 9.72×10^{-03} (Table 2)
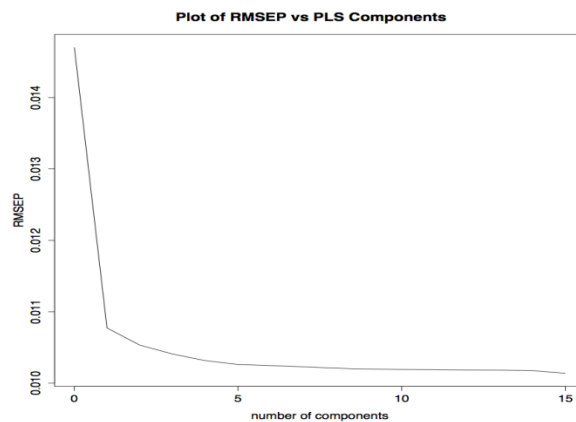

Plot of RMSEP vs PLS Components

Figure 1. Screen plot of PLS regression components. Five PLS regression components were selected and used to build the model.

Table 2. Mean absolute error of MLR, PLS, SVM (linear, radial, and polynomial)

Method	Mean absolute error
MLR	9.17×10^{-03}
PLS regression	9.72×10^{-03}
SVM linear	9.64×10^{-03}
SVM radial	9.12×10^{-03}
SVM polynomial	5.46×10^{-02}

2.4.3. SVM

SVM was trained using three kernels (linear, radial, and polynomial). A grid search was used for selecting the best parameters (gamma and cost) for linear, radial, and polynomial kernels. The training dataset was split into two sets; the first set for training and second set for validation. Table 3 shows the results of grid search for linear, radial, and polynomial kernels, respectively.

The best gamma parameters for linear, gamma, and polynomial were 1×10^{-02}, 1.0×10^{-04}, and 1.2×10^{-03}, respectively; and the best cost parameters for linear, gamma, and polynomial were 1, 1×10^{05}, and 100, respectively.

The mean square errors for predictions were 1.02×10^{-04}, 5.01×10^{-05}, and 1.9×10^{-03} for linear, radial, and polynomial kernels, respectively.

The best parameters obtained from tuning the kernels were used for training the methods. The mean absolute errors of SVM trained using the three kernels were 9.64×10^{-03}, 9.12×10^{-03}, and 5.46×10^{-02} for linear, radial, and polynomial, respectively (Table 2).

Table 3. Results of obtained from tuning SVM three kernels (linear, radial, and polynomial)

Kernel	Gamma	Cost	Mean squared error
SVM linear	1×10^{-02}	1	1.02×10^{-04}
SVM radial	1.0×10^{-04}	1×10^{05}	5.01×10^{-05}
SVM polynomial	1.2×10^{-03}	1×10^{02}	1.9×10^{-03}

3. RESULTS AND DISCUSSIONS

Table 2 and Figure 2 show the performances of all the three methods (MLR, PLS regression, SVMs) on our test dataset.
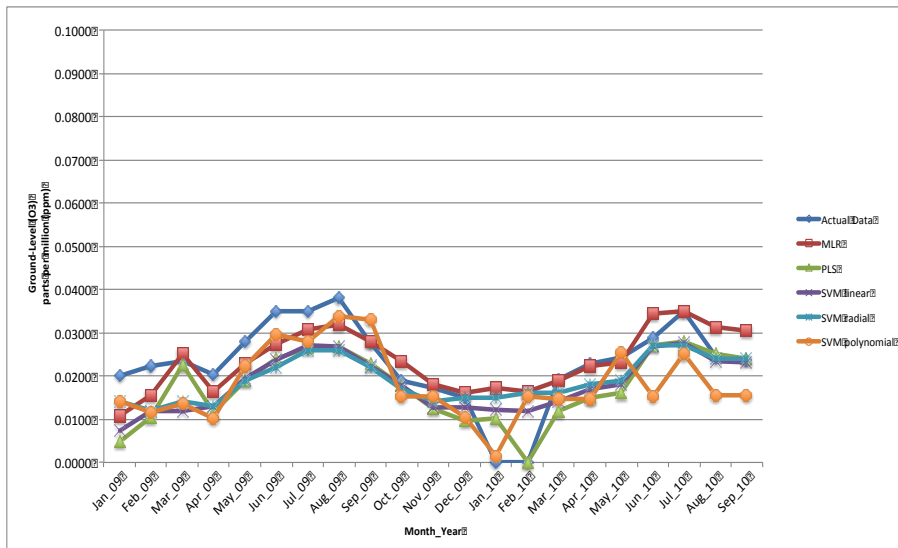

Figure 2. Monthly values of actual, predicted MLR, PLS, SVM linear, SVM radial, and SVM polynomial values of ground-level ozone. MLR: Multiple-least square regression; PLS regression; SVM linear: SVM using linear kernel; SVM radial: SVM using radial kernel; and SVM polynomial: SVM using polynomial kernel

The MLR trained model was used to predict ground-level ozone concentrations using test dataset (2009 and 2010). The model mean absolute error was 9.17x 10-03(Table 2).

MLR, PLS, and SVM with linear and radial kernels performed better than SVM with polynomial kernel in predicting the test dataset as indicated by low MAE (mean absolute error). Additionally, our results show a strong linear relationship between the ground-level ozone concentrations and the predictor variables.

The monthly value of actual data was found to be 0.02 parts per million at the month of January 2009. The SVM using polynomial kernel was below the actual data value and lie between multiple least square regression and actual data. This is the same position where SVM e radial was situated. The position of Multiple least square regression was at 0.01 parts per million.

In the month of February the actual data shows a decreasing trend. While Multiple Least squares Regression is showing a rapid increasing trend , the SVM polynomial is decreased to 0.01 parts per million and at the same point the SVM radial was also present. The PLS regression in February also depicts the position at 0.01 parts per million.

In the month of March the most rapid increasing trend was depicted by Multiple Least squares Regression and PLS. The Actual data is still in decreasing trend but at a slow rate. The SVM polynomial is also increased than the month of February. Here SVM radial is depicting almost the same level of ground ozone that is shown by SVM polynomial. The March shows a decreased trend in SVM linear.

With the start of April the actual data retains its position which it had in the month of January, i.e. 0.0200 parts per million of ground level ozone. The Multiple Least squares Regression also shows a decreased trend. PLS regression in this month has reached the 0.01 ppm level which is far less than the previous month i.e. more than 0.02 ppm. SVM linear has slightly increased whereas the SVM radial has a very slight decrease than last month. SVM polynomial has reached the 0.01 ppm level of ground ozone in this month.

From now onwards there is an increased trend in all the three methods and the predictor variables till the month of August. Actual data also attains the highest position in 18 months after the month of august. From August the trend in every predictor tends to be decreasing especially actual data. In the meanwhile only SVM polynomial is sustained and depicting only a very less fluctuation. After the month of September the trend falls very rapidly specially SVM polynomial. The concentration of ground level ozone decreases in every predictor till the month of November. Multiple Least squares Regression is positioned at the highest point among other methods in this month with 0.02 ppm ground level ozone.

In the month of November the lowest position is of SVM linear. The trend of decrease does not stop here but continues with some fluctuations in different variables and methods till the month of January. January shows the lowest trend of concentration of ground level ozone which ousts it from the periodic season. The SVM radial reaches the lowest of all and is 0 ppm in this month. The SVM polynomial does not depict a different condition than it.

There is only a bit difference where it lies with the SVM radial i.e. near 0 ppm. From January the PLS falls rapidly and reaches 0 ppm in the month of February. Only SVM polynomial is increased at a higher pitch than others and reaches the 0.018 ppm in this month. With the start of March the ground level ozone started increasing. Only SVM radial is at a sustaining position than others and there is very less difference in its readings from previous 3 to 4 months. This depicts the start of season of higher concentration of ozone ground level. From now onwards every method and variable depicts and increasing trend with slight fluctuations. This trend is sustained till the mid of May. After May SVM polynomial has a fluctuating trend among others while SVM polynomial and SVM radial are rather stable. This situation continues till the month of September.

Studies compared the SVM prediction performance in all the branches of atmospheric sciences, such as meterology, atmospheric physics and chemistry in addition to weather forecasting. These Studies support our findings, where SVM demonstrated a robust tool for prediction, examples of these studies could be found in [35-39].

4.   CONCLUSIONS AND SUGGESTIONS

The adverse effects of ozone are vulnerable and can be spread to very long distances and up to a wide range, depending on the direction and wind speed and can be the reason of various bad effects on the health of inhabitants.

In this report three methods (MLR, PLS regression, and SVMs) were effective in predicting ground-level ozone concentrations ozone concentrations at Al Jahra station in Kuwait. MLR, PLS regression, and SVM using linear and radial kernels performed better than SVM using polynomial kernel. SVM tuning is computationally expensive, especially tuning radial and polynomial kernels.

From the analysis it is found that the concentration of ground level ozone (O3) varies with season. This depicts the impact of predictor variables on the concentration. Usually, in the starting 4 months the concentration of O3 is low in air. This trend increases from the month of April and continues till the months of July and August. From then the concentration becomes decreasing till the January and February. In the study the concentration of ozone is determined by the multiple least squares regression, PLS regression, and SVM.

The data for research was collected for eighteen consecutive months starting from January 2009 to September 2010. With the help of the statistical methods used for the forecast of ground level ozone concentration it is found that the concentration is variable on monthly basis. Ozone concentration has a direct impact due to the minor changes in the rainfall, humidity or temperature. It is observed from the study that different methods although give different prediction readings for ozone concentration, some of them are varied a lot in their results. The present research on the specified site of prediction that is Al Jahra station in Kuwait is identified and the work will also be helpful for the future forecast of ozone concentration.

REFERENCES

[1] Abdul-Wahab S, Bouhamra W, Ettouney H, Sowerby B, Crittenden BD, "Analysis of ozone pollution in the Shuaiba industrial area in Kuwait," *International journal of environmental studies*, 2000 Feb 1;57(2):207-24.

[2] Tilton BE, "Health effects of tropospheric ozone," *Environmental Science & Technology*, 1989 Mar;23(3):257-63.

[3] Lefohn AS, Edwards PJ, Adams MB, "The characterization of ozone exposures in rural West Virginia and Virginia," Air & Waste: *Journal of the Air & Waste Management Association*, 1994 Nov;44(11):1276-83.

[4] Jallad KN, Espada-Jallad C, "Analysis of ambient ozone and precursor monitoring data in a densely populated residential area of Kuwait," *Journal of Saudi Chemical Society*, 2010 Oct 31;14(4):363-72.

[5] Gery MW, Whitten GZ, Killus JP, Dodge MC, "A photochemical kinetics mechanism for urban and regional scale computer modelling," *Journal of Geophysical Research*: Atmospheres, 1989 Sep 20;94(D10):12925-56.

[6] Seinfeld JH, "Urban air pollution: state of the science," *Science,* 1989 Feb 10;243(4892):745.

[7] Seinfeld JH, Pandis SN, Noone K, "Atmospheric chemistry and physics: from air pollution to climate change".

[8] Kim SE, Kumar A, "Accounting seasonal nonstationarity in time series models for short-term ozone level forecast," Stochastic Environmental *Research and Risk Assessment*, 2005 Oct 1;19(4):241-8.

[9] Warsito, Agung, Abdul Syakur, and Galuh Susilowati, "An Ozone Reactor Design with Various Electrod Configurations," *International Journal of Electrical and Computer Engineering1*.2 (2011): 93.

[10] Rao ST, Sistla G, Henry R, "Statistical analysis of trends in urban ozone air quality," *Journal of the Air & Waste Management* Association, 1992 Sep 1;42(9):1204-11.

[11] Ettouney RS, Abdul-Wahab S, Elkilani AS, "Emissions inventory, ISCST, and neural network modelling of air pollution in Kuwait," *International Journal of Environmental Studies*, 2009 Apr 1;66(2):193-206.

[12] Ettouney RS, Mjalli FS, Zaki JG, El-Rifai MA, Ettouney HM, 'Forecasting of ozone pollution using artificial neural networks," *Management of Environmental Quality: An International Journal,* 2009 Sep 25;20(6):668-83.

[13] Feister U, Balzer K, "Surface ozone and meteorological predictors on a subregional scale," Atmospheric Environment. Part A. General Topics. 1991 Jan 1;25(9):1781-90.

[14] Korsog PE, Wolff GT, "An examination of urban ozone trends in the northeastern US (1973–1983) using a robust statistical method. Atmospheric Environment," Part B. Urban Atmosphere, 1991 Jan 1;25(1):47-57.

[15] Fiore AM, Jacob DJ, Logan JA, Yin JH, "Long-term trends in ground level ozone over the contiguous United States, 1980-1995," *Journal of Geophysical Research*, 1998 Jan 20;103(D1):1471-80.

[16] Katsoulis BD, "The relationship between synoptic, mesoscale and microscale meteorological parameters during poor air quality events in Athens, Greece," *Science of the total environment*, 1996 Mar 1;181(1):13-24.

[17] Galbally IE, Miller AJ, Hoy RD, Ahmet S, Joynt RC, Attwood D, "Surface ozone at rural sites in the Latrobe Valley and Cape Grim, Australia," *Atmospheric Environment* (1967), 1986 Jan 1;20(12):2403-22.

[18] Al-Shammari ET, "Public warning systems for forecasting ambient ozone pollution in Kuwait," *Environmental Systems Research*, 2013 Feb 1;2(1):2.

[19] Bloomfield P, Royle JA, Steinberg LJ, Yang Q, "Accounting for meteorological effects in measuring urban ozone levels and trends," *Atmospheric Environment*, 1996 Sep 1;30(17):3067-77.

[20] Wolff GT, Lioy PJ, "An empirical model for forecasting maximum daily ozone levels in the northeastern US," *Journal of the Air Pollution Control Association*, 1978 Oct 1;28(10):1034-8.

[21] Raftery AE, Madigan D, Hoeting JA, "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 1997 Mar 1;92(437):179-91.

[22] Pride KR, Peel JL, Robinson BF, Busacker A, Grandpre J, Bisgard KM, Yip FY, Murphy TD, "Association of short-term exposure to ground-level ozone and respiratory outpatient clinic visits in a rural location–Sublette County, Wyoming, 2008–2011," *Environmental research*, 2015 Feb 28;137:1-7.

[23] Dawes R, Lolur P, Li A, Jiang B, Guo H, "Communication: An accurate global potential energy surface for the ground electronic state of ozone".

[24] Pavon-Dominguez P, Jimenez-Hornero FJ, de Rave EG, "Multifractal analysis of ground–level ozone concentrations at urban, suburban and rural background monitoring sites in Southwestern Iberian Peninsula," *Atmospheric Pollution Research*, 2013 Apr 30;4(2):229-37.

[25] Borges AS, Andrade MD, Guardani R, "Ground-level ozone prediction using a neural network model based on meteorological variables and applied to the metropolitan area of São Paulo," *International Journal of Environment and Pollution*, 2012 Jan 1;49(1-2):1-5.

[26] Lu WZ, Wang D, "Learning machines: Rationale and application in ground-level ozone prediction," *Applied Soft Computing*, 2014 Nov 30;24:135-41.

[27] Sundaramoorthi D, "A data-integrated simulation model to forecast ground-level ozone concentration," *Annals of operations research*, 2014 May 1;216(1):53-69.

[28] Robichaud A, Ménard R, "Multi-year objective analyses of warm season ground-level ozone and PM 2.5 over North America using real-time observations and Canadian operational air quality models," *Atmospheric Chemistry and Physics*, 2014 Feb 17;14(4):1769-800.

[29] Sicardi V, Ortiz J, Rincón A, Jorba O, Pay MT, Gassó S, Baldasano JM, "Assessment of Kalman filter bias-adjustment technique to improve the simulation of ground-level ozone over Spain," *Science of the total environment*, 2012 Feb 1;416:329-42.

[30] Kroeger T, Escobedo FJ, Hernandez JL, Varela S, Delphin S, Fisher JR, Waldron J, *"Reforestation as a novel abatement and compliance measure for ground-level ozone,"* Proceedings of the National Academy of Sciences, 2014 Oct 7;111(40):E4204-13.

[31]   Hsu WC, Rosenberger JM, Chen VC, Sattler ML, Sule NV, "Optimizing ground-level ozone control strategies with mixed integer nonlinear programming," COSMOS technical report; 2013.
[32]   Mahmood MA, Al-Shammari ET, El-Bendary N, Hassanien AE, Hefny HA, *"Recommender system for ground-level Ozone predictions in Kuwait,"* InComputer Science and Information Systems (FedCSIS), 2013 Federated Conference on 2013 Sep 8 (pp. 107-110). IEEE.
[33]   Mei Z, Fang C, Guihua L, "Dynamic Error Analysis of CMM Based on Variance Analysis and Improved PLSR," *Indonesian Journal of Electrical Engineering and Computer Science*, 2014 Jul 1;12(7):5342-9.
[34]   Vasanth PC, Nataraj KR, "Facial Expression Recognition Using SVM Classifier,' *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, 2015 Mar 1;3(1):16-20.
[35]   Yao XJ, Panaye A, Doucet JP, Zhang RS, Chen HF, Liu MC, Hu ZD, Fan BT. Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. *Journal of chemical information and computer sciences*, 2004 Jul 26;44(4):1257-66.
[36]   Hamidi O, Poorolajal J, Sadeghifar M, Abbasi H, Maryanaji Z, Faridi HR, Tapak L, "A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran," *Theoretical and Applied Climatology*, 2015 Feb 1;119(3-4):723-31.
[37]   Luna AS, Paredes ML, de Oliveira GC, Correa SM, "Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil," *Atmospheric Environment*, 2014 Dec 31;98:98-104.
[38]   Gong B, Ordieres-Meré J, "Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong," *Environmental Modelling & Software*, 2016 Oct 31;84:290-303.
[39]   Chaiyakhan K, Chujai P, Kerdprasop N, Kerdprasop K. Hourly, *"Ground-level Ozone Concentration Prediction using Support Vector Regression,"* InProceedings of the International MultiConference of Engineers and Computer Scientists 2017 (Vol. 1).

BIOGRAPHY OF AUTHOR

Eiman Tamah Al-Shammari is an Assistant Professor in the Department of ISC at Kuwait University. She received her PhD from George Mason University. Where she carries two master degrees one in Computer engineering from Kuwait University and another one in Maagement of Information Systems from Walton school of Business. In 2013 she was awarded a patent in the field of Arabic Natural Language Processing. Where she focused in Data and text mining during her doctorate studies. Her research interests shifter later to include embedding of information systems issues in education and real life problems. Her interest in educational technologies started when she decided to fight for her autistic nephew educational rights. She is the author and co-author of many refereed articles. She has various books in social media . She is also a music composer, a poet and a movie director. Later she gained a spot in Guinness World record , after creating an Anti-bullying campaign.