

Application of Support Vector Machine and Gene Expression Programming on Tropospheric ozone Prognosticating for Tehran Metropolitan

Vahid Mehdipou^{a*}, Mahsa Memarianfard^b

^a M.Sc. Environmental Civil Engineering student at K. N. Toosi University of Technology, Tehran, Iran.

^b Assistant Prof. Civil Engineering Faculty, K. N. Toosi University of Technology, Tehran, Iran.

Received 18 July 2017; Accepted 25 August 2017

Abstract

Air pollution became fatal issue for humanity and all environment and developed countries unanimously allocated vast investments on monitoring and researches about air pollutants. Soft computing as a novel way for pollutants prediction can be used for measurement tools calibration which can coincidentally decrease the expenditures and enhance their ability to adapt quickly. In this paper support vector machine (SVM) and gene expression programming (GEP) as two powerful approaches with reliable results in previous studies, used to predict tropospheric ozone in Tehran metropolitan by using the photochemical precursors and meteorological parameters as predictors. In a comparison between the two approaches, the best model of SVM gave superior results as it depicted the RMSE= 0.0774 and R= 0.8459 while these results of gene expression programming, respectively, are 0.0883 and 0.7938. Sensitivity of O₃ against photochemical precursors and meteorological parameters and also for every input parameter, has been analysed discreetly and the gained results imply that PM_{2.5}, PM₁₀, temperature, CO and NO₂ are the most effective parameters for O₃ values tolerances. For SVM, several kernel tricks used and the best appropriate kernel selected due to its result. Nonetheless, gamma and sin2 values varied for every kernel and in the last radial basis function kernel opted as the best trick in this study. Finally, the best model of both applications revealed, and the resulted models evaluated as reliable and acceptable.

Keywords: Gene Expression Programming; Support Vector Machine; Tropospheric Ozone; Air Pollution; Tehran.

1. Introduction

In a worldwide study in 2012 by world health organization (WHO) three million people lose their life because of air pollutions in the whole world which shows the severe and hazardous effects of pollutants on human and environment [1]. In this report Iran introduced as one of the effected countries by air contaminations. The vast impacts of the air pollution made measurement and studies inevitable where in the all metropolitan cities, there are daily and even hourly measurement facilities [2, 3], these measurements demand huge amount of expenditure [4]. With innovative approaches, the measuring site instruments can be calibrated, and also the errors can be decreased coincidentally with the expenditure reductions. Ozone as a constitutive of Troposphere layer or ground level layer, announced as bad ozone or Tropospheric ozone [5] has been scrutinized in this article. Another type of ozone exists in Stratosphere layer which is named as ozone layer that prevents the entrance of harmful sun's rays. In this paper the Tropospheric ozone aimed to analysis and predict. Accumulation of large amounts of ozone in the low attitudes can cause deep pulmonary and respiratory difficulties such as asthma, lung cancer and chronic coughs and death [6-8].

* Corresponding author: vahid.mehdipour1992@gmail.com

➤ This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights.

There are numerous studies about the health effects of ground level ozone where in a comprehend study about 95 cities of United states, results showed that the all cities suffer from high level ozone volume and one-third decrease of ozone in these cities prevent 4000 yearly deaths in the country[9]. Another similar research about European countries depicted that 22000 individuals lose their life because of direct ozone poisoning[10]. These statistics illustrate emergency of studies about ozone and pragmatic suppositions presentations. Hence, in this paper the ground level of volume of ozone in Tehran, Iran, predicted and modelled with soft computing methods with self-education ability. These statistics illustrate emergency of studies about ozone and pragmatic suppositions presentations. Hence, in this paper the ground level of volume of ozone in Tehran, Iran, predicted and modelled with soft computing methods with self-educable ability.

Lu and Wang [11], tried to predict the clean days of Hong Kong that they evaluated the used methods as reasonable, also in one more paper from these researchers [12], they used multilayer perceptron (MLP) and support vector machine (SVM) to predict the ozone concentration and they concluded SVM shows lower errors and is more reliable in comparison with MLP. Schlink et al, scrutinized the 15 different approaches for ozone prediction which they used 8 predictors. Datasets collected from Germany, Italy, England and Czech Republic. In the last, the artificial neural network depicted the best result among the others [13]. Kisi et al, used three methods to predict the sulphur dioxide in 3 zones of Delhi, India and the least square support vector machine introduced as the superior method. Comparison of the results of these methods, with previous studies, depicted that all three methods had better outcomes [14]. As one of the GEP uses in the environmental engineering, Mehdipour et al, used gene expression programming to predict the dissolved oxygen of Eymir lake, Ankara, Turkey, and the best model of the method gave reliable answers [15]. Noori et al, applied ANN and ANFIS methods to predict the daily concentration of carbon monoxide in the Tehran, Iran [16]. They used gamma test and forward selection (FS) to collect the datasets and finally the results depicted that using pre-run selections of datasets causes error dwindling and also time saving. The hybrid approach of FS-ANN evolved a model with the least error and the highest correlation coefficient. Support vector machine and gene expression programming and all other soft computing approaches are useful methods in several environmental agendas and numerous paper published in this area [17-21]. All cited papers are evidences of the importance of numerical and soft computational methods application to environmental problems, and specifically of the air pollutions. Numerical methods should develop in the future to replace with the present expensive and massive facilities

2. Methodology

2.1. Gene Expression Programming (GEP)

For the first time, Ferreira evolved the gene expression programming as a developed method with the base of genetic algorithms (GA) and has been used extensively in the latest studies [22]. GEP approach has a few advantages: (a) simplicity of chromosomes for manipulation (b) The expression tree of GEP is exclusive for related chromosomes (c) gives explicit formulation for predicting the parameter (d) there is no constraints for the complexity of chromosomes structures [23]. The main difference between GA and GEP is in the existence of individuals where in GA, individuals have linear threads and static lengths which are called chromosomes, but in GPA the threads can be nonlinear with variable lengths. First step of this program to solve any problem is to produce initial population, which happens with random births of chromosomes and in the next, the chromosomes transform to expression trees (ETs) that is assessed by evaluation criteria to depict the solubility of produced ETs. If the results satisfy the evaluation criteria, population producing stops, and if the results are not sufficient, system reproduces with modification to make new generation with better quality and this process happen till to find demanded criteria [24, 25]. Figure 1. [22] shows the GEP process. The procedure to make a model for O_3 prediction (as dependent) using 8 different input combinations (as independents), is: 1-opting a fitness function 2-create chromosomes 3- chromosome structure 4-linking functions 5-genetic functions [26]. In the present study 70% of datasets defined as training data and residual datasets have been used as testing data. GEP uses training values to provide models and uses the test datasets to assess the compatibility of structured models. Over-training or over-fitting is one of the prevalent problems of the GEPs which is under appropriate control in this paper. As the result of model for testing data is better than the results for training data, the system suffers over-training which is unacceptable. At every step of the programming for O_3 prediction, results have been controlled discreetly due to avoid the cited defection. Table 1 illustrates the structure of evaluating GEP for ozone prediction. In this programming, the number of chromosomes was constant for all steps and assumed 30 while, the head size presumed as 6, 7 and 8 randomly among the programs. Gene numbers also varied randomly and they assumed as 4 and 5. + opted as linking function for generated ETs as it ostensible in equations 6 to 9. Initial production of random genes in the first step of expression programming conducted by 4 main functions (+ / - \times) in this programming. However these functions caused lengthiness in the final equation of Ozone forecasting, it made the function more tangibile.

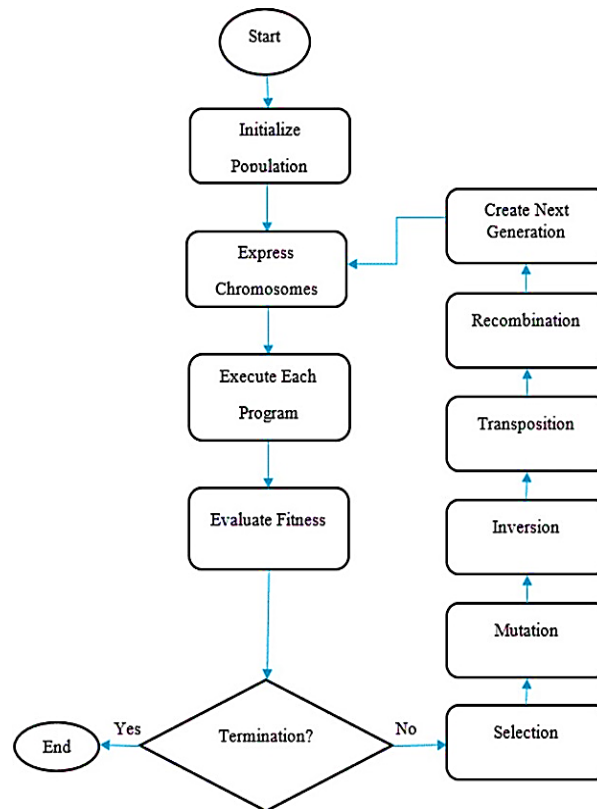


Figure 1. GEP process flowchart

Table 1. GEP architecture for ozone prediction

Number of chromosomes	30
Head size	6,7,8
Number of genes	4,5
Linking function	+
Fitness function	RMSE
Mutation	0.00206
Inversion	0.00546
Training samples	70%
Testing samples	30%
Generation functions	+ / - ×

2.2. Support Vector Machine (SVM)

The first version of vector machines invented by Cortes & Vapnik [27], in which the machine was able to classify and analysis the input datasets. Data classification is the common duty of methods which learns from training data and test their selves by testing datasets. When in a problem solution, datasets are collected from one or several classes, the SVM classifies the datasets for an easier solution. If the datasets occur on a 2-D, same as the Figure 2. [29], boarder lines can separate each class and it's obvious that innumerable lines can separate classes, while the best line with the maximum equal distance from different classes is desirable [28]. Equations 1 and 2 relating to the border lines:

$$\vec{w} \cdot \vec{x} - b = 1 \quad (1)$$

$$\vec{w} \cdot \vec{x} - b = -1 \quad (2)$$

Where geometrically, the super line function is $2/\|\mathbf{w}\|$. As the linear classifiers are not able to separate the datasets, Kernel function transforms the all datasets into N dimension space to categorize. Several types of Kernel function are available such as: linear, polynomial and radial basis function (RBF) and using the right kernel is critical issue of problem solution [29]. Radial basis function because its capability on previous environmental researches with many independent parameters [30] and due to the results of Table 2, has been chosen as a kernel function in this paper. The gamma and \sin_2 amounts for this kernel has been analyzed to distinguish optimum amount and gamma= 1 and $\sin_2= 0.2$ used for final programming steps. 70% of collected datasets defined as trainer and 15% are designated for testing

and 15% for validating. Prediction of O_3 in the Tehran metropolitan area, using the carbon monoxide, nitrogen dioxide, Sulphur dioxide, particulate matters ($PM_{2.5}$ and PM_{10}), air temperature, relative humidity and wind speed makes extreme complexity for vector machine, nonetheless, the machine could evolve 286 classifier lines where 97 of these lines, were boarder lines. Categorizing the 8 predictor in a multifarious dimensional space occurs in this research, while plethora number for each parameter, causes more complexity during the program running, but then SVM is a capable method for handling the large amounts of datasets.

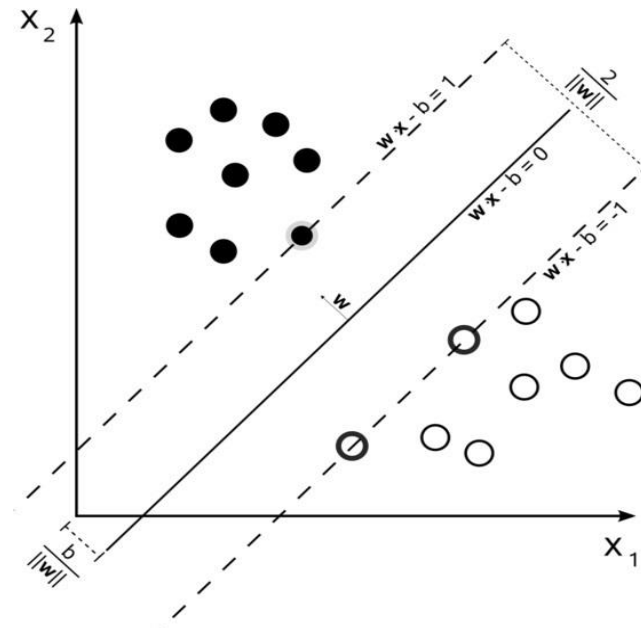


Figure 2. Boarder lines and super line for data classification

2.3. Evaluation and Comparison Criteria

In the present article the two evolved methods are compared to each other by root mean square error (RMSE) and correlation coefficient (R). Each of which approaches, evaluated by these assessments, where R and RMSE put in use to discover the similarity between observed and simulated data. Equation 3 for correlation coefficient and Equation 4 for root mean square error, illustrates that lower RMSE (not lower than 0) and higher R (not more than 1) represents more accuracy of simulation, where Y_m and Y_p are observed and predicted O_3 , respectively and also \bar{y}_m and \bar{y}_p are the average of all observed and predicted values of ozone. N in these equations represents the total input data number of ground level ozone and it is equal to 730. The reasons why just these two criteria used for this research is that, these two evaluation criteria have been used in several soft computing studies and showed reliable performances [31, 32] and in a comparison with other criteria they have tangible supremacies [33]. Nonetheless, there are numerous criteria for statistical assessment but using most powerful and uncomplicated methods seem to be more acceptable.

$$R = \frac{\sum_{i=1}^N (Y_m - \bar{y}_m) \times (Y_p - \bar{y}_p)}{\sqrt{\sum_{i=1}^N (Y_m - \bar{y}_m)^2} \times \sqrt{\sum_{i=1}^N (Y_p - \bar{y}_p)^2}} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_m - Y_p)^2}{N}} \quad (4)$$

3. Study Area

Tehran in heart of Middle East, in 29th biggest city in the world with an economy based on trade and produce by manifold type of factories of numerous industries. Approximately, all factories of Tehran and vicinity cities aren't equipped by air pollutants filters [34, 35]. Feeble public transportation of Tehran, which contains expensive taxis, crowded metro-buses and limited metro lines, are causes of inhabitants privilege to use their personal cars which is the trigger point of traffic jams [36]. Tehran has 1274 km² area which is divided into 22 districts and centre of this city has 51° E longitude and 35° N latitude where the attitude of Tehran varies from 1830 m to 900 m above seas with a slope from north to south [37]. The Figure 3. depicts Iran country and Tehran County with the borders of districts.

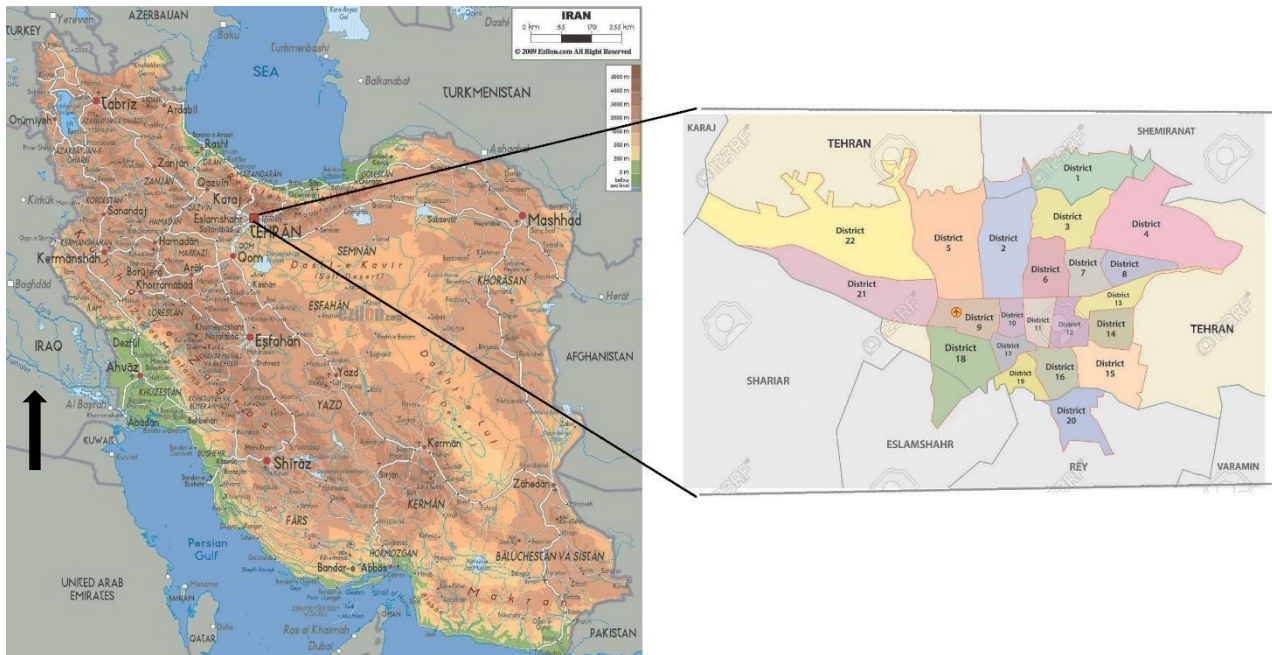


Figure 3. Iran & Tehran map including Tehran districts

3.1. Datasets

During March 2014 to March 2016 datasets collected from 22 stations of photochemical precursors measuring. Ground level ozone (O_3), Sulphur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), particulate matter with size of 2.5 micrometre ($PM_{2.5}$) and particulate matter with size of 10 micrometre (PM_{10}) collected from each station in each district. The maximum daily amount of every precursor in all 22 stations selected as that precursor's value of that day, i.e. for CO which is monitored by 22 stations in the first day, the maximum value of monitored CO designated for the first day. Meteorological parameters are also directly impact on every photochemical precursor and thereby, ambient air temperature, humidity and wind speed opted for ground level ozone prediction. The daily average value of each monitored meteorological parameters, selected. On the other hand, 730 days monitored for 9 parameters while the 8 parameters of collected datasets used to prognosticate the O_3 concentration of Tehran, Iran.

3.2. Data Normalization

Monitored datasets have different units, i.e. wind speed defining by km/h while the humidity expressing by % in this difference make it impossible to use datasets with together without pre-processes, hence in this paper the all collected datasets converted to [0-1] by applying the Equation 5. Being in the same scale and normalization of parameters provides the datasets for further processes [38]. X_{min} and X_{max} are the minimum and maximum amount of each parameters and X_i is value that normalization occurs on it.

$$X = 0.1 + 0.9 \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad (5)$$

4. Results and Discussions

In this paper, aimed to forecast the O_3 concentration of Tehran, Iran using the photochemical precursors and meteorological parameters and in this section of paper, all variables have been calculated. The RMSE and R values of gene expression programming and support vector machine, assessed and the superior method is introduced.

4.1. Results of GEP

Multifarious running happened to give the best possible results and in the next step the overtraining is controlled by comparing the test and train results. The results, showed that, the RMSE and R for training data are, 0.08835 and 0.8066 while these values for testing data are, 0.08836 and 0.7938 and higher correlation coefficient by lower root mean square error of training data depict that the system isn't defected by overtraining. Testing data sets. Figure 4 illustrates the linear regression between predicted and observed ozone and R^2 relate the ETSS value is 0.6301 and subsequently, the correlation coefficient is $R = 0.7938$. Figure 5. shows the in testing data (218 days of total monitoring) how the acquired model follows the observed O_3 . As it mentioned earlier, GEP provides a specific formulation for O_3 prediction by independent parameters. This approach, expressed four expression trees (ETS) to

solve this problem and because of opted linking function between ETSs, +, the expression trees should be plus one by one to achieve complete formulation. Equations 6, 7, 8 and 9 relate the ETSs.

$$Y_1 = (((d(PM_{2.5}) * d(PM_{2.5})) + d(NO_2)) * (d(NO_2) * d(PM_{2.5}))) * ((d(TEMP) + d(SO_2)) / (-6.0094 - d(WIND))) \quad (6)$$

$$Y_2 = Y_1 + ((d(PM_{10}) * ((d(TEMP) + d(HUM)) * d(WIND))) + (d(TEMP) * d(NO_2))) \quad (7)$$

$$Y_3 = Y_2 + (((1.3724 / -9.1990) / (d(WIND) + 9.1990)) * (1.3724 - (-9.1990 - d(WIND)))) \quad (8)$$

$$Y_4 = Y_3 + (d(CO) * d(NO_2) / (((d(CO) / d(SO_2)) * d(TEMP)) + -4.8973)) \quad (9)$$

Y_4 is the complete formulation for the O_3 calculation.

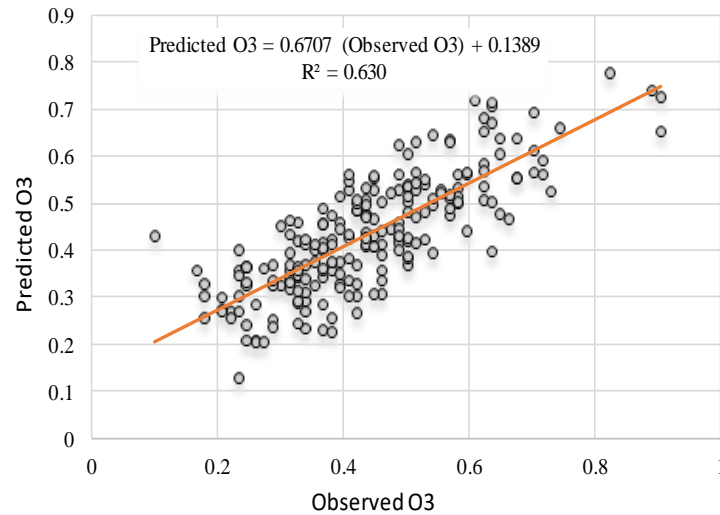


Figure 4. Linear regression for observed and predicted O_3 by GEP

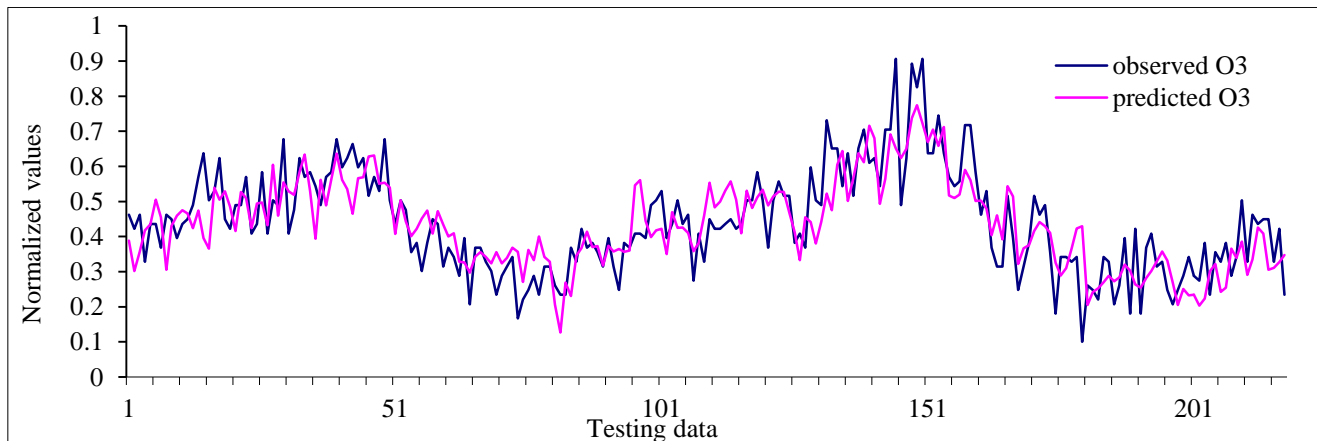


Figure 5. Predicted and observed O_3 by GEP

4.2. Results of SVM

At end of running process, same as the GEP, the overtraining needed to be controlled. Results for training datasets, are RMSE= 0.0771 and R= 0.8460 and for the testing datasets are, 0.0774 and 0.8459 respectively, which are the reasons to show that system didn't encounter with over fitting. Figure 6, demonstrates the linear regression of predicted and observe O_3 . Furthermore, other prevalent kernel functions applied to this prediction to evaluate their performance for this issue. However, the kernel parameter (sin2) and gamma varied by kernel types in the Figure 5. As radial basis function showed better results, more tolerances of gamma and sin2 analyzed for RBF kernel and finally, gamma = 1 and sin2= 0.2 gave the best answer. One remarkable point in these analyses is that, none of these kernel functions with various gamma and sin2, could not acquire results better than training datasets results.

Table 2. Kernel functions performances

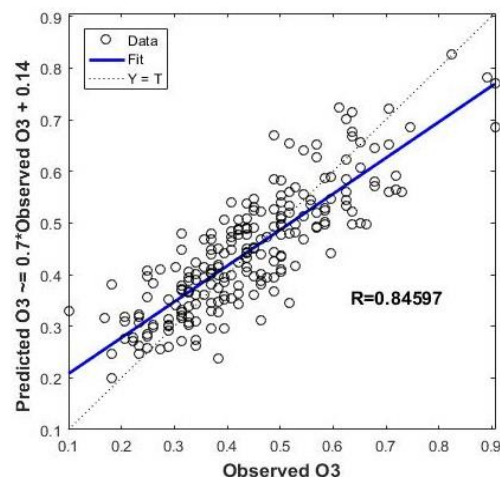
Kernel type	Gamma	Sin2	MSE	R
Linear	1	0.2	0.0076	0.7994
	0.1	1	0.0076	0.7993
	10	0.2	0.0076	0.7994
	100	2	0.0074	0.7994
polynomial	10	0.2	0.0068	0.8224
	0.1	0.002	0.0068	0.8440
	1	2	0.0068	0.8443
	0.5	0.002	0.0062	0.8403
Multi-layer perceptron	10	0.2	0.0088	0.7774
	1	0.2	0.0084	0.7775
	10	2	0.0086	0.7744
	0.5	0.002	0.0088	0.7774
Radial basis	10	0.2	0.0060	0.8457
	10	0.5	0.0060	0.8456
	10	2	0.0060	0.8451
	1	0.2	0.0060	0.8459
	1	0.02	0.0060	0.8458
	1	0.002	0.0060	0.8456
	100	0.2	0.0060	0.8456
	100	2	0.0060	0.8455
	0.01	0.2	0.0060	0.8458
	0.1	0.2	0.0060	0.8456

Considering Table 2, results that the distribution of the datasets (predictors and target parameter) are in classifiable shape by used kernel tricks with respect to the MSE and R values, nonetheless, the accuracy and errors are the dominant factors to choose a pragmatic kernel trick.

4.3. O₃ Sensitivity Analysis

4.3.1. Sensitivity Analysis for Photochemical Precursors and Meteorological Parameters

Due to measuring the sensitivity of ground level ozone to photochemical precursors and meteorological parameters, all 8 independent variables set in different groups to use as input combinations. In the first group, named A, just photochemical precursors used to predict the O₃ and in the second group (B), only meteorological parameters put in use to prognosticate the target parameter. Figure 7. and Figure 8. showing the regression line of models with photochemical precursors and meteorological parameters. The correlation coefficient is derivable from figures, hence, R of group A. is: 0.6944 and for group B: 0.7827 while the RMSE value for group A is: 0.1044 and for group B is: 0.0900 Figure 9. and Figure 10. also representing that how modelled dataset follow the observed datasets. Comparing the Figure 6. with Figure 7. and Figure 8. shows that meteorological parameters, (ambient air temperature, relative humidity and wind speed) are more likely to be effective than the photochemical precursors. Both groups could not satisfy the R value, which is already acquired by all variables.

Figure 6. Linear regression for O₃ prediction by all independents

4.3.2. Analysis to Each Parameter

In the next step, parameters added one by one to the system to analysis each added variable impact. For this aim, two types of central parameters have been chosen: (i) CO and NO₂ (ii) PM_{2.5} and humidity. And then, other 5 parameters added one by one to cores. Group C assumed for input combinations for core (i) and group D designated for core (ii). This process shows a clear and understandable image of any factor effectiveness. SVM, since exhibited more capability, and superior results in comparison with GEP, used to manipulate the analysis processes. Table 3. shows the arrangement of input combinations. Table 2, resulted that in this case, the best answer, happens when the $\sin 2$ is 0.2 and the gamma is equal to 1 while radial basis function (RBF) opted as kernel trick. Henceforth, for all analysed input combination into SVM, the mentioned circumstances applied. In the following, Table 4. displays that, results which acquired by running each input combinations, and correspondingly the impact of every parameter is understandable.

As there is no accurate formula for ozone prediction or calculating, every selection of parameters as cores and groups as input combinations will not be precise enough. CO and NO₂ as two main compounds of the ozone constitution [39, 40] and PM_{2.5} and humidity as two dominant parameters on solar radiation alters, selected as cores of groups. Solar radiation stimulates the tropospheric ozone construction processes and therefore, the particulate matters density blocks the solar radiation receiving into the earth's surface.

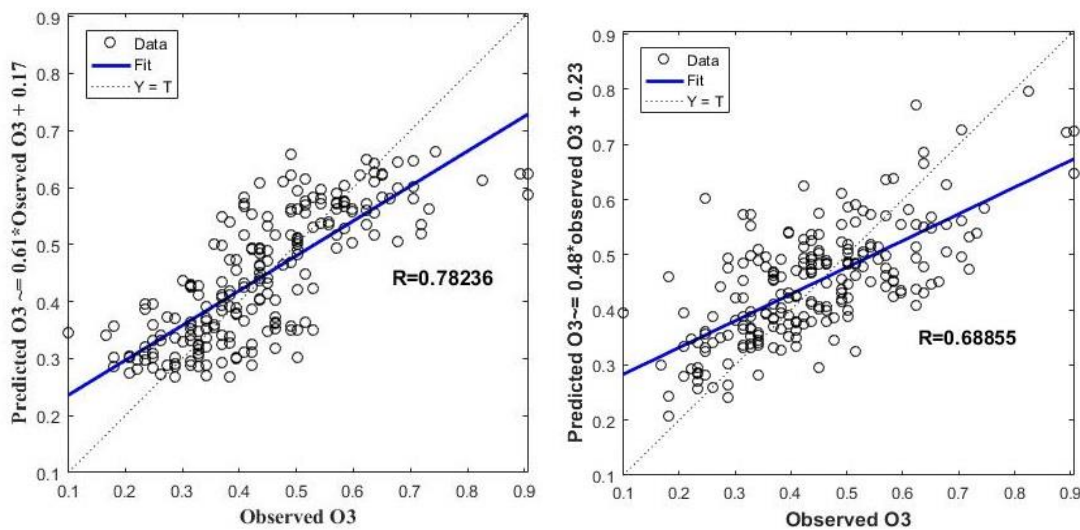


Figure 7. Linear regression for O₃ prediction considering photochemical precursors

Figure 1. Linear regression for O₃ prediction considering meteorological parameters

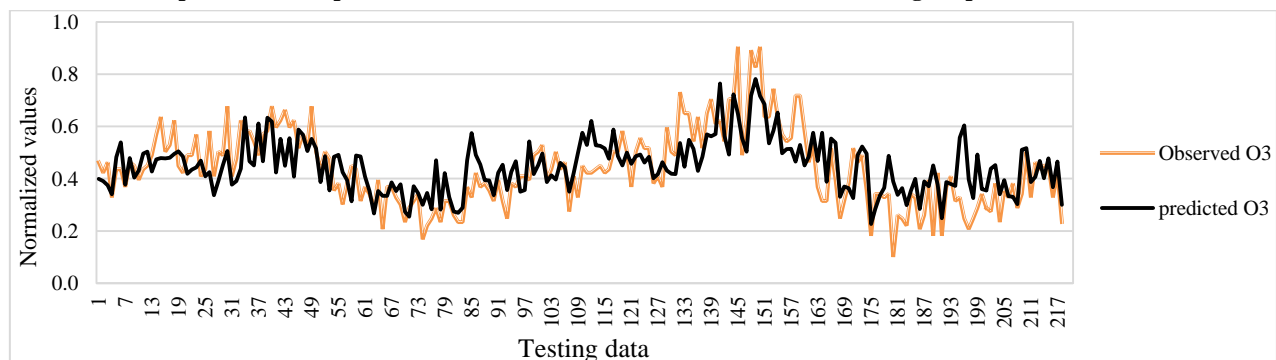


Figure 2. Observed and predicted O₃ considering photochemical precursors as input combinations

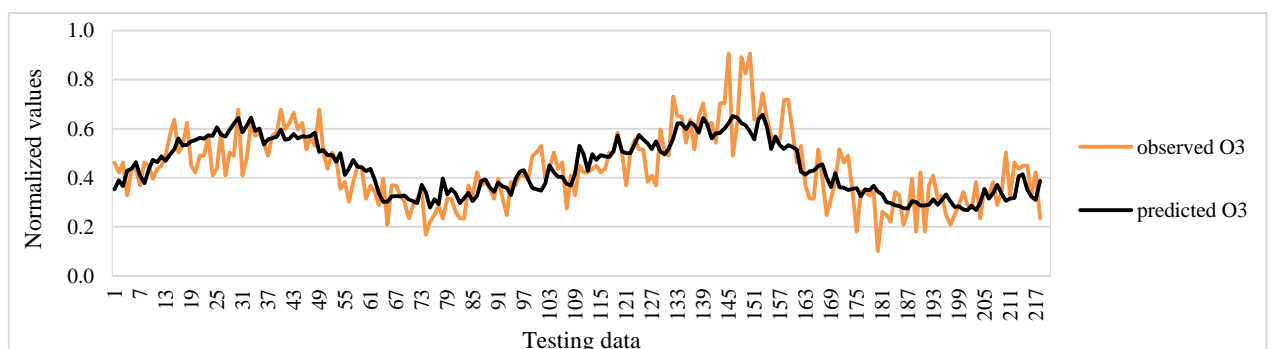


Figure 3. Observed and predicted O₃ considering meteorological parameters as input combinations

Table 3. Input combinations and their component

Model name	Input parameters
Input combination C0	CO, NO ₂
Input combination C1	CO, NO ₂ , PM _{2.5}
Input combination C2	CO, NO ₂ , PM _{2.5} , PM ₁₀
Input combination C3	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂
Input combination C4	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂ , TEMP
Input combination C5	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂ , temp, hum
Input combination C6	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂ , temp, hum, wind
Input combination D0	PM _{2.5} , hum
Input combination D1	PM _{2.5} , hum, PM ₁₀
Input combination D2	PM _{2.5} , hum, PM ₁₀ , SO ₂
Input combination D3	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp
Input combination D4	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp, CO
Input combination D5	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp, CO, NO ₂
Input combination D6	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp, CO, NO ₂ , wind

Table 4. O₃ sensitivity analysis's results for all parameters

Input combination named	Components	RMSE	R
Input combination C0	CO, NO ₂	0.1700	0.1688
Input combination C1	CO, NO ₂ , PM _{2.5}	0.1260	0.4902
Input combination C2	CO, NO ₂ , PM _{2.5} , PM ₁₀	0.1104	0.6480
Input combination C3	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂	0.1048	0.6884
Input combination C4	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂ , TEMP	0.0793	0.8359
Input combination C5	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂ , temp, hum	0.0793	0.8360
Input combination C6	CO, NO ₂ , PM _{2.5} , PM ₁₀ , SO ₂ , temp, hum, wind	0.0774	0.8443
Input combination D0	PM _{2.5} , hum	0.1288	0.4575
Input combination D1	PM _{2.5} , hum, PM ₁₀	0.1208	0.5541
Input combination D2	PM _{2.5} , hum, PM ₁₀ , SO ₂	0.1104	0.6540
Input combination D3	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp	0.0830	0.8197
Input combination D4	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp, CO	0.0824	0.8240
Input combination D5	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp, CO, NO ₂	0.0800	0.8342
Input combination D6	PM _{2.5} , hum, PM ₁₀ , SO ₂ , temp, CO, NO ₂ , wind	0.0774	0.8459

5. Conclusion

In the present paper, the O₃ value predicted by GEP and SVM methods as two and potent novel approaches while using R and RMSE for comparing the acquired results from modelling. Datasets divided into two section, one for training the methods and other for testing their capability. Moreover, for analyzing the O₃ sensitivity against other input parameters, datasets once classified as two classes, meteorological and photochemical as group A and B, and once, two cores selected for as input combinations C and D and in the next steps other parameters added one by one. Also, few prevalent kernel tricks applied for SVM to figure the best kernel out for studied datasets and RBF kernel revealed as the best among the tested kernels. With respect to the R= 0.7938 and RMSE= 0.0883 for GEP and R= 0.8459 and 0.0774 for SVM, support vector machine introduced as the better method in comparison with gene expression programming, however, the results of GEP are reliable and acceptable. On the other hand, in this study, classifying the datasets had superior outcomes in comparison with producing the chromosome and genes to solve the O₃ prediction problems. Furthermore, sensitivity analysis showed that PM_{2.5}, PM₁₀ are the most effective parameters and also CO, NO₂ and ambient air temperature are effective factors for ozone prognosticating while the humidity and SO₂ had less impact on target parameter.

However, the wind speed assumed to be one effective feature due to its effects on particulate matter movement, had a trifle influence on the ozone. This would be nebulous result for other case studies while, in the Tehran, the low value of average wind speed had not pragmatic effects on ground level ozone tolerances. Low wind speed is the result of sky scrapers constructions around the city and morphology of the case. As particulate matters halt the sun rise, which is

essential for O_3 producing by CO and NO_2 , the attained results are acceptable and authors notify that the sun radiation would be very important to be analyzed in the next studies.

6. Acknowledgements

Authors show their special respects to Dr. Jalal Shiri, assistant professor of Tabriz university for his undeniable scientific contributions and also regards toward Ph.D. candidate at Amirkabir university of technology, dear Farzin Homayounfar, and many thanks for dear professor Ian Waite from USGS, Dr. Albert Saldin, Dr. Mitchel Sami, Dr. Ramin Giahi from Iowa state university and other precious scientific collaborators who helped us during this study.

5. References

- [1] WHO, "Health Aspects of Air Pollution with Particulate Matter , Ozone and Nitrogen Dioxide," Rep. a WHO Work. Gr. Bonn, Ger. 13–15 January 2003, no. January, p. 98, 2003.
- [2] M. Bešter-Rogač and D. Habe, "Modern advances in electrical conductivity measurements of solutions," *Acta Chim. Slov.*, vol. 53, no. 3, pp. 391–395, 2006.
- [3] Z. Z. Liu, J. Y. Wang, and H. T. Wang, "Ozone Measurements at Some Work Places for Environment Protection and Career Health Care," *Ozone Sci. Eng.*, vol. 0, no. 0, pp. 1–6, 2017.
- [4] X. Basagana et al., "Effect of the number of measurement sites on land use regression models in estimating local air pollution," *Atmos. Environ.*, vol. 54, no. 1, pp. 634–642, 2012.
- [5] J. Zheng, L. Zhong, T. Wang, P. K. K. Louie, and Z. Li, "Ground-level ozone in the Pearl River Delta region: Analysis of data from a recently established regional air quality monitoring network," *Atmos. Environ.*, vol. 44, no. 6, pp. 814–823, 2010.
- [6] Ö. Çapraz, B. Efe, and A. Deniz, "Study on the association between air pollution and mortality in İstanbul, 2007–2012," *Atmos. Pollut. Res.*, vol. 7, no. 1, pp. 147–154, 2016.
- [7] U. Schlink et al., "Statistical models to assess the health effects and to forecast ground-level ozone," *Environ. Model. Softw.*, vol. 21, no. 4, pp. 547–558, 2006.
- [8] M. L. Bell, A. McDermott, S. L. Zeger, and J. M. Samet, "in 95 US Urban Communities , 1987-2000," *Forestry*, vol. 292, no. 19, pp. 2372–2378, 2004.
- [9] R. Gauld et al., "The World Health report 2008 - Primary healthcare: How wide is the gap between its agenda and implementation in 12 high-income health systems?," 2008.
- [10] W.-Z. Lu and D. Wang, "Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme," *Sci. Total Environ.*, vol. 395, no. 2–3, pp. 109–116, 2008.
- [11] W. Lu and D. Wang, "Learning machines : Rationale and application in ground-level ozone prediction," *Appl. Soft Comput. J.*, vol. 24, pp. 135–141, 2014.
- [12] U. Schlink et al., "A rigorous inter-comparison of ground-level ozone predictions," *Atmos. Environ.*, vol. 37, no. 23, pp. 3237–3253, 2003.
- [13] O. Kisi, K. S. Parmar, K. Soni, and V. Demir, "Modeling of air pollutants using least square support vector regression, multivariate adaptive regression spline, and M5 model tree models," *Air Qual. Atmos. {&} Heal.*, Apr. 2017.
- [14] V. Mehdipour, M. Memarianfard, and F. Homayounfar, "Application of Gene Expression Programming to water dissolved oxygen concentration prediction," *Int. J. Hum. Cap. Urban Manag.*, vol. 2, no. 1, pp. 1–10, 2017.
- [15] R. Noori, G. Hoshyaripour, K. Ashrafi, and B. N. Araabi, "Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration," *Atmos. Environ.*, vol. 44, no. 4, pp. 476–482, 2010.
- [16] J. Shiri, P. Marti, and V. P. Singh, "Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning," *Hydrol. Process.*, vol. 28, no. 3, pp. 1215–1225, 2014.
- [17] J. Shiri, "Evaluation of FAO56-PM, empirical, semi-empirical and gene expression programming approaches for estimating daily reference evapotranspiration in hyper-arid regions of Iran," *Agric. Water Manag.*, vol. 188, pp. 101–114, 2017.
- [18] M. Memarianfard, A. Hatami, and M. Memarianfard, "Artificial neural network forecast application for fine particulate matter concentration using meteorological data," *Glob. J. Environ. Sci. Manag.*, vol. 3, no. 3, pp. 333–340, 2017.
- [19] V. Mehdipour, "Temporal modeling of tropospheric ozone and analysis of its relationship with photochemical precursors considering meteorological parameters," K. N. Toosi University of Technology, 2017.
- [20] C. Ferreira, "Gene Expression Programming : A New Adaptive Algorithm for Solving Problems," *Complex Syst.*, vol. 13, no. 2, pp. 87–129, 2001.
- [21] A. Guven and Ö. Kisi, "Daily pan evaporation modeling using linear genetic programming technique," *Irrig. Sci.*, vol. 29, no. 2, pp. 135–145, Mar. 2011.
- [22] C. Ferreira, "Designing Neural Networks Using Gene Expression Programming," *Appl. Soft Comput. Technol. Chall. Complex.*, pp. 517–536, 2006.
- [23] C. Ferreira, "Automatically defined functions in gene expression programming," *Stud. Comput. Intell.*, vol. 13, pp. 21–56, 2006.

- [24] S. Kim, J. Shiri, V. P. Singh, O. Kisi, and G. Landaras, "Predicting daily pan evaporation by soft computing models with limited climatic data," *Hydrol. Sci. J.*, vol. 60, no. 6, pp. 1120–1136, 2015.
- [25] C. Cortes and V. Vapnik, "Support Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [26] S. Canu and A. Rakotomamonjy, "Ozone peak and pollution forecasting using support vectors," 2001.
- [27] M. G. Genton, "Classes of Kernels for Machine Learning: A Statistics Perspective," *J. Mach. Learn. Res.*, vol. 2, no. 4, pp. 299–312, 2001.
- [28] S. Theodoridis, *Pattern Recognition*, 4th editio. Burlington: Academic Press, 2008.
- [29] K. Roushangar and F. Homayounfar, "Prediction of Flow Friction Coefficient using GEP and ANN Methods," *Int. J. Artif. Intell. Mechatronics*, vol. 4, no. 2, pp. 65–68, 2015.
- [30] Y. Feng, W. Zhang, D. Sun, and L. Zhang, "Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification," *Atmos. Environ.*, vol. 45, no. 11, pp. 1979–1985, 2011.
- [31] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014.
- [32] A. R. Hosseinpour, M. H. Forouzanfar, M. Yunesian, F. Asghari, K. H. Naieni, and D. Farhood, "Air pollution and hospitalization due to angina pectoris in Tehran, Iran: A time-series study," *Environ. Res.*, vol. 99, no. 1, pp. 126–131, 2005.
- [33] H. Vafa-arani, S. Jahani, H. Dashti, J. Heydari, and S. Moazen, "A system dynamics modeling for urban air pollution : A case study of Tehran , Iran," *Transp. Res. PART D*, vol. 31, pp. 21–36, 2014.
- [34] S. Seyedabrishami and A. Mamdoohi, "Impact of Carpooling on Fuel Saving in Urban Transportation : Case Study of Tehran," vol. 54, pp. 323–331, 2012.
- [35] N. Bagha, M. Arian, M. Ghorashi, M. Pourkermani, R. El, and A. Solgi, "Geomorphology Evaluation of relative tectonic activity in the Tehran basin , central Alborz , northern Iran," *Geomorphology*, vol. 29, no. 2, pp. 135–145, 2014.
- [36] V. Simeonov, J. Einax, I. Stanimirova, and J. Kraft, "Environmetric modeling and interpretation of river water monitoring data," *Anal. Bioanal. Chem.*, vol. 374, no. 5, pp. 898–905, 2002.
- [37] B. J. Finlayson-Pitts and J. N. Pitts, "Atmospheric chemistry of tropospheric ozone formation: Scientific and regulatory implications," *J. Air Waste Manag. Assoc.*, vol. 43, no. 8, pp. 1091–1100, 1993.
- [38] Y. Wang, D. J. Jacob, and J. a. Logan, "Global simulation of tropospheric O₃ -NO_x -hydrocarbon chemistry: 3. Origin of tropospheric ozone and effects of nonmethane hydrocarbons," *J. Geophys. Res. Atmos.*, vol. 103, no. D9, pp. 10757–10767, 1998.
- [39] P. Dey and A. K. Das, "A utilization of GEP (gene expression programming) metamodel and PSO (particle swarm optimization) tool to predict and optimize the forced convection around a cylinder," *Energy*, vol. 95, no. 1, pp. 447–458, 2016.
- [40] O. Ivanciuc, "Applications of Support Vector Machines in Chemistry," in *Reviews in Computational Chemistry*, vol. 23, 2007, pp. 291–400.