Integration of GIS and Artificial Neural Network for prediction of Ozone Concentration in Semi-rural areas of Rawalpindi and Islamabad

Sheikh Saeed Ahmad[1]*, Neelam Aziz[1], Madiha Ejaz[2] and Muhammad Tahir Ali Shah[3]

[1]Department of Environmental Sciences, Fatima Jinnah Women University, Rawalpindi, Pakistan
[2]Department of Computer Sciences, Fatima Jinnah Women University, Rawalpindi, Pakistan
[3]Department of Agriculture, Allama Iqbal Open University, Islamabad, Pakistan

Abstract:

Ozone is one of the most effective pollutants in lower atmosphere. Concentration of ozone in atmosphere reveals its impact on plants, human and on other organic materials. Many techniques had been used in past to calculate the concentration of ozone with the help of other environmental factors like wind, humidity, rainfall, temperature and etc. Prediction models like artificial neural network have gained much reputation in calculating accurate results with learning data. This paper shows a study of integration of predicted ozone concentration from neural network and GIS. The study initiated with data collection from the study area. The collected data is then fed to neural network as training data to get the concentrations of ozone with input variables temperature, humidity and rainfall. The study shows the great dependence of ozone concentration upon environmental factors. Finally the results are spatially interpolated with the help of GIS.

Keywords: Artificial Neural Network Model (ANN), Climatic Variables, Geographic Information System (GIS), Inverse Distance Weighted (IDW), Rawalpindi and Islamabad, Pakistan, Semi-rural areas, Tropospheric ozone concentration,

1. Introduction

The presence of ozone molecule in ozone layer saves earth life from harmful damage by absorbing 97-99% of harmful radiations from sun. Whereas same ozone molecule present in lower atmosphere of earth is declared harmful air pollutant. Ozone in lower atmosphere has major impact on plants, vegetation, animals and most of human respiratory diseases. Many studies have shown a significant correlation between ozone levels and respiratory illness [1]. The emission of the pollutant is directly proportional to nature of industrial and urban activities. So relationship between different other pollutants and meteorological factors can be used to determine ozone level. Many approaches have been made for such purpose. Such approaches can be linear or non linear. Multiple linear regression, principal component regression, quantile regression, among others, are a few examples of linear models [2][3] and, on the other hand, artificial neural networks are the nonlinear models most commonly used [4][5]. One idealistic approach would be to make a deterministic model that depends on chemical processes and atmospheric movements. Other approach is Empirical Kinetic Modeling Approach EKMA that is used for designing emission control strategies [6]. Abdul-Wahab developed a statistical empirical model from measured ambient air quality data, which was based on obtaining the functional relationship between the ozone level and various independent variables by using a stepwise multiple regression procedure [7]. Multiple linear regression is used to obtain a linear input output model for a given data set. If the functional relationship is non-linear, then one or more other inputs can be transformed first into a non-linear form. There are many non-linear multivariate statistical methods which are used to approximate any non-linear relationship; but the assumption of functional dependency is a serious drawback of such methods. [8] Artificial neural network is the so far the strongest approach that is known for developing predictive model without any assumptions and are able to handle multivariable problems. Predictions from ANN were proven more accurate when contrasted with applied linear regression analysis techniques [9] and regression models [10]. Artificial neural networks learn even a complex relationship between different independent variable while the processing on neurons took place [11] and have ability to develop non-linear relationships [12]. ANNs are widely used in environmental problems and their accuracy of prediction has been proven [13, 14, 15, 16]. The multiplayer perceptron, first application of ANN was developed in Slovenia and was aimed at predicting SO2 levels. [14]. ANN can also be applied to predict ozone concentration [17, 18]. Ruiz-Suarez used neural network models for short-term forecasting of ozone in a selected spot of Mexico City [19]. Yi and Prybutok used feed-forward neural network model for predicting ozone concentrations in an urban area with four vehicle emitted pollutants variables carbon dioxide, nitric oxide, nitrogen dioxide, and oxides of nitrogen and three meteorological variables temperature, wind speed, and wind direction[20] . Abdul-Wahab and Al-Alawi used neural network for the Assessment and prediction of tropospheric ozone concentration Levels in Khaldiya region of Kuwait. Thirteen variables CH_4, NMHC, CO, CO_2, NO, NO_2, SO_2, temperature, relative humidity, wind speed, wind direction, solar radiation and dust were selected as input and Ozone as output [21]. Study

of Elkamel illustrated the successful use of a neural network to predict ozone concentrations using both meteorological and chemical data near an industrial area in Kuwait [22]. Hence artificial neural network proved best for prediction of ozone concentration. Detailed descriptions on the use of ANNs in environmental modeling can be found in Maier and Dandy [23]

1.1. What is Artificial Neural Network

An artificial neural network is a system based on the operation of biological neural networks i.e. emulation of biological neural system. The basic constituent of neural network is neutrons that are arranged in different layers known as input, hidden and output layer also known as architecture of neural network. It is a directed graph where a vertex corresponds to a neuron and an edge to a synapse.

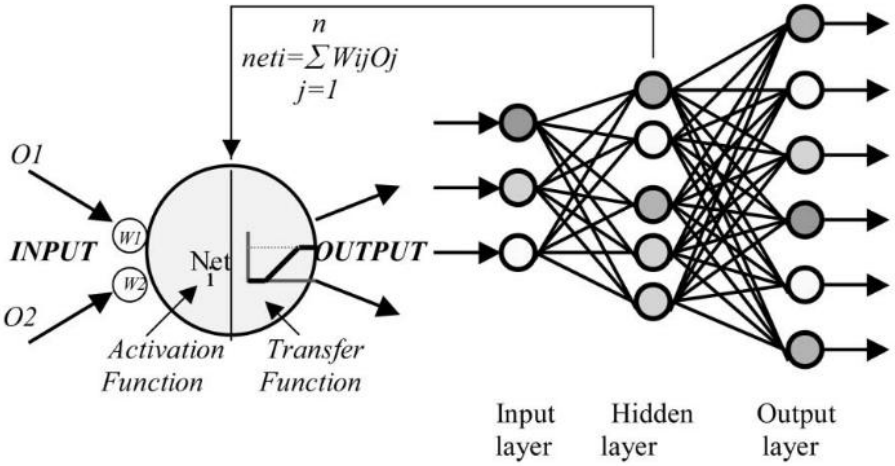

Figure 1 Feed forward Artificial Neural Network Architecture [18]

Each layer processes the data and sends the result to next layer; the output layer shows the predicted output for the inputs given in input layer. A neural network is a parallel system, capable of resolving paradigms that linear computing cannot. As a biological predecessor neural network is adaptive in nature which means during the training of neural network each parameter is changed and is used for solving the problem. During the operation of ANN the dataset is split down into three sets training, validation and testing. Training set is used to train the artificial neural network, i.e. to adjust the artificial neural network weights to maximize the artificial neural network's predictive ability and minimize its forecasting error. Validation set is used to tune the artificial neural network topology or parameters other than weights. Also used for automatic comparison of alternative artificial neural networks. Testing set is used only to test the accuracy of the predictions made by the artificial neural network on new data. Network properties contain network activation and error functions. An activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1. The properties are applied to all networks tested by the architecture search method as well as to a manually selected network.

2. Materials and methods

2.1. Study area

Pakistan is badly plagued with environmental problems due to unlimited population growth and unchecked vehicular emissions. Bad effects of ozone can spread in a wide range and up to very long distances, depending on wind speed and direction and can badly affect the health of inhabitants and agricultural entities in adjacent rural areas. Therefore, there is an urgent need for air pollution impact assessment and prediction studies on crops in the developing areas of Pakistan.

The study area of this research encompasses twin cities of Pakistan, Islamabad and Rawalpindi. Islamabad is capital of Pakistan and Rawalpindi is 15Km apart from it. Twin cities are located at 33°40′N 73°10′E with almost 4.5 million population. The study area is divided into 5 different zones with zero point reference because it is a major gateway between twin cities for many routes. Boundaries of each zone are defined within the 40 km of reference point. 40 km distance is chosen based on previous studies in this domain. Zones are as follows in figure 1.

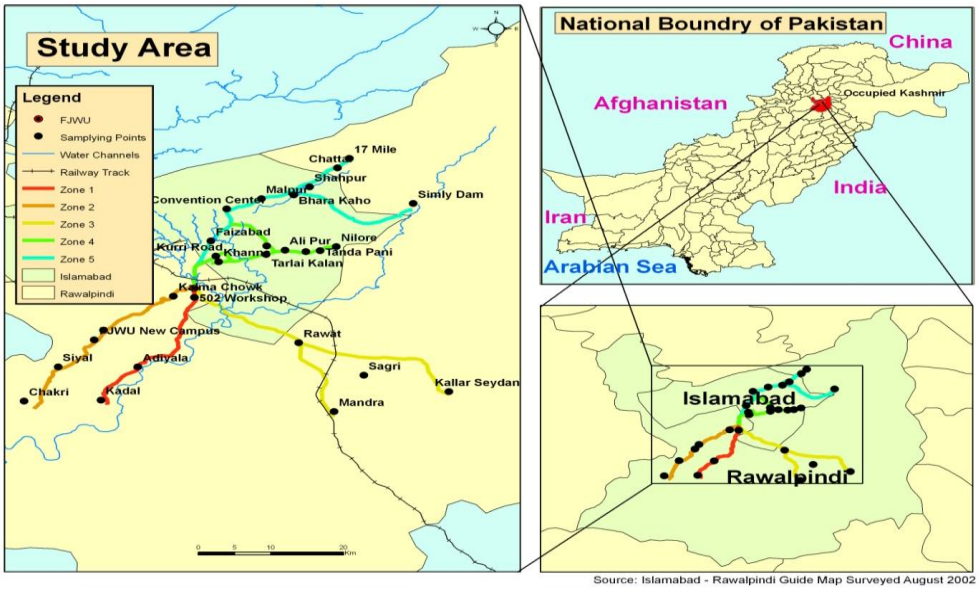

Figure 1 Base Map

Zone 1: 502 workshop, Adyala village, Kadal

Zone 2: Kalma Chowk, Pir Mehr Ali Shah Town, FJWU New Campus, Siyal and Chakri

Zone 3: Rawat, Kallar, Sagri and Mandra

Zone 4: Kuri, Khanna, Tarlahi, Alipur, Jhangi Syedan, Thanda Pani, Nelore and Chakshahzad

Zone 5: Convention Center, Malpur, Bara khau, Shahpur Village, Chattar, 17 Mile and Samli Dam

2.2. Dataset preparation

The ozone concentration was determined by The Model 400E ozone analyzer. The Model 400E ozone analyzer is a microprocessor-controlled analyzer that determines the concentration of Ozone (O_3) in a sample gas drawn through the instrument. The microprocessor uses calibration values, the UV absorption measurements made on the Sample Gas in the absorption tube along with data regarding the current temperature and pressure of the gas to calculate a final O_3 concentration, stores in one of the unit's Internal Data Acquisition System and shows result to the user via front panel.

With the determination of concentration of ozone other environmental factors were also measured and saved along with respective concentration of ozone that includes temperature (˚C), humidity(%) and rainfall(mm) in particular season. The data is collected for 18 consecutive months from november 2009 to march 2011 from 28 different sites of the research.

3. Results and discussion

3.1. Network Development

The data of database was split into three, mutually exclusive, portions of 70%, 15%, 15% respectively for training, validation and testing datasets. The design of the neural network (architecture) contains 4 hidden layers, layer 1 has 20 neurons, layer 2 contains15 neurons, layer 3 contains 10 neurons and layer 4 contains 5 neurons.

The training pattern for the neural network contains four inputs which are season_id, temperature, humidity and rainfall and the output is concentration of respective area for given inputs. Neural network is trained by using quick propagation algorithm. The Quick propagation algorithm calculates the weight change by using the quadratic function $f(x) = x^2$. Quick propagation coefficient was set to 1.75, learning rate was 0.1 and iterations were 500.

Logistic function is used for input and output activation function which is sigmoid curve and is calculated using the following formula: $F(x) = 1/ (1+e^{-x})$. Its output range is [0.1]. Sum-of-Squares is the most common error function and is used as output error function. The error is the sum of the squared differences between the actual value (target column value) and neural network output.

3.2. Model Results

The data is tested and results are produced for all sets i.e.: training set, validation set, test set. Actual vs. Output graph displays a line graph of the actual and network output values for records displayed in the Actual vs. Output Table. Horizontal axis displays the selected input column values and vertical axis displays the network output. The visualization for actual vs. output is shown in figure 2.

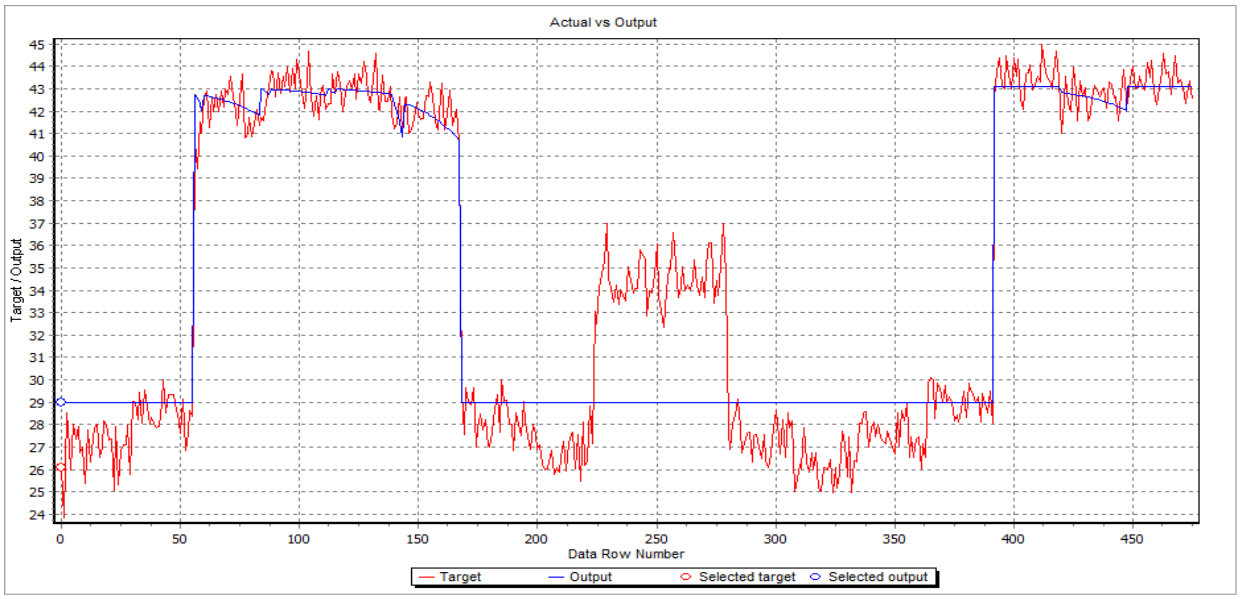

Figure 2 Actual verses Output

The environmental variables which are temperature, humidity and rainfall are very important in prediction of ozone. The information of relationship between climatic variables and ozone concentration is necessary for assembling of database.

ANN along with concentration of ozone also assesses the dependence and importance of input variables which are environmental variables in this case. Figure 3 depicts that the ozone concentration level is negatively related to the relative humidity (RH in %) and rainfall, and positively related to temperature (in C^0). Periods with higher temperature and lower humidity usually lead to the higher level of ozone in the atmosphere and vice versa. Similar changes in ozone concentration levels with respect to the climatic variables have already been described in literature [24] [25]

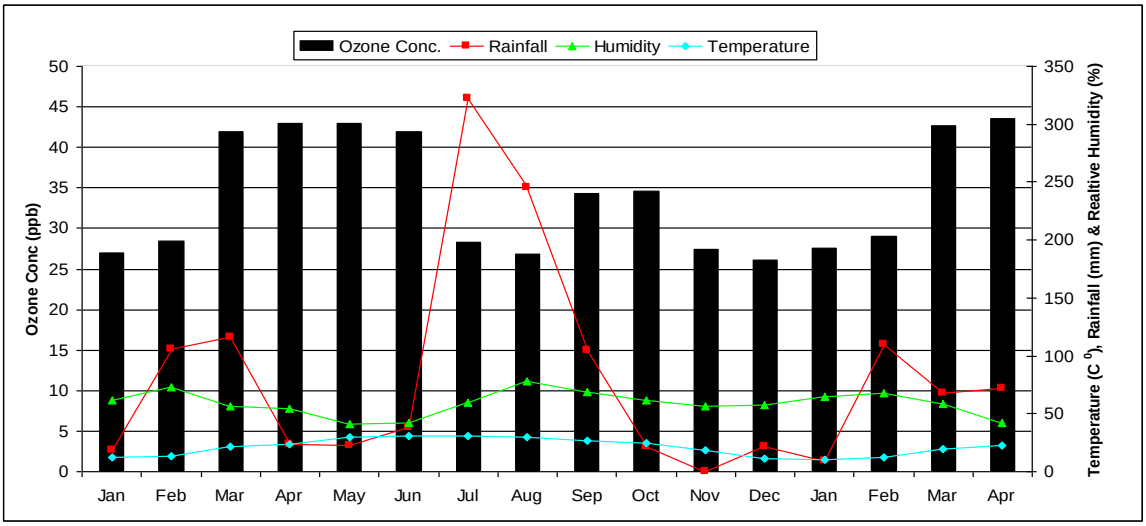

Figure 3 Relationship between O_3 concentration and climatic variables

Fig 4 shows Error dependence graph that displays the network error dependence on values of numerical input columns. The Error Dependence Graph allows analyzing that which ranges of the selected input column tend to produce bigger or smaller network errors.

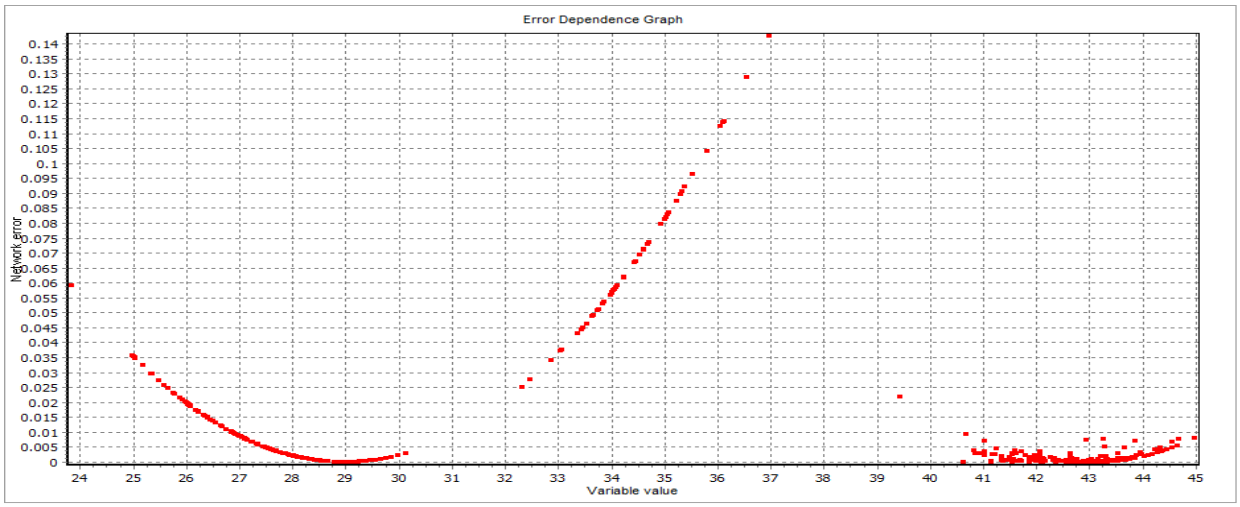

Figure 4 Error Dependence Graph

The neural network is queried after it is trained and tested. The targeted value for the neural network is concentration so querying inputs is season_id, temperature, humidity, rainfall and area_id. The results of the concentrations from neural network are then integrated with GIS.

3.3. Integration of ANN with GIS

GIS is powerful tool that is used to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts. The results of predicted concentrations of Ozone were mapped on GIS. The results of predicted concentration value of ozone is stored in different excel sheets along with area identifiers. Different excel sheets were used to distinguish between different values of temperature, humidity and rainfall, i.e. the excel sheet contains columns named as area_id, temperature, rainfall, humidity, concentration, longitude and latitude. The values of temperature, rainfall and humidity of same excel sheet are similar with different values of concentration, area_id, longitude and latitude. The excel sheets were imported to ArcMap 9.2 version. The data of excel sheets were mapped by using longitudes and latitudes present in the excel sheet. Inverse distance weighting is the simplest interpolation method and is used for interpolating concentration values on map. Figure 5 and 6 shows two interpolated maps for spring season when temperature is 12 degrees centigrade, humidity is 42 and rainfall is 68 millimeter and for autumn season when temperature is 14 degrees centigrade, humidity is 62 and rainfall is 22 millimeter respectively.

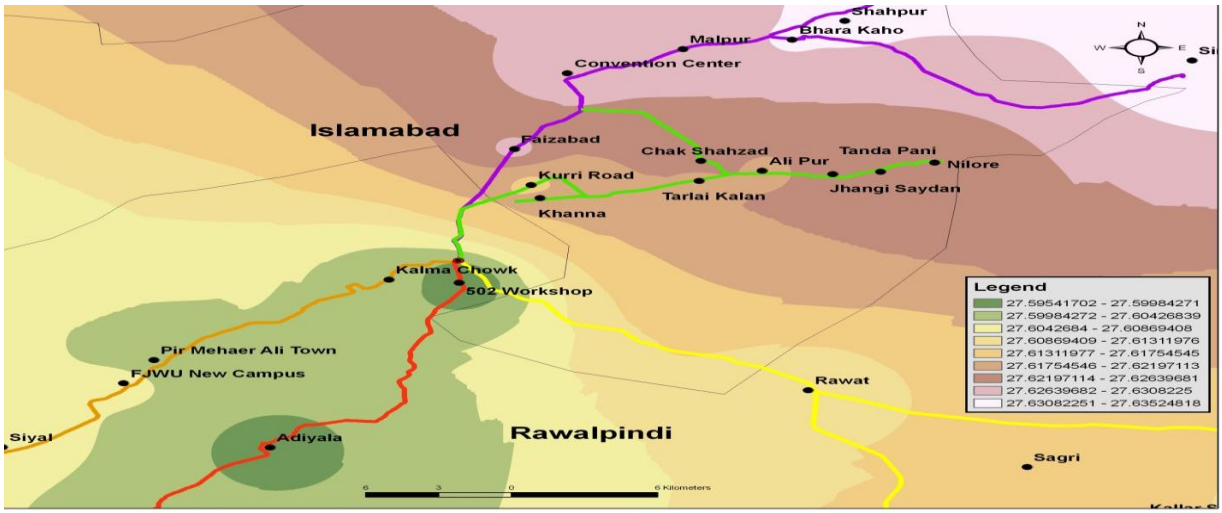

Figure 5 Predicted ozone concentration in spring season

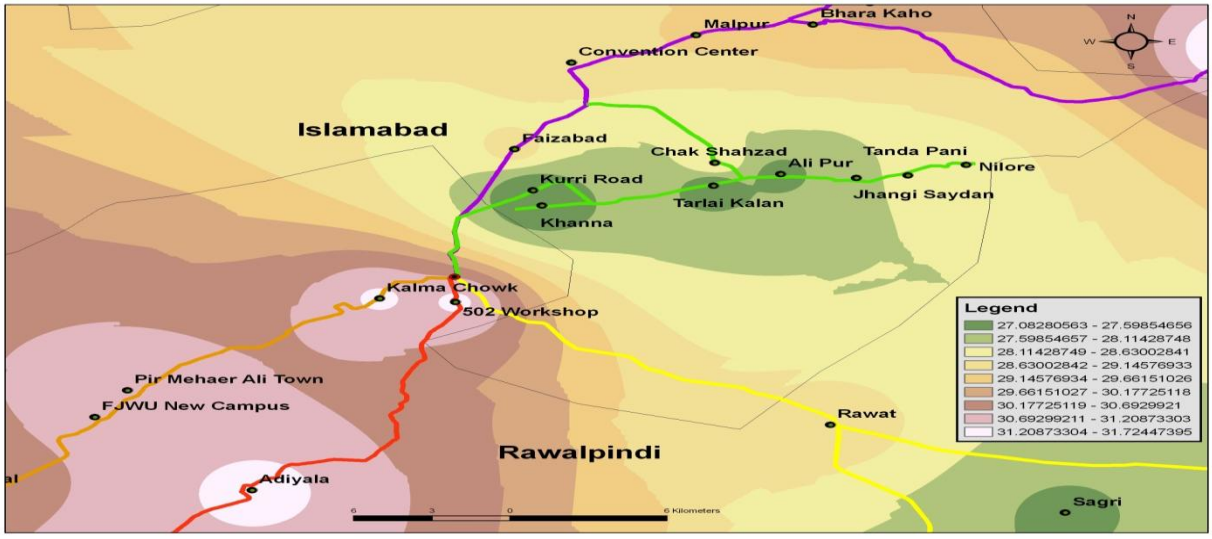

Figure 6 Predicted ozone concentration in autumn season

4. Conclusion

The present research study highlights the importance of air quality management. With the help of artificial neural network it is concluded that O_3 concentration has been increasing day by day due to variable factors. A minor change in temperature, humidity or rainfall directly impacts the concentration of O_3. With the study it is observed that the O_3 concentration was more in Rawalpindi as compared to Islamabad. Spatial interpolation showed variations graphically in O3 concentration at different sampling sites of twin cities in different months. Through present research work hot spots of O3 in twin city have been identified and this work will be helpful in future fort calculation of O_3 in atmosphere and for drawing spatial interpolation of O3. Moreover study highlights the areas of high concentration to control it before it increase above alarming levels.

5. Acknowledgment: Thanks to Higher Education Commission (HEC) of Pakistan for providing the research grant for the study under NRPU Program.

6. References

[1] M. Jerrett, R.T. Burnett, C.A. Pope III, K. Ito, G. Thurston, S. Krewski, Y. Shi, E. Calle, and M. Thun. "Long-Term Ozone Exposure and Mortality", The New England journal o f Medicine, Vol. 360, Issue 11, 2009, pp. 1085–1095.

[2] S.I.V. Sousa, J.C.M. Pires, M.C. Pereira, M.C.M. Alvim-Ferraz, and F.G. Martins. "Potentialities of quantile regression to predict ozone concentrations", Environmetrics, Vol. 20, 2009, pp. 147–158.

[3] J.CM. Pires, and F.G. Martins. "Correction methods for statistical models in tropospheric ozone forecasting", Atmospheric Environment, Vol. 45, 2011, pp. 2413–2417.

[4] M. Gardner and S. Dorling. "Artificial neural network-derived trends in daily maximum surface ozone concentrations". Journal of Air and Waste Management, 51, 2011, pp. 1202–1210.

[5] G. Latini, R.C. Grifoni, and G. Passerini. "The importance of meteorology in determining surface ozone concentrations—a neural network approach", **Coastal Environment: Environmental Problems in Coastal Regions IV**: Ecology and the Environment, Vol. 58, 2002, pp. 405–414

[6] J.R. Kinosian. "Ozone-precursor relationship from EKMA diagrams". Environmental Science and Technology, Vol. 16, Issue 12, 1982, pp. 880-883.

[7] A. Abdul-Wahab, W. Bouhamra, H. Ettouney, B. Sowerby, and B.D. Crittenden. " Predicting ozone levels: a statistical model for predicting ozone levels in the Shuaiba Industrial Area, Kuwait", Environmental Science and Pollution Research, Vol. 3, Issue 4, 1996, pp. 195-204.

[8] U. Utojo, and B.R. Bakshi. "Connection between neural networks and multivariate statistical methods: an overview", In: D.R. Baughman, Y.A. Liu (eds.), Neural Networks in Bioprocessing and Chemical, 1995.

[9] B. Chelani Asha, C.V. Chalapati rao, K.M. Phadke, and M.Z. Hasan. "Prediction of Sulphur dioxide concentration using artificial neural networks", Environmental Modeling and Software, Vol. 17, 2002, pp. 161-178.

[10] M.W. Gardner and S.R. Dorling. "Neural network modeling and prediction of hourly NO_x and NO_2 concentrations on urban air in London", Atmospheric Environment, Vol. 33, 1999, pp. 709-719.

[11] D.R. Baughman, and Y.A. Liu. "Neural Networks in Bioprocessing and Chemical Engineering", Academic Press, New York, 1990, pp. 10-13.

[12] S. Chen, and S.A. Billings. "Neural network for non linear dynamic system modeling and identification", International Journal of Control, Vol. 56, 1992, pp. 319-346.

[13] J. Schumuller. "Neural networks and environmental applications", Expert systems for environmental applications, In: H. Hushou (ed.), ACS Symposium Series, 431. ACS, Miami Beach, Florida, 1988, pp. 5-8.

[14] M. Boznar, M. Lesjak, and P. Mslakar. "A network-based method for short-term predictions of ambient SO2 concentrations in highly polluted industrial areas of complex terrain", Atmospheric Environment, Vol. 27B, 1993, pp. 221-230.

[15] A. Pasini, and S. Potesta. "Neural Network Modelling: Perspectives of Application for Monitoring and Forecasting Physical-Chemical Variables in the Boundary Layer, in Urban Air Pollution", I. Allegrini and F. De Santis (eds.)), Springer-Verlag, 1996, pp.329-340.

[16] O. Mayora-Ibarra, L. Motta, and F. Curatelli. "Forecasting of atmospheric pollution levels with neural networks". Neural Networks and their Applications, Marseilles, Vol. 11, 1998, pp. 323-328.

[17] **C.** Paoli, G. Notton, M.L. Nivet, M. Padovani, and J.L. Savelli. "A Neural Network model forecasting for prediction of hourly ozone concentration in Corsica," 10th International Conference on Environment and Electrical Engineering, , 2011, pp.1-4, 8-11 May 2011, Rome , Italie.

[18] M. Cai, Y. Yafeng , and X. Min. "Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach", Transportation Research Part D: Transport and Environment, Vol. 14, Issue 1, 2009, pp. 32-41.

[19] J.C. Ruiz-Suarez, O.A. Mayora-Ibarra, and J. Torres-Jimenez. "Short-term ozone forecasting by artificial neural networks", Advance in Engineering Software, Vol. 23, 1995, pp. 143-149.

[20] J. Yi, and R. Prybutok. "A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area", Environmental Pollution, Vol. 92, Issue 3, 1996, pp. 349-357.

[21] A. Elkamel, S. Abdul-Wahab, W. Bouhamra, and E. Alper. "Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach", Advances in Environmental Research, Vol. 5, 2001, pp. 47-59.

[22] A. Elkamel. "An artificial neural network for predicting and optimizing immiscible flood performance in heterogeneous reservoirs", Computers and Chemical Engineering, Vol. 22, 1998, pp. 1699-1709.

[23] H. Mayer. "Air pollution in cities". Atmospheric Environment, Vol. 33, 1999, pp. 3029-4037.

[24] A.T. Chan, E.S.P. So, and S.C. Samad. "Strategic guidelines for street canyon geometry to achieve sustainable street quality", Atmospheric Environment, Vol. 35, 2001, pp. 5681-5691.

[25] P. Martin, B. Cabañas, F. Villanueva, M.P. Gallego, I. Colmenar, and S. Salgado. "Ozone and Nitrogen Dioxide Levels Monitored in an Urban Area (Ciudad Real) in Central-southern Spain", Water Air Soil Pollution, Vol. 208, Issue (1-4), 2010, pp. 305-316.