

FORECASTING OF CRITERIA AIR POLLUTANTS' CONCENTRATIONS WITH STATISTICAL MODELS

Utkan Özdemir, Simge Çankaya

Kocaeli University Department of Environmental Engineering, 41380, Turkey

Abstract

The adverse effects of air pollutants have become a common problem in many developed or developing countries because of the major risks to human and environmental health. Particulate matter, ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides, and lead were defined as criteria air pollutants by Environmental Protection Agency (EPA) such as chronic respiratory problems, eye irritation, difficulty in breathing, and pulmonary and cardiovascular diseases, including cancers, it is significant to forecast the concentrations of pollutants. For this purpose, there are numerous statistical models such as artificial neural networks (ANN), multiple linear regressions (MLR) and Taguchi which are modeled the pollutant concentrations assessed. The Taguchi method developed by Genichi Taguchi is an effective design of experiments (DOE) approach and the impacts of several factors can be determined simultaneously with this method. ANN has also been used for predict the concentrations of pollutants depends on several input parameters such as wind direction, relative humidity, temperature, pressure. In the literature, ANN modeling studies generally compare with MLR because of the higher sensitivity of models. In this review, we investigated the statistical models that are used for forecasting air pollutant concentrations and compared the advantages, limitations and performance indicators of those models.

Key words: statistical models, air pollution, design of experiments

1. INTRODUCTION

Air pollution is one of the important problems in new world design. The adverse effects of these pollutants have become a common problem in environmental science area (Chan & Jian 2013). Most of them affect human or environment health (Kurt et al. 2008). Chronic respiratory problems, pulmonary and cardiovascular diseases, cancer (especially lung cancer), eye irritation etc arises from higher dispersion of air pollutants (Taner & Özdemir 2014). So, it is very important to monitoring criteria of air pollutants with forecasting models (Gardner & Dorling 1999). Box models, Gaussian plume models, persistence regression models are commonly used for this purpose (Russo et al. 2013). In last decades, different model studies are using for monitoring and predicting environmental data. Artificial neural network (ANN), multiple linear regressions (MLR), Taguchi, Box-Behnken, fuzzy systems etc are promising alternative to conventional models in this area (Özdemir et al. 2014).

In this short-review study, some models were investigated and compared the advantages, limitations and performances indicators of these models. The other aim of this study is showing the differences between optimization capacities of models and prediction process in air pollution data. The main of this study is indicates the importance of model studies in environmental area.

1.1 Artificial neural networks (ANN)

ANN can model the highly linear non-linear functions and can be trained to accurately with unseen data. Most of scientist are used ANN for wastewater treatment control, air pollution control and waste management system (Chakraborty et al. 2013). Most of neural network studies indicate the prediction-forecasting of future trends of data given current and previous conditions (Russo et al. 2013). In air pollution research, ANN can be used the estimate air pollution concentration levels such as NO_x (Perez & Trier 2001), SO_x (Mok & Tam 1998), CO (Cai et al. 2009), particulate matter (PM) (Taner & Özdemir 2014), O₃ (Haje & Olej 2012). Figure 1 shows the general procedure of ANN.

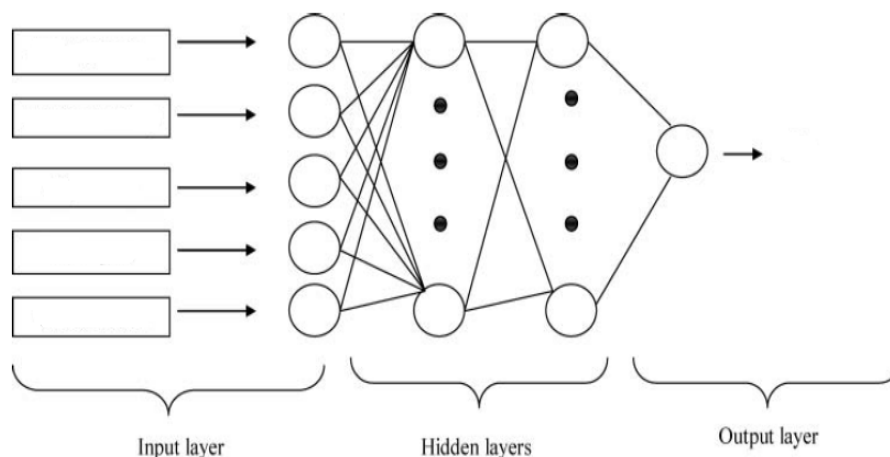


Fig 1. The general scheme of ANN

1.2. Taguchi optimization model

Taguchi statistical design method developed by Genichi Taguchi can also be applied to environmental or biological data for optimizing process and predicting future aspects easily and cheaper (Al et al. 2013). On the other hand, Taguchi is an effective and innovator method in different areas such as material design, energy, coal-oil agglomeration etc. Furthermore, Taguchi gives better graphical visualization compared the other optimization models. So, it can be used easily in environmental science (Özdemir et al. 2014).

Taguchi model has different orthogonal array such as L_8 , L_{16} , L_{32} for determine the optimum condition for system. Orthogonal arrays showed the regular dispersion for better prediction and monitoring (Al et al. 2013). The other important parameter is signal to noise ratio for Taguchi. Signal to noise ratio (S/N) is required to measure the quality characteristics (Özdemir et al. 2014). There are three different types of S/N ratios (larger is better, smaller is better and nominal is better) in Taguchi applications (Zolfaghari et al. 2011). These types are related with characteristic of output data. Figure 2 shows an example about main effects for S/N ratios (L_{32} orthogonal array, larger is better characteristic).

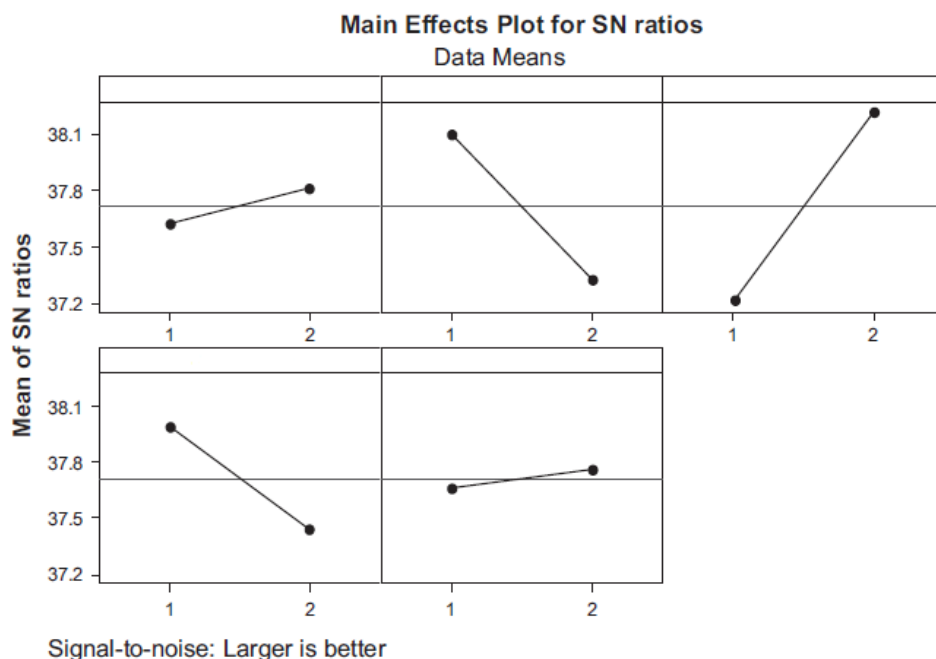


Fig 2. Main effects plot for S/N ratios (example).

1.3. Multiple linear regression (MLR)

Multiple linear regression (MLR) models are used with the aim of determining the effects of independent factors on the dependent variable (Abdul-Wahab et al. 2005). The MLR model can also be represented as sensitivities (Cannon & McKnedry 2002). The general MLR formula can be expressed below:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n + \varepsilon \quad (1)$$

In this equation, Y is the dependent variable, X_i is the independent variable, a_i regression coefficients and ε is the conjectural error of the model.

2. AREAS OF USAGE OF THE MODELS

In this section, areas of usage of the some models (ANN, Taguchi and MLR) were compared with a short review.

2.1. Importance of the usage area

In environmental science, especially air quality monitoring, usage area is more important factor for future scenario. The researchers should use the available model in environmental media for better prediction.

Table 1. Usage areas of the models.

Model	Usage Area	Reference
Taguchi	Adsorption (wastewater)	Zolfaghari et al. 2011
Taguchi	Toxicology	Al et al. 2013
Taguchi	Adsorption (wastewater)	Özdemir et al. 2014
ANN	Meteorological factors on PM	Taner & Özdemir 2014
ANN	Ozone prediction	Haje & Olej 2012
ANN	Air pollution forecasting	Kurt et al. 2008
MLR	Meteorological factors on PM	Taner & Özdemir 2014
MLR	Ground-level ozone	Abdul-Wahab et al. 2005

As seen from the table, ANN models are commonly using in air pollution studies. Environmental data were supporting with MLR applications in ANN model studies. These studies showed that ANN model systems are more efficient than MLR for prediction and sensitivity of environmental data sets. Taguchi is a design of experiment system. So, it generally uses in wastewater treatment process, erosion studies etc. in environmental studies. Furthermore, it can be use in air quality monitoring with comprehensive data analyses.

ANN models have been used to investigate the complex relationship between parameters. So, it can be investigate impacts of air pollutants easily. On the other hand, predicting capacities are higher and mean squared error (MSE) values are lower. ANN's general behaviors are similar to black-box method (Viotti et al. 2002).

3. EXAMPLE STUDY FOR ANN

Taner and Özdemir (2014) were studied with backpropagation feed-forward network type for determined the meteorological factors on PM10 pollution in urban and industrial areas.

In example study, urban and industrial areas were chosen for better evaluating of different pollutant sources. With this aim, the effects of temperature, relative humidity, air pressure, wind speed, and direction were investigated. MLR analyses show the relationship between predicted and observed PM10 with correlation coefficients. R^2 values are 0.74 and 0.36 for the urban and industrial areas respectively. In ANN analyses R^2 values are 0.87 and 0.49 for urban and industrial areas.

Table 2. Some using functions for example study.

Training function	Adaptation function	Tranfer function
Trainr	Learngdm	Tansig
Trainr	Learngdm	Logsig
Trainlm	Learngdm	Logsig
Trainlm	Learngdm	Tansig
Trainlm	Learngdm	Tansig
Trainlm	Learngdm	Tansig
Trainlm	Learngd	Tansig

Trainlm	Learngd	Logsig
Trainr	Learngd	Tansig
Trainr	Learngd	Logsig



Fig 3. Applications area for example study.

4. CONCLUSION

In this study, effects of some models in environmental systems, especially air pollution were investigated. This short-review study indicates ANN is the most effective model for predicting and explaining complex relationship between pollution data. On the other hand, Taguchi, which is innovative method, can also be used in this system. But Taguchi is a design of experiment model. So, comprehensive data analyses were needed for this system. MLR models are generally used as supporter after ANN analyses. Lots of model studies showed that MSE values of ANN are lower than the MSE values of MLR. This situation showed that ANN is the most effective model for predicting and determining air pollution data.

REFERENCES

- Abdul-Wahab, SA, Bakheit, CS, & Al-Alawi, SM 2005, 'Principal component and multiple regression analysis in modelling of groundlevel ozone and factors affecting its concentrations', *Environmental Modelling and Software*, vol. 20 pp.1263–1271.
- Al, G, Özdemir, U, & Aksoy, Ö 2013, 'Cytotoxic effects of Reactive Blue 33 on *Allium cepa* determined using Taguchi's L_8 orthogonal array', *Ecotoxicology and Environmental Safety*, vol. 98 pp. 36-40.
- Cai, M, Yin, Y, & Xie, M 2009, 'Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach', *Transportation Research Part D: Transport and Environment*, vol.14 pp. 32–41.
- Cannon, AJ & McKendry, IG 2002, 'A graphical sensitivity analysis for statistical climate models: Application to Indian monsoon rainfall prediction by artificial neural networks and multiple linear regression models', *International Journal of Climatology*, vol. 22 pp.1687–1708.

- Chakraborty, S, Chowdhury, S, & Saha, PD 2013, 'Artificial neural network (ANN) modeling of dynamic adsorption of crystal violet from aqueous solution using citric-acid-modified rice (*Oryza sativa*) straw as adsorbent', *Clean Technologies and Environmental Policy*, vol.15 pp.255–264.
- Chan, KY & Jian, L 2013, 'Identification of significant factors for air pollution levels using a neural network based knowledge discovery system', *Neurocomputing*, vol. 99, pp. 564-569.
- Gardner, MW & Dorling, SR 1999, 'Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London', *Atmospheric Environment*, vol.33 pp.709–719.
- Haje, P & Olej, V 2012, 'Ozone prediction on the basis of neural networks, support vector regression and methods with uncertainty', *Ecological Informatics*, vol.12 pp.31–42.
- Kurt, A, Gulbagci, B, Karaca, F, & Alagha, O 2008, 'An online air pollution forecasting system using neural networks', *Environmental International*, vol. 34 pp. 592–598.
- Mok, KM & Tam, SC 1998, 'Short-term prediction of SO₂ concentration in Macau with artificial neural networks', *Energy and Buildings*, vol.28 pp.279–286.
- Özdemir, U, Özbay, B, Özbay, İ, & Veli, S 2014, 'Application of Taguchi L32 orthogonal array design to optimize copper biosorption by using *Spaghnum moss*', *Ecotoxicolog and Environmental Safety*, vol. 107 pp. 229-235.
- Perez, P, & Trier, A 2001, 'Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago, Chile', *Atmospheric Environment*, vol.35 pp. 1783–1789.
- Russo, A, Raischel, F, & Lind, PG 2013, 'Air quality prediction using optimal neural networks with stochastic variables', *Atmospheric Environment*, vol. 79 pp. 822–830.
- Taner, S & Özdemir, U 2014, 'Impacts of meteorological factors on PM₁₀: artificial neural networks (ANN) and multiple linear regression (MLR) approaches', *Environmental Forensics*, vol.15 pp. 329-336.
- Viotti, P, Liuti, G, & Genova PD 2002, 'Atmospheric urban pollution: Applications of an artificial neural network (ANN) to the city of Perugia', *Ecological Modelling*, vol. 148pp. 27–46.
- Zolfaghari, G, Esmaili-Sari, A, Anbia, M, Younesi, H, Amirmahmoodi, S, & Ghafari-Nazari, A 2011, 'Taguchi optimization approach for Pb(II) and Hg (II) removal from aqueous solution using modified mesoporous carbon', *Journal of Hazardous Materials*, vol.192 pp.1046-1055.