Scientific
Research
Publishing

Analysis of Tropospheric Ozone by Artificial Neural Network Approach in Beijing

Muhammad Azher Hassan, Zhaomin Dong*

School of Space and Environment, Beihang University, Beijing, China
Email: *dongzm@buaa.edu.cn

Abstract

Higher concentration of tropospheric ozone in atmosphere reveals its adverse effects on human health, plants, and on environment. So, there is a need for atmospheric pollutants analysis and their concentration variation, which is a key factor for air quality management in urban areas. The Beijing Olympic center site was used as area of study and five recorded meteorological parameters temperature, dew point, wind speed, pressure, and relative humidity were employed as inputs imputes. Nitrogen Dioxide (NO_2) and hour of day are also considered as input parameters for modeling of tropospheric ozone concentrations. Several deterministic methods are available for local air quality forecasting and prediction. But, in this study, multilayer perceptron (MLP) and generalized regression neural model (GRNM) were considered for prediction of ozone ground level concentration. The root mean squared errors (RMSE) and mean absolute error (MAE) value for MLP model were lower, which confirms its fitness for forecasting purpose. Regression coefficient for MLP in this study was calculated 0.91 and for GRNM model provides 0.76 value. The dew point and relative humidity were the most dominant input imputes found by model, which results in higher concentration of tropospheric ozone.

Keywords

Neural Network, Surface Ozone, Air Pollution, NN Modeling, Ozone in Beijing

1. Introduction

Air pollution has become critical issue for health exposure and it is increasing by due to excessive emission from vehicles and industries. The traffic emission from diesel incorporated vehicles mainly consist of brown toxic nitrogen dioxide

(NO$_2$) gas and carbon monoxide (CO), which are the main contributor for the formation of secondary pollutants like tropospheric ozone (O$_3$) (Charron & Harrison, 2003). As numerous study confirmed that human health and ecosystem are adversely affected due to exposure of elevated concentration level of tropospheric ozone. So, there is an urgent need for the understanding the precursors and controlling ozone concentration in order to minimize the effects of tropospheric ozone on human health and environment.

In order to ensure a good air quality of Beijing city, continuous monitoring of anthropogenic pollutants emissions is recorded by regional Air Quality monitoring concern authority and real-time air quality can be accessed by China National Environmental Monitoring Center website (http://106.37.208.233:20035/). Recorded measurement data from these concern authorities revealed that an increase in ozone concentration trend was noticed which was attributed to the growth of emission of its precursors (Bureau, 2017). The Beijing city is surrounded by plenty of industries, the direct emission from chimneys into air without any proper treatment applied from these industries as well as transport vehicles emission become a cause of pollution plum in surrounding area (Wang et al., 2018). The transport of these pollutants directly affected by metrological conditions (He et al., 2017). So, this work primarily focus on multilayer based forecasting neural network model to understand relationship of the air quality in Beijing city on the behalf of encountered meteorological conditions in different season throughout the year.

Different techniques has been implemented for the assessment of tropospheric ozone (O$_3$) with support of significant interrelated environmental factors like temperature, dew point, wind speed, and humidity. Commonly, neural network models are considered as best alternatives of highly sophisticated deterministic model due to their ability to skip the all redundant investigation of physical parameters happened in background during transformation and dispersion process of the pollutants. The goal of neural networks (NN) is to generalize or approximate the mathematical relationship of input vectors for dimensional output variable.

Among possible statistical neural network models, generalized regression neural model (GRNM) and Multilayer Perceptron (MLP) model are mostly used for machine learning purpose of complex problems. These both models have ability to judge a smooth measureable relationship in between predictand and predictors variables due to distinguished approximate capabilities. For example, in the past decade, neural network machine learning has been employed in order to forecast hourly ozone level for period of 8 hours (Corani, 2005; Ibarra-Berastegi et al., 2008). Furthermore, recently carried out works have confirmed that that MLP neural network architecture structure assist to obtain accurate hourly prediction on day ahead (Coman et al., 2008; Hrust et al., 2009). Even artificial neural network prediction approach has predicted a model for 1 hour ahead concentration level for Corsica Island (Paoli et al., 2010). The neural network (NN) model can be applied no matter the linear problem or non-linear problem, and can also predict continuous values after supervised training based

on the predictor parameters.

In this study, two NN architecture structure model MLP and GRNM were applied in order to forecast hourly ozone levels for case study site in Beijing, China. The performance indicators values confirm the best fitness of MLP model when compared with GRNM model. In addition, effect of meteorological, NO_2 and time as input factors analyzedplus seasonal variations were also measured in correlation with tropospheric ozone. Overall, this work is useful for predicting the concentration of ozone pollutant based meteorological input factors and as well as helpful for the purpose of air quality management and mitigation throughout the year over the selected location.

2. Materials and Methodology

2.1. Site and Data Description

The Olympic center of Beijing located at (39.98˚N, 116.39˚E) is downtown area of populated city. Air pollutant and meteorological hourly observations for one year (January, 2016 to December, 2016) were used in the model for this NN study. The data was gathered through National Meteorological Information Center official website (http://data.cma.cn). About 4.8% (336 out of 8784) of measured values of tropospheric ozone were missing which was filled by first order linear interpolation method to make the data continuous. Data of NO_2 and other metrological parameters were also fetched from mentioned site. **Table 1** conveys the input and outputs used in this study with commonly used statistics parameters. The total data set has 8784 data points, over the span of whole year 2016 hourly observations. During forecasting, the 70% of random data points used as training purposes and 15% of said data points retained for each of training and testing for the both models we used in this study.

Data was first collected and preprocessed before being served to train a neural network (NN) model. Stepwise linear regression technique was adopted for search strategy of feature selection that is based on forward selection of predictors to pre-check the correlation with the target output. Flowchart of methodology, adopted for prediction of surface ozone is depicted in (**Figure 1**).

Table 1. Descriptive statistics of metrological parameters and air pollutant.

Input/Output	Observer	Unit	Values Range	Average	STD
Input parameters	Temperature	˚C	−16 to 37	13.19	12.16
	Dew point	˚C	−34 to 27	2.84	13.91
	Wind speed	Km/h	0 to 50	10.76	7.92
	humidity	%	8 to 100	55.76	25.48
	Pressure	hPa	989.9 to 1041.7	1012.6	10.55
	NO_2	μg/m³	2.0 to 210	50.47	32.83
	Time	Hours	0 to 23	-	-
Output Observer	Ozone	μg/m³	2.0 to 342.0	57.78	55.10
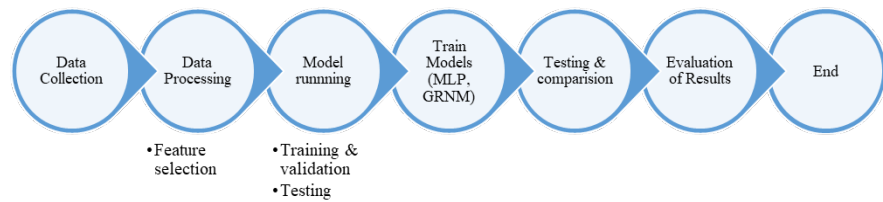
Figure 1. Flow chart of modelling methodology.

2.2. Selection of Metrological Precursors

The inputs variables for these both models were nitrogen dioxide (NO_2), time, temperature, dew point, wind speed, pressure, and percentage humidity with the final objective for hourly ozone concentration assessment. Dataset was divided into 3 parts: training data set, validation data set, and testing data. Training includes 6148 (70%) data points which were used to test and validate 1318 (15%) models.

2.3. Construction of Model

2.3.1. Multilayer Perceptron (MLP)

The MLP generally applied model for the type of back propagation neural network for prediction of multiple outputs or intake of multiple parameters. Its architecture structure is involved the combination of processing elements and connections (Coman et al., 2008; Cabaneros et al., 2017). The processing elements of MLP model termed as neurons are organized in 3 layers; 1) input layer, 2) hidden layers, 3) output layer. The input signal from an input layer is delivered to next layer called as hidden layer shown in (**Figure 2**). The each unit involved in the hidden layer zone sums its applied input and then further processes it with a transfer function and finally forwarded the result to the output layer. Mathematically, it can be presented as:

$$Y = \partial(\sum_{i=1}^{n} w_i X_i + b) = \partial(w^K X = b) \tag{1}$$

where n is number of nodes, w donated the weights of vectors, t is target at time, b is the bias, X is the input of vectors, and ∂ are the sigmoid activation function which is given in Equation (2):

$$f(u) = \frac{1}{1 + e^{-u}} \tag{2}$$

The MLP model used in this study was trained using Levenberg-Marquardt backpropagation, which is excellent supervised algorithm technique with 20 hidden layers.

2.3.2. Generalized Regression Neural Model (GRNM)

Generalized Regression Neural Model (GRNM) model working principle is generally based on kernel regression neural network. GRNM can generalize any complex function among input and output vectors, with approximation of the function from training dataset. In GRNM reliability with the training set is relatively high, estimation error become lower when data size becomes large. A
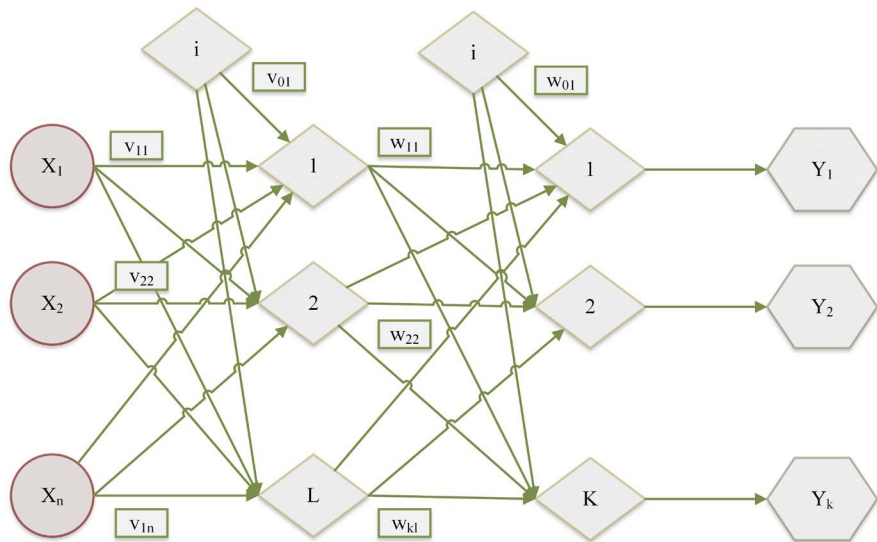
Figure 2. Multilayer perceptron structure.

GRNM comprises of 4 layers; 1) input layer, 2) pattern layer, 3) summation layer and 4) output layer. Input layer consist of number of inputs and that first layer connected with pattern layer. Third layer is summation layer which achieve normalization of output set together with output layer. Simplest diagram of GRNM is presented in (**Figure 3**). Every pattern layer is joined to the summation layer through two neurons, i.e. U and W summation neurons. Summation neuron U consists of the un-weighted factors of the pattern layer, and W is consist of weighted output. Output predicted value (Y) generated through output layer by simply dividing the output of each W weighted neuron by U un-weighted summation neuron.

$$Y' = \frac{\sum\limits_{i=1}^{n} Y_i \exp - G(X, X_i)}{\sum\limits_{i=1}^{n} \exp - G(X, X_i)_i} \tag{3}$$

$$G(X, X_i) = \sum\limits_{k=1}^{n} (\frac{X_i - X_{ik}}{\partial})^2 \tag{4}$$

where, Y_i donates weighting of ith neuron between the pattern and summation layer, n represents number of patterns, G is Gaussian function.

3. Results of the Model Scenarios

3.1. Multilayer Perceptron (MLP)

The inputs variables for MLP model to forecast hourly ozone concentration, were used with 3 different ways i.e. 1) all predictors 2) 6 predictors (without NO_2) 3) 5 predictors (without NO_2 and time) as shown in (**Table 2**). Results of MLP model using all predictors involved in this study were exhibited through scatter plots for training, testing and validation test between actual versus predicted hourly ozone concentration in (**Figure 4**). The validation of the MLP
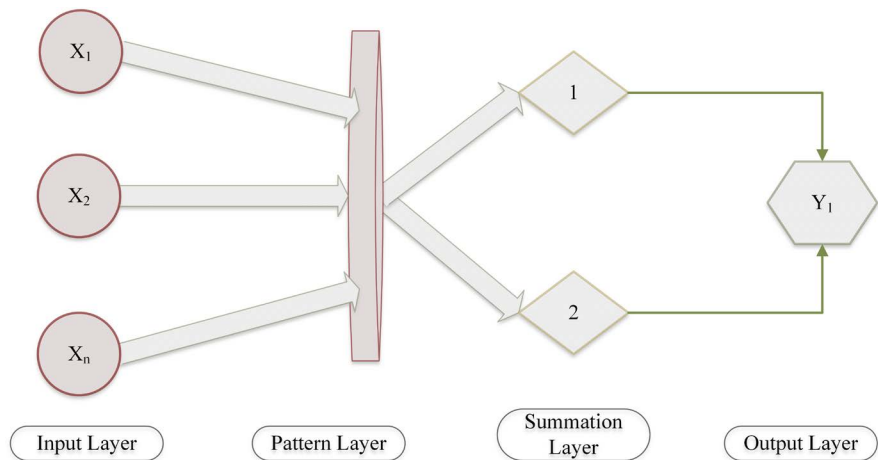
Figure 3. GRNM structural network.
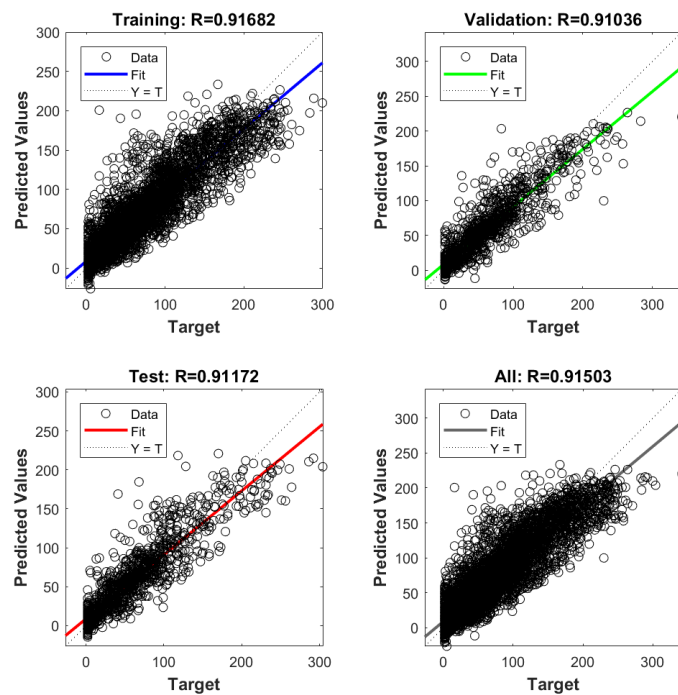
Figure 4. Regression plots of MLP Model for all predictors.

Table 2. MLP model output correlation coefficients (R).

Predictors	Training (R)	Validation (R)	Test (R)
All Predictors	0.91	0.91	0.91
6 predictors (without NO_2)	0.88	0.87	0.89
5 predictors (without NO_2 and time)	0.80	0.808	0.805

simplified network were also in agreement with training and testing data. In case of all predictors involved for prediction purpose the quite higher value of R = 0.91 was noticed. In addition, with the removal of input imputes of air pollutant (NO_2) and time there was no significant differences in performance indicators

observed. So, MLP model with temperature, dew point, wind speed, pressure, and percentage humidity (only metrological factors) shows agreement of R = 0.80 for prediction of hourly tropospheric ozone. Predicted values of MLP model vs observed values of hourly tropospheric ozone are shown in (**Figure 5**).

3.2. Generalized Regression Neural Model (GRNM)

The GRNM model was also employed for the prediction of ozone concentration over the study area. The results exhibited that in case of GRNM model applied for forecasting purpose, the values for performance indicators RMSE (26.94) and MAE (16.42) were relatively higher than MLP model RMSE (21.74) and MAE (14.22) values.

This depicts that GRNM model gives higher error values which are inimitable when compared with originated low forecast error for ANN models. Results of RNGM model using all predictors used in this study exhibited correlation coefficient (R = 0.76) shown in (**Figure 6**) and graph between actual versus predicted hourly ozone concentrations are drawn in (**Figure 7**).
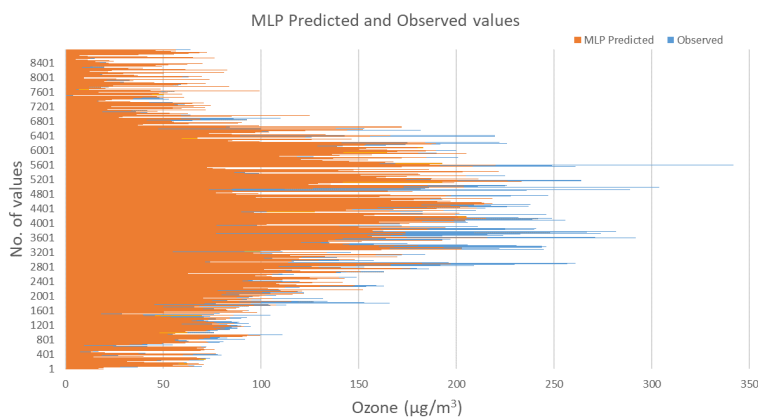

Figure 5. Observations vs Predicted values for MLP.
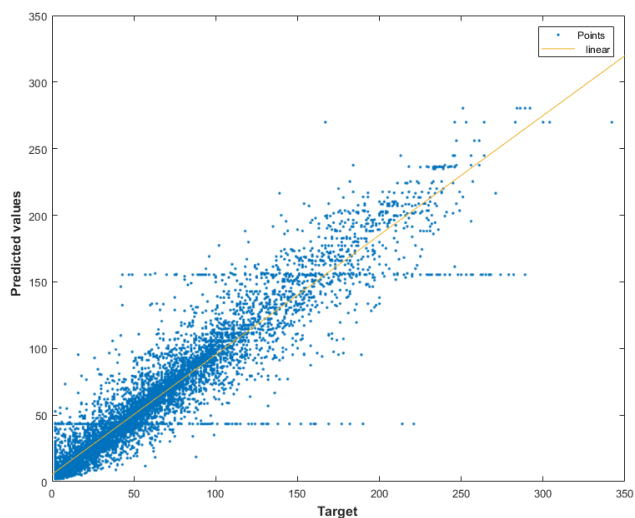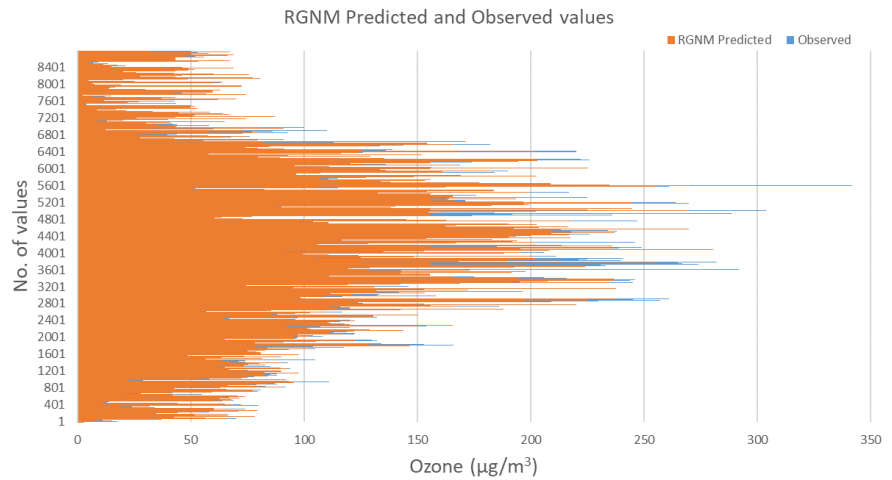

Figure 6. Regression plot of RGNM.

Figure 7. Observations vs Predicted values for RGNM.

3.3. Evaluation of Models

Both the root mean square (RMSE) and mean absolute error (MAE) values were considered in order to evaluate the performance of applied models on the behalf of amount of error originated in predicting the concentration of ground level ozone. Correlation coefficient (R) is also calculated as evaluation parameter to check the fitness of model. The performance of MLP and RGNM models are judged between observed/measure values and model predicted values applied during ANN approaches shown in (**Table 3**). The values for RMSE and MAE were determined by applying the following RMSE (5) and MAE (6) equations:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{obs} - Y_{pre})^2} \qquad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_{obs} - Y_{pre}| \qquad (6)$$

4. Seasonal Variation

The fluctuations in hourly ozone concentration level are noticed in Beijing city throughout the year. This fluctuation could be referred to variation in temperature, wind speed, dew point, and relative humidity prevailing conditions during different seasons. The recorded seasonal based values of temperature, wind speed, dew point, relative humidity, and O_3 are given in (**Table 4**).

For different seasons, the dominant variation for hourly tropospheric ozone concentration was depicted after the end of winter season. It was also observed that ozone concentration level reached to maximum value during summer and monsoon seasons. **Table 5** represents that relatively higher concentration of O_3 (292 - 342 μg/m³) was noticed during the summer and monsoon season and at the same time peak values for dew point and humidity recorded as well as compared to other seasons. Our findings through this analysis revealed that the studied meteorological parameters have key importance or interrelationship with O_3 concentration. In addition, the relative humidity and dew point were observed

Table 3. MLP and GRNM results comparison.

Models	RMSE	MAE	R
Model-1: MLP	21.74	14.22	0.91
Model-2: GRNM	26.94	16.42	0.76

Table 4. Statics of input data over the seasons.

Time	Parameter	Temp	Dew point	Wind speed	humidity	Pressure	NO$_2$	O$_3$
Quarter 1 (Jan-Mar)	Min	−16	−34	0	8	999	4	2
	Max	23	7	50	100	1042	191	166
	Avg.	1.42	−12.73	13.36	39.66	1021.57	46.92	42.14
	STD	7.88	7.20	9.62	20.56	7.68	33.38	30.88
Quarter 2 (Apr-Jun)	Min	4	−19	0	8	990	2	2
	Max	37	25	47	100	1021	182	292
	Avg.	20.76	7.65	11.97	49.14	1004.54	41.35	93.00
	STD	6.49	8.55	7.97	24.40	5.72	25.98	59.08
Quarter 3 (July-Sep)	Min	6	−4	0	9	990	4	2
	Max	37	27	50	100	1025	143	342
	Avg.	25.32	18.31	8.84	68.82	1004.21	42.35	78.51
	STD	5.10	5.41	5.14	20.45	5.92	22.95	60.15
Quarter 4 (Oct-Dec)	Min	−12	−21	0	11	1002	6	2
	Max	28	18	47	100	1037	210	169
	Avg.	5.22	−1.99	8.93	65.20	1020.17	70.78	17.70
	STD	7.40	9.54	7.30	24.46	6.74	37.30	23.74

as the most dominant meteorological factors for predicting the concentration of ozone.

5. Conclusion

In this study, neural network approach was employed for forecasting the ground-level ozone concentration in Beijing city in China. Both MLP and GRNM were applied with metrological parameters and NO$_2$ for hourly prediction of surface ozone. MLP was best suited model as depicted by lower value of performance assessment parameters RMSE and MAE. It was deduced from model results that dew point and relative humidity were the most prominent factors to predict the ground-level ozone concentration. The obtained results of forecasting of predicted surface ozone concentration might be valuable for the purpose of air quality management for whole year over the selected location.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Bureau, B. E. P. (2017). Air Quality Monthly Report.
http://bjepb.gov.cn/bjhrb/xxgk/ywdt/hjzlzk/dqhjzl/index.html

Cabaneros, S. M. S., Calautit, J. K. S., & Hughes, B. R. (2017). Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO_2 Pollution. *Energy Procedia, 142,* 3524-3530. https://doi.org/10.1016/j.egypro.2017.12.240

Charron, A., & Harrison, R. M. (2003). Primary Particle Formation from Vehicle Emissions during Exhaust Dilution in the Roadside Atmosphere. *Atmospheric Environment, 37,* 4109-4119. https://doi.org/10.1016/S1352-2310(03)00510-7

Coman, A., Ionescu, A., & Candau, Y. (2008). Hourly Ozone Prediction for a 24-h Horizon Using Neural Networks. *Environmental Modelling & Software, 23,* 1407-1421. https://doi.org/10.1016/j.envsoft.2008.04.004

Corani, G. (2005). Air Quality Prediction in Milan: Feed-Forward Neural Networks, Pruned Neural Networks and Lazy Learning. *Ecological Modelling, 185,* 513-529. https://doi.org/10.1016/j.ecolmodel.2005.01.008

He, J., Gong, S., Yu, Y., Yu, L., Wu, L., Mao, H., & Li, R. (2017). Air Pollution Characteristics and Their Relation to Meteorological Conditions during 2014-2015 in Major Chinese Cities. *Environmental Pollution, 223,* 484-496. https://doi.org/10.1016/j.envpol.2017.01.050

Hrust, L., Klaić, Z. B., Križan, J., Antonić, O., & Hercog, P. (2009). Neural Network Forecasting of Air Pollutants Hourly Concentrations Using Optimised Temporal Averages of Meteorological Variables and Pollutant Concentrations. *Atmospheric Environment, 43,* 5588-5596. https://doi.org/10.1016/j.atmosenv.2009.07.048

Ibarra-Berastegi, G., Elias, A., Barona, A., Saenz, J., Ezcurra, A., & Diaz de Argandoña, J. (2008). From Diagnosis to Prognosis for Forecasting Air Pollution Using Neural Networks: Air Pollution Monitoring in Bilbao. *Environmental Modelling & Software, 23,* 622-637. https://doi.org/10.1016/j.envsoft.2007.09.003

Paoli, C., Voyant, C., Muselli, M., & Nivet, M., (2010) A Neural Network Model Forecasting for Prediction of Hourly Ozone Concentration in Corsica. *10th International Conference on Environment and Electrical Engneering (EEEIC)*, IEEE.

Wang, L., Li, P., Yu, S., Mehmood, K., Li, Z., Chang, S., & Seinfeld, J. H. (2018). Predicted Impact of Thermal Power Generation Emission Control Measures in the Beijing-Tianjin-Hebei Region on Air Pollution over Beijing, China. *Scientific Reports, 8,* 934. https://doi.org/10.1038/s41598-018-19481-0