CHEMICAL ENGINEERING TRANSACTIONS
VOL. 61, 2017

A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet

Guest Editors: Petar S Varbanov, Rongxin Su, Hon Loong Lam, Xia Liu, Jiří J Klemeš

ISBN 978-88-95608-51-8; **ISSN** 2283-9216

DOI: 10.3303/CET1761289

Correlation Analysis between Ozone and Other Atmosphere Characteristics

Yuhan Ding[a,*], Guohai Liu[a], Xianqi Guo[b], Congli Mei[a], Peisuo Yang[a]

[a]Jiangsu University, Zhenjiang, 212013, Jiangsu, China
[b]Putuo Environmental Monitoring Station, Zhoushan, 316100, Zhejiang, China
yhding@ujs.edu.cn

The correlation between ozone and other atmosphere characteristics is important for air pollution analysis. It is complicated to model and analyse. In this paper, a correlation analysis method based on neural network and mean impact value (NN-MIV) is proposed. In NN-MIV correlation analysis method, the external correlation of key variable (ozone) and auxiliary variables (other atmosphere characteristics) is calculated by MIV and internal correlation is calculated by NN. By this means, stable correlations are obtained. The correlation value shows that ozone is most related to NO_2, followed by humidity, $PM_{2.5}$, temperature and pressure. The correlation result is then analysed and some valuable conclusions are presented. Finally, an NN correlation model (or soft sensing model) of ozone is constructed by the 5 most correlated variables and a good soft sensing result is obtained.

1. Introduction

The tropospheric ozone is a severe secondary pollutant and a kind of greenhouse gas in the atmosphere. The tropospheric ozone is mainly produced by the photochemical reaction from the first pollutant, such as NO_x, SO_2 and volatile organic compound, emitted by motor vehicles and factories (Jenkin et al., 2000). With the increasing consumption of the mineral fuel, the emission of the first pollutant is increasing rapidly, resulting in the increasing of the ozone in the atmosphere. High density of the ozone will influence the human health and creature growth, bring serious harm to the ecosystem (Wang et al., 2001).

Researches indicate that ozone is not only related to the other pollutant gas, but also the ultraviolet, atmosphere temperature, wind speed and inhalable particles (Sadanaga et al., 2017). Therefore, it is necessary to learn the correlation between these atmosphere characteristics (including the pollutant gases) and the ozone. Finding the correlation model is helpful to predict and control the air pollution (Sillman, 1999).

There are quite a few researches reported in this area. For instance, Saitoa et al. (2002) studied the relationship between O_3, and its precursors (NO_x and NMHC), but they did not research the relationship between O_3 and other atmosphere characteristics. Nishanth et al. (2014) analysed the correlation between O_3 and its precursors as well as investigated the influence of PM_{10} on surface O_3. Santurtún et al. (2015) discussed the temporal variations of surface ozone concentrations and its link with atmospheric pattern. However, the analysis method is the simple classification and no numerical result of the relationship is presented.

In this paper, a correlation analysis method based on neural network and mean impact value (NN-MIV) is proposed. In NN-MIV correlation analysis method, the external correlation of key variable (ozone) and auxiliary variables (other atmosphere characteristics) is calculated by MIV method (Sun et al., 2012) and internal correlation is calculated by NN method (Du Jardin et al., 2010). A stable correlation is established by combining the both methods. In the meantime, an NN correlation model (or soft sensing model) of ozone is constructed using the 5 most correlated variables. The soft sensing result is expected to be used as real measurement value in some special circumstances such as ozone sensor failure.

Please cite this article as: Ding Y., Liu G., Guo X., Mei C., Yang P., 2017, Correlation analysis between ozone and other atmosphere characteristics, Chemical Engineering Transactions, 61, 1747-1752 DOI:10.3303/CET1761289

2. Data source and analytical method

The data used to analyse the correlation between the ozone concentration and atmosphere characteristics are collected from 2013 to 2015 with interval being 1 h in Donggang, Putuo, Zhoushan, China. As Zhoushan is an island city, the ozone and other atmosphere data are little affected by the industry and traffic of other cities, and the relation between ozone and another atmosphere variable is relatively stable and easy to research. The sampling equipment is on the top of a seven-story building in 29.57° North Latitude and 12.18° East Longitude. The collected data include SO_2, NO_2, CO, $PM_{2.5}$, PM_{10}, temperature (T), atmosphere pressure (P), humidity (H), wind speed (WS) and O_3. Table 1 shows some typical data of each variable.
Then the collected data are analysed and processed, forming several batches of valid data. After that, the correlation between variables can be researched.

Table 1: Typical data

O_3 (mg/m^3)	SO_2 (mg/m^3)	NO_2 (mg/m^3)	CO (mg/m^3)	$PM_{2.5}$ (mg/m^3)	PM_{10} (mg/m^3)	T (℃)	P (kPa)	H (%)	WS (m/s)
135	9	11	0.9	27	38	10.183	101.896	65.177	0.464
120	8	14	0.9	26	37	9.803	101.837	66.168	0.173
109	7	17	0.9	25	36	9.54	101.766	65.144	0.121
109	7	14	0.9	23	35	9.621	101.692	62.988	0.236
109	6	12	0.9	24	36	9.468	101.695	64.764	0.166
109	5	13	0.9	22	33	9.435	101.74	68.077	0.158
81	6	40	1	25	39	9.86	101.795	69.415	0.159
66	11	56	1.1	34	53	12.56	101.828	55.844	0.123
84	12	39	0.9	27	42	14.557	101.877	52.767	0.256
107	13	24	0.9	24	39	15.457	101.936	42.814	0.556

3. NN-MIV correlation analysing method

3.1 Mean impact value (MIV) correlation analysis

Consider an independent input variable vector which contains *p* variables, observe it *m* times to get the variable space $\boldsymbol{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_m]$, and each dependent output variable corresponding to the sample point can be written as $\boldsymbol{Y} = [\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \cdots \quad \boldsymbol{y}_m]$. Taking independent variable vector ***X*** including *m* samples as input, the corresponding output vector ***Y*** as output, an initial neural network is trained and saved. Give a 10 % increase and 10 % decrease to a single independent variable at one time. In this way, *2p* ($i = 1, \cdots p$) new variable spaces can be obtained.

$$\boldsymbol{X}_i^{(1)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{i1}(1+10\%) & x_{i2}(1+10\%) & \cdots & x_{im}(1+10\%) \\ \vdots & \vdots & \cdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pm} \end{bmatrix} \quad (1)$$

$$\boldsymbol{X}_i^{(2)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{i1}(1-10\%) & x_{i2}(1-10\%) & \cdots & x_{im}(1-10\%) \\ \vdots & \vdots & \cdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pm} \end{bmatrix} \quad (2)$$

Take the newly-constructed variable spaces one by one as the input of the neural network model trained previously, and *2p* groups of output vector are obtained through the network:

$$\boldsymbol{Y}_i^{(1)} = \left[y_{i_1}{}^{(1)} \quad y_{i_2}{}^{(1)} \quad \cdots \quad y_{i_m}{}^{(1)} \right] \quad (3)$$

$$\mathbf{Y}_i^{(2)} = \left[y_{i_1}{}^{(2)} \quad y_{i_2}{}^{(2)} \quad \cdots \quad y_{i_m}{}^{(2)} \right] \tag{4}$$

Each group is corresponding to the sample point whose $i-th$ ($i=1,2\cdots p$) variable index is changed. Calculate the difference of Eq(3) and Eq(4) to obtain the impact value by the equation $IV_i = Y_i^{(1)} - Y_i^{(2)}$ when the *i-th* variable indicator is changed. And the mean impact value can be calculated as following:

$$MIV_i = \sum_{j=1}^{m} IV_i(j)/m\,,\; i=1,2,\cdots p \tag{5}$$

The symbol of MIV_i indicates the contribution of the independent input variable to the dependent output variable, whose value represents a relation between the input variable and the output variable. Therefore, the correlation of x_i and y can be calculated as:

$$\alpha_{\mathrm{M}_i} = \frac{MIV_i}{\sum_{i=1}^{p} |MIV_i|} \tag{6}$$

This method calculates the correlation of input variables to the output by the changes of external input, hence it can be defined as external correlation.

3.2 Neural network (NN) correlation analysis

The neural network (Souza et al., 2015) used in the method is a kind of single hidden layer feedforward neural network as shown in Figure 1. The structure of the NN can be described by the following expression:

$$y = \beta_0 + \sum_{j=1}^{q} F(\omega_{j0} + \sum_{i=1}^{p} \omega_{ji} \mathbf{x}_i)\beta_j \tag{7}$$

In Eq(7), $x = \left[x_1 \quad x_2 \quad \cdots \quad x_p \right]^T$, p refers to the number of input variables, $\beta_j (j=0,1,\cdots,q)$ are the weights from hidden layer to output layer, $\omega_{ji} = (\omega_{j0}, \omega_{j1}, \cdots \omega_{jp})$ are the weights from input layer to hidden layer. Define $Z_j = \omega_{j0} + \sum_{k=1}^{p} \omega_{ji} x_i$, and Eq(7) can be transformed to $y = \beta_0 + \sum_{j=1}^{q} F(\mathbf{Z}_j)\beta_j$. On the basis of neural interpretation diagram (NID) (Özesmi and Özesmi, 1999), the correlation of input variables can be calculated according to the correlation coefficient and the covariance and the correlation of input x_i to hidden layer o_j can be calculated as:

$$u_{ji} = \frac{Cov(Z_j, x_i)}{Var(Z_j)} \omega_{ji}, j=1,\ldots,q; i=1,\ldots,p \tag{8}$$

and the correlation of hidden layer o_j and the output y can be calculated as:

$$v_j = \frac{Cov(F(Z_j), y)}{Var(y)} \beta_j, j=1,\ldots,q \tag{9}$$

The overall correlation value of the $i-th$ input x_i and the output y is:

$$C_i = \sum_{j=1}^{q} v_j u_{ji}, i=1,\ldots,p \tag{10}$$

or:

$$\alpha_{N_i} = \frac{C_i}{\sum_{i=1}^{n} |C_i|} \tag{11}$$

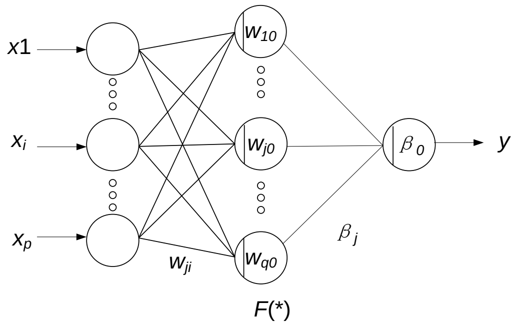

Figure 1: Single hidden layer feedforward neural network

3.3 Variable selection method based on NN-MIV

MIV correlation represents model's character on the changes over external input and the calculated correlation is the external one. Meanwhile, correlation analysing method based on NN calculate the correlation according to the weights from input layer to hidden layer and from hidden layer to output layer, hence this kind of correlation is called internal correlation. In order to make use of the advantages of both methods to obtain the most optimal correlation, the authors propose a neural network variable selection method based on MIV (NN-MIV), where the integrated correlation is defined as

$$C_{NM_i} = \left|\alpha_{M_i}\alpha_{N_i}\right| = (C_1,\cdots,C_p)^T \tag{12}$$

In normalized way, the overall integrated correlation of x_k and output can be expressed as:

$$\alpha_{MN_i} = \frac{C_{NM_i}}{\sum_{i=1}^{p}\left|C_{NM_i}\right|} \tag{13}$$

4. Experiment and discussion

Based on the NN-MIV method proposed in the previous section, the authors researched the relationship between the ozone concentration (denoted as y) and other atmosphere characteristics, including SO_2 concentration, NO_2 concentration, CO concentration, $PM_{2.5}$ concentration, PM_{10} concentration, temperature T, atmosphere pressure P, humidity H, and wind speed WS (denoted as x_1, x_2, …, x_9).

First, a 9-15-1 feedforward neural network was constructed, with activated function of hidden layer neurons being “tansig” and the one of output layer being “purelin”. Then the NN was trained with Levenberg-Marquardt training algorithm (Lera and Pinzolas, 2002) for 500 times and the training error got less than 10^{-4}.

Based on the trained network, MIV correlation analysing method was applied to calculate the external correlation according to Eq(1) - Eq(6), and then NN correlation analysing method was applied to calculate the internal correlation according to Eq(7) - Eq(11).

The integrated correlations were obtained according to Eq(12) and Eq(13). The correlation calculated is very stable, and only minor digit number varies, which does not change the correlation sequence as shown in Table 2. The calculated correlation values are also shown graphically in Figure 2.

Table 2: The correlation between O_3 and other variables

Denotation	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Variable	SO_2	NO_2	CO	$PM_{2.5}$	PM_{10}	T	P	H	WS
Correlation	0.0072	0.50	0.011	0.079	0.015	0.045	0.040	0.28	0.016
Sequence	9	1	8	3	7	4	5	2	6

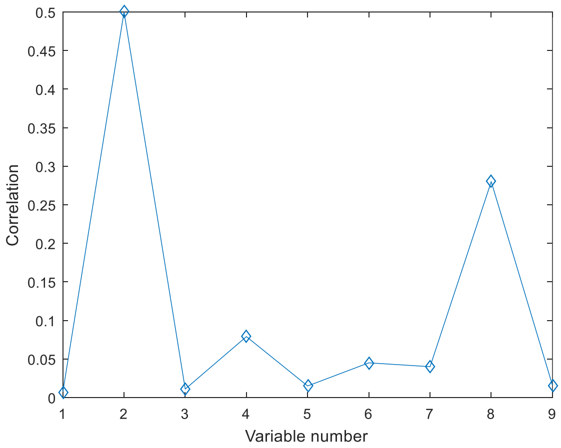

Figure 2: The correlation graph between O_3 and other variables

From the table and figure, the relationship between ozone concentration O_3 and other atmosphere variables has the following features:

(1) O_3 has the maximum correlation with NO_2 and has small correlation with SO_2 and CO, which indicates that O_3 is transformed from NO_2 by photochemical reaction, not SO_2 and CO.

(2) T, H, and P have high correlation with O_3, indicating that the temperature, humidity and atmosphere pressure are all related to the transformation from NO_2 to O_3. Although T, H, P will not directly change NO_2 to O_3, they will influence the ultraviolet intensity which is the key factor to the photochemical reaction, which is reflected by the correlation between T, H, P and O_3.

(3) $PM_{2.5}$ and PM_{10} also have relative high correlation with O_3. It is because the extinction effect of the particles will decrease the sun radiation and influence the photochemical reaction level, and change the amount of O_3 transformed from NO_2. Since correlation between $PM_{2.5}$ and O_3 are bigger than PM_{10} and O_3, we believe that the extinction effect of $PM_{2.5}$ is greater than the one of PM_{10}.

(4) O_3 is also related to WS. Since wind can take O_3 to other place and bring it from other place, the O_3 concentration should have some relation to the wind speed. However, the distribution of O_3 is very complex, and the relation is very complex, too. This complex relation will make the calculated correlation value varies, and lead to a result neither too big nor too small if the data scale is large, as appeared here in this paper.

Furthermore, a 5-10-1 neural network structure was constructed by the 5 most correlated variables, which is the so-called NN soft sensing model. This soft sensing model was trained by the data of year 2013 and 2014, and tested by the data of year 2015. Figure 3 presents a period of time of soft sensing result. It can be seen from the figure that, the soft sensing model constructed by the correlation analysis can provide O_3 soft sensing result close to the real data and it can be used as the real value in some special circumstances such as ozone sensor failure.

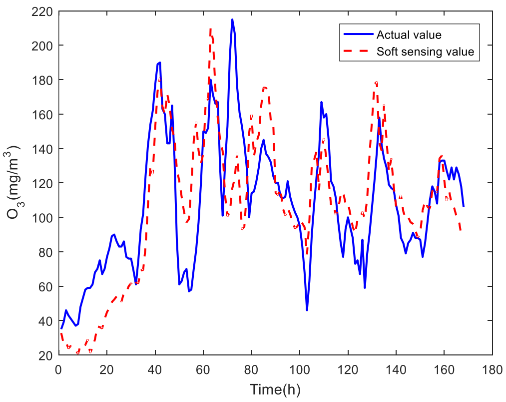

Figure 3: Soft sensing result

5. Conclusion

In this paper, an NN-MIV correlation analysing method is proposed to obtain the correlation between ozone concentration and other atmosphere characteristics. The method analyses the internal correlation by NN method and external correlation by MIV method and obtain a stable correlation value. The deep meaning of the correlation is then analysed and discussed. An NN soft sensing model is constructed by the 5 most correlated variables. Experimental result shows that this soft sensing model can provide a soft sensing result very close to the real data, indicating it can be used in some special circumstances to represent the real value such as ozone sensor failure. The relations between variables from different areas should be further developed and the variables from neighbouring area will be added to the soft sensing model to make the soft sensing result more accurate.

Acknowledgments

This work is supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD [2011]6), and Open Research Foundation of Key Laboratory of Modern Agricultural Equipment and Technology in Jiangsu University (NZ201301).

References

Du J.P., 2010, Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy, Neurocomputing, 13 (2), 32-37.

Jenkin M.E., Clemitshaw K.C., 2000, Ozone and other secondary photochemical pollutants: chemical processes governing their formation in the planetary boundary layer, Atmospheric Environment, 34(16), 2499-2527.

Lera G., Pinzolas M., 2002, Neighborhood based Levenberg – Marquardt algorithm for neural network training, IEEE Transactions on Neural Networks, 13(5), 1200-1203.

Nishanth T., Praseed K.M., Kumar M.K.S., Valsaraj K.T., 2014, Influence of ozone precursors and PM_{10} on the variation of surface O_3 over Kannur, India, Atmospheric Research, 138(3), 112-124.

Sadanaga Y., Kawasaki S., Tanaka Y., Kajii Y., Bandow H., 2017, New system for measuring the photochemical ozone production rate in the atmosphere, Environmental Science & Technology, 51(5), 2871-2878.

Saitoa S., Nagao I., Tanaka H., 2002, Relationship of NO_X and NMHC to photochemical O_3 production in a coastal and metropolitan areas of Japan, Atmospheric Environment, 36(8), 1277-1286.

Santurtún A., González-Hidalgo J.C., Sanchez-Lorenzo A., Zarrabeitia M.T., 2015, Surface ozone concentration trends and its relationship with weather type in Spain (2001-2010), Atmospheric Environment, 101(1), 10-22.

Sillman S., 1999, The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments, Atmospheric Environment, 33(12), 1821-1845.

Souza R.M.S., Coelho G.P., Da Silva A.E.A., Pozza S.A., 2015, Using ensembles of artificial neural networks to improve PM10 forecasts, Chemical Engineering Transactions, 43, 2161-2166.

Sun W., Liu X., Wang H., 2012, Weight analysis of cast blasting effective factors based on MIV method, Journal of China University of Mining and Technology, 41(6), 993-998.

Wang T., Cheung V.T.F., Anson M., Li Y.S., 2001, Ozone and related gaseous pollutants in the boundary layer of Eastern China: Overview of the recent measurements at a rural site, Geophysical Research Letters, 28(12), 2373- 2376.

Özesmi S.L., Özesmi U., 1999, An artificial neural network approach to spatial habitat modeling with interspecific interaction, Ecological Modelling, 116(1), 15-31.