Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2014.1002283

57 (2016) 1322–1326
January

Taylor & Francis
Taylor & Francis Group

Assessment of prediction accuracy in autonomous air quality models

Szymon Hoffman

Department of Chemistry, Water and Wastewater Technology, Czestochowa University of Technology, 69 Dabrowskiego St., 42-200 Czestochowa, Poland, email: szymon@is.pcz.czest.pl

Received 15 October 2014; Accepted 2 December 2014

ABSTRACT

In the study, a long-term set of data collected at the air monitoring station located in Lodz (central Poland) was analysed. Two air pollutants—O_3 and CO—were chosen in order to carry out the prediction procedure. The prediction was performed using regression neural networks. The modelled concentrations were compared to the actual ones in order to assess the prediction accuracy. Approximation errors were calculated for the entire range of concentrations and also separately for several concentration sub ranges. The following measures of error were considered: the mean absolute error, the mean squared error, the root mean squared error, the mean absolute relative error, Pearson's correlation coefficient and Willmott's indexes of agreement. Values of errors and their variabilities in different ranges were analysed. It was stated that only some error measures properly reflect the difficulties in modelling concentrations in the entire range of concentrations as well as in different sub ranges of concentrations. The use of a single error measure may lead to incorrect interpretation.

Keywords: Air pollutants; Air monitoring; Concentration modelling; Approximation error; Neural models

1. Introduction

Concentrations of pollutants measured at air monitoring stations can be approximated with modelling techniques based on the regression analysis, the time-series analysis or other statistical methods [1,2]. The methods which exploit the knowledge collected in historical data, at the same site or at the neighbouring monitoring stations, without recourse to any data coming from the out of the monitoring system, were called autonomous models [3]. In these models, concentrations of other pollutants measured at the same monitoring stations, meteorological data as well as time of measurement were used as predictors of specified pollutant concentration. The essential problem is how to assess the quality of obtained autonomous models. Time-series recorded at air quality monitoring stations are characterized by a specific variation and a specific noise level. Thus, criteria of air quality models performance should be adapted to the specificity of modelling data.

When the real data are available, the quality of approximation is estimated by comparing the modelled values with the actual values in order to evaluate the modelling error. There are many various measures

Presented at the 12th Scientific Conference on Microcontaminants in Human Environment 25–27 September 2014, Czestochowa, Poland

of this error, each of whom results from a specific statistical approach [4,5]. Some of them are universal, such as the mean absolute error, the mean squared error, the root mean squared error, the mean absolute relative error and Pearson's correlation coefficient. They are widely applied in various fields of engineering and science. Others, like Willmott's indexes of agreement, are specially recommended to atmospheric sciences [6–8]. Many different error measures are usually presented in issued papers [9–12]. However, sometimes only a single kind of error is given [13].

In this study, some commonly used error measures were compared in order to recommend procedures that enable to assess appropriately the model quality. The results of tests allowing for comparing the variability of various errors, when changing the level of the modelled concentrations, were presented and discussed in this study.

2. Materials and methods

The analysed data-set derived from the air monitoring station Widzew in the city of Lodz (Central Poland). The data were gathered in the period of 2004–2010. The data collection was obtained from Voivodeship Inspectorate for Environmental Protection in Lodz. The examined time-series involved hourly concentrations of main air pollutants: O_3, NO, NO_2, PM_{10}, SO_2, CO, as well as hourly averages of some meteorological data: Temperature, wind speed, wind direction, relative humidity and solar radiation.

Two air pollutants were chosen in order to carry out the prediction procedure—O_3 and CO. For each of these pollutants, five independent variables were selected from the other time-series measured at the air monitoring station. The prediction of O_3 concentration was performed with a regression neural network that used NO_2 concentration, temperature, relative humidity, day and hour as explanatory variables. In the case of CO, the regression neural network explored the knowledge hidden in the following explanatory variables: NO concentration, PM_{10} concentration, temperature, day and hour. Both neural models used a perceptron with five neurons placed in a single hidden layer. Such a relatively simple neural network structure allows for effective exploration of knowledge hidden in data [14].

The analysis was carried out using the programme Statistica Data Mining. In each neural network, the analysed set of data was randomly divided into three different subsets: The training subset (50% of cases), the verification subset (25% of cases) and the test subset (25% of cases). Back propagation as well as Levenberg–Marquardt algorithms were used for neural networks training. The modelling of O_3 concentration was performed only once for entire set of concentrations with one neural regression network. A similar procedure was used to model the concentration of CO—the entire set of predicted concentration was prepared using a single neural regression model. Approximation errors were calculated for the entire range of concentrations but also separately for several concentration sub ranges. Values of errors and their variability in different sub ranges were analysed.

The values of the model errors were estimated basing on divergences between the model output y_i (predicted concentrations) and the observed concentration values x_i. Seven different categories of prediction error were calculated with the formulas given below, where n—number of observations; $\overline{y}$—mean value in the set of predicted concentrations; $\overline{x}$—mean value in the set of observed concentrations:

MAE—mean absolute error:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{1}$$

MSE—mean squared error:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \tag{2}$$

RMSE—root mean squared error:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{3}$$

MARE—mean absolute relative error:

$$\mathrm{MARE} = \frac{\sum_{i=1}^{n}\left|\frac{y_i - x_i}{x_i}\right|}{n} \tag{4}$$

r—Pearson's correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

d—Willmott's index of agreement:

$$d = 1 - \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{\sum_{i=1}^{n} (|y_i - \bar{x}| + |x_i - \bar{x}|)^2} \quad (6)$$

d_1—modified Willmott's index of agreement:

$$d_1 = 1 - \frac{\sum_{i=1}^{n} |y_i - x_i|}{\sum_{i=1}^{n} (|y_i - \bar{x}| + |x_i - \bar{x}|)} \quad (7)$$

3. Results and discussion

Seven different measures of the fitting of predicted concentrations to observed ones were considered in the analysis. Modelling errors were calculated for the entire range of actual O_3 and CO concentrations. Then, similar calculations were made separately for specified sub ranges of observed concentrations of both pollutants, $O_{3,obs}$ and CO_{obs}. Error values were shown in Table 1 for O_3, and in Table 2 for CO, respectively.

The results indicate that the modelling error values may vary significantly, depending on the range of concentrations of both pollutants.

Measures based on differences between actual and predicted concentrations, such as MAE, MSE and RMSE, behave similarly. They reach the minimum values for the third O_3 concentration sub range (40–60 $\mu g/m^3$) and the third CO concentration sub range (300–400 $\mu g/m^3$), that is for the most numerous sub ranges. This result is understandable because these sub ranges are also the most represented in the learning process of both modelling networks. A neural network adaptation process is dominated by the cases contained in the most numerous concentration sub ranges. Moving to higher and higher concentration sub ranges the modelling accuracy decreases due to fewer training cases and more extreme ones. MSE is often treated as a measure of the model accuracy, since its formula is based on the sum of squared errors of the respective cases. This sum is usually assumed as the objective function which minimization runs during neural network training. RMSE is calculated as the square root of MSE. The advantage of RMSE use is its physical dimension ($\mu g/m^3$), consistent with the dimension of the modelled variable. Considered measures (MAE, MSE, RMSE) properly reflect the difficulties in modelling concentrations in the entire range of concentrations as well as in different sub ranges of concentrations.

The average relative error MARE is extremely high in the first sub range of actual concentrations for either O_3 or CO. In subsequent sub ranges, the value of MARE decreases rapidly. This error measure is unstable. Its value can be especially misleading in the case of very small concentrations of pollutants, close to 0 $\mu g/m^3$. For concentrations equal to 0 $\mu g/m^3$, the calculation error is not feasible. Since most of measuring concentrations of main air pollutants can reach values close to 0 $\mu g/m^3$, this measure cannot be recommended for assessing the accuracy of models.

Measures such as r, d, d_1, which refer to averages in the formulas, have much lower values in the specified intervals than in the entire concentration range. The results show that we should be careful in drawing conclusions regarding the accuracy of the models. Comparing the error value for the full range model

Table 1
Values of specified prediction errors calculated for different sub ranges and for the entire range of $O_{3,obs}$ concentrations (hourly data, Lodz-Widzew 2004–2010)

Ranges of $O_{3,obs}$ concentrations ($\mu g/m^3$)	Number of observations	MAE ($\mu g/m^3$)	MSE $(\mu g/m^3)^2$	RMSE ($\mu g/m^3$)	MARE	R	d	d_1
0–20	6,214	10.1	159	12.6	1.348	0.533	0.513	0.362
20–40	9,685	9.3	139	11.8	0.318	0.396	0.515	0.367
40–60	12,057	9.2	134	11.6	0.187	0.342	0.525	0.371
60–80	9,753	11.3	199	14.1	0.163	0.337	0.475	0.338
80–100	6,085	12.6	270	16.4	0.142	0.300	0.409	0.303
100–120	2,866	13.2	291	17.1	0.121	0.268	0.384	0.286
120–140	1,135	16.3	436	20.9	0.126	0.141	0.294	0.225
140–160	315	24.2	757	27.5	0.164	0.174	0.243	0.159
160–180	46	34.1	1,300	36.1	0.202	0.066	0.213	0.127
180–200	9	56.1	3,273	57.2	0.300	0.328	0.124	0.068
0–200	48,165	10.7	1,901	13.8	0.347	0.895	0.942	0.773

Table 2
Values of specified prediction errors calculated for different sub ranges and for the entire range of CO_{obs} concentrations (hourly data, Lodz-Widzew 2004–2010)

Ranges of CO_{obs} concentrations ($\mu g/m^3$)	Number of observations	MAE ($\mu g/m^3$)	MSE $(\mu g/m^3)^2$	RMSE ($\mu g/m^3$)	MARE	R	d	d_1
100–200	1,467	122.3	17,043	130.5	0.715	0.093	0.185	0.112
200–300	9,030	60.8	6,339	79.6	0.247	0.114	0.322	0.245
300–400	12,718	54.2	5,061	71.1	0.156	0.294	0.446	0.337
400–500	8,547	65.0	6,833	82.7	0.146	0.279	0.413	0.298
500–600	5,012	78.0	9,594	97.9	0.143	0.227	0.356	0.257
600–800	5,028	114.1	20,615	143.6	0.166	0.327	0.451	0.328
800–1,000	1,936	157.8	38,503	196.2	0.179	0.285	0.358	0.256
1,000–2,000	1,446	187.1	56,267	237.2	0.153	0.671	0.762	0.555
2,000–6,000	95	368.5	215,716	464.5	0.144	0.862	0.897	0.679
0–6,000	45,279	78.4	11,775	108.5	0.191	0.898	0.944	0.750

with the corresponding error values for sub ranges can lead to false conclusions. In any concentration sub ranges, the model does not achieve values of the coefficients r, d, d_1 obtained for the whole range model. The values of these measures for O_3 gradually decrease when moving to higher and higher concentrations of the sub ranges. However, in the case of CO, the values of r, d and d_1 may increase with the level of concentration sub ranges. Measures of this type can be used for a comparison of the accuracies of autonomous models created for different air pollutants provided that they refer to the entire range of concentrations.

4. Conclusions

Based on the results of the analysis the following conclusions have been formulated:

(1) To assess the quality of the prediction of air pollutant concentrations several different measures of error of approximation should be used. The use of a single measure may lead to incorrect interpretation.
(2) MAE, MSE, RMSE properly reflect the difficulties in modelling concentrations in the entire range of concentrations as well as in different sub ranges of concentrations.
(3) The average relative errors like MARE cannot be recommended for assessing the accuracies of autonomous models.
(4) Measurement errors depending on the average concentration values (r, d, d_1) show worse performance when the assessment is referred to a selected range of concentrations in comparison with the values obtained for the entire range of the modelled concentrations. Measures of this type can be used for comparing the accuracies of autonomous models created for different air pollutants provided that they refer to the entire range of concentrations.

Acknowledgements

This study was carried out within a research project of Czestochowa University of Technology BS-PB-402-301/11.

References

[1] A.E. Milionis, T.D. Davies, Regression and stochastic models for air pollution—I. Review, comments and suggestions, Atmos. Environ. 28 (1994) 2801–2810.
[2] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences, Atmos. Environ. 32 (1999) 2627–2636.
[3] S. Hoffman, Treating missing data at air monitoring stations, in: L. Pawłowski, M.R. Dudzińska, A. Pawłowskieds (Eds.), Environmental Engineering, Taylor & Francis Group, London, 2007, pp. 349–353.
[4] A. Venkatram, Computing and displaying model performance statistics, Atmos. Environ. 42 (2008) 6862–6868.
[5] A.M. Mouton, B. De Baets, P.L.M. Goethals, Ecological relevance of performance criteria for species distribution models, Ecol. Modell. 221 (2010) 1995–2002.
[6] C.J. Willmott, Some comments on the evaluation of model performance, Bull. Am. Meteorol. Soc. 63 (1982) 1309–1313.
[7] C.J. Willmott, S.G. Ackleson, R.E. Davis, J.J. Feddema, K.M. Klink, D.R. Legates, J. O'Donnell, C.M. Rowe, Statistics for the evaluation and comparison of models, J. Geophys. Res. 90 (1985) 8995–9005.
[8] C.J. Willmott, S.M. Robeson, K. Matsuura, A refined index of model performance, Int. J. Climatol. 32 (2012) 2088–2094.

[9] S.R. Dorling, M.W. Gardner, Statistical surface ozone models: An improved methodology to account for non-linear behaviour, Atmos. Environ. 34 (2000) 21–34.
[10] A. Karppinen, J. Kukkonen, T. Elolahde, M. Konttinen, T. Koskentalo, A modelling system for predicting urban air pollution: Comparison of model predictions with the data of an urban measurement network in Helsinki, Atmos. Environ. 34 (2000) 3735–3743.
[11] S.S.M. Nagendra, M. Khare, Modelling urban air quality using artificial neural network, Clean Technol. Environ. Policy 7 (2005) 116–126.
[12] A. Plaia, A.L. Bondi, Single imputation method of missing values in environmental pollution data sets, Atmos. Environ. 40 (2006) 7316–7330.
[13] W.-Z. Lu, W.-J. Wang, Potential assessment of the "support vector machine" method in forecasting ambient air pollutant trends, Chemosphere 59 (2005) 693–701.
[14] S. Hoffman, Application of Neural Networks in Regression Modelling of Air Pollution Concentrations, Wydawnictwa Politechniki Częstochowskiej, Częstochowa, 2004 (in Polish).