Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 142 (2017) 3524–3530

Energy Procedia

www.elsevier.com/locate/procedia

9th International Conference on Applied Energy, ICAE2017, 21-24 August 2017, Cardiff, UK

Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO_2 Pollution

Sheen Mclean S. Cabaneros[a,] *, John Kaiser S. Calautit[b], Ben Richard Hughes[a]

[a]*Department of Mechanical Engineering, University of Sheffield, Sheffield, S10 2TN, UK*
[b]*Department of Architecture and Built Environment, University of Nottingham, NG7 2RD, UK*

Abstract

Traffic-related air pollution has been a serious concern amongst policy-makers and the public due to its physiological and environmental impacts. An early warning system based on accurate forecasting tools must therefore be implemented to circumvent the adverse effects of exposure to major air pollutants. A multilayer perceptron neural network was trained and developed using air pollution and meteorological data over a two-year period from a monitoring site in Marylebone Road, Central London to predict roadside concentration values of NO_2 24 hours ahead. Several hybrid models were also developed by applying feature selection techniques such as stepwise regression, principal component analysis, and Classification and Regression Trees to the neural network model. Most roadside pollutant variables, e.g., oxides of nitrogen, were found to be significant in predicting NO_2. The statistical results reveal overall prediction superiority of the hybrid models to the standalone neural network model.

Peer-review under responsibility of the scientific committee of the 9th International Conference on Applied Energy.

Keywords: Artificial neural network; Multilayer Perceptron; Air pollution forecasting; NO_2; Artificial intelligence

1. Introduction and Literature Review

Traffic-related air pollution has been one of the major concerns amongst researchers and legislators due to its consequent impacts on human health and environment. In 2012, one-eighth of the total number of deaths worldwide was reportedly attributed to air pollution [1]. Air pollution poses a huge challenge in most metropolitan areas, such

* Corresponding author. Tel.: +447704-693-423
E-mail address:smscabaneros1@sheffield.ac.uk

1876-6102
Peer-review under responsibility of the scientific committee of the 9th International Conference on Applied Energy.
10.1016/j.egypro.2017.12.240

as Central London which recently breached the European Union (EU) legal limit of 40 μg/m^3 of nitrogen dioxide (NO_2) on average per year. A brown toxic gas resulting from road traffic emission, NO_2 has been linked to several respiratory illnesses, including asthma in children [2]. Immediate action must therefore be made to manage and minimise pollution-related mortalities in the future. While the installation of monitoring devices on pollution hotspots for mitigation and policy-making purposes is necessary, a prognostic approach is needed to take preventive actions when pollution concentrations exceed imposed limits. Hence, the development of tools for accurate short-term forecasting and control of air pollution is of paramount importance.

Air pollution models can be generally categorised into two types: namely, deterministic (or mechanistic) and statistical (or data-driven) approaches. Deterministic models rely on the mathematical representation of various physical transport and chemical reactions of pollutants, making them time-consuming and computationally expensive [3]. Because of these limitations, statistical models have become popular alternatives to the said approach due to their ability to establish a relationship between the predictors and the output variables without scrutinising all the physical parameters behind the transformation and dispersion processes of the pollutants.

Amongst these statistical models, artificial neural networks (ANNs) have often been employed in forecasting. ANNs are mathematical models that can mimic the learning processes of a biological brain. Compared to its conventional linear counterparts such as multiple linear regression (MLR) and partial-least squares regression (PLS) models, ANNs are capable of modelling complex and nonlinear relationships between input and output variables. Several forms of ANNs have been recently used in the field of atmospheric modelling, including the multilayer perceptron (MLP), radial basis function (RBF), and Elman neural networks [4]–[6].

Due to the black box nature of the modelling scheme of ANNs, neural networks are, however, unable to detail the underlying physical or chemical processes of air pollutants. This leaves the modellers the dilemma of choosing the more significant variables for a specific forecasting application, thus relying on a large set of available data. However, model complexity tends to increase as the number input dimensionality increases, resulting to poor model performance. Previous works highlighted the need to apply the process called feature selection to rectify this issue [7]-[10]. Feature selection identifies an optimal set of predictors to minimise model complexity and consequently improve the performance of any model.

ANN modelling tends to be site-specific, i.e., the developed ANN models are applicable only for the area where the data were obtained [11]. In the context of the successful implementation of several feature selection techniques in past case studies, this work will further investigate the effectiveness of the said technique in the development of forecasting tool that can estimate the hourly levels of NO_2 at a different location. This is achieved by developing several MLP models based on Classification and Regression Trees (CART), stepwise regression, and principal component analysis (PCA) techniques.

2. Methodology

The modelling methodology of the study is illustrated in Fig.1 (a). Data was first gathered and preprocessed before being fed to train a neural network model. Several variants of the model were then developed based on the various sets of model predictors identified by some feature selection techniques. Lastly, several metrics were used to measure and compare the performance of the models. The following sections shall describe the aforementioned procedures in details.

2.1 Data collection and preprocessing

The site selected in this case study is Marylebone Road, London, which is a busy road comprising of three lanes of traffic in each direction and carrying approximately 80,000 vehicles per weekday. Hourly air pollution and meteorological data from January 2008 to December 2009 were collected from two monitoring sites, namely, London Marylebone Road (LMR) (latitude/longitude: 51.522530, -0.154611; altitude: 35m) and London Bloomsbury (LB) (latitude/longitude: 51.522290, -0.125889; altitude: 20m), both belonging to Automatic Urban and Rural Network (AURN) of the UK. Fig. 1 (b) depicts the location and the urban layout of LMR station. LMR is a kerbside site while LB is an urban centre site in a small square surrounded by less busy roads 2 km to the east [12]. The latter accounts for the urban background for the former site. Table 1 shows a summary of collected hourly

pollutant data and meteorological values from both urban and background settings. The variable names of predictors that were collected from the background site are followed by the code 'bg'.

Variables with large magnitudes have the tendency to mask those with small ones which, in effect, would create some discrepancies in the model results [4]. Hence, all input data were initially normalised into values ranging from 0 to 1 based on the maximum and minimum of each variable. Moreover, time steps with at least a missing parameter value were excluded from the input in order to avoid any estimation error for the lacking data.

Fig. 1. (a) Flowchart of the modelling methodology; (b) London Marylebone Road monitoring site [13].

Table 1. Descriptive statistics of hourly air pollutant and meteorological data for the period January 2008 to December 2009

Variable	Units	Range	Mean	Std. Dev.
NO	ug/m^3	[1, 713]	135.70	106.55
NO_X	ug/m^3	[10, 1398]	322.33	215.47
CO	ug/m^3	[0.1, 2.9]	0.72	0.39
PM_{10}	ug/m^3	[0, 255]	35.11	17.39
$PM_{2.5}$	ug/m^3	[-4, 211]	21.46	10.85
O_3	ug/m^3	[0, 104]	14.52	14.79
NO_bg	ug/m^3	[0, 334]	21.41	32.62
NO_2_bg	ug/m^3	[8, 176]	51.05	22.30
NO_X_bg	ug/m^3	[10, 686]	83.62	68.26
(continued…)				
CO_bg	mg/m^3	[0.1, 1.6]	0.26	0.17
PM_{10}_bg	ug/m^3	[-2, 146]	18.60	11.55
$PM_{2.5}$_bg	ug/m^3	[-3, 71]	13.04	7.04
O_3_bg	ug/m^3	[0, 124]	28.59	20.05
Relative Humidity	%	[19, 94]	66.90	14.08
Temperature	ºC	[-2.3, 30.1]	14.14	5.61
Wind Direction	ºN	[0, 359]	-	-
Wind Speed	m/s	[0, 4.6]	0.61	0.48
Solar Radiation	W/m^2	[0, 802]	34.14	83.83

2.2 Multilayer Perceptron (MLP) network

2.2.1 MLP Architecture

A multilayer perceptron (MLP) neural network was selected due to its popularity to model highly nonlinear functions [9][14]. An MLP network is composed of interconnected neurons or nodes, namely, the input, output and hidden layers. The number of nodes in the input layer depends on the number of input variables, while the output layer consists of a single node resembling the target variable, e.g. NO_2 concentration. On the other hand, a network can have more than one hidden layer, each having multiple neurons. An approximation result by Kolmogorov was used to bypass this uncertainty [15]. Figure 2(a) illustrates an MLP with three layers. The network operates by feeding a set of values, i.e., x_1, x_2,…, x_N, through the nodes of the input layer, and transmitting the sum of their weighted values to the hidden layer through some nonlinear activation function. Mathematically, the process can be expressed as

$$\hat{x}_j = f\left(\sum_{i=1}^{N} w_{j,i}x_i + b_j\right),\ j = 1,2,\ldots,k \tag{1}$$

where k is the number of nodes in the hidden layer, N is the number of nodes in the input layer, $\hat{x}_j$ is the output value of node j, $w_{j,1}, w_{j,2}, \ldots, w_{j,N}$ and b_j are, respectively, the weighting and bias factors of node j in the hidden layer, and f is the logistic sigmoid activation function given by

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

where $e \approx 2.71828$. Given the weighting and bias factors w_j and b, the single node in the output layer computes its output y_t, the target at time t, in the same way:

$$y_t = f\left(\sum_{j=1}^{k} w_j \hat{x}_{j,t} + b\right). \tag{3}$$

2.2.2 Model training, validation and testing

Out of the 7566 collected temporal data points, 5296 points (70%) were allocated for training, while 1135 points (15%) were used for both validation and training sets. The partitioning was done randomly to ensure that every element of the subsets represents the entire dataset. The training set was used to determine the optimal values of the individual weights of the MLP network. The network was trained using the Levenberg-Marquardt backpropagation algorithm, which adjusts the values of weights between interconnecting neurons based on some error function of the model and target values to minimise the overall error [16]. The algorithm is based on a non-linear optimization method called the gradient descent. The simplified outline of the algorithm is depicted in Figure 2(b). The trained network was then applied to the validation and test data set to estimate the target output using the lagged values for 24 hours of the aforementioned pollutant and meteorological data. All of these methods were implemented in MATLAB R2017a software.

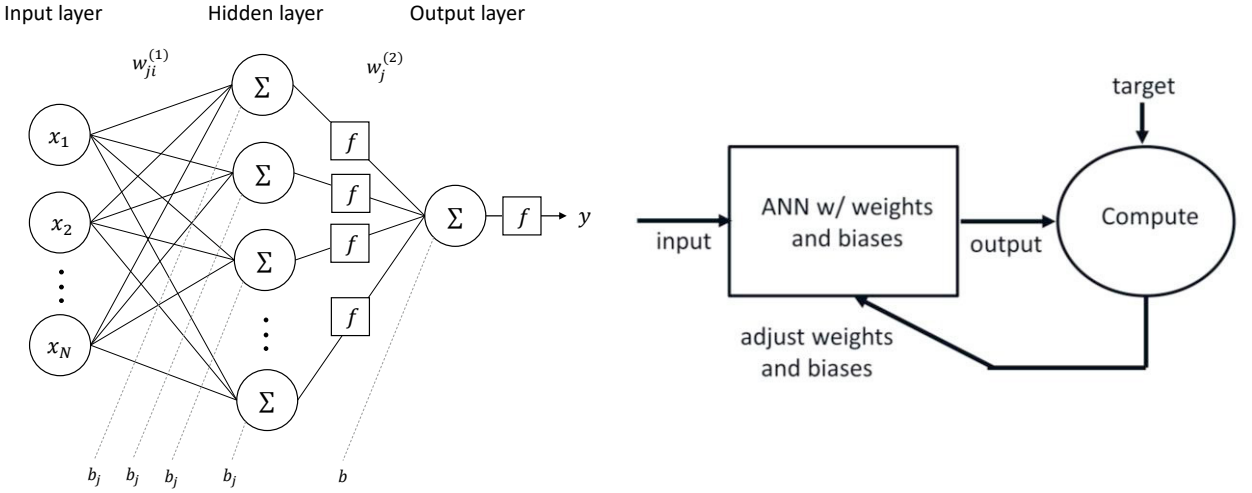

Fig. 2. (a) The MLP model architecture; (b) The backpropagation algorithm

2.3 Feature Selection

The three feature selection methods considered in the study are the following:

- Stepwise regression is a linear search strategy that is based on two known predictor selection techniques, namely, forward selection and backward elimination methods. Forward selection begins with no predictors in the model and subsequently adds the one that improves the model performance, e.g., having the highest correlation with the target output. Backward elimination commences with all inputs and iteratively eliminates one that provides the least increase in the squared error.
- PCA operates by transforming the original input space via singular value decomposition into a set of orthogonal vectors, called principal components (PCs). The PCs would represent the predictors that provide the maximum value of variation in the input space.

- CART is a machine learning technique capable of building regression and categorisation models based on several input data. CART works by examining each predictor and partitioning the set of inputs into two classes (or child nodes) based on the predictor value that maximises the variance between the two groups.

These methods are fully covered in [17][18]. The said techniques were then applied to the original dataset to generate new sets of input variables, thus creating three hybrid variants of the MLP model using such sets. All computations were implemented in MATLAB R2017a software.

2.4 The Performance Indicators

To assess the performance of the models, the following statistical descriptors were calculated:
a) the root mean squared error (RMSE) which serves as the error function during the network testing, given by

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_S}\left[\sum_{t=1}^{N_S}\left(\hat{y}_t - y_t\right)^2\right]} \tag{4}$$

b) the coefficient of determination (r^2) which describes the association between the model predicted and actual values, given by

$$r^2 = \left[\frac{1}{N_S - 1}\sum_{t=1}^{N_S}\left(\frac{\hat{y}_t - \hat{\mu}}{\hat{\sigma}}\right)\left(\frac{y_t - \mu}{\sigma}\right)\right]^2 \tag{5}$$

and, c) the fractional bias (FB) which measures the tendency of the model to over or under predict, given by

$$\mathrm{FB} = 2\left(\frac{\mu - \hat{\mu}}{\mu + \hat{\mu}}\right) \tag{6}$$

where $\hat{y}_t$ and y_t denote the t-th predicted and actual pollutant concentration values, respectively, N_S is the number of samples, and $\hat{\mu}$, μ, $\hat{\sigma}$ and σ are the overall mean and standard deviation of the predicted and the actual pollutant values, respectively.

3. Results and Discussion

The of predictors and number of hidden nodes of the developed models are summarised in Table 2. The MLP model that did not undergo feature selection is represented by Model 1.

The table reveals the success of the three techniques in reducing the complexity of the input space by removing almost more than half of the original set of input variables. The common variables that were retained in all the models are on-site measurements of NO and NO_x, indicating which variables are significant in the characterization of NO_2. The second most common predictors selected were temperature, hour of the day, the particulate matters, and CO. This finding seems to account for the close association between the said predictors and NO_2 in the atmosphere [19].

The number of hidden nodes, N_H, was determined by considering values from ($2N$+1) to 45, N being the number of predictors. The performance error of each model incorporating the various estimates of N_H were then recorded. The process was repeated three times to account for the random initialization of weights between neurons. The N_H associated with the model yielding the least average MSE results was considered the optimum value.

The performance metrics of each model are shown in Table 3, where the labels (1) and (0) indicate the ideal values for r^2 and FB, respectively. Overall, the hybrid models performed better than the stand-alone MLP model. The given RMSE values indicate that the feature selection techniques are effective in improving the forecasting accuracy of the MLP model. Based on the RMSE and r^2 values, Model 3 gave the best prediction results. On the other hand, the FB values suggest that Model 4 has the least tendency to overestimate or underestimate NO_2 measurements. Lastly, Figure 3 illustrates the predicted NO_2 values by Model 3 against the actual pollutant readings on the test set. The time series clearly agrees with the satisfactory overall performance of the model as it captured the actual hourly fluctuations of NO_2.

Table 2. List of MLP models with their respective predictors and number of nodes in the hidden layer

Model Name	Predictors	N_H
Model 1: MLP	All	39
Model 2: MLP + CART	Hour, NO, NO_x, PM_{10}, O_3_bg, NO_bg, Temperature	15
Model 3: MLP + PCA	Hour, Month, O_3, NO, NO_x, CO, PM_{10}, $PM_{2.5}$, NO_2_bg, Temperature, Wind Direction	23
Model 4: MLP + Stepwise regression	NO, NO_x, CO, $PM_{2.5}$, Wind Speed	11

Table 3. Model Performance Statistics

	RMSE (ugm-3)	r^2 (1)	FB (0)
Model 1: MLP	23.4584	0.9004	-0.0901
Model 2: MLP + CART	22.1656	0.9163	-0.0168
Model 3: MLP + PCA	22.0562	0.9209	-0.0432
Model 4: MLP + Stepwise regression	23.2929	0.9070	-0.0033

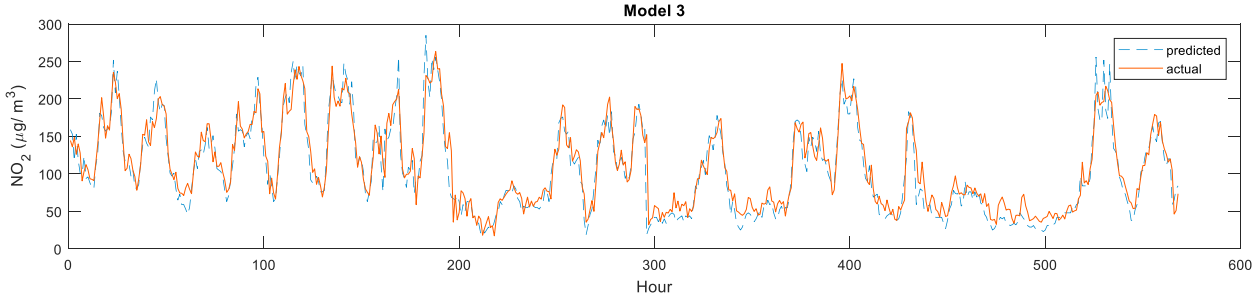

Fig.3. Predicted (Model 3) vs. actual NO_2 values

4. Conclusions

The study investigated the effect of implementing feature selection techniques on neural network models for the prediction of roadside NO_2 concentrations. Hybrid ANN models based on stepwise regression, principal component analysis, and Classification and Regression Trees were developed and tested.

Results showed that the most significant variables in predicting roadside NO_2 concentrations are roadside and background pollutant variables, e.g., oxides of nitrogen, and O_3. This is in accordance with the strong underlying chemical relationship between the said atmospheric species and NO_2 in ambient environment. Furthermore, a few meteorological variables were found to be vital in predicting NO_2. Since most of these variables are atmospherically related to each other, applying the input selection techniques would seem practical to avoid redundancy without losing much information of the input space. Other feature selection techniques can be employed to further investigate the most important pollutant and meteorological predictors of NO_2.

The study reveals the overall prediction superiority of the hybrid models to the plain neural network model. The results clearly demonstrate the ability of the said models to provide accurate NO_2 readings while employing only a few elements in the input space. In general, the current work reveals the potential of hybrid neural network models to be implemented in real-time forecasting of NO_2 concentrations in urban settings, and confirms the results of the past case studies that also dealt with ANN modelling. These prediction tools coupled with smart implementation of air quality management plans could minimise pollution-related mortalities in the future.

Acknowledgements

The authors like to thank the British Council and the Commission on Higher Education (CHED) of the Republic of the Philippines for funding this research.

References:

[1] World Health Organization, “Burden of disease from Household Air Pollution for 2012,” 2014. [Online]. Available: http://www.who.int/phe/health_topics/outdoorair/databases/FINAL_HAP_AAP_BoD_24March2014.pdf. [Accessed: 05-Aug-2017].
[2] Department of Environment Food and Rural Affairs, “AIR QUALITY EXPERT GROUP Nitrogen Dioxide in the United Kingdom,”

2004. [Online]. Available: www.defra.gov.uk/environment/airquality. [Accessed: 24-Apr-2017].
[3] A. L. Dutot, J. Rynkiewicz, F. E. Steiner, and J. Rude, "A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions," *Environ. Model. Softw.*, vol. 22, no. 9, pp. 1261–1269, 2007.
[4] N. Sharma, K. K. Chaudhry, and C. V. C. Rao, "Vehicular pollution modeling using artificial neural network technique : A review," vol. 64, no. September, pp. 637–647, 2005.
[5] S. Samarasinghe, "Neural Networks for Applied Sciences and Engineering," p. 15, 2006.
[6] H. T. Shahraiyni and S. Sodoudi, "Statistical modeling approaches for pm10 prediction in urban areas; A review of 21st-century studies," *Atmosphere (Basel).*, vol. 7, no. 2, pp. 10–13, 2016.
[7] R. M. Durao, M. T. Mendes, and M. Joao Pereira, "Forecasting O3 levels in industrial area surroundings up to 24??h in advance, combining classification trees and MLP models," *Atmos. Pollut. Res.*, vol. 7, no. 6, pp. 961–970, 2016.
[8] M. S. Alam and A. McNabola, "Exploring the modeling of spatiotemporal variations in ambient air pollution within the land use regression framework: Estimation of PM_{10} concentrations on a daily basis," *J. Air Waste Manage. Assoc.*, vol. 65, no. 5, pp. 628–640, 2015.
[9] A. Suleiman, M. R. Tight, and A. D. Quinn, "Hybrid Neural Networks and Boosted Regression Tree Models for Predicting Roadside Particulate Matter," *Environ. Model. Assess.*, vol. 21, no. 6, pp. 731–750, 2016.
[10] Y. Bai, Y. Li, X. Wang, J. Xie, and C. Li, "Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions," *Atmos. Pollut. Res.*, vol. 7, no. 3, pp. 557–566, 2016.
[11] S. Nejadkoorki, F., & Baroutian, "Forecasting extreme PM10 concentrations using artificial neural networks," *Int. J. Environ. Res*, vol. 6, no. 1, pp. 277–284, 2012.
[12] Department for Environment Food and Rural Affairs, "Marylebone Road, London - Non-Automatic Data - Defra, UK." [Online]. Available: https://uk-air.defra.gov.uk/data/maryleboneroad. [Accessed: 23-Apr-2017].
[13] "Google Maps." [Online]. Available: https://www.google.co.uk/maps/@53.3868266,-1.4831339,15z. [Accessed: 25-Apr-2017].
[14] M. . Gardner and S. . Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, no. 14–15, pp. 2627–2636, 1998.
[15] A. Kolmogorov, "On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition.," *Dokl. Akad. Nauk SSSR, 144, 679-681. Am. Math. Soc. Transl.*, no. 28, pp. 55–59, 1957.
[16] M. T. Hagan, H. B. Demuth, and M. H. Beale, "Neural Network Design," *Bost. Massachusetts PWS*, vol. 2, p. 734, 1995.
[17] L. Breiman, J. H Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees," *Chapman & Hall*, p. 12, 1984.
[18] I.T. Jolliffe, "Principal Component Analysis," *Springer Series in Statistics*, 2nd ed., 2002.
[19] J. Colls, *Air Pollution 2nd Edition*. 29 West 35th Street, New York, NY 10001: Spon Press, 2001.

Biography

The author is a current PhD researcher for the Energy 2050 group in the University of Sheffield, United Kingdom. His primary area of research has focused on machine learning techniques and environmental modelling. He enjoys hiking and swimming in his spare time.