

SHORT- AND LONG-TERM FORECASTING OF AMBIENT AIR POLLUTION LEVELS USING WAVELET-BASED NON-LINEAR AUTOREGRESSIVE ARTIFICIAL NEURAL NETWORKS WITH EXOGENOUS INPUTS

SHEEN MCLEAN CABANEROS¹, JOHN KAISER CALAUTIT², & BEN HUGHES¹

¹Department of Mechanical and Aerospace Engineering, University of Strathclyde, Glasgow, UK

²Department of Architecture and Built Environment, University of Nottingham, UK

ABSTRACT

Roadside air pollution is a major issue due to its adverse effects on human health and the environment. This highlights the need for parsimonious and robust forecasting tools that help vulnerable members of the public reduce their exposure to harmful air pollutants. Recent results in air pollution forecasting applications include the use of hybrid models based on non-linear autoregressive artificial neural networks (ANN) with exogenous multi-variable inputs (NARX) and wavelet decomposition techniques. However, attempts employing both methods into one hybrid modelling system have not been widely made. Hence, this work further investigates the utilisation of wavelet-based NARX-ANN models in the short- and long-term prediction of hourly NO₂ concentration levels. The models were trained using emissions and meteorological data collected from a busy roadside site in Central London, United Kingdom from January to December 2015. A discrete wavelet transformation technique was then implemented to address the highly variable characteristic of the collected NO₂ concentration data. Overall results exhibit the superiority of the wavelet-based NARX-ANN models improving the accuracy of the benchmark NARX-ANN model results by up to 6% in terms of explained variance. The proposed models also provide fairly accurate long-term forecasts, explaining 68–76% of the variance of actual NO₂ data. In conclusion, the findings of this study demonstrate the high potential of wavelet-based NARX-ANN models as alternative tools in short- and long-term forecasting of air pollutants in urban environments.

Keywords: *artificial neural networks, air pollution, air pollution forecasting, NARX, wavelet transform.*

1 INTRODUCTION

Roadside air pollution continues to attract special attention from both decision-making and scientific communities as it is being linked to premature mortalities and chronic illnesses among individuals residing in densely populated areas [1]–[3]. This highlights the need for air pollution monitoring and early-warning systems. Automated monitoring sites measure the concentrations of several key air pollutants and meteorological variables, contributing to the development of a time series database. Through the collected data, modelling tools are trained to model air pollution evolution and spatiotemporal trends. Air quality forecasts can assist legislators and urban city planners in making informed protection measures to manage air pollution and traffic [1], [2]. They also provide the public with early-warning updates that influence the daily behaviour of the public during potential peak pollution events.

Artificial neural networks (ANNs) are among the most popular black-box tools in air pollution forecasting applications [5]–[7]. Inspired by the information-processing mechanisms of biological neurons, ANNs have been shown to be robust tools capable of nonlinear mapping and self-adaptation [4]. However, ANN models have difficulties dealing with extreme levels of air pollutant concentrations [8], [9]. This can be explained by the limited continuous observations of the extreme pollutant levels leading to fewer representative training

data, e.g. the imbalance data problem, and the highly variable concentration levels at a local scale. Due to the data-driven nature of ANN models, the manner the inputs are represented has a direct influence on the performance of ANN models. [10], [11].

Wavelet transformation is a technique applied to decompose a given original function, e.g. an air quality time series, into several subseries with lesser variability. The utilisation of wavelets with ANNs in the context of air pollution forecasting has been proposed in recent years. For instance, a wavelet-based Support Vector Machine model was applied to predict CO levels at various locations in Warsaw, Poland. A wavelet transformation with an ensemble of ANN models was applied to predict daily average levels of PM_{10} in Warsaw, Poland [12]. A wavelet-feedforward ANN model was employed to predict hourly levels of O_3 at three urban sites in Oltenia, Romania [2]. Finally, wavelet transformation was applied with ANN models to predict daily PM_{10} levels at an urban site in Chongqing, China [13].

However, the results pertaining to the effectiveness of wavelet transformation in improving the performance of ANN models in air pollution forecasting are still limited. The effect of wavelet transformation on the performance of models should be further investigated to ensure that the said method is practical when implemented in rapid air quality forecasting schemes.

Therefore, this study presents the use of wavelet-based non-linear autoregressive ANN with exogenous multi-variable inputs (NARX) models, or NARX-ANN models, in forecasting roadside NO_2 levels. The proposed methodology is tested using the data collected from a busy street in Central London, United Kingdom. Furthermore, benchmark NARX-ANN models are employed to test the effectiveness of the proposed hybrid approach. The rest of the paper is structured as follows. The area description, data analysis and pre-processing methods, and the modelling techniques implemented are described in Section 2. The numerical results are presented in Section 3, while Section 4 concludes this paper.

2 MATERIALS AND METHODS

2.1 Area of study

The data was collected from an air quality monitoring station located in Marylebone Road, Central London, and was provided online by the Automatic Urban and Rural Network online resource [14]. The location was selected because it is of urban type and has experienced several threshold level exceedances in the past [15]. The Marylebone Road monitoring station is located next to a busy road comprising of three lanes of traffic in each direction and carrying approximately 80,000 vehicles per weekday. The cabin housing the air quality monitors is located on the south side of the road in a street canyon aligned on an axis of 75° to 255° in Central London.

2.2 Data collected

The collected data includes NO_2 , O_3 , PM_{10} , barometric pressure (BP), temperature (T), wind speed (WS) and wind direction (WD) measured from January to December 2015. The emissions variables were selected as they are highly correlated with NO_2 [16]. On the other hand, the meteorological variables were selected based on their availability, e.g. least number of

Table 1: Descriptive statistics of the collected emissions and meteorological variables collected in 2015 within the study area.

Variables	Mean	Min	Max	Standard deviation
Hourly NO ₂ level ($\mu\text{g}/\text{m}^3$)	88.5	10.1	290.4	40.6
Hourly of O ₃ level ($\mu\text{g}/\text{m}^3$)	15.1	0	69.8	12.4
Hourly of PM ₁₀ level ($\mu\text{g}/\text{m}^3$)	24.1	0	117.1	12.9
Hourly temperature (°C)	9.8	-6.7	29.6	5.5
Hourly wind speed (m/s)	3.6	0.1	12.1	1.7
Hourly wind direction (°)	200.6	0	359.8	95.8
Hourly barometric pressure (mbar)	1,012	972	1,035	9.01

missing values, with respect to the study period chosen. For missing data gaps less than or equal to 8 h, the average value of the succeeding and preceding six intervals of 1 h was used [17]. Otherwise, a slight modification of the hour mean method [18] was implemented, i.e. to replace the missing hourly value with the mean of all known hourly observations of the same season. Table 1 shows the main statistics describing the set of data collected within the study location after the imputation process.

Temporal details such as hour of the day (HoD) and month of the year (MoY) were also considered in this study to account for the cyclic nature of the air pollutant concentration levels. For model development purposes, the wind-related variables were transformed into two components, namely, $W_x = WS \cos(WD)$ and $W_y = -WS \sin(WD)$, to account for their cyclic characteristic and avoid sudden jumps of values. Similarly, the time-scale components were transformed into sinusoidal variables. For instance, HoD was split into two variables, namely, $HoD_x = \cos(2\pi d / D)$ and $W_x = WS \cos(WD)$, where d is the ordinal number referring to the day of the week, i.e. 1 corresponding to Sunday, and D is the total number of days. The variable MoY was pre-processed in a similar way.

2.3 Analysis of the collected data

As shown in Table 1, the average value of the collected hourly concentrations was high, e.g. $88.50 \mu\text{g}/\text{m}^3$. Additionally, the EU threshold of $200 \mu\text{g}/\text{m}^3$ was exceeded 60 times during 2015, which breaches the legal limit of only 18 times in a year. It is apparent that models capable of providing reliable short- and long-term forecasts are needed to facilitate urban traffic managers in providing interventions. As depicted in Fig. 1, the collected hourly NO₂ data exhibits high variability. To further quantify this observation, the ratio of the standard deviation and mean value of each air pollution time series, e.g. $(std. / \bar{x})$, as well as the signal-to-noise ratio (SNR), defined in decibels as $SNR = 20 \log(\bar{x} / std.)$, were calculated. The computed $(std. / \bar{x})$ ratio of the collected NO₂ concentration data is 0.82 which is a high value, while the SNR is 6.76. These characteristics highlight the difficult and complex nature of the prediction task. This

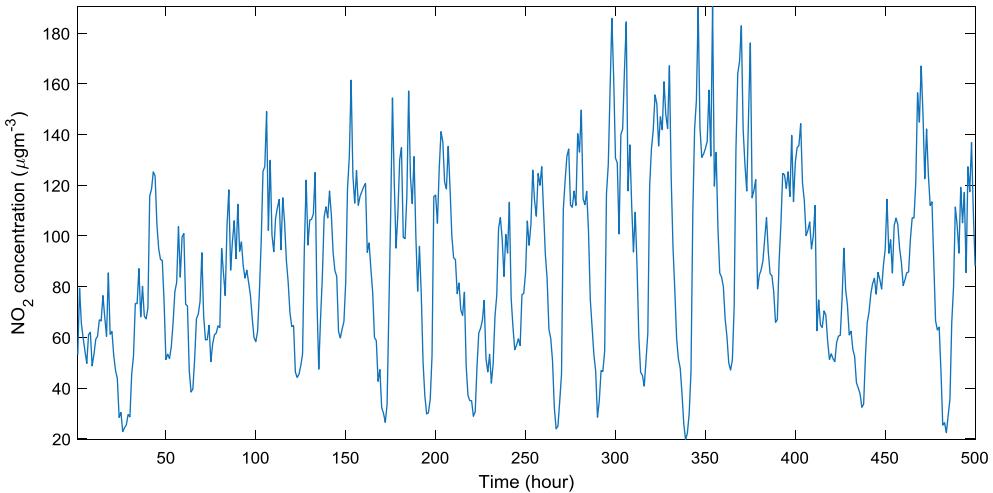


Figure 1: Plot of the first 500 samples of hourly NO_2 values collected at Marylebone Road, Central London from January to December 2015.

justifies the decision of the authors to consider a wavelet decomposition technique to address the said issue of variability.

The effect of lags on the autocorrelation function of the collected NO_2 data is shown in Fig. 2. It can be revealed that the autocorrelation scores degrade as the lag increases, revealing that the current hourly NO_2 concentration is highly dependent to its previous values. However, the behaviour of the autocorrelation scores across different lags of a multiple of 24 h indicates a cyclic pattern. This indicates the influence of seasonal parameters on the NO_2 concentration levels on an hourly scale. This also suggests that the lag analysis is needed to ensure optimal model performance.

2.4 Non-linear Autoregressive Neural Network with Multiple Exogenous Variables (NARX-ANN) model

NARX-ANN models are one of the most popular tools in nonlinear black-box modelling applications. A NARX-ANN model is described as a discrete time input-output recursive equation:

$$\hat{y}(t) = F \left[x(t-1), \dots, x(t-n_x), y(t-1), \dots, y(n_y) \right], \quad (1)$$

where both n_x and n_y denote the maximum lags of the exogenous and endogenous variables x and y , respectively, and $\hat{y}(t)$ denotes the one-step ahead prediction of the actual value $y(t)$. The function $F(\cdot)$ is represented by a feedforward ANN.

Feedforward ANNs implement a non-linear parametrised mapping from an input \mathbf{x} to an output \mathbf{y} ,

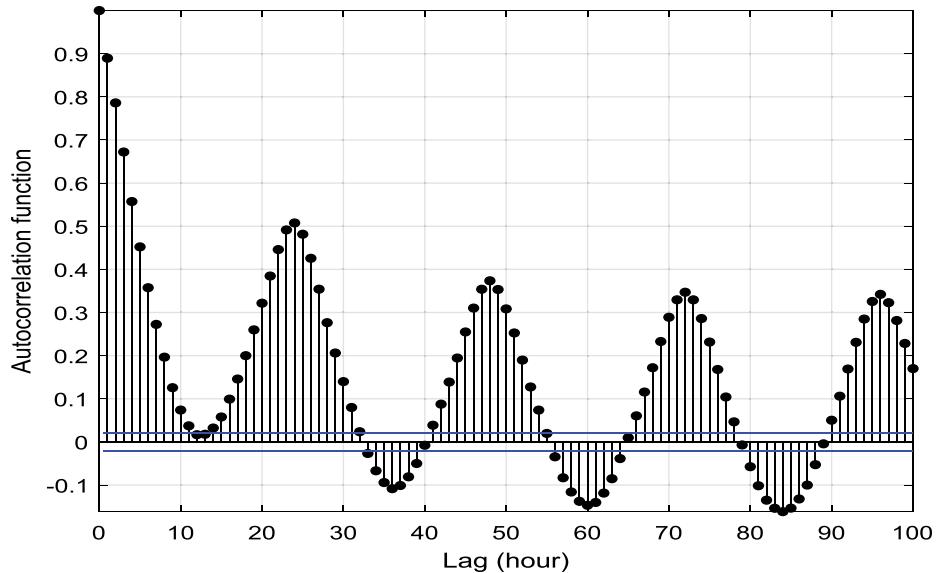


Figure 2: Autocorrelation function for NO_2 with maximum lag at 100 h.

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}), \quad (2)$$

where \mathbf{w} represents the weights and biases of the network, and f is a continuous real mapping, commonly referred to as the activation function. ANNs consist of single input and output layers, and one or more hidden layers, each of which has a varying number of interconnected neurons. The output of each node is scaled by the weights and fed forward to the nodes of succeeding layer:

$$y_i = f \left(\sum_{k=1}^N x_k w_{k,i} + b_i \right), \quad (3)$$

where N is the number of nodes of the preceding layer. ANNs are trained using a paired dataset $D = \{x^{(m)}, t^{(m)}\}$, where t is the target value, and $m \in [1, N_s]$, where N_s is the number of training samples, by adjusting \mathbf{w} so as to minimise difference between the input and target values. The training process is carried out repeatedly according to a gradient descent algorithm until a stopping criterion is met. A more detailed discussion about ANNs can be found in the literature (see [4], [11]).

2.5 Wavelet-based ANN model

Discrete wavelet transformation (DWT) is a technique that decomposes a given time series into subseries at various scales to reveal important feature characteristics and reduce randomness. In more detail, DWT transforms a given time series $s(t)$ into a finite summation of shifted wavelets at different scales according to the following expansion:

$$s(t) = \sum_j \sum_k c_{jk} \Psi(2^j t - k), \quad (4)$$

where c_{jk} is a set of wavelet coefficients, and $\Psi(2^j t - k)$ denotes the wavelet on j th scale shifted by k samples [19]. A J -level DWT decomposes a time series into detailed wavelet coefficients $D_j(t)$ of the proper time shifts t at various scales, $j = 1, 2, \dots, J$, and the approximation coefficient $A_J(k)$. The original time series can then be represented by the sum

$$s(t) = \sum_{i=1}^J D_i(t) + A_J(t). \quad (5)$$

The 4-level decomposition of the first one thousand hourly observations of the collected NO_2 data is shown in Fig. 3.

The idea behind this scheme is to let the NARX-ANN network estimate the values of D_i and A_J using the lagged values of the predictors. That is, at each scale n the neural network estimator, F_n , is implemented to forecast the t th wavelet coefficients of the n th scale based

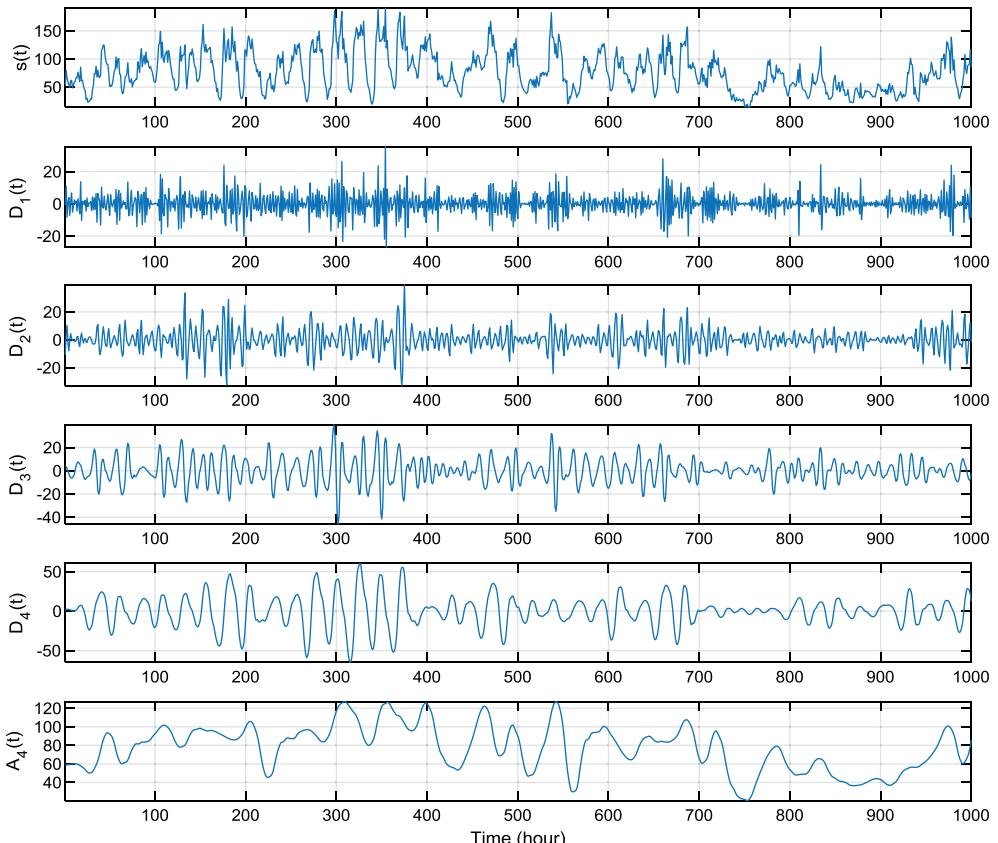


Figure 3: Wavelet decomposition of a subset of the collected NO_2 time series data, $s(t)$; $D_1 - D_4$ represent the detailed coefficients, and A_4 the coarse approximation of $s(t)$ on the fifth level.

on the set of lagged exogenous variables, M , and wavelet coefficients, as described in Eq. (6) and (7):

$$\hat{D}_n(t) = F_k(M_i(t-\Delta t), \dots, M_i(t-n\Delta t), D_n(t-\Delta t), \dots, D_n(t-m\Delta t)), \quad (6)$$

$$\hat{A}_J(t) = F_{k+1}(M_i(t-\Delta t), \dots, M_i(t-n\Delta t), D_n(t-\Delta t), \dots, D_n(t-m\Delta t)), \quad (7)$$

where $k=1, 2, \dots, J$; $i=1, 2, \dots, N_{pred}$; N_{pred} denotes the number of predictors utilised by the model; and m and n is the number of lags for the target and exogenous variables, respectively. Finally, the original form of the predicted NO_2 concentration can be retrieved using the following expression:

$$\hat{s}(t) = \sum_{i=1}^J \hat{D}_i(t) + \hat{A}_J(t). \quad (8)$$

In this paper, the Daubechies wavelets Db5 [20] were chosen to implement the decomposition process as this provided the lowest variability of the signals at each level after a series of trial-and-error procedure missing period. Additionally, the selection for the value of J is usually based on the ratio $\text{std}(A_j)/\text{std}(s)$, e.g. the standard deviation of A_j must be substantially smaller than that of $s(t)$. However, choosing a larger value of J also increases the number of terms in Eq. (5), thus accumulating more approximation errors when Eq. (6) and (7) are carried out via NARX-ANN [21]. As such, J was chosen to be 5. The general modelling scheme of the wavelet-based approach is outlined in Fig. 4. Various scales of the original time series of NO_2 levels were initially generated via DWT. The set of exogenous variables was then combined with the wavelet coefficients, e.g. D_i for $i=1, 2, 3, 4, 5$ and A_5 to form the final set of predictors which are then fed to $(J+1)$ NARX-ANN models. Lastly, the predicted values of wavelet coefficients at various time scales, e.g. \hat{D}_i and \hat{A}_J , were then reconstructed using Eq. (8).

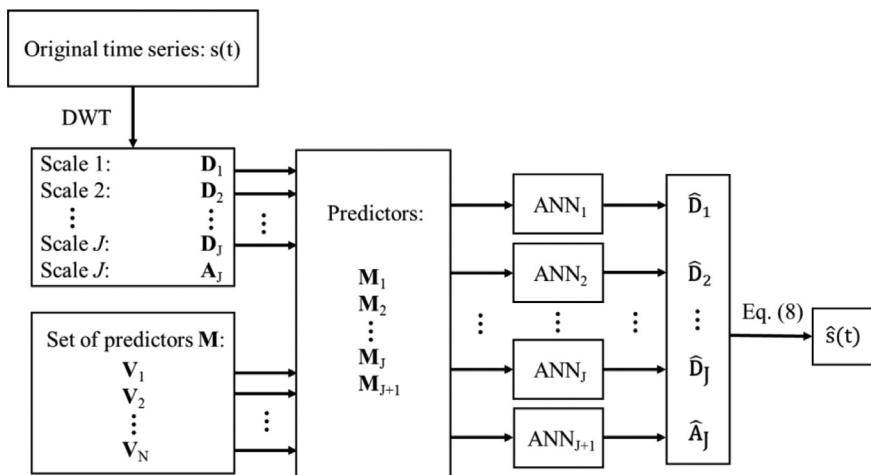


Figure 4: General modelling scheme of the wavelet-based NARX-ANN approach.

2.5 Model development

All collected data have been initially normalised using the max-min normalisation scheme to ensure that all values are in the same range [11]. The data were then split into three sets of which 70%, 15% and 15% of the total 8,760 hourly samples of each variable were allocated for the training, validation and testing sets, respectively. The partitioning was carried out randomly to ensure that every element of the subset of data represents the entire dataset. Additionally, the first four lags of the exogenous variables, e.g. $(x(t-1), x(t-2), x(t-3), x(t-4))$, and the endogenous variable, e.g. $(y(t-1), y(t-2), y(t-3), y(t-4))$, are defined as inputs, and $y(t)$ as the target, see Eq. (1).

Logistic sigmoid and linear activation functions were utilised in the hidden and output layers, respectively. Furthermore, only one hidden layer was employed in the network as it has been found to be sufficient in approximating any smooth measurable mapping between input and output variables [22]. The network weights and biases were initialised using the Nguyen-Widrow algorithm. Finally, the optimal number of hidden neurons was determined by a trial-and-error procedure which involves training multiple models across different prediction horizons and predictors via the Levenberg-Marquardt backpropagation algorithm. The process was repeated ten times to account for the sensitivity of the initial weights per run. The number of hidden neuron associated with the configuration that yielded the least average mean absolute error of the validation set was then selected.

In summary, two models were built in the study, namely, the plain NARX-ANN and the hybrid wavelet-based NARX-ANN approaches. Additionally, a variant of the models was trained for each prediction horizon, e.g. 1-h and 24-h ahead. Lastly, different subsets of the predictors were utilised to train the said models. Specifically, a model was developed using variables of only emissions, weather, time-scale and their combinations. To avoid bias, the models were run 100 times to account for the random initial values of the weights each time the model is run. The average of the results was then selected as the final outputs. Finally, the results of each model were assessed using root mean square error (RMSE), fractional bias (FB) and coefficient of determination (r^2) between the observed and predicted values. All algorithms were written and implemented in MATLAB R2018b software [23].

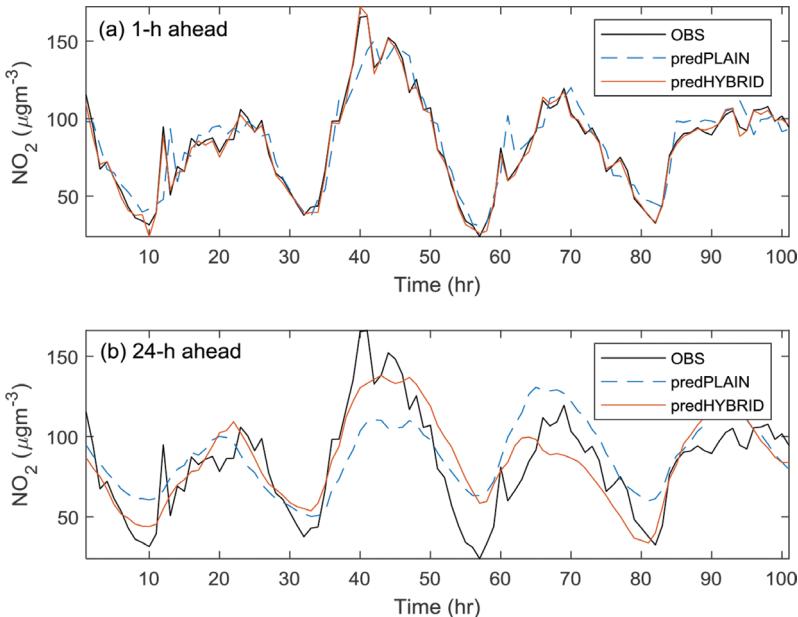
3 RESULTS AND DISCUSSION

Table 2 provides the performance of the developed models.

In the case of 1-h ahead forecasting of NO_2 , the best results were obtained by the wavelet-based NARX-ANN model trained by all predictors, with lowest RMSE score of $13.023 \mu\text{g}/\text{m}^3$. This highlights the importance of emissions, weather and time-scale predictors in approximating NO_2 using data-driven models. The model can also explain more than 95% of the variance of the actual NO_2 data, improving the performance of the plain NARX-ANN models by up to 6%. The wavelet-based models also have lesser tendencies in underestimating or overestimating the actual data in general. Fig. 5 shows the plots of the observed NO_2 observations for the last 100 h the test period (year 2015) and the predicted NO_2 values of the best performing plain and hybrid NARX models. It can be seen in Fig. 5(a) that the 1-h ahead estimates of the wavelet-based model coincide very well with the values of the actual data. This observation is in accordance with the distribution of forecasting error histograms shown in Fig. 6(a).

Table 2: The performance results of different models for hourly prediction of NO₂.

Exogenous Predictors	Δt	Model type					
		NARX-ANN			Wavelet-based NARX-ANN		
		RMSE	FB	r^2	RMSE	FB	r^2
O ₃ , PM ₁₀	1	18.592	0.0221	0.889	13.591	0.0089	0.951
T, BP, WS, WD	1	18.898	0.0298	0.890	14.044	0.0091	0.943
HoD, MoY	1	17.276	0.0189	0.907	16.100	-0.0109	0.932
All predictors	1	17.029	0.0102	0.913	13.023	0.0083	0.955
O ₃ , PM ₁₀	24	35.406	0.1056	0.568	32.351	0.1221	0.678
T, BP, WS, WD	24	36.322	0.1401	0.594	31.701	0.1032	0.680
HoD, MoY	24	31.771	0.1031	0.658	28.622	0.0970	0.733
All predictors	24	31.702	0.0624	0.675	28.043	0.0454	0.756

Figure 5: Comparison of the observed and predicted (a) 1-h ahead and (b) 24-h ahead values of NO₂ using NARX-ANN (predPLAIN) and wavelet-based NARX-ANN (predHYBRID) models.

However, it is evident that the models fail to accurately approximate extreme values. This may be explained by the relatively fewer available extreme values in the training data set [9]. It is also worth noting that the wavelet-based model trained only using O₃ and PM₁₀ data also provided significantly accurate results, suggesting that emissions data can be enough to train models in locations where meteorological variables are not available or missing. The worst performance was exhibited by the model trained only by monthly and hourly cycles. Among

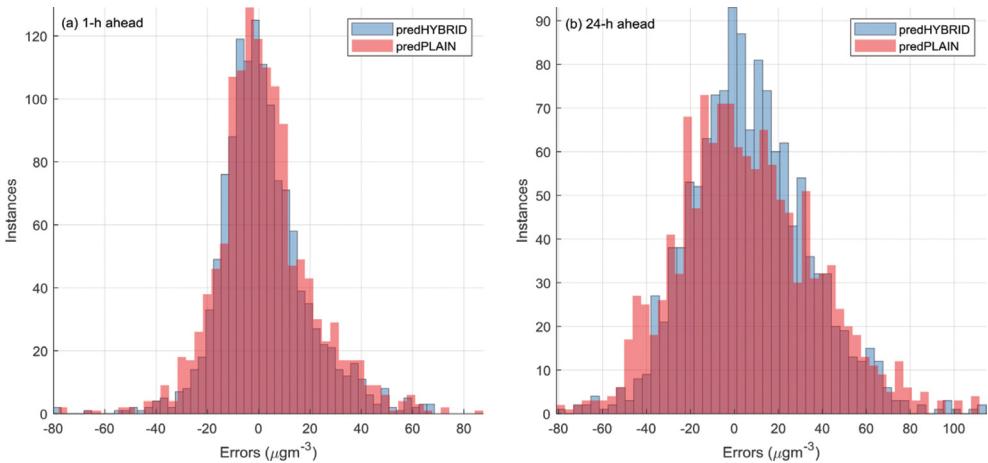


Figure 6: Error histograms of the (a) 1-h ahead and (b) 24-h ahead forecasting results using NARX-ANN (predPLAIN) and wavelet-based NARX-ANN (predHYBRID) models.

the benchmark models, the best results were obtained by the model that also included all predictors, while the worst one was achieved by only utilising emissions data.

In the case of 24-h ahead forecasting of NO_2 , the best results were still obtained by the wavelet-based model that utilised all predictors, reducing the RMSE and FB scores of its counterpart plain models by $3.7 \mu\text{g}/\text{m}^3$ and 0.017, respectively. Furthermore, it is apparent that the accuracy of all model results declined as the prediction horizon jumped from 1 to 24 h. This is consistent with the results of previous case studies and almost all forms of forecasting [24], [25]. Nonetheless, the results of the best performing wavelet-based model was able to explain approximately 76% of the variability of the observed data, which is 17% more accurate than those attained by the worst-performing benchmarks model. Consistent with quantitative results in Table 2, both plain and wavelet-based NARX-ANN models encountered difficulties in following the hourly trends of the actual NO_2 values 24-h in advance, see Fig. 5(b). It can also be depicted in Fig. 6(a) that both models suffered severe magnitudes of error for the 24-h prediction task. The models tend to underestimate or overestimate the extreme concentration levels in many occasions. Overall, the results indicate that the wavelet decomposition process can significantly improve the performance of the plain NARX-ANN models.

4 CONCLUSIONS

In this paper, hybrid models based on wavelet decomposition and NARX-ANNs were developed to provide 1-h and 24-h ahead forecasts of NO_2 concentration levels. The models were trained using hourly emissions, meteorological and time-scale variables collected from a busy roadside location in Central London. Plain NARX-ANN models were built to serve as benchmark models. The overall results highlight the superior performance of the wavelet-based NARX-ANN models when compared to those from the plain NARX-ANN models. The results of this study confirms the effectiveness of the wavelet decomposition method in reducing the variability of the NO_2 time series, thus improving the performance of the

benchmark models. The results also suggest that the inclusion of emissions and weather variables is beneficial in developing accurate data-driven models, although the models using only time-scale data as exogenous variables also yield results with a relatively acceptable level of accuracy. Additionally, the results reveal that the hybrid models can generate fairly accurate 24-h forecasts of NO₂. In conclusion, this study demonstrates the high potential of wavelet-based NARX-ANN models as robust and parsimonious tools for real-world air quality management and forecasting applications.

ACKNOWLEDGEMENTS

The authors would like to thank the British Council Philippines and the Commission on Higher Education (CHED) of the Republic of the Philippines (grant no. 261810845), for funding this research, and UK-AIR, DEFRA for making the air quality data available online.

REFERENCES

- [1] Baklanov, A. et al., Integrated systems for forecasting urban meteorology, air pollution and population exposure. *Atmospheric Chemistry and Physics*, **7**, pp. 855–874, 2007.
- [2] Dunea, D., Pohoata, A., & Iordache, S., Using wavelet-feedforward neural networks to improve air pollution forecasting in urban environments. *Environmental Monitoring and Assessment*, **187**(7), 2015.
- [3] Moustris, K. P., Ziomas, I. C. & Paliatsos, A. G., 3-day-ahead forecasting of regional pollution index for the pollutants NO₂, CO, SO₂, and O₃ using artificial neural networks in Athens, Greece. *Water Air Soil Pollution*, **209**(1–4), pp. 29–43, 2010.
- [4] Bishop, C. M., *Neural Networks for Pattern Recognition*. New York, NY: Oxford University Press, Inc., 1995.
- [5] Ordieres, J. B., Vergara, E. P., Capuz, R. S. & Salazar, R. E., Neural network prediction model for fine particulate matter (PM2.5) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environmental Model Software*, **20**(5), pp. 547–559, 2005.
- [6] Capilla, C., Multilayer perceptron and regression modelling to forecast hourly nitrogen dioxide concentrations. *WIT Transactions on Ecology and the Environment*, **183**, pp. 39–48, 2014.
- [7] Cabaneros, S. M. S., Calautit, J. K. S. & Hughes, B. R., Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO₂ Pollution. *Energy Procedia*, **142**, pp. 3524–3530, 2017.
- [8] Catalano, M., Galatioto, F., Bell, M., Namdeo, A. & Bergantino, A. S., Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environmental Science & Policy*, **60**, pp. 69–83, 2016.
- [9] Gong, B. & Ordieres-Meré, J., Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environmental Model Software*, **84**, pp. 290–303, 2016.
- [10] Li, C., Liang, M. & Wang, T., Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals. *Mechanical Systems and Signal Processing*, **64–65**, pp. 132–148, 2015.
- [11] Hagan, M. T., Demuth, H. B. & Beale, M. H., *Neural Network Design*, Vol. 2. PWS: Boston, MA, p. 734, 1995.
- [12] Siwek, K. & Osowski, S., Engineering Applications of artificial intelligence improving the accuracy of prediction of PM 10 pollution by the wavelet transformation and an en-

- semble of neural predictors. *Engineering Applications of Artificial Intelligence*, **25(6)**, pp. 1246–1258, 2012.
- [13] Bai, Y., Li, Y., Wang, X., Xie, J. & Li, C., Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, **7(3)**, pp. 557–566, 2016.
- [14] F. & R. A. (Defra) webmaster@defra.gsi.gov.u. Department for Environment. Automatic Urban and Rural Network (AURN) – Defra, UK. 2017.
- [15] The Guardian. London breaches annual air pollution limit for 2017 in just five days | Environment | The Guardian. 2017. [Online]. Available at <https://www.theguardian.com/environment/2017/jan/06/london-breaches-toxic-air-pollution-limit-for-2017-in-just-five-days>, (accessed 26 March 2019).
- [16] World Health Organization. Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide Report on a WHO Working Group Ozone-adverse effects Nitrogen Dioxide-adverse effects Air Pollutants, Environmental-adverse effects Meta-analysis Air-standards Guidelines, 2003.
- [17] Santos, P. J. Martins, A. G. & Pires, A. J., Designing the input vector to ANN-based models for short-term load forecast in electricity distribution systems. *Electrical Power & Energy Systems*, **29**, pp. 338–347, 2007.
- [18] Li, K. H., Le, N. H. U. D., Sun, L. I. & Zidek, J. V. Spatial-temporal models for ambient hourly PM 10 in Vancouver. *Environmetrics*, **338**, 1999.
- [19] Mallat, S. G., A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11(7)**, pp. 674–693, 1989.
- [20] Daubechies, I., *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM Press, 1988.
- [21] Osowski, S. & Garanty, K., Forecasting of the daily meteorological pollution using wavelets and support vector machine. *Engineering Applications of Artificial Intelligence*, **20(6)**, pp. 745–755, 2007.
- [22] Hornik, K., Stinchcombe, M. & White, H., Multilayer feedforward networks are universal approximators. *Neural Networks*, **2(5)**, pp. 359–366, 1989.
- [23] The MathWorks, I., MATLAB Documentation – MathWorks United Kingdom. [Online]. Available at <https://uk.mathworks.com/help/matlab/index.html> (accessed 27 February 2019).
- [24] Turias, I. J., González, F. J., Martin, M. L. & Galindo, P. L., Prediction models of CO, SPM and SO₂ concentrations in the Campo de Gibraltar Region, Spain: a multiple comparison strategy. *Environmental Monitoring and Assessment*, **143(1–3)**, pp. 131–146, 2008.
- [25] Solaiman, T. A., Coulibaly, P., & Kanaroglou, P., Ground-level ozone forecasting using data-driven methods. *Air Quality, Atmosphere and Health*, **1(4)**, pp. 179–193, 2008.