



Improving the prediction of air pollution peak episodes generated by urban transport networks



Mario Catalano^{a,*}, Fabio Galatioto^{b,1}, Margaret Bell^b, Anil Namdeo^b, Angela S. Bergantino^c

^a Department of Civil, Environmental, Aerospace, Materials Engineering (DICAM), University of Palermo, Italy

^b Transport Operations Research Group (TORG), School of Civil Engineering and Geosciences, Newcastle University, United Kingdom

^c Department of Economics, Management and Law (DISAG), University of Bari, Italy

ARTICLE INFO

Article history:

Received 30 July 2015

Received in revised form 19 March 2016

Accepted 20 March 2016

Available online 26 March 2016

Keywords:

Air quality forecasting

Exceedances of pollutant concentration limits

Nitrogen dioxide

Artificial neural network

ARIMAX model

Ensemble techniques

ABSTRACT

This paper illustrates the early results of ongoing research developing novel methods to analyse and simulate the relationship between transport-related air pollutant concentrations and easily accessible explanatory variables. The final scope is to integrate the new models in traditional traffic management support systems for a sustainable mobility of road vehicles in urban areas.

This first stage concerns the relationship between the hourly mean concentration of nitrogen dioxide (NO₂) and explanatory factors reflecting the NO₂ mean level one hour back, along with traffic and weather conditions. Particular attention is given to the prediction of pollution peaks, defined as exceedances of normative concentration limits. Two model frameworks are explored: the Artificial Neural Network approach and the ARIMAX model. Furthermore, the benefit of a synergic use of both models for air quality forecasting is investigated.

The analysis of findings points out that the prediction of extreme concentrations is best performed by integrating the two models into an ensemble. The neural network is outperformed by the ARIMAX model in foreseeing peaks, but gives a more realistic representation of the concentration's dependency upon wind characteristics. So, the Neural Network can be exploited to highlight the involved functional forms and improve the ARIMAX model specification. In the end, the study shows that the ability to forecast exceedances of legal pollution limits can be enhanced by requiring traffic management actions when the predicted concentration exceeds a lower threshold than the normative one.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Research context and problem definition

Air pollution in urban areas is mainly due to the intense use of motorized transport for travelling, in particular private cars and heavy goods vehicles. This is a priority issue for transportation planners and public authorities, given the harmful effects of pollution to human health and the environment (Migliore et al., 2012; Bergantino et al., 2013).

Numerous studies (Heinrich et al., 2005; Zhang et al., 2012a) argue that acute exposure to air pollutants may cause serious health concerns such as eye irritation, breathing difficulty, cardiovascular problems, while chronic exposure may lead to damage to the body's immune, neurological, reproductive and respiratory

systems, cancer and even premature death. In November 2014, the British Committee on the Medical Effects of Air Pollutants reported that air pollution may be responsible for as many as 60,000 early deaths in Britain each year. This follows the study presented in 2010 by the World Health Organization (WHO) Regional Office for Europe and the Organization for Economic Co-operation and Development (OECD), covering the whole European region, including non-EU states such as Norway and Switzerland. For this macro-area, 600,000 premature deaths each year are estimated as a consequence of air pollution from small particles (produced by the exhausts of diesel vehicles) and nitrogen dioxide, NO₂ (WHO, 2010). Also the environment is affected in terms of global climate change and adverse effects for plants and eco-systems (Seinfeld and Pandis, 2006; Zhang et al., 2012a).

Various national contexts throughout the world have issued guidelines and regulations to protect human health and the environment. The United States Environmental Protection Agency (EPA) has set national air quality standards for six pollutants:

* Corresponding author.

E-mail addresses: mariocat73@gmail.com (M. Catalano), fabio.galatioto@ts.catapult.org.uk (F. Galatioto).

¹ Transport Systems Catapult, UK.

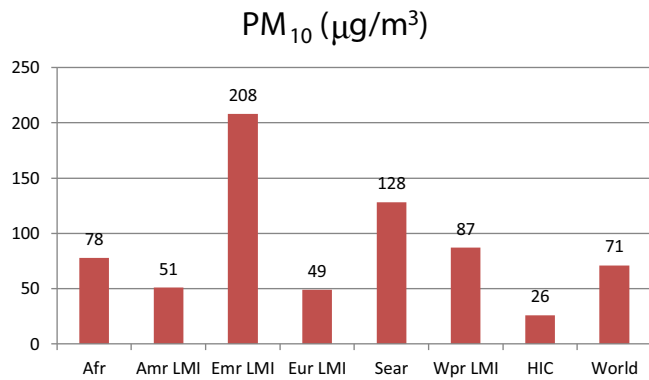


Fig. 1. PM₁₀ levels by region, for the last available year in the period 2008–2012 (WHO, 2014). Afr: Africa, Amr: America, Emr: Eastern Mediterranean, LMI: Low and middle-income, Sear: South-East Asia, Wpr: Western Pacific, HIC: high-income countries.

sulphur dioxide (SO₂), NO₂, carbon monoxide (CO), ozone (O₃), lead (Pb) and particulate matter (Seinfeld and Pandis, 2006). In Europe, over the last decades, the European Union (EU) has adopted a range of environmental measures to improve the quality of life for the Community's citizens. The final step of this legislative process is the Directive 2008/50/EC (EU, 2008), which has integrated an extensive body of laws establishing health-based concentration standards for a number of pollutants in outdoor ambient air. The European Commission has the task of ensuring that the environmental law is applied by the Member States through infringement procedures.

Long-term measures like mode switch policies in favour of mass transit and public regulation on road use are pretty effective in abating atmospheric pollution in cities (Allen et al., 2011), but pollution peaks and the consequent exceedance of regulative concentration thresholds are often caused by substantial fluctuations of mobility patterns and weather conditions around their expected behaviours. Hence, air quality protection needs to be fine-tuned through the introduction, in the local policy portfolio, of further tools and actions for predicting extreme pollution events and managing traffic in real time, in order to prevent the predicted concentration peaks.

Given the above, this research focuses on the investigation of traffic-related air pollution with the final aim of developing enhanced mathematical models to support short-term decisions for sustainable mobility of road vehicles in urban areas. In more detail, the study addresses the challenge of forecasting accurately the concentration of NO₂, which is subject to an hourly standard of concentration, to enable local authorities to mitigate or even prevent exceedances of concentration limits through real-time traffic management.

By achieving the identified objective, this research will be of strategic importance in many national contexts. The worldwide scale of atmospheric pollution problems has been acknowledged in the 2014 version of the WHO Ambient Air Pollution database consisting mainly of urban air quality data (WHO, 2014). In the report, annual concentration means of PM₁₀ and PM_{2.5},² for about 1600 cities of 91 countries in the 2008–2013 period, have been calculated. As can be seen in Fig. 1, the world's annual mean levels of PM₁₀ by region range from 26 to 208 µg/m³; in addition, the world's average is 71 µg/m³ against the recommended value of 20 µg/m³. Particular concern is associated to the East side of the planet, where countries like China, India, Nepal, Bangladesh, Mongolia and, in the Mediterranean Area, Egypt, Iran, Jordan,

Afghanistan, Pakistan far exceed the world's yearly mean concentration of PM₁₀.

The 2014 Air Quality in Europe Report (European Environment Agency, 2014) states that, in EU cities, exposure to atmospheric pollution levels exceeding the WHO concentration limits (in general stricter than the EU standards) is significantly widespread for various chemical agents. In 2012, above limit exposure to PM₁₀ and PM_{2.5} respectively involved 64% and 92% of the total EU-28 urban population. Moreover, in the case of O₃, in the same year, the exposure incidence rose to 98% of people living in towns. Despite the clear decreasing trend of NO₂ yearly mean concentration over the recent years,³ in 2011, 21 European countries still recorded exceedances of the limit values at one or more monitoring stations. Specifically, in the United Kingdom, the NO₂ levels have exceeded the WHO and EU target values persistently. This is confirmed by the fact that, in the early part of 2014, the European Commission launched legal proceedings against the UK for its failure to cut excessive levels of NO₂ (EU Press Release Database, 2014). Lastly, while exposure of Europeans to CO concentrations above the EU and WHO thresholds is negligible, in the case of benzene (C₆ H₆), around 10% of the EU-28 urban population is subject to pollution above the WHO levels and, in the case of SO₂, the value is 37%.

This paper presents the early stage of ongoing research on air quality modelling, which refers to NO₂, a toxic gas emitted by road vehicles, industry and households, which, even in the case of short-term exposures (from 30 min to 24 h), may irritate the eyes, nose, throat and lungs, while, in the long-term, may affect lung function permanently. Furthermore, it is the main precursor for ground-level ozone, that is very harmful to human health.

For NO₂, the EU environmental legislation (EU, 2008) sets two types of standard: the hourly mean concentration which cannot go beyond the level of 200 µg/m³ more than 18 times each calendar year; whilst, the annual average of hourly concentrations is not allowed to exceed 40 µg/m³. Moreover, the 2008 Air Quality Directive also defines an 'alert' threshold value of 400 µg/m³. When this threshold is exceeded over three consecutive hours in areas of at least 100 km² or an entire air quality management zone (whichever is the smaller), authorities have to implement short-term action plans.

In the research presented here, the relationship has been modelled to explore NO₂ hourly concentration in terms of explanatory variables which relate to urban transport and influence emissions, along with weather conditions, that are responsible for dispersion and transformation of pollutants.

2. Review of literature and research gaps

Few studies on real-time air quality forecasting near urban roads have appeared in the relevant scientific literature. Amongst some that are particularly interesting for this work, since they investigate the relationship between nitrogen oxides' levels and meteorological as well as transport-related variables, include Kukkonen et al. (2003); Cai et al. (2009); Galatioto and Bell (2013); Nagendra and Khare (2006); Perez and Trier (2001); Viotti et al. (2002). The leitmotiv of these studies is to consider the Artificial Neural Network (ANN), from the domain of Artificial Intelligence science, found to be the most effective tool to predict air quality in urban areas. In some cases, this methodology is compared with other approaches, but these are usually linear regression models or deterministic methods simulating the physical processes involved.

³ Between 2003 and 2012, in EU-28, the NO₂ annual mean concentration on average fell by 18%.

² Particles with diameter smaller than 10 and 2.5 microns, respectively.

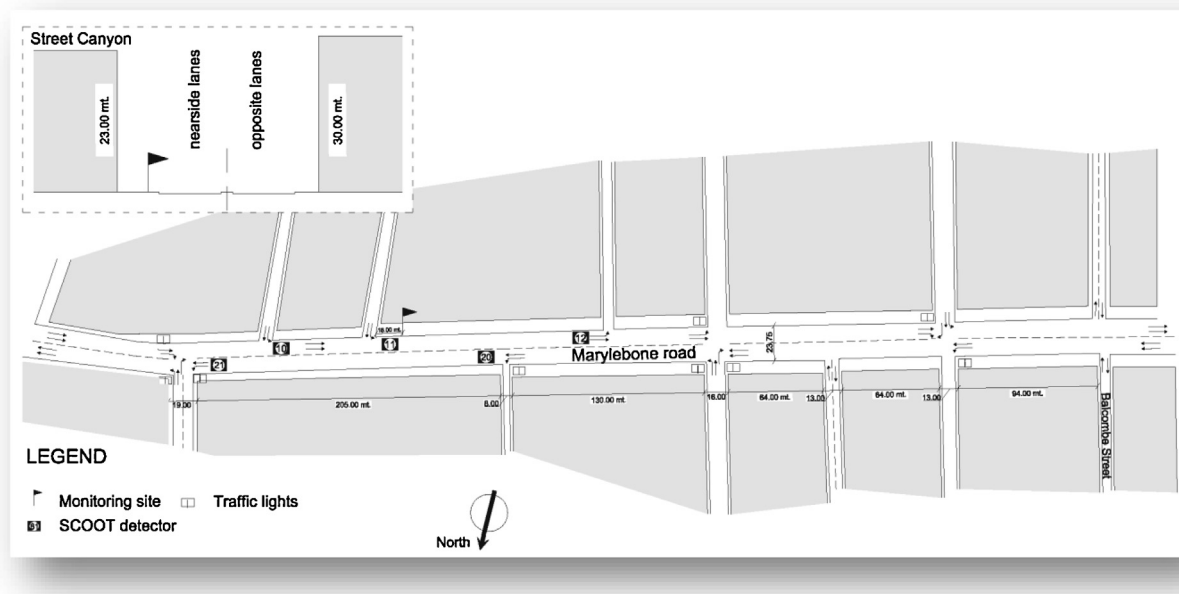


Fig. 2. The study area, Marylebone road in London, indicating the location of the air quality monitoring site.

Further statistical techniques have been adopted in relation to other pollutants. For example, Baur et al. (2004) compared the performance of quantile regression with multiple linear regression for predicting ozone concentrations. Kaushik and Melwani (2007) adopted the Seasonal Autoregressive Integrated Moving Average (ARIMA) model to forecast the daily levels of sulphur dioxide, nitrogen dioxide and suspended particles. Sayegh et al. (2014) evaluated different alternatives for the prediction of particulate matter concentration, namely multiple linear regression, quantile regression, generalised additive models and regression trees.

A promising area of research concerns the integration of the statistical approach into a fuzzy logic framework to handle uncertainty in environmental decision-making, as argued by Fisher (2006) for air quality forecasting.

For short-term predictions of pollution concentration, statistical methods are traditionally more suitable than deterministic models (physically-based), which are computationally intensive. Notwithstanding, one of the most important and current limitations of the statistical methods, including the ANN, is that, even when their overall performance is very good, they cannot accurately estimate concentration peaks (Zhang et al., 2012a; Galatioto and Zito, 2009; Zito et al., 2008). A solution to this problem may derive from ensemble forecasting, a multi-model approach integrating the predictions of diverse models,⁴ that is usually employed to reduce estimation bias without increasing sample variance significantly (Zhang et al., 2012b). The use of this technique to improve the accuracy of extreme pollution events has not been comprehensively investigated yet.

With respect to the research gaps motivating this research, it is worth stressing that few previous studies have compared and integrated the popular ANN with other complex statistical models, especially for the prediction of severe pollution episodes caused by urban transport. The work of Nunnari et al. (1998) tackled the problem of predicting concentrations of O₃ and NO_x in a zone with a high density of industrial plants, the Priolo-Melilli-Augusta petrochemical area in Sicily, showing that the ANN outperforms

the ARIMA model overall. Furthermore, Prybutok et al. (2000) worked on forecasting daily maximum O₃ concentration at Houston and showed that an ANN was more accurate than the ARIMA and linear regression models at forecasting extreme values concentration, mainly because the data presented clear non-linear patterns. Díaz-Robles et al. (2008) compared and combined the ARIMA and ANN models to improve the forecast accuracy of PM₁₀ concentration, with particular regard to alert episodes, for an area in Chile, where residential wood burning is a major pollution source during cold winters. Sánchez et al. (2013) used a time series of SO₂ concentrations, registered in a control station in the vicinity of a coal-fired power station in northern Spain, to develop and compare neural networks, ARIMA models and a hybrid method combining both which proved the best at predicting concentration peaks.

As suggested by the review of literature, further empirical evidence is needed to quantify the benefit of the proposed multi-model (ensemble) approach for extreme concentration forecasting, especially if the main source of pollution is road traffic. Therefore, in an attempt to address the identified research gaps, this research explores the popular neural network and compares it with a sophisticated parametric method, namely the Auto-Regressive Integrated Moving Averages with eXogenous inputs (ARIMAX) model (Hamilton, 1994). This is based on a set of indicators for missed exceedances and false alarms, in addition to the traditional metrics to measure the difference and the covariance between observed and simulated concentrations.

The ensemble approach has shown promise in its successful application to air quality modelling of monitored data (Bell et al., 2015; Díaz-Robles et al., 2008; Sánchez et al., 2013; Zhang et al., 2012b) and in other fields such as astronomy, astrophysics, early diagnosis of certain diseases, etc. (see Re and Valentini, 2012 and VV. AA., 2008). Hence, the effectiveness of this approach for predicting exceedances of the NO₂ hourly concentration threshold set by the EU is assessed.

3. Case study

This section provides details of the study area including geometric characteristics and the technologies used to collect

⁴ For example, taking a weighted average or other function of the different models' predictions.

Table 1

Descriptive statistics for the dataset of air quality, transport and weather collected in 2007 within the study area.

Variables	Observations	Average	Min	Max	Standard Deviation
Hourly mean concentration of NO ₂ (µg/m ³)	8584	102.5	4	329	53.1
Hourly traffic volume ^a (pcu/hour)	8734	3505.7	303.7	5161.5	1173.2
Hourly mean wind speed (km/h)	7993	2.8	0.1	15.1	2.1
Hourly mean wind direction (degrees North)	7993	203.9	2.9	358.1	82.4
Hourly mean temperature (degrees centigrade)	7993	13.5	−0.8	29.7	5.2

^a The original data on traffic were disaggregated by six vehicle classes: Motorcycle; Car or Light Van (length < 5.2 m); Car and trailer; Rigid Lorry, Heavy Van (length ≥ 5.2 m) or mini-bus; Articulated Lorry; Bus or Coach. The measured flows for each category in passenger car units were obtained by following scaling factors (Lavecchia et al., 2007): 1 for motorcycles, cars and light vans; 1.5 for light vehicles including mini-buses (5 m < length < 7.5 m); 2 for Rigid lorries, buses and heavy vans (7.5 m < length < 12.5 m) and 3 for articulated lorries (length > 12.5 m).

air quality and traffic data. The descriptive statistics and time series analysis techniques employed are presented along with results.

3.1. Air quality monitoring site

The study area is Marylebone road in the City of London (see Fig. 2). Marylebone Road has three lanes each way, with the nearside lanes in both directions reserved for buses and taxis. The traffic along the corridor is controlled by a demand responsive signal control system, namely SCOOT, Split Cycle Offset Optimization Technique (Hunt et al., 1981). The cabin housing precision air quality monitors is located on the south side of the road, at a point where the road is characterised by a canyon height to Width (H/W) ratio of 0.86⁵ (see inset in Fig. 2). The reach dataset available over a ten-year period (1998–2007) contains traffic (flow and speed for each lane), meteorological conditions and air quality data at an hourly resolution.

3.2. Descriptive analysis of the dataset

The analysis seeks to establish the relationship between the hourly mean concentration of NO₂ and explanatory factors related to traffic and meteorological conditions in a central corridor of London, Marylebone Road, throughout the year 2007. Table 1 shows the main statistics describing the set of data collected in 2007 within the study-area. The average of all hourly NO₂ concentrations was high (102.5 µg/m³); furthermore, the 200 µg/m³ EU threshold was exceeded 457 times during 2007, against the legal limit of 18 times in a year. Such conditions may arise due to two main reasons:

- the high hourly traffic volumes which during the period of 7:00am to 9:00pm are more than 4000 passenger car units (pcu) per hour (about 76% of the road capacity) for 72% of time;
- the quasi-canyon at the monitoring site with an average building H/W ratio of around 0.9.⁶

This severe NO₂ pollution problem requires long-term policy to realise sustainable transport as well as effective air quality prediction models to support real-time traffic management interventions in the short-term.

Fig. 3 shows the relationships between the hourly concentration (µg/m³) of NO₂ at Marylebone road in 2007 and the explanatory variables taken into account: the hourly traffic volume (pcu/hour) and the hourly averages of wind speed (km/h), wind direction (degrees North) and temperature (degrees centigrade). In detail, it represents the relationship between the dependent variable and each independent attribute after controlling for the other explanatory factors. The various scatter graphs are based upon a

preliminary regression of NO₂ concentration on the above transport and meteorological conditions related variables and, more precisely, plot each regressor against the relative partial residual, which is obtained by excluding the effect of the regressor itself from the estimated general regression residual (Mallows, 1986). The traffic flow has a positive impact on concentration, while wind speed is inversely related to pollutant concentration; moreover, the relationship between pollution and wind direction is seen to be complex and needs further investigation. On the other hand, the temperature effects appear small and, in the case of wind variables specifically, the influence on concentration appears to follow a non-linear behaviour.

3.3. Time series analysis

The time series of the considered variables have been analysed to check their stationarity, a statistical property of the series-generating stochastic process which requires its mean and autocovariance to be date-independent. It is a regularity condition which enables the development of a statistical model to explain specific variable with its previous values and a set of exogenous predictors (Hamilton, 1994). Whether this property holds in our case-study has been checked through graphical inspections and formal statistical tests.

As Fig. 4 displays, the time mean and variance of the variables of interest are quite stable through the year, except for the hourly averages of temperature and vehicle speed. Temperature, consistent with summer/winter cycles, shows an increasing trend in the first half of the year and a decreasing one in the second half. In the vehicle speed case, there is a turning point in the time series behaviour in the early part of the year. This is consistent with a change in the speed limit.

As such the non-stationary input variables speed and temperature cannot be used to explain NO₂ hourly concentration, which is stationary.

The stationary nature of NO₂ concentration, traffic volume, wind speed and direction, clearly illustrated in their respective time series plots, has been confirmed by the Augmented Dickey-Fuller (ADF) unit root test (Hamilton, 1994). On one hand, under the most general null hypothesis, this test assumes that the model representing the true behaviour of a given time series y_t can be formulated as a random walk with drift:

$$y_t = \alpha + \rho (= 1) \times y_{t-1} + \varepsilon_t \quad (1)$$

where,

y_t : dependent variable (in our case, the hourly mean concentration of NO₂) at time t , assumed to depend upon its first order lagged value y_{t-1} by coefficient ρ equal to one (unit root), which implies non stationarity;

α : drift term which makes y_t as the sum of a linear time trend (linear function of time t) and a series of random impulses

$$(y_t = \alpha \times t + \sum_{i=0}^t \varepsilon_i);$$

⁵ The ratio of buildings' average height to road width, used to measure the extent to which a road can be considered a street canyon (Tartaglia, 1999).

⁶ Tartaglia (1999) associates the street canyon condition to an H/W ratio of 1–1.5.

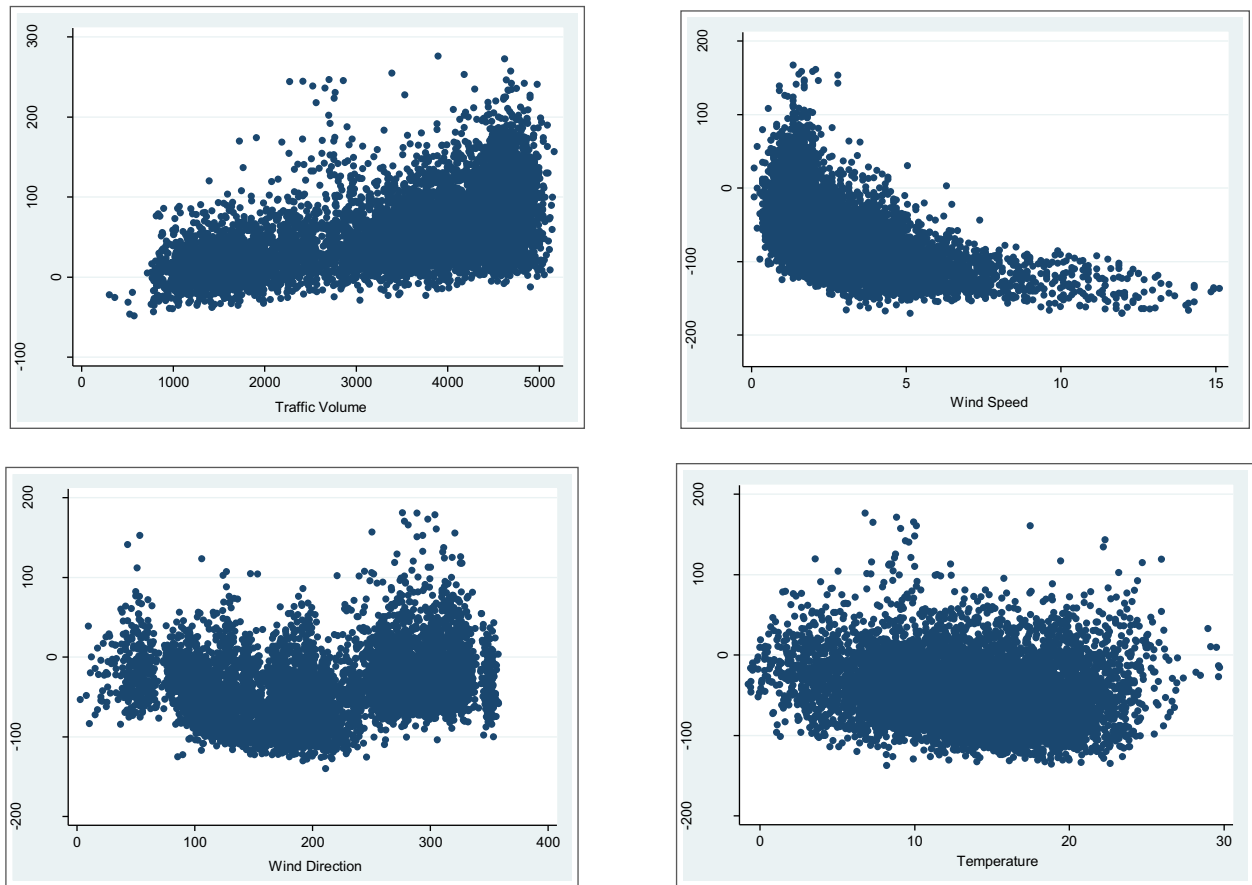


Fig. 3. Scatter plots depicting, for the set of hourly observations recorded in 2007 at Marylebone road, the relationships between the NO₂ mean concentration ($\mu\text{g}/\text{m}^3$) relative partial residuals and the following explanatory variables: traffic volume (pcu/hour), wind speed (km/h), wind direction (degrees North) and temperature (degrees centigrade).

ε_t : independently and identically distributed error terms with zero mean (white noise).

On the other hand, the most general alternative hypothesis of the test is the stationary behaviour of the time series around a deterministic linear time trend:

$$y_t = \mu + \delta \times t + \rho(< 1) \times y_{t-1} + \varepsilon_t \quad (2)$$

where,

$(\mu + \delta \times t)$ is the trend, specifically a linear function of time t ;
 ρ is smaller than one, thus making the fluctuations of the series around its trend be the manifestation of a first-order autoregressive model.

Since model (2) might not capture fully the underlying serial correlation, given that the autocorrelation could be second order level or above, the ADF test fits a transformation of model (2) through the Ordinary Least Square technique as follows:

$$\Delta y_t = \mu + \delta \times t + \beta \times y_{t-1} + \gamma_1 \times \Delta y_{t-1} + \dots + \gamma_k \times \Delta y_{t-k} + \varepsilon_t \quad (3)$$

where,

$\Delta y_t = y_t - y_{t-1}$: first difference of variable y_t ;

$\gamma_j \times \Delta y_{t-j} = \gamma_j \times (y_{t-j} - y_{t-j-1})$;

$y_j, j = 1, 2, \dots, k$: additional elements to capture the likely presence of further serial correlation in

Testing $\beta = 0$ amounts to testing $\rho = 1$ (Hamilton, 1994) or, equivalently, that y_t follows a unit root process. This means that the time series is not stationary and, hence, should be made stationary.

This is achieved by subtracting its first-order lagged value from each element of the series.

The test has been performed for the traffic volume and the hourly averages of NO₂ concentration, wind speed and direction. Several values of lag k ($k = 1-24$) have been tested and in each case the null hypothesis ($\beta = 0$) has been rejected, thus pointing out stationarity of the above variables, as suggested by the preliminary graphical inspection.

Table 2 presents the test results for lag $k=24$, which can take account of possible forms of autocorrelation till a day before the hour of interest. Since the series under consideration do not show a trend, but have a nonzero mean, under the null hypothesis, the value of the drift (α) has been set at zero, while, under the alternative hypothesis, only the $(\delta \times t)$ term of regression (3) has been dropped. Table 2 shows that, in each instance, the test statistic is far smaller than the critical values at 1%, 5% and 10% significance levels, which makes the assumption of stationarity highly plausible.

In order to specify a model for the time series of NO₂ hourly concentration, it is important to study its autocorrelation. This has been achieved through the visual analysis of the total and partial⁷ autocorrelation functions. In Fig. 5, the behaviour of total autocorrelation reveals a cyclical or seasonal effect: in fact, as the lag extends, the positive value of autocorrelation declines, but

⁷ The partial autocorrelation of variable x_t at lag k measures the correlation between x_t and x_{t+k} after the effects of correlations with $x_{t+1}, \dots, x_{t+k-1}$ have been removed, while the total autocorrelation incorporates such correlation chain effects.

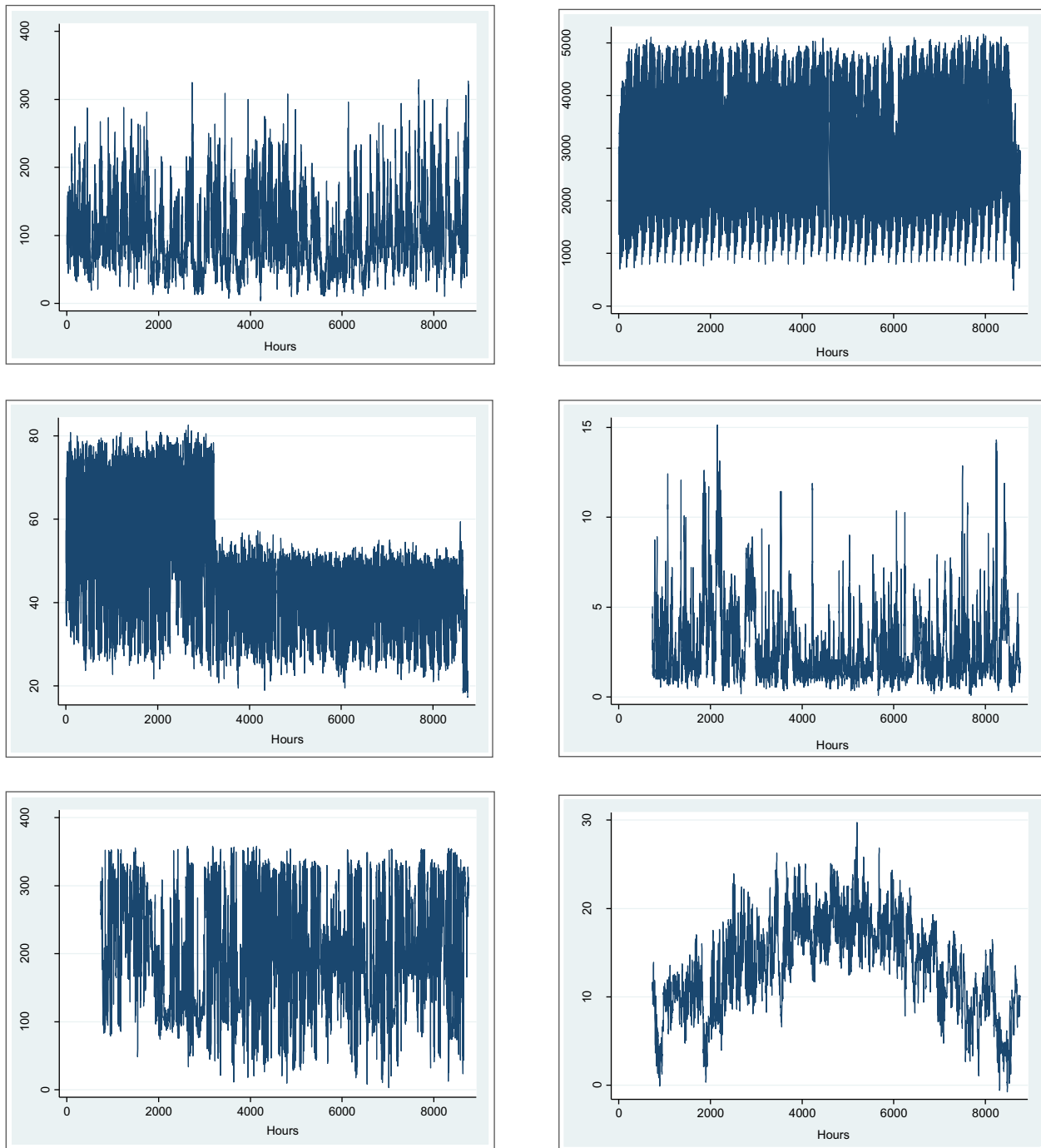


Fig. 4. Time series plots showing the hourly averages of NO₂ concentration ($\mu\text{g}/\text{m}^3$), vehicle mean speed (km/h), wind speed (km/h), wind direction (degrees North), temperature (degrees centigrade) and the hourly traffic volume (pcu/hour), in 2007 at Marylebone road.

Table 2

Augmented Dickey-Fuller unit root test results if lag $k=24$; $H_0: \beta=0$ and $\alpha=0$; $H_1: \delta=0$.

Variables	Test Statistic	Critical Value 1%	Critical Value 5%	Critical Value 10%
Hourly mean concentration of NO ₂ ($\mu\text{g}/\text{m}^3$)	-9.090	-3.430	-2.860	-2.570
Total hourly traffic volume (pcu/hour)	-8.905			
Hourly mean wind speed (km/h)	-9.143			
Hourly mean wind direction (degrees North)	-9.075			

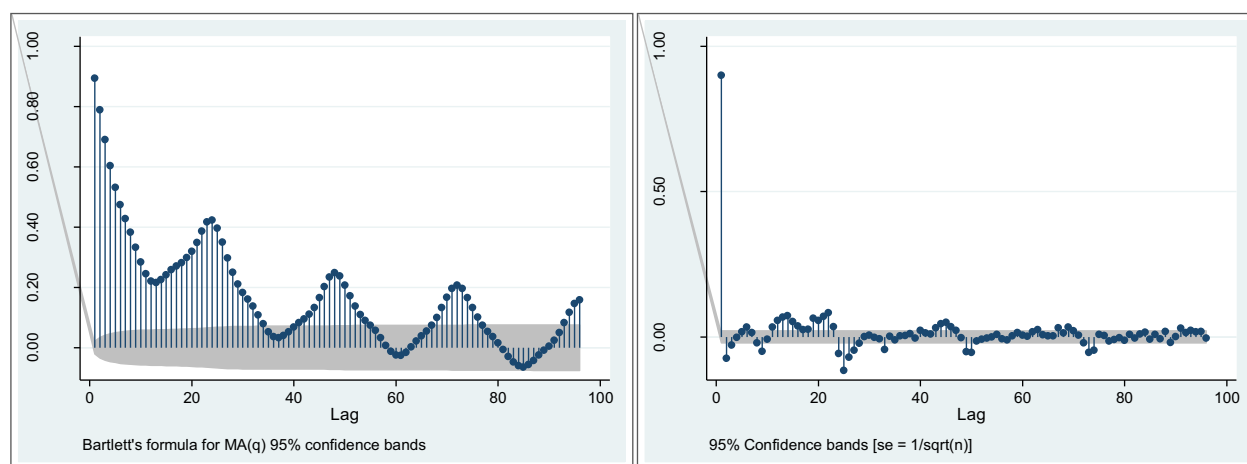


Fig. 5. Graphs of the total and partial autocorrelations for the hourly average of NO_2 concentration ($\mu\text{g}/\text{m}^3$) in 2007 at Marylebone road.

with a reversal occurring whenever the time interval reaches a multiple of twenty-four hours. Importantly, the partial autocorrelation highlights a strong dependence of concentration upon its previous hourly value, whereas the direct relationships with past concentrations are negligible.

The cyclical effect can be interpreted considering that road mobility follows a cyclical daily pattern, which stems from people's routine and, hence, is incorporated into the behaviour over time of all the pollutants strictly related to transport, such as NO_2 . The positive correlation with the mean concentration of the previous hour, instead, is due to the fact that the pollutant emitted during an hour requires a certain time to be dissipated, so a fraction of it represents the background for the next hour. In fact, depending on meteorological conditions (namely wind speed) pollution can build up over longer periods, but their effects are second order and are in part stochastic over long period measurements.

Moreover, the background pollution in an hour contains not only what has been emitted and not dispersed during the preceding hour, but also what is emitted in the hour of interest by sources different from transport, including mainly residences and offices in the study-area of this work. The impact of changes in this element of the background on concentration can be likewise taken into account by the behaviour of average concentration in the previous hour. In fact, variations in the emission pattern of residential activities and offices are likely to span at least one day.⁸

4. Modelling approaches explored

This section details the modelling approaches in forecasting NO_2 concentration, namely the Artificial Neural Network and the ARIMAX (Autoregressive Integrated Moving Average with Explanatory variable) model, and presents the estimation results before illustrating the comparative analysis in Section 4. Both models belong to the statistics domain and use mathematical functions built on empirical data to predict concentration. Unlike the ARIMAX methodology, the ANN does not require specific assumptions on the mathematical relationships between the pollutant concentration and its explanatory factors. There is very little literature on this type of comparative analysis, especially for the prediction of extreme levels of air pollution due to road traffic in urban areas.

⁸ For example, think about the possibility of an exceptional day in spring time, in which temperature is significantly lower than the seasonal mean, thus requiring an unusual use of heating by people and producing, throughout the day, higher levels of NO_2 background concentration than expected.

The comparison is based on the dataset of NO_2 hourly mean concentrations ($\mu\text{g}/\text{m}^3$) measured throughout the year 2006 at Marylebone road in London. The variables to explain the state of concentration are representative of weather and traffic conditions during the hour of interest: traffic volume (pcu/hour), hourly mean of wind speed (km/h), hourly mean of wind direction (degrees North). Besides, either models embody the double form of autocorrelation emerging from the analysis of correlograms (Fig. 5), which is the dependency of concentration upon the previous one hour and one day values. Further investigation will test the possibility of multiple step ahead forecasting of concentration (e.g. two or three hours in advance), since pollution measurements need to be scrutinised by air quality authorities before being used to initiate abatement measures which cannot be fully effective if implemented shortly before an anticipated pollutant event.

As a consequence, to use these models for one step ahead concentration forecasting, traffic flow and wind characteristics have to be foreseen on an hourly basis. For the former to be predicted, assignment of historical hourly origin-destination matrices improved with real-time traffic counts can be used to estimate the future road network conditions. In the case of wind, instead, short-term predictions are largely available and can be obtained by the providers of weather forecast (Met Offices), that regularly process the meteorological data collected across a country.

The following subsections illustrate the concepts underpinning the two approaches and the frameworks of the specific models derived from the data.

4.1. Neural network

The neural network is an artificial intelligence technique that mimics the human brain behaviour in a learning process. The important feature of an ANN is its adaptive nature and “learning by examples” is the logic used by this method to accomplish classification and regression tasks. This makes neural networks virtually applicable to every situation in which the relationships between a response variable and its predictors are very complex and cannot be easily outlined based on a priori knowledge and theoretical considerations (Migliore and Catalano, 2007).

The Multilayer Perceptron (MLP) architecture has proved to be the most suitable class of neural networks for air quality forecasting by previous studies (Nagendra and Khare, 2002). It consists of a system of layered and interconnected nodes or neurons. Nodes within the input layer and one or more hidden

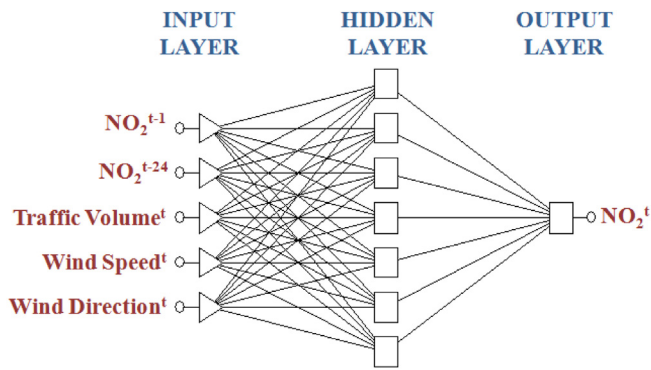


Fig. 6. Neural Architecture to forecast the hourly average of NO_2 concentration ($\mu\text{g}/\text{m}^3$) in 2007 at Marylebone road.

layers are connected to all nodes in neighbouring layers (Bishop, 1995). The input neurons provide signals to the hidden layer, where each node sums the inputs, processes the result with a non-linear transfer or activation function (logistic or hyperbolic tangent) and then distributes it to the output layer. The latter computes the dependent variable value in a similar manner. Each neuron-to-neuron connection is associated with a specific weight.

MLP has the ability to learn through training, which requires a series of input vectors and associated outputs. During training, the output from the MLP is compared with the desired value, an error signal is propagated back through the network and the magnitude of this error is used to adjust the weights iteratively until a stopping criterion is met. Besides a set of observations for training (the training set), an additional set of new data, namely the validation or selection set, is used to limit the network complexity in favour of its generalization performance, which is the prediction accuracy with respect to unknown cases. In fact, during a typical network learning session, the training set error generally decreases as a function of the number of iterations. On the contrary, the error measured in relation to independent data (validation set) often shows a decrease at first, followed by an increase as the network starts to over-fit. Halting training at the point of smallest error within the validation set should achieve the best degree of the network generalization.

It is common practice in the application of neural networks to train many different candidate architectures and then the best is selected on the basis of the performance obtained with the validation/selection set. However, a selection based only on the validation set could be misleading, because the network performance has a random component due to noise in the data. Therefore, the choice of a specific network structure should be confirmed by its performance within a third independent set of data often refer to as the test set.

Fig. 6 illustrates conceptually the neural network framework that has been selected to simulate the hourly average of NO_2 concentration ($\mu\text{g}/\text{m}^3$) in 2007 at Marylebone road: an MLP with five inputs, seven hidden neurons and one output. The independent variables are the previous one hour and one day mean concentrations along with non-lagged predictors: the traffic volume (pcu/hour) and the hourly averages of wind speed (km/hour) and wind direction (degrees North). The model design has been performed through a number of trials with different architectures (three and four layer MLP networks) and the network with the best selection set performance was chosen. At each learning round, a resampling method was used to form the training and selection sets in order to make the candidate networks as diverse as possible. More specifically, whilst fixing the test subset throughout the training procedure, the Monte Carlo technique was adopted to draw at random the elements of training

and selection subsets from the whole database. Weight regularization (Bishop, 1995) was performed to decide the number of hidden units adding an extra term to the error function for penalizing and reducing those weights that give a small contribution to the network performance. This procedure prunes away entire hidden units when their fanning-out weights are below a fixed threshold, thus limiting the curvature of sigmoid activation functions and improving the network generalization skill.

Table 3 reports the Pearson-R correlation coefficient for the best model predictions and the performance value for each subset of data (train, selection, test). The high correlation reveals the ability of the neural model to reproduce the historical behaviour of pollutant concentration levels. Every subset-specific performance indicator is computed as the ratio of standard deviation of the prediction error to that of the specific subsample of data. A ratio substantially below 1 indicates that the network has performed far better than a simple mean estimator and its use in forecasting is then justified. As Table 3 illustrates, for the chosen network, the ratio for each sub-component of the total dataset is quite low, which indicates a good level of performance. In addition, the selection performance is better than that for the training set, which means that the network has not over-learned.

Fig. 7 displays the response surfaces plotted in relation to two different pairs of input variables. The response surface is constructed to show that the two selected inputs vary from their observed minimum to their observed maximum value, while all other inputs are held at fixed values (their averages). In one case, the graph (left) visualizes the dependency of the NO_2 hourly concentration upon its lagged values showing a strong positive impact of the mean concentration in the preceding hour and a slight effect of the corresponding hourly concentration in the previous day.

In the other graph (right), the relationship between concentration and wind characteristics is clearly non-linear: as the wind speed increases, concentration falls at a decreasing marginal rate, whereas the dependency upon the wind direction is represented by a U-shaped curve.

This quantifies the influence of the quasi-canyon form of Marylebone road. The part of the road section just off the monitoring device presents an H/W ratio⁹ close to 1, which according to Tartaglia (1999) indicates the possibility of a street canyon behaviour of transport-related pollution. This means that, when the wind speed is at least 1 m/s and its direction with the road normal axis has an angle in the -45° to $+45^\circ$ range, a vortex is generated inside the street canyon and the pollutant concentration on the windward side is far higher than that on the leeward side. Marylebone road has a Southwest-Northeast alignment and the pollution receptors are located on its southern boundary (see Fig. 2). It follows that, when the wind comes from the South, the set of sensors is on the windward side. Therefore, as the wind direction moves closer to 360° (or zero) becoming more northerly, an increasing pollution concentration is observed. On the contrary, if the wind comes from North, the sensors are on the leeward side, hence the measured concentration should decrease as the wind moves closer to 180° North. The fact that the trained neural network is able to simulate and therefore assess the influence of this complex behaviour of pollution adds credibility to the ANN as a valid and valuable statistical tool to understand the effects of traffic related air pollution in our urban areas.

In the end, Fig. 8 shows that increasing traffic volumes, as expected, have a positive impact on concentration according to a mathematical relationship slightly deviating from the linear

⁹ Tartaglia (1999) states that, when the ratio of buildings' average height to road width is 1 or greater, street canyon effects appear.

Table 3
Performance of the best MLP neural network.

Observations	Train Performance	Selection Performance	Test Performance	Observations-Predictions Correlation coefficient
7563	0.403	0.365	0.400	0.920

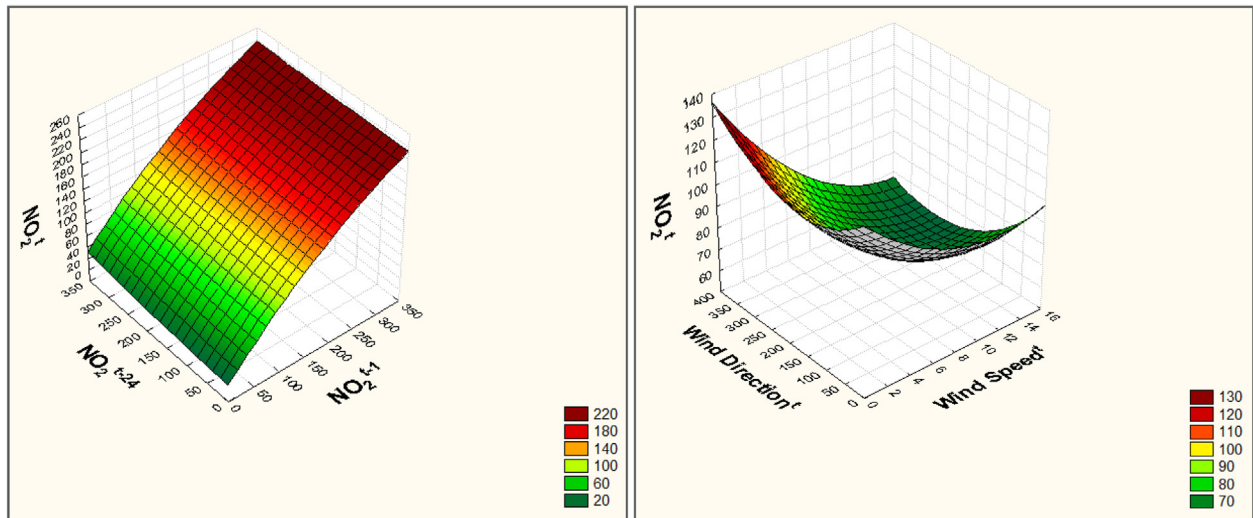


Fig. 7. Response surfaces describing how the trained MLP neural network simulates the dependency of the NO₂ hourly concentration (µg/m³) upon its lagged values one hour and one day back (left) and the wind characteristics, namely speed in km/hour and direction in degrees North (right).

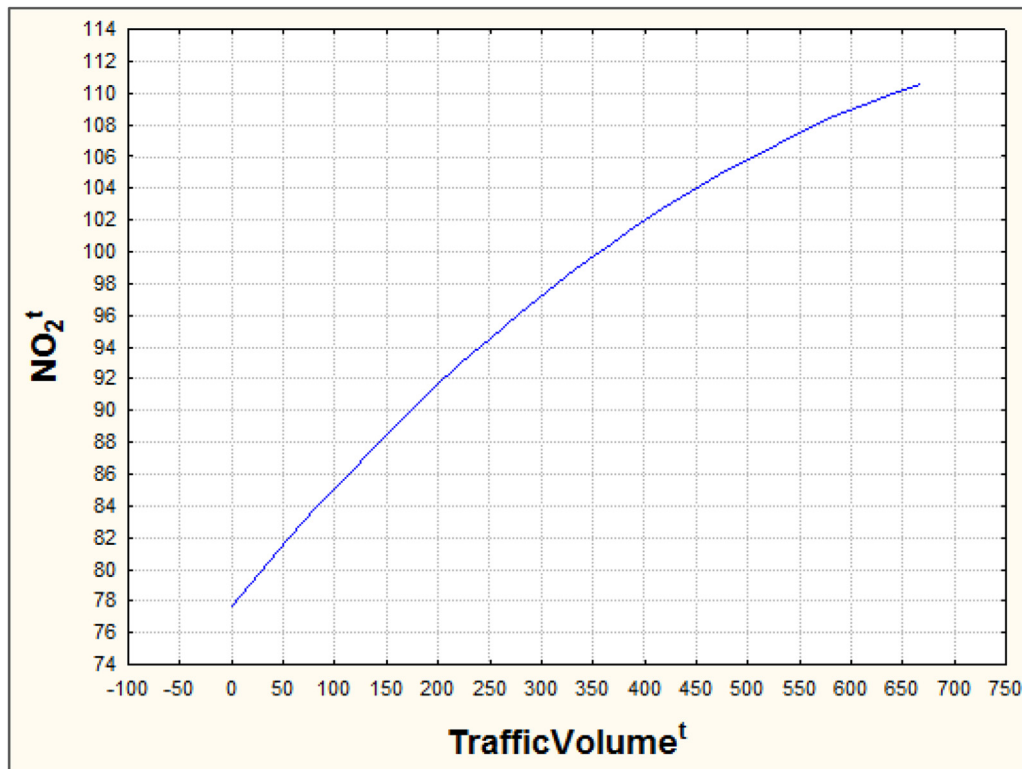


Fig. 8. Response line describing how the trained MLP neural network simulates the dependency of the NO₂ hourly concentration (µg/m³) upon traffic volume (pcu/hour).

function.

4.2. Statistical model

The ARIMAX model was developed independently as an alternative approach to predict NO₂ hourly concentration at Marylebone road. This postulates an underlying stochastic data-generating process and represents it as the integration of two main components: one captures the relationship between the dependent variable and its past manifestations (autoregressive part), the other incorporates the effect on the dependent variable of a finite series of random impulses (moving average part). Formally, an ARIMAX model, with p autoregressive (AR) elements and q moving average (MA) terms, can be written as follows:

$$\begin{aligned} & (1 - \rho_1 L^1 - \rho_2 L^2 - \dots - \rho_p L^p) \times (y_t - \mathbf{x}'_t \times \boldsymbol{\beta}) \\ & = (1 + \theta_1 L^1 + \theta_2 L^2 + \dots + \theta_q L^q) \times \varepsilon_t \end{aligned} \quad (4)$$

or, more succinctly:

$$\rho(L^p)(y_t - \mathbf{x}'_t \times \boldsymbol{\beta}) = \theta(L^q)\varepsilon_t \quad (5)$$

where,

ρ_j : coefficient of the j^{th} -order AR element;

θ_j : coefficient of the j^{th} -order MA element;

L^j : lag operator transforming a variable at time t into its j^{th} -order lagged manifestation ($L^j z_t = z_{t-j}$);

y_t : dependent variable at time t ;

\mathbf{x}'_t : vector of exogenous covariates at time t ;

$\boldsymbol{\beta}$: vector of coefficients;

$\varepsilon_t \sim N(0, \sigma^2)$, meaning that it is a white noise disturbance;

$\rho(L^p) = (1 - \rho_1 L^1 - \rho_2 L^2 - \dots - \rho_p L^p)$;

$\theta(L^q) = (1 + \theta_1 L^1 + \theta_2 L^2 + \dots + \theta_q L^q)$.

If the considered series shows a cyclical behaviour, model (5) can be turned into a seasonal ARIMAX or SARIMAX model by the introduction of multiplicative terms both for the AR part and for the MA component (La Franca et al., 2010):

$$\rho(L^p)\rho_s(L^p)(y_t - \mathbf{x}'_t \times \boldsymbol{\beta}) = \theta(L^q)\theta_s(L^q)\varepsilon_t \quad (6)$$

where, the seasonal symbols (those with subscript s) have the same meaning as their non-seasonal counterparts, but apply to the series every s units of time (hours in our case).

The SARIMAX model for the 2007 time series of NO₂ concentrations at Marylebone road has one AR term and one MA element both for the non-seasonal part and for the seasonal one, the latter being based on a periodicity of 24 h (recall subsection 2.3). The exogenous regressors are again the traffic volume (pcu/hour), the wind speed (km/hour) and the wind direction (degrees North). Under the base formulation, no prior knowledge on the shape of their influence on concentration is

postulated and, therefore, they are assumed to be linearly related to the dependent variable.

In view of the above, formula (6) can be specified as follows:

$$\rho(L^1)\rho_{24}(L^1)(y_t - \mathbf{x}'_t \times \boldsymbol{\beta}) = \theta(L^1)\theta_{24}(L^1)\varepsilon_t \quad (7)$$

Table 4 exhibits the results of the SARIMAX model estimation, conducted by the maximum likelihood technique and using the Kalman filter via the prediction error decomposition (Hamilton, 1994). This approach handles missing data (for the dependent variable and/or its covariates) by continuing the state-updating recursions of the Kalman filter even if the contribution to estimation from the sample is partial or even null. All the estimated parameters have correct signs with high statistical significance except for the constant. The main limitation of the estimated SARIMAX model is its inability to reproduce the non-linear relationship between concentration and the wind characteristics, with due consideration of wind direction. Therefore, the resulting linear impact on pollution is very far from the more plausible parabolic behaviour revealed with the neural network and explained as a street canyon effect.

In line with the Box and Jenkins' guidelines (Box et al., 2008), the SARIMAX residuals have been analysed to check whether they can be reasonably considered Gaussian and independently distributed as required by the theory. Fig. 9 shows the frequency distribution of residuals along with their autocorrelation function for the first 24 lags. As can be observed, the assumption of a white noise behaviour of residuals is supported by the empirical evidence.

5. Comparative analysis

This section compares the two forecasting models in relation to a sample of observations which have not been involved in the building process of each of the above described models. The ANN with the SARIMAX comparison is based on three criteria, represented through statistical and categorical metrics: the extent to which observed and simulated concentrations covary; the global forecasting accuracy and the ability to alert in advance against the exceedances of the legal concentration threshold. Finally, to explore the potential advantage of ensemble forecasting, the predictions of the two models have been combined and the resulting performance has been evaluated.

5.1. The dataset for comparison

The ANN and SARIMAX modelling approaches are compared based on out-of-estimation sample data, collected in 2006. In this period, the annual average of NO₂ hourly concentrations was very close to that calculated for the year 2007 (110.6 versus 102.5), but the number of cases in which the 200 $\mu\text{g}/\text{m}^3$ concentration was exceeded is 1.5 times greater (686 versus 457). As emerges from Table 5, in comparison to Table 1, the base statistics describing the

Table 4
The SARIMAX model for the 2007 series of NO₂ hourly concentrations at Marylebone road.

Regressors	Coefficient	Standard Error	z	P > z	[Conf. Int. 95%]
Traffic Volume (pcu/h)	0.074	0.007	10.530	0.000	0.061 0.088
Hourly Mean Wind Speed (km/h)	-8.805	0.348	-25.330	0.000	-9.486 -8.123
Hourly Mean Wind Direction (degrees North)	0.025	0.005	4.990	0.000	0.015 0.035
Constant	-0.118	0.123	-0.960	0.336	-0.358 0.122
NO ₂ Hourly Mean Concentration 1Hour Lagged ($\mu\text{g}/\text{m}^3$)	0.873	0.009	98.050	0.000	0.856 0.890
NO ₂ Hourly Mean Concentration 1 Day Lagged ($\mu\text{g}/\text{m}^3$)	0.074	0.016	4.670	0.000	0.043 0.104
Error Term 1Hour Lagged	-0.098	0.020	-4.810	0.000	-0.138 -0.058
Error Term 1 Day Lagged	-0.950	0.007	-130.530	0.000	-0.965 -0.936
Error Standard Deviation	20.378	0.288	70.830	0.000	19.814 20.942

LN (Likelihood) = -33783.07; N. observations = 7563; Wald $\chi^2(7) = 31839.73$

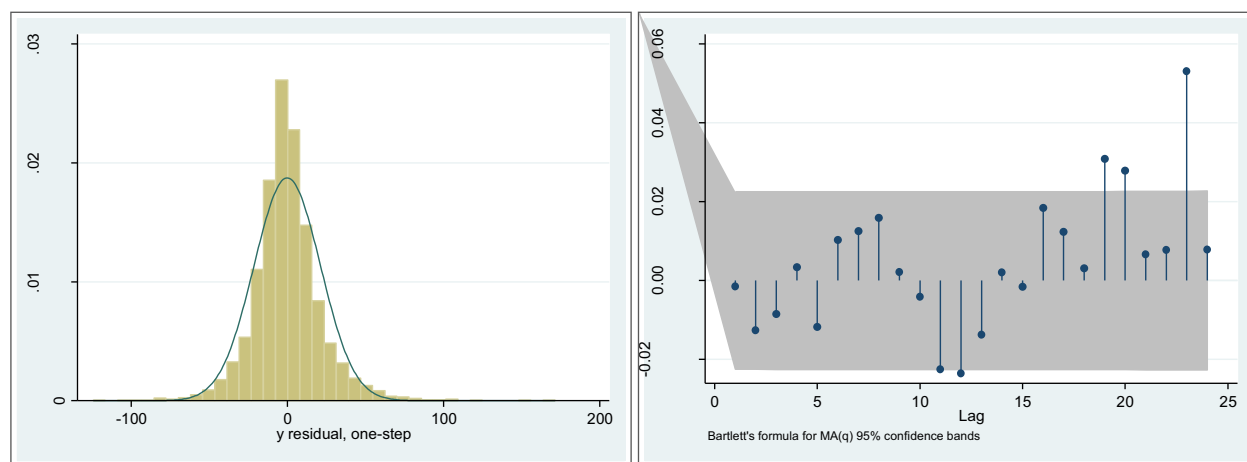


Fig. 9. Analysis of the SARIMAX model's residuals: distribution and autocorrelation.

Table 5

Descriptive statistics for the set of air quality, transport and weather data collected in 2006 within the study-area. In parentheses, the same statistics for the year 2007.

Variables	Observations	Average	Min	Max	Standard Deviation
Hourly mean concentration of NO ₂ (μg/m ³)	8518 (8584)	110.6 (102.5)	2 (4)	403 (329)	56.6 (53.1)
Total traffic volume (pcu/hour)	5637 (8734)	3377.2 (3505.7)	721.7 (303.7)	4899.2 (5161.5)	1158.1 (1173.2)
Hourly mean wind speed (km/h)	6871 (7993)	3.0 (2.8)	0.2 (0.1)	17.6 (15.1)	2.4 (2.1)
Hourly mean wind direction (degrees North)	6871 (7993)	217.9 (203.9)	1.0 (2.9)	358.7 (358.1)	85.8 (82.4)
Hourly mean temperature (degrees centigrade)	6865 (7993)	13.9 (13.5)	-2.1 (-0.8)	37.3 (29.7)	7.1 (5.2)

2006 set of observations do not differ substantially when compared to those of 2007.

5.2. Evaluation of model forecasting performance and sensitivity analysis

The ANN and the SARIMAX models have been studied with respect to the ability to predict the NO₂ concentrations recorded at Marylebone road in 2006. In order to provide initial conclusions on the potential benefit of ensemble forecasting, the two models have been integrated with a mixed method taking, for each case, the maximum of the concentrations predicted by both of them. The choice of the *max* functional form for the ensemble is motivated by the tendency, for both models, to underestimate pollution peaks. Indeed, within the set of cases in which the EU threshold hourly concentration (200 μg/m³) is exceeded, the mean ratio of the predicted concentration to the actual one is 0.89 if the SARIMAX is applied and 0.86 if the ANN is used; it becomes slightly higher (0.92) if the two models are merged into an ensemble.

Fig. 10 visually outlines the distributions of measured and simulated concentrations. The graphic tool employed is the box plot, which represents the 25th and 75th percentiles with the bottom and top lines of the box, respectively; the difference between these two percentiles is termed interquartile range. In addition, the line intersecting the box identifies the median of the distribution, while the two whiskers provide a measure of the data's spread.¹⁰ The isolated points outside of the spread represent outliers. As can be seen in the picture, compared to the measured all models reproduce the observed dispersion of concentrations

quite well. Also, as displayed by the upper whiskers, all models clearly underestimate the high levels of concentration, even though the ensemble performs best.

Table 6 compares all models through five performance indicators. Two statistical indices measure the linear correlation and the mean absolute percent difference (MAPE) between observed and estimated concentrations. The values of these indicators show that, overall, the three models do not differ significantly and all can be deemed fairly effective. Moreover, two categorical indices are used to evaluate the models' ability to alert in advance against the exceedance level of the 200 μg/m³ limit. One indicator is the percentage of the observed illegal

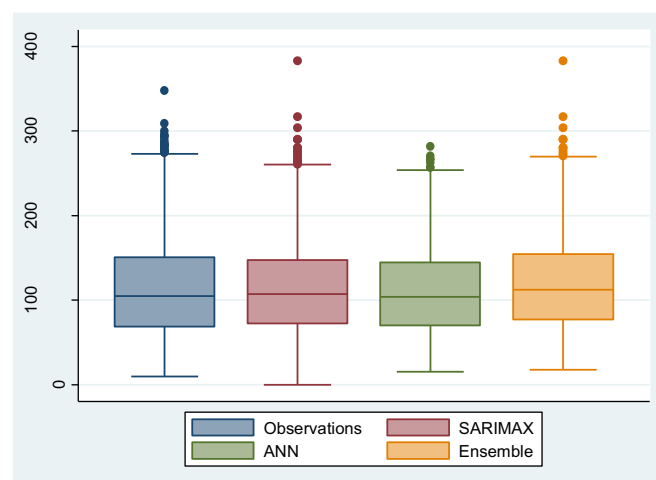


Fig. 10. Box Plot of NO₂ concentration time series for measured and modelled in 2006 at Marylebone road.

¹⁰ This ranges from 1.5 times the interquartile difference above the 75th percentile to 1.5 times the interquartile difference below the 25th percentile.

Table 6Evaluation of the forecasting performance when the alarm threshold is set at 200 $\mu\text{g}/\text{m}^3$.

Modelling Approach	Correlation	MAPE (%)	Share of actual illegal concentrations estimated over the alarm threshold (%)	False Alarm Ratio (%)	Mean Concentration at False Alarms ($\mu\text{g}/\text{m}^3$)
SARIMAX model	0.91	18.62	54.61	33.86	177.85
Artificial Neural Network	0.91	16.53	49.38	30.40	180.59
Ensemble: max [Pred _{SARIMAX} , Pred _{ANN}]	0.92	19.32	61.95	36.77	177.66

Table 7Evaluation of forecasting performance when the alarm threshold is set at 180 $\mu\text{g}/\text{m}^3$.

Modelling Approach	Correlation	MAPE (%)	Share of actual illegal concentrations estimated over the alarm threshold (%)	False Alarm Ratio (%)	Mean Concentration at False Alarms ($\mu\text{g}/\text{m}^3$)
SARIMAX model	0.91	18.62	76.97	46.45	172.14
Artificial Neural Network	0.91	16.53	71.56	45.48	173.14
Ensemble: max [Pred _{SARIMAX} , Pred _{ANN}]	0.92	19.32	82.15	50.91	170.38

concentrations forecasted by the models as concentrations over an alarm threshold (just 200 $\mu\text{g}/\text{m}^3$). From this point of view, the ensemble approach yields a rather better performance, since the *max* operator lowers the tendency to underestimate extreme values of pollution.

The fourth indicator is the share of the predicted concentrations beyond the chosen alarm bound (alarms) that do not correspond to real concentrations above the legal limit of 200 $\mu\text{g}/\text{m}^3$ (false alarm ratio). All models present good levels of this performance measure with a percentage of false alarms varying in the 30–37% range.

The fifth indicator is the actual level of mean concentration in the case of false alerts, which can be viewed as a measure of their social benefit. For each model, Table 6 shows that, when a false alarm is generated, the corresponding real concentration is quite high on average (around 180 $\mu\text{g}/\text{m}^3$), which makes such an alarm useful anyway from an air quality protection perspective. Another advantage of these wrong alerts is their contribution to the lowering of annual mean concentration, which is far beyond the EU target of 40 $\mu\text{g}/\text{m}^3$ in the study-area.

Since the models underestimate pollution peaks, their ability to foresee the need for traffic management could be enhanced by establishing an alarm threshold smaller than the legal limit of 200 $\mu\text{g}/\text{m}^3$, but sufficiently high, say 180 $\mu\text{g}/\text{m}^3$, to initiate a traffic control intervention. In this way, under the new approach, the traffic measures are triggered in the 180–200 $\mu\text{g}/\text{m}^3$ range of predicted concentrations, building-in a degree of leeway to ensure that the target (concentration < 200 $\mu\text{g}/\text{m}^3$) is respected.

Table 7 endorses this assumption showing that the percentage of actual illegal concentrations foreseen by the models as “alarming” concentrations hugely improve by as much as 22 points, at the expense of an expected, but acceptable, rise in the incidence of false alarms (from 30 to 37% to 45–51%). What is more, the mean concentration when a false alarm is issued remains high (170–173 $\mu\text{g}/\text{m}^3$), which guarantees a useful traffic management intervention.

In terms of the share of actual exceedances captured by the models, results could be improved further by lowering the alarm threshold further to the detriment of the false alarm ratio; however, the value of 180 $\mu\text{g}/\text{m}^3$ leads to a satisfactory balance between the two conflicting evaluation indices.

Figs. 11–13 compare the behaviours of measured and simulated concentrations during the validation period for each of the three models in turn. The scatter plots show similar and high levels of

linear correlation between observations and forecasts for all models, while the line plots (on the left) show that the ANN has the worst performance in predicting pollution peaks. This can be explained considering that, on one hand, the non-linear patterns in the dataset (the influences of wind speed and direction, in particular) have not probably as strong an impact on pollutant concentration compared to the other predictors (previous hours concentrations and traffic volume), which are linearly related to NO₂ pollution.¹¹ On the other hand, the training of a neural network does not use the entire available database, because the test subset of data (containing in this application as many as 1712 observations) is not employed in the learning process. The probability of occurrence of higher pollution events is smaller within the smaller sample.

A sensitivity analysis was performed on the influential factors, namely traffic volume, wind speed and wind direction, in order to check and compare the robustness of the ANN and SARIMAX forecasting models. As these are likely to be subject to measurement and prediction errors, the models' forecasting performance has been evaluated under two scenarios:

1. a “*Low Scenario*” perturbing the explanatory variables' levels, observed in 2006 at Marylebone road, sampling randomly from a uniformly distributed error that varies from –5% to +5%;
2. a “*High Scenario*” based on a –15% to +15% uniformly distributed error.

The error range for each scenario has been set considering that short-term predictions of traffic volume and wind characteristics are fairly reliable in the current applications of traffic and weather forecast.

Table 8 compares the performance indicators for each scenario with those derived from the original dataset considering an alarm concentration threshold set at 180 $\mu\text{g}/\text{m}^3$. Given the small differences in the performance measures, the two models are robust to the measurement and prediction errors that are likely to occur in real world.

¹¹ This is confirmed by the limited range of variation of the U-shaped function in Fig. 7.

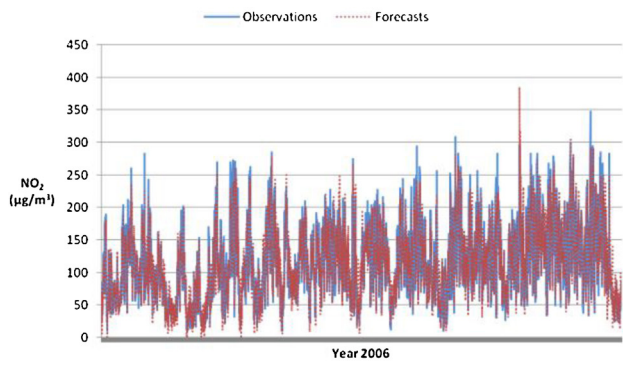


Fig. 11. NO₂ concentrations observed in 2006 at Marylebone road versus SARIMAX forecasts.

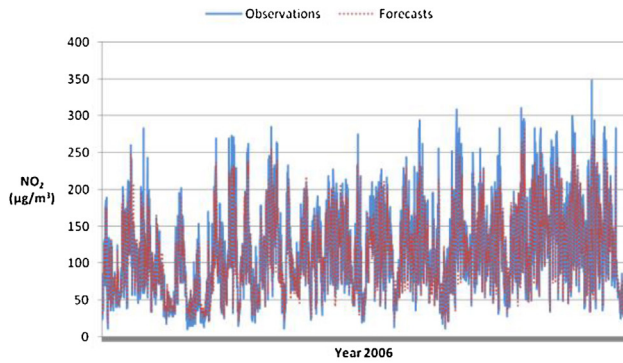
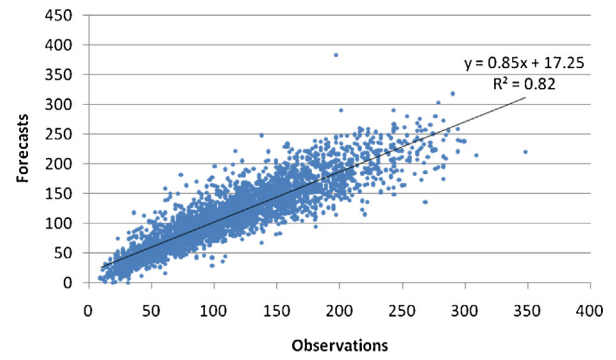


Fig. 12. NO₂ concentrations observed in 2006 at Marylebone road versus Artificial Neural Network forecasts.

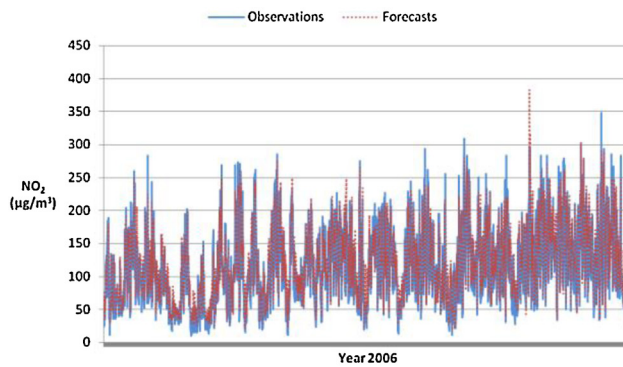
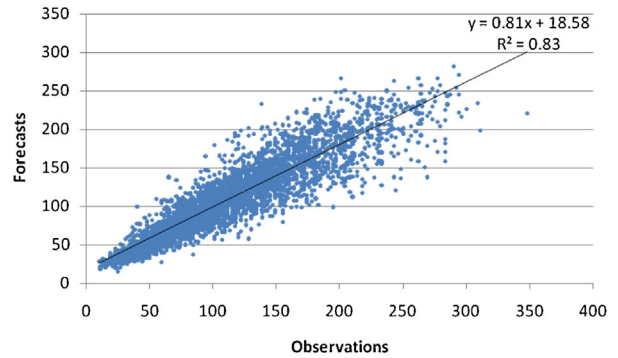


Fig. 13. NO₂ concentrations observed in 2006 at Marylebone road versus Ensemble forecasts.

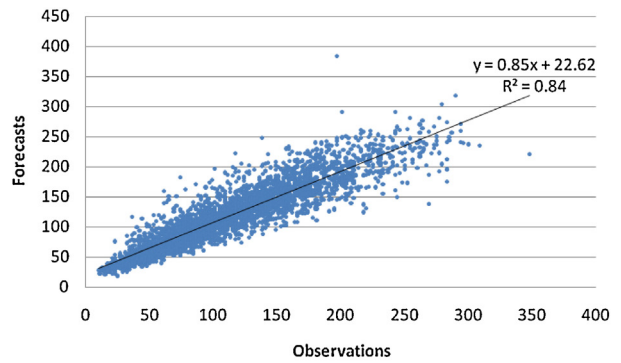


Table 8

Sensitivity analysis results when the alarm threshold is set at 180 µg/m³.

Modelling Approach	Correlation	MAPE (%)	Predicted Exceedances (%)	False Alarm Ratio (%)	Mean Concentration at False Alarms (µg/m ³)
SARIMAX model	0.91	18.62	76.97	46.45	172.14
SARIMAX model-Low Scenario	0.91	18.87	76.64	46.19	172.51
SARIMAX model-High Scenario	0.90	20.17	76.32	46.54	172.85
Artificial Neural Network	0.91	16.53	71.56	45.48	173.14
Artificial Neural Network-Low Scenario	0.91	16.60	70.94	45.30	172.85
Artificial Neural Network-High Scenario	0.91	16.71	72.50	45.79	172.77

6. Conclusions and future steps

This research investigated the relationship between the hourly mean concentration of NO₂ and explanatory factors reflecting the NO₂ mean level one hour and 24 h back, along with traffic and meteorological conditions. The final scope was to develop novel and more effective models for hourly pollution peak event forecasting as a statistical analysis method to be incorporated into traditional traffic management decision support system to deliver sustainable mobility in urban areas.

The results demonstrated that the neural network can represent the non-linear relationships between hourly concentration and meteorological conditions although its ability in predicting exceedances of the legal concentration limit was less promising. Therefore, ANN has an important role to play as an effective tool to analyse the non-linear processes and support the specification of other and more accurate statistical models such as the SARIMAX model, whose specification requires prior knowledge or theoretical assumptions.

Also, this initial study demonstrated that the ensemble approach is very promising for the prediction of extreme pollution events. It represents an interesting area of research that deserves further investigation. In particular, an appealing issue is the influence on the performance in forecasting pollution peaks of the number and type of models in the ensemble.

Finally, the ability of the models to alert against exceedances of the legal concentration limit by fixing an alarm threshold lower than 200 µg/m³ was demonstrated. This improvement occurs because such an alarm bound compensates for the tendency of models to underestimate extreme levels of NO₂ pollution. This additional alarm threshold was shown to be a good compromise between warning performance and incidence of false alarms.

Further research will regard the prediction of low frequency concentration peaks as well as the high levels of pollution deriving from exceptional conditions, such as city events (football matches or exhibitions) that attract many visitors and modify the patterns of urban mobility significantly. In such cases, previous studies (Stockwell et al., 2002) have highlighted the limitation of traditional site-specific statistical models.

Acknowledgements

The authors acknowledge with gratitude the Engineering and Physical Sciences Research Council (EPSRC, United Kingdom) who sponsored the research project SECURE, Self Conserving Urban Environment in which the research presented in the paper was conducted. Also, the authors thank the COST (European Cooperation in Science and Technology) for supporting Dr Catalano short term scientific visits (STSM) to Newcastle University.

References

Allen, R.W., Amram, O., Wheeler, A.J., Brauer, M., 2011. The transferability of NO and NO₂ land use regression models between cities and pollutants. *Atmos. Environ.* 45 (2), 369–378.

Baur, D., Saisana, M., Schulze, N., 2004. Modelling the effects of meteorological variables on ozone concentration—a quantile regression approach. *Atmos. Environ.* 38, 4689–4699.

Bell, Margaret, Bergantino, Angela S., Catalano, Mario, Galatioto, Fabio, 2015. Prediction of air pollution peaks generated by urban transport networks. Working papers SIET, Trieste, pp. 1–19 (ISSN 1973–3208).

Bergantino, A.S., Bierlaire, M., Catalano, M., Migliore, M., Amoroso, S., 2013. Taste heterogeneity and latent preferences in the choice behaviour of freight transport operators. *Transp. Policy* 30, 77–91.

Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*. Oxford University Press.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 2008. *Time Series Analysis: Forecasting and Control*, 4th ed. Wiley, Hoboken, NJ.

Cai, M., Yin, Y., Xie, M., 2009. Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transp. Res. Part D* 14, 32–41.

Díaz-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G., Moncada-Herrera, J.A., 2008. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile. *Atmos. Environ.* 42, 8331–8340.

EU Press Release Database, 2014. Environment: Commission Takes Action Against UK for Persistent Air Pollution Problems. http://europa.eu/rapid/press-release_IP-14-154_en.htm.

European Union (2008), Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. European Environment Agency of the European Union, 2014. *Air Quality in Europe—2014 Report*. Publications Office of the European Union, Luxembourg.

Fisher, B.E.A., 2006. Fuzzy approaches to environmental decisions: application to air quality. *Environ. Sci. Policy* 9, 22–31.

Galatioto, F., Zito, P., 2009. Traffic parameters estimation to predict road side pollutant concentrations using neural networks. *Environ. Model. Assess.* 14 (3), 365–374.

Galatioto, F., Bell, M.C., 2013. Exploring the processes governing roadside pollutant concentrations in urban street canyon. *Environ. Sci. Pollut. Res.* 20 (7), 4750–4765.

Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton, New Jersey.

Heinrich, J., et al., 2005. Studies on health effects of transport-related air pollution. In: Krzyzanowski, M., Kuna-Dibbert, B., Schneider, J. (Eds.), *Health Effects of Transport-Related Air Pollution*. World Health Organization-Europe, Denmark.

Hunt, P.B., et al. (1981), SCOOT-A Traffic Responsive Method of Coordinating Signals TRRL report LR1014. Crowthorne.

Kaushik, I., Melwani, R., 2007. Time series analysis of ambient air quality at its intersection in Delhi (India). *J. Environ. Res. Dev.* 2, 268–272.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in Central Helsinki. *Atmos. Environ.* 37, 4539–4550.

La Franca, L., Catalano, M., Castelluccio, F., Montano, F., 2010. Travelling towards and from minor islands through non-conventional air transport: demand and cost analysis, in Island Sustainability. In: Favro, S., Brebbia, C.A., Querini, G. (Eds.), *WIT Transactions on Ecology and the Environment Series*, volume 130. Wit Press, United Kingdom, pp. 251–266.

Lavecchia, C., Pilati, S., Angelino, E., Fossati, G., 2007. Analisi dei dati di traffico esistenti per la definizione dei profili temporali: metodologia ed esempio di applicazione. XIII Expert Panel Emissioni Da Trasporto Su Strada, Roma, 4 October 2007.

Mallows, C.L., 1986. Augmented partial residuals. *Technometrics* 28, 313–319.

Migliore, M., Catalano, M., 2007. Urban public transport optimization by Bus ways: a neural network-based methodology, in *Urban Transport XIII: Urban Transport and the Environment in the 21st Century*. In: Brebbia, C.A. (Ed.), *WIT Transactions on the Built Environment Series*, volume 96. Wit Press, United Kingdom, pp. 347–356.

Migliore, M., Catalano, M., Lo Burgio, A., Maritano, L., 2012. The analysis of urban travellers' latent preferences to explain their mode choice behaviour, in *Environmental Impact*. In: Brebbia, C.A., Aravossis, K., Brooks, C.N., Chon, T.-S. (Eds.), *WIT Transactions on Ecology and the Environment Series*, volume 162. Wit Press, United Kingdom, pp. 193–203.

Nagendra, S.M.S., Khare, M., 2006. Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecol. Modell.* 190, 99–115.

Nunnari, G., Nucifora, A.F.M., Randieri, C., 1998. The application of neural techniques to the modelling of time-series of atmospheric pollution data. *Ecol. Modell.* 111, 187–205.

Perez, P., Trier, A., 2001. Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago, Chile. *Atmos. Environ.* 35, 1783–1789.

Prybutok, V.R., Yi, J.S., Mitchell, D., 2000. Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *Eur. J. Oper. Res.* 122, 31–40.

Re, M., Valentini, G. (2012) Ensemble methods: a review, in *Advances in Machine Learning and Data Mining for Astronomy*, Chapman & Hall Data Mining and Knowledge Discovery Series, Chap. 26, pp. 563–594.

Sánchez, A.B., Ordóñez, C., Sánchez Lasheras, F., de Cos Juez, F.J., Roca-Pardiñas, J., 2013. Forecasting SO₂ pollution incidents by means of elman artificial neural networks and ARIMA models. *Abst. Appl. Anal.* 1–6.

Sayegh, A.S., Munir, S., Habeebullah, T.M., 2014. Comparing the performance of statistical models for predicting PM₁₀ concentrations. *Aerosol Air Qual. Res.* 14, 653–665.

Seinfeld, J.H., Pandis, S.N., 2006. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons, Inc.

Stockwell, W.R., et al., 2002. The scientific basis of NOAA's air quality forecast program. *Environ. Manager* 8, 20–27.

Tartaglia, M., (1999). *L'Inquinamento dell'Aria da Traffico Stradale*. Ed. Bios.

VV. AA, 2008. Special issue on applications of ensemble methods. *Inf. Fusion* 9 (1).

- Viotti, P., Liuti, G., Genova, P.D., 2002. Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia. *Ecol. Modell.* 148, 27–46.
- World Health Organization, 2014. Ambient (outdoor) Air Pollution in Cities Database 2014. . <http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/>.
- World Health Organization Regional Office for Europe, 2010. Millennium Development Goals in the WHO European Region: A Situational Analysis at the Eve of the Five-year Countdown. World Health Organization Regional Office for Europe, Copenhagen.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012a. Real-time air quality forecasting, part I: history, techniques, and current status. *Atmos. Environ.* 60, 632–655.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., Baklanov, A., 2012b. Real-time air quality forecasting, part II: state of the science, current research needs, and future prospects. *Atmos. Environ.* 60, 656–676.
- Zito, P., Chen, H., Bell, M.C., 2008. Predicting real-time roadside CO and NO₂ concentrations using neural networks. *IEEE Trans. Intell. Transp. Syst.* 9(3), 514–522.