Atmospheric Environment 127 (2016) 163–175

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Understanding how roadside concentrations of NO_x are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees

Arwa Sayegh*, James E. Tate, Karl Ropkins

Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK

HIGHLIGHTS

- Air quality at urban, motorway, and tunnel sites has been studied and compared.
- New method has been developed for splitting traffic data to four traffic states.
- Deriving traffic influence on roadside NO_x depends on the quality of background NO_x.
- Different traffic states have been shown to have different influence on roadside NO_x.
- Roadside NO_x appears to reach a minimum at around 22 °C of ambient air temperature.

ARTICLE INFO

Article history:
Received 16 May 2015
Received in revised form
4 December 2015
Accepted 8 December 2015
Available online 12 December 2015

Keywords:
Air quality
NO_x
Air temperature
Wind conditions
Continuous monitoring
Vehicle emissions

ABSTRACT

Oxides of Nitrogen (NO_x) is a major component of photochemical smog and its constituents are considered principal traffic-related pollutants affecting human health. This study investigates the influence of background concentrations of NO_x, traffic density, and prevailing meteorological conditions on roadside concentrations of NO_x at UK urban, open motorway, and motorway tunnel sites using the statistical approach Boosted Regression Trees (BRT). BRT models have been fitted using hourly concentration, traffic, and meteorological data for each site. The models predict, rank, and visualise the relationship between model variables and roadside NO_x concentrations. A strong relationship between roadside NO_x and monitored local background concentrations is demonstrated. Relationships between roadside NO_x and other model variables have been shown to be strongly influenced by the quality and resolution of background concentrations of NO_x, i.e. if it were based on monitored data or modelled prediction. The paper proposes a direct method of using site-specific fundamental diagrams for splitting traffic data into four traffic states: free-flow, busy-flow, congested, and severely congested. Using BRT models, the density of traffic (vehicles per kilometre) was observed to have a proportional influence on the concentrations of roadside NO_x, with different fitted regression line slopes for the different traffic states. When other influences are conditioned out, the relationship between roadside concentrations and ambient air temperature suggests NO_x concentrations reach a minimum at around 22 °C with high concentrations at low ambient air temperatures which could be associated to restricted atmospheric dispersion and/or to changes in road traffic exhaust emission characteristics at low ambient air temperatures. This paper uses BRT models to study how different critical factors, and their relative importance, influence the variation of roadside NO_x concentrations. The paper highlights the importance of either setting up local background continuous monitors or improving the quality and resolution of modelled UK background maps and the need to further investigate the influence of ambient air temperature on NO_x emissions and roadside NO_x concentrations.

1. Introduction

Road transport has become a worldwide major source of emissions. In 2010, 40.5% of Oxides of Nitrogen (NO_x) emissions came from road transport in the European Economic Area countries (EEA,

* Corresponding author.
E-mail address: ts11ass@leeds.ac.uk (A. Sayegh).

http://dx.doi.org/10.1016/j.atmosenv.2015.12.024
1352-2310/

2011). The major traffic-related air quality pollutants of concern in the UK are NO_x and particulate matter. Air quality pollutants with concentrations higher than certain limit values are associated with weakened human health and a life expectancy reduction of each person (DEFRA, 2007). As a result, air quality is an ongoing research issue in the UK and worldwide.

Describing the complete pathway from the generation of emissions, dispersion, and chemical transformations that results in ambient air pollution concentrations is extremely challenging because of the highly variable temporal and spatial processes at work (Mayer, 1999). Road transport is now a significant source of emissions. Statistics suggest road transport accounts for 32% (almost one third) of NO_x emissions in the UK (DEFRA, 2014). Factors such as the level of traffic flow and vehicle fleet composition (fuel type, age, presence and performance of emission control technologies) influence the source intensity. Road type also influences these factors as a result of, for example, different road widths, road lanes, road gradients, speed limits, traffic management restrictions, trips characteristics, and driving conditions. The transport of road traffic emissions to ambient air pollution is driven initially by the turbulent wake following a vehicle (Uhrner et al., 2007) and the movement of preceding vehicles. Subsequently, atmospheric processes influenced by the surrounding topography and meteorological parameters dilute, disperse, and transform pollutants through chemical reactions (Mudakavi, 2010). Meteorological parameters such as wind speed and wind direction, air temperature, humidity, solar radiation, and boundary layer height are often correlated; for example an increase in air temperature is generally associated with an increase in boundary layer height and vice versa (Mudakavi, 2010).

Experimental, dispersion, and statistical modelling techniques often based on linear or non-linear regression models which relate traffic volume and meteorological variables to concentration measurements at a specific site are used to predict and forecast the temporal and spatial variations in air quality (Aldrin and Haff, 2005). Traditional statistical models such as linear regression are often relatively poor in terms of predictions (De'ath, 2007). Modern statistical modelling techniques (e.g. neural networks, independent component analysis, boosting, and random forests) are being developed and used both as diagnostic models to explore the relationship between responses and influential factors and/or as predictive models. Difficulty in doing so becomes more apparent when data is dominated by features with nonlinearities and/or complex interactions.

Classification and regression tree methods are now put forward as methods to both explain interactions and offer predictions in a number of fields. Boosted Regression Trees (BRT) is one classification and regression method. BRT is a type of additive regression model in which its individual terms are simple decision trees (Elith et al., 2008). The method combines the strengths of two algorithms: regression trees that relate a response to its predictors by recursive binary splits and a boosting technique which combines many simple models to give improved predictive performance (Elith et al., 2008). BRT is able to cope with many types of responses and predictors such as numeric or categorical, and loss functions such as Gaussian, Laplace, Bernoulli, and Poisson (Ridgeway, 2012). The method has been applied in ecological studies by, e.g., Cappo et al. (2005) and Elith et al. (2008) as well as Moisen et al. (2006) who have compared Generalised Additive Models (GAM) to tree-based methods and have suggested the use of GAM as a supplement to classification and regression trees. In the field of ecological modelling, De'ath (2007) has shown that BRT, unlike many other regression methods, can be used for both prediction and explanation of the underlying relationships between response and predictors. The two main outputs of a BRT model which are the partial dependence plots and the variable relative influence measure can be used together for interpretation. Friedman (2001) has suggested that while these might not provide a complete description, they can at least give an insight of the relation between the response and the predictors in question.

Carslaw and Taylor (2009) have developed and applied the BRT in the air quality field, studying hourly concentrations of NO_x close to the international Heathrow airport in the UK, in an attempt to understand the complex interactions between the different source types, particularly to distinguish between aircraft and road transport emission sources. They have also inferred that the outcomes of the applied BRT are reasonable by comparing it with results that include an independent analysis of source emissions. Specifically, they have compared the result of the mean contribution of heavy aircraft to NO_x concentrations derived from the BRT model with that derived from an independent aircraft plume sampling study and have found them to be in good agreement.

This paper investigates how background concentrations of NO_x, traffic density, and meteorological conditions influence roadside NO_x concentrations monitored at three different sites. The sites are adjacent to different road types, which exhibit contrasting traffic flow characteristics, i.e. speed and intensity of more polluting acceleration events, as well as different local dispersion conditions due to the presence or not of obstructions such as buildings that may restrict air-flows. The paper utilises a coordinated database of traffic flow and meteorological measurements, background air pollution measurements and/or predictions, throughout at least one calendar year, to investigate the influence of the different parameters on the monitored concentrations of NO_x adjacent to an urban road, open motorway, and motorway tunnel. The paper contrasts the results obtained for each of the three road types and offers explanation of the results, notably the relative importance of parameters and their interactions in explaining the variation of roadside NO_x concentrations.

2. Method

2.1. Study sites and data used

Three sites have been selected for study, each with different road, traffic, and meteorological characteristics. These are an ‘urban site’ on the A660 Headingley in Leeds, UK an ‘open motorway site’ on the M25 Staines, UK, and a ‘motorway tunnel site’ on the M25 Bell Common, UK. Air quality monitoring stations are located on each site and their measurements are subject to Automatic Urban and Rural Network (AURN) quality control procedures (AEAT, 2003). Fig. 1 shows the geographic locations of each of the local monitoring stations used in this study and Table 1 provides detailed description of the road characteristics and roadside monitoring station of each site.

The urban study site has an air quality monitoring station at the kerbside of the A660 arterial in Headingley, Leeds, UK. Headingley is approximately two miles to the North-West of Leeds city centre and comprises residential and retail premises, a sporting stadium, and light industrial properties. The area is heavily trafficked with commuting through traffic and demand for local services which regularly exceeds the capacity of the network during morning and evening peak-periods. Loop detector 2008 traffic data near the monitoring station shows an average bi-directional traffic flow of 775 veh h^{-1} and average traffic speed of 44 km h^{-1}. The monitoring station is located on the A660 Otley road which has buildings on both sides that form a symmetrical regular street canyon environment (Vardoulakis et al., 2003) with an aspect ratio (height to width) of around 0.8. The equipment is situated 100 m south of the A660/B6157/Wood lane junction, 15 m south of the Shire Oak minor

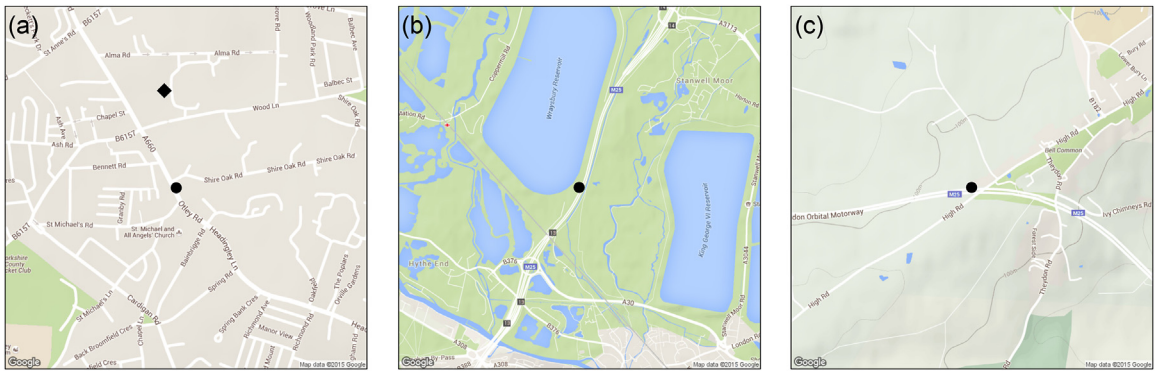

Fig. 1. Circles indicate geographic locations of each of the (a) urban, (b) open motorway, and (c) motorway tunnel local roadside monitoring stations. 'Diamond' in (a) indicates the location of the local background monitoring station at the urban site, 0.25 km away from the urban local roadside monitoring station.

Table 1
Site and roadside monitoring stations description for each of the three sites.

Description \ site		Urban	Open motorway	Motorway tunnel
Road characteristics	Road type	Arterial urban road	Rural motorway	Suburban motorway
	Number of lanes per direction	1	5	3
	Speed limit (km h^{-1})	48	Variable: 80/96/112	112
Roadside air quality monitoring station	Location/grid reference (WGS84)[a]	Southbound: (427987, 436047)	Clockwise J13- J14: (502806, 173570)	Clockwise J26- J27: inside tunnel
	Distance from traffic lane edge (m)	1.75	5.5	Not available
	Installation date	September 1998	June 1995	January 2007
	Relevant measured pollutants[b]	NO, NO_2, and NO_x	NO, NO_2, NO_x, and O_3	NO, NO_2, NO_x, O_3
	Measurement method for NO, NO_x	Chemiluminescence	Chemiluminescence	Chemiluminescence
	Measurement method for O_3	Not measured	Ultra violet absorption	Ultra violet absorption
	Study period	Jan 2008–Jun 2009 (18 months)	Jan 2006–Dec 2010 (48 months)	Jan 2007–Sep 2008 (21 months)
	Number of hourly measurements	13,127	43,824	15,335
	Data collection rate for NO_x (%)	96%	97%	99%

[a] All WGS84 grid references in the form of (Easting, Northing).
[b] NO_x [ppb] measurements are converted to NO_x[μg m^{-3}] using: $NO_x[\mu g\ m^{-3}] = \{(NO[ppb] + NO_2[ppb]) * 12.187 * M\}/(273.15 + T)$ where M is the NO_2 molecular weight (46.0055 g mol^{-1}); T is the hourly ambient air temperature (°C) measured at each site; atmospheric pressure of 1 atm (1013 mb) is assumed.

road, and 30 m north of the T-junction with St. Michael's Road. These junctions provide breaks in the street canyon walls that influence air-flows and the dispersion of pollutants but also disturb the traffic flow and the emissions generated. The A660/B6157/Wood lane junction also operates an Urban Traffic Management and Control (UTMC) application where traffic signal time alters depending on time of day and traffic demands. The University of Leeds has established a local background monitoring station approximately 0.25 km away from the roadside monitoring station which measures concentrations accumulating from emission sources other than the road traffic. For this urban site, coordinated measurements of roadside air quality, background concentrations, traffic conditions, and prevailing meteorological conditions are available between years 2008 and 2009. Further details on collected data are provided in Table 2.

The other two monitoring sites were operated by the Transport Research Laboratory (TRL) on behalf of the UK Highways Agency (HA) and are subject to quality assurance and quality control procedures used for AURN (TRL, 2011). The open motorway M25 Staines monitoring station is located on the western side of the London orbital motorway. Encircling greater London, the motorway is one of the busiest motorways in UK. Loop detector 2010 traffic data near the monitoring station shows an average bi-directional traffic flow of 8374 veh h^{-1} and average traffic speed of 100 km h^{-1} (HATRIS, 2007). The open motorway M25 Staines section has been upgraded to a controlled motorway with Variable Speed Limits (VSL) in November 2005 (Highways Agency, 2014) with the hard shoulder running only in case of emergency. VSL imposes different speed limits at different times in response to traffic data. Coordinated measurements are available for this site between 2006 and 2010.

The third motorway tunnel site is located on the clockwise bore of the M25 Bell Common. Clockwise loop detector 2008 traffic data near the monitoring station shows an average traffic flow of 2476 veh h^{-1} and average traffic speed of 102 km h^{-1} (HATRIS, 2007). The curved tunnel is longitudinally ventilated by the 'piston effect' of the movement of traffic. Its traffic bounds are separated by a continuous curtain wall (TRL, 2007). Coordinated measurements are available for this site between 2007 and 2008. Further details of the two motorway monitoring stations including site photographs can be obtained from TRL (2011).

Due to the absence of continuous background monitors near the two motorway sites, background NO_x are instead extracted from the annual modelled 1 × 1 km grid background maps reported by the Department of Environment, Food, and Rural Affairs (DEFRA) based on UK Pollution Climate Mapping (PCM) (DEFRA, 2012) with the emission sector from the M25 motorway removed to avoid double counting. In order to account for the diurnal variability in the background concentrations, variable background profile factors derived by the HA are applied to the annual averages. In the absence of background monitoring stations, low spatial and temporal resolution of modelled background maps are the typical alternative source of background data used in local air quality review and assessment. To investigate the extent of the limitation of such data, the modelled background map for the urban site is also extracted in order to be compared and analysed with monitored background NO_x.

For the open motorway and motorway tunnel sites, traffic flow

Table 2
Background data, traffic data, and meteorological data description for each of the three sites.

Description\site		Urban	Open motorway	Motorway tunnel
Background concentration data	Data source	Monitored: local background monitoring station Modelled: DEFRA's background maps	DEFRA's background maps	DEFRA's background maps
	Grid reference (WGS84)[a]	Monitored: (427956, 436293) Modelled: (427500, 436500)	Modelled: (501500, 173500)	Modelled: (544500, 201500)
Traffic data	Data source	Leeds City Council	TRADS and JTDB databases	TRADS and JTDB databases
	Inductive loops location/grid reference (WGS84)[a]	Southbound: (428331, 435753) Northbound: (428372, 435734)	Clockwise: (502939, 174014) Anticlockwise: (502972, 174051)	Between J26 and J27
	Two-way annual average daily traffic (vehs/day) (DfT, 2014)	17,000 (2008 counts)	186,000 (2010 counts)	115,000 (2008 counts)
Meteorological data	Data source	Central Leeds met station	■ Relative humidity and air temperature: roadside air quality monitoring station ■ Wind speed and wind direction: Met Office MIDAS stations	
	Station location/grid reference (WGS84)[a]	(430296, 432323)	(506296, 157956)	(518537, 177257)
	Met station direct distance from air quality monitoring station	4.5 km	16 km	35 km
	Direction from air quality monitoring station	South-East	South-West	South-West
	Relative humidity and air temperature sensor height above ground level (m)	8	2.1	2.1
	Met station wind vane and anemometer height above ground level (m)	12	Unobstructed: 10 obstructed: >10	Unobstructed: 10 obstructed: >10

[a] All WGS84 grid references in the form of (Easting, Northing).

and traffic speed are extracted from the Traffic Flow Data System (TRADS) and Journey Time Database (JTDB) databases known collectively as the HA Traffic Information System (HATRIS, 2007). As for the meteorological data, relative humidity and air temperature monitored at the M25 Staines monitoring station are used for both sites since these are not recorded by the M25 Bell Common one and the two sites, which are 50 km distant, are considered within the same topographical area of the City of London. Hourly wind speed and wind direction measured to the nearest 10° are extracted from the closest available meteorological stations from the British Atmospheric Data Centre using the UK Met office Integrated Data Archive System (MIDAS) Land Surface Stations database (Met Office, 2012). As detailed in Table 2, two different meteorological stations are selected for each of the open motorway and motorway tunnel sites; the selection is based not only on their proximity to the monitoring stations but also on the data availability throughout the period of air quality monitoring data which is different between the two sites.

2.2. Data summary

The distribution of hourly Ozone (O_3), Nitric Oxide (NO), roadside NO_x, and background NO_x for each study site as well as the prevailing wind speed recorded by the selected meteorological station are summarised in Table 3.

The diurnal roadside and background NO_x profiles alongside traffic density for both traffic directions are presented in Fig. 2 to illustrate the relationship between traffic activity and ambient concentrations. Hourly traffic density is defined as the amount of traffic occupying one kilometre of the road link. It is a function of hourly traffic flow and speed and is a useful, single parameter for describing vehicle activity. Throughout the paper, 'near-bound' refers to the traffic flow direction adjacent to the monitoring station and 'distant-bound' denotes the opposite traffic flow on the other side of the road. Fig. 2 highlights the difference in NO_x concentrations between the urban, open motorway, and motorway tunnel sites and compares the diurnal profile of NO_x concentrations with the diurnal profiles of traffic densities at both the 'distant-bound' and 'near-bound' of each site. For the tunnel site, only the 'near-bound' data are highlighted since the tunnel is separated by a continuous curtain wall limiting the influence of 'distant-bound' traffic on the in-tunnel concentrations.

Hourly-averaged roadside NO_x are extremely high inside the motorway tunnel reaching concentrations greater than 1000 μg m^{-3} during the morning peak. Ventilation in unidirectional road tunnels is typically assisted by the piston effect of moving vehicles where vehicles induce their own airflow in the same direction of the tunnel (NHMRC, 2008). Dispersion of in-tunnel emissions is largely restricted and the dilution of airborne pollution is limited. Chemical reactions are limited as well due to the absence of sunlight inside the tunnel restricting the formation of O_3.

Slightly higher concentrations of roadside NO_x on the open motorway site are observed compared to the urban site as a result of higher traffic flows, traffic densities, and proportions of Heavy Good Vehicles (HGVs) on motorways compared to urban roads. Concentrations on the urban site are still considered relatively high despite a maximum hourly average traffic density of only 24 veh km^{-1} (Fig. 2a). The influence of buildings on both sides of the road is expected to restrict dispersing airflows, resulting in the relatively high concentrations of roadside NO_x (Wood et al., 2009). The increase in traffic densities at all sites during daylight hours leads to an increase in roadside concentrations of NO_x. However, morning peaks tend to always have higher values even when density values are lower compared to the afternoon peak. The atmosphere is more stable, with typically lower wind speeds in the morning compared to the evening period. This can also be associated with the air temperature profile and boundary layer height, which rises with increases in air temperature (Mudakavi, 2010).

The traffic flow-speed fundamental diagrams of the three sites are shown in Fig. 3. Bell et al. (2006) have classified urban traffic data into four traffic states using the k-means algorithm. This study, however, proposes a new traffic state identification method which categorises traffic into four traffic states using site-specific traffic flow, traffic speed, and traffic density data. The three site-specific speed-density, flow-density, and flow-speed diagrams are used as

Table 3
Summary of hourly data for each of the study sites.

Site	Statistic	O_3 [μg m^{-3}]	NO [μg m^{-3}]	NO_x [μg m^{-3}]	Background NO_x [μg m^{-3}]		Wind speed [m s^{-1}]
					Monitored	Modelled	
Urban	Minimum	–	0.0	2.6	0.3	14.6	0.2
	Mean	–	49.0	129.3	39.0	27.5	3.1
	Median	–	34.0	101.5	25.4	26.0	2.7
	Maximum	–	593.6	1258.8	729.7	50.0	12.1
Open motorway	Minimum	0.1	0.0	0.1	–	14.8	0.0
	Mean	26.7	72.3	173.5	–	25.2	2.6
	Median	21.7	44.3	123.8	–	24.3	2.1
	Maximum	139.1	878.4	1731.0	–	36.1	13.4
Motorway tunnel	Minimum	0.1	4.4	24.0	–	15.3	0.0
	Mean	1.5	533.1	985.1	–	21.4	2.4
	Median	1.3	493.0	908.6	–	22.5	2.1
	Maximum	14.9	4172.0	7071.0	–	27.5	9.3

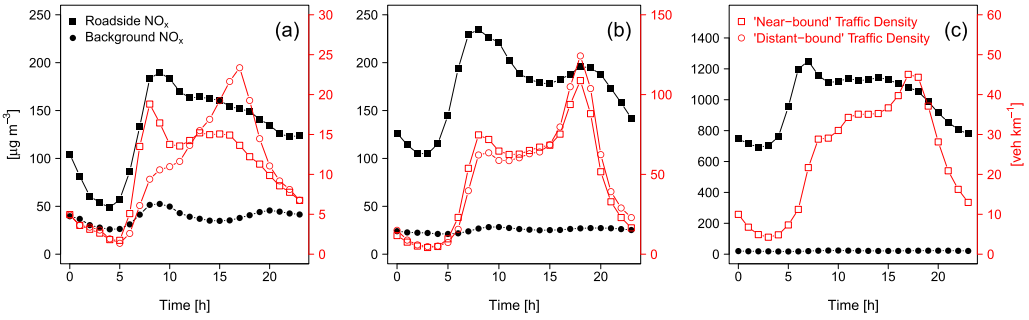

Fig. 2. 'Near-bound' and 'Distant-bound' traffic density [veh km^{-1}] (right y-axes) and diurnal profile of roadside concentrations of NO_x and background concentrations of NO_x [μg m^{-3}] (left y-axes) at the (a) urban, (b) open motorway, and (c) motorway tunnel sites.

a basis to splitting data into the different traffic states. Approximate critical points are identified based on site-specific three fundamental diagrams which signal the start of a new traffic state. The four traffic states are defined as follows.

- Free-flow is characterised by very low traffic densities and high traffic speeds and is identified using the speed-density fundamental diagram (Fig. A.1a, Fig. A.2a, Fig. A.3a). Approximate traffic density at which traffic speed starts to drop is defined as the critical traffic density signalling the start of the busy-flow traffic state.
- Busy-flow is characterised by flows reaching or approaching the capacity of the link along with relatively high traffic speeds and is identified using the flow-density fundamental diagram (Fig. A.1b, Fig. A.2b, Fig. A.3b). Approximate traffic density at which traffic flow starts to drop is defined as the critical traffic density signalling the start of the congested-flow traffic state.
- Congested-flow is described by a drop in traffic flows after reaching its maximum and is identified using the flow-speed fundamental diagram (Fig. A.1c, Fig. A.2c, Fig. A.3c). Approximate traffic speed at which traffic flow starts to decrease is defined as the critical point signalling the start of the severely congested state.
- Severely congested-flow is described by high traffic densities and low traffic speeds i.e. points below the critical traffic speed identified in the congested-flow traffic state.

The final four traffic states obtained for each site are shown in Fig. 3 but Appendix A describes in details how the traffic data is used to identify them, based on the above method, at each of the three sites. The urban and open motorway sites exhibit special features due to the adaptive traffic management systems in-place,

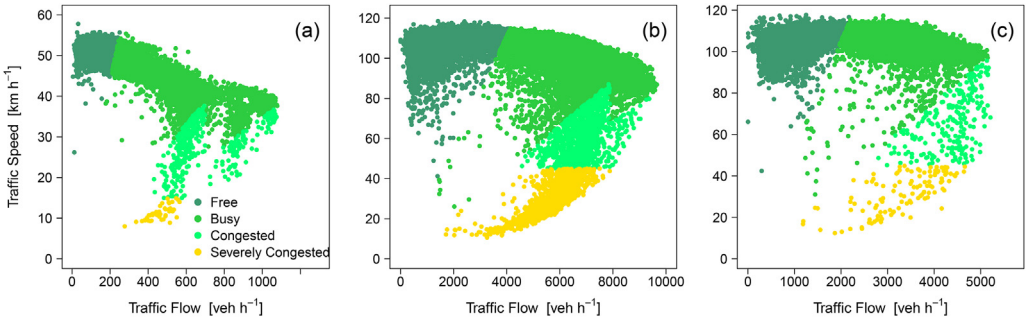

Fig. 3. Traffic flow [veh h^{-1}] vs. traffic speed [km h^{-1}] scatterplots on the 'near-bound' of the (a) urban, (b) open motorway, and (c) motorway tunnel sites. Methodological details of the traffic state splits at each site are available in Appendix A.

namely the UTMC signals in the vicinity of the urban junction and the VSL operation on the open motorway section, respectively. The operation of both traffic management technologies varies the road capacities at different times, giving rise to the jagged cuts observed between busy and congested states for data collected at the urban and open motorway sites.

In calculating traffic densities and deriving traffic states, it has been observed that traffic density data points between the 5th and 95th percentile are restricted to free-flow and busy-flow traffic states. The analysis of the influence of traffic density on the variation of roadside NO_x is thus restricted to these two traffic states within this paper.

2.3. BRT model development

In order to investigate how background concentrations of NO_x, traffic density, and meteorological conditions influence roadside NO_x concentrations, a BRT model has been fitted to the hourly roadside NO_x at each of the three study sites. This section starts by highlighting the predictors that have been included in all the three site-specific BRT models as well as the reason behind their inclusion. Variable/predictor selection is inherently achieved in BRT because the algorithm largely ignores irrelevant or non-informative predictors (Elith et al., 2008). Although variable selection methods exist such as forward stepwise selection or backward elimination, this paper makes use of the BRT model's ability to select important variables and minimise or even discard irrelevant ones. The first predictor is the background NO_x. Rather than subtracting it from the roadside NO_x, it is included as a predictor (Richmond-Bryant et al., 2009; Pan et al., 2013) to account for background levels, domestic and/or industrial sources as well as neighbouring transport emission sources. This will allow for a better understanding of the relationship between background and roadside NO_x and will highlight the difference between using monitored background for the urban site and modelled background maps for the open motorway and motorway tunnel sites. This will also provide an understanding of how studying the traffic contribution to roadside NO_x can be influenced by the quality of background NO_x data.

As described in section 2.2, traffic density can replace both traffic flow and traffic speed. Concentrations are known to decrease with the increase of distance between source-receptor (Hitchins et al., 2000). Hence, the impact of each traffic lane is dependent on its proximity to the monitoring equipment. Since traffic counts are only available for each traffic bound rather than each lane, traffic density for each bound is separately included in all models except for the motorway tunnel model where the monitoring station is located inside the clockwise tunnel bore which is wall-separated from the anti-clockwise bore and hence the anti-clockwise traffic is assumed to not influence the quality of air on the clockwise direction. In order to capture the variation of traffic emission strength caused by different vehicle types and/or engine sizes, hour of the day and day of the week have been integrated as two predictors in each of the three models.

Fluctuations in meteorology have a significant impact on air quality through its interference with source emissions. Wind speed, wind direction, relative humidity, and air temperature are considered as the meteorological predictors in the models. Meteorological parameters are also considered as predictors for the motorway tunnel site in order to understand the extent of their influence on in-tunnel air quality and whether these can be neglected when, for instance, deriving emission factors from in-tunnel air quality data. Day of the year has also been integrated in the model to capture any seasonal meteorological variation not accounted for in the meteorological predictors such as boundary layer height and solar radiation.

A BRT model is developed for each site under study where the response or the dependent variable is the hourly-averaged roadside NO_x concentrations in $\mu g\ m^{-3}$ and the predictors or the independent variables are those related to background NO_x, traffic density, and meteorological parameters as shown in Table 4 below.

For the urban site, two BRT models are developed. The first one uses local monitored background NO_x while the second model uses modelled background maps extracted for the same site from DEFRA's model similar to those used for the open motorway and motorway tunnel sites. These two models are compared to investigate the relative influences of high resolution monitored background NO_x and low resolution modelled background NO_x maps.

By selecting the response and predictor variables, three model parameters explained in details in Ridgeway (2012) need to be specified for model fitting purposes: the learning rate 'lr', interaction depth 'tc', and number of trees 'nt'. The learning rate is a shrinkage parameter applied to each tree during the expansion process to shrink the contribution of each tree as it is added to the model, while the interaction depth is the maximum depth of variable interactions controlling the size of the trees. For instance, an interaction depth of 1 implies an additive model, 2 implies a model with up to 2-way interactions and so on. Based on these two parameters, the tree number required for optimal prediction is determined (Elith et al., 2008) using one of three methods explained in Ridgeway (2007). In this study, the volume of data (18, 48, and 21 months of hourly data at each of the urban, open motorway, and motorway tunnel sites) allows the use of Cross-Validation (CV) method to develop and test a model and the test method to evaluate the model predictive performance of different models with varying interaction depth parameter. Briefly, the CV method has the advantage of using all the data for both training and validation by repeating the process v-times on different combination of subsamples and calculating the mean performance of the v-models. The v-fold CV method partitions the data into v-subsets where v-models are built based on the 'v − 1' subsets and the model performance is tested based on the last remaining subset (Ridgeway, 2007); this is repeated several times depending on the number of trees model parameter and the optimum number of trees with the minimal error is determined.

All models are fitted in R using the 'gbm' or 'generalised boosted machine' package version 1.6–3.2 (R Core Team, 2012; Ridgeway, 2012). The full dataset used for each site is first divided into two where 70% is used as the training set to develop the model and 30% is used as the independent testing set. The latter is mainly used to determine the optimum interaction depth which is best identified using data not involved in the model development process (Elith et al., 2008). Also, a random component is injected in order to improve the prediction performance. This is achieved using a bag fraction of 0.5 or 50% of the training set observations randomly selected to fit each consequent tree (Friedman, 2002).

For each site, multiple 'gbm' models are fitted using different combinations of learning rates and interaction depths and the optimum number of trees via the 5-fold CV method is determined. In line with Carslaw and Taylor (2009), a Gaussian error distribution is assumed for all models. Learning rates of 0.1 and 0.001 are tested with different interaction depths while taking into consideration that a slower learning rate requires an increase in the interaction depth (Elith et al., 2008). Candidate models with the optimum number of trees are then evaluated using the independent set of observations and a range of evaluation statistics such as the mean gross error, root mean squared error, and correlation coefficient. BRT models with an optimal interaction depth of 3 have been used for the urban and open motorway sites and a BRT model with an interaction depth of 2 has been used for the motorway tunnel site.

Table 4
Predictors used to model hourly roadside NO_x concentrations [$\mu g\ m^{-3}$].

Predictors	Description	Variable type
$[NO_x]_{bkgd}$	Background NO_x concentrations in $\mu g\ m^{-3}$	Continuous
$K_{near\text{-}bound}$	Hourly traffic densities on the bound nearer to the monitoring station in veh km^{-1}	Continuous
$K_{distant\text{-}bound}$	Hourly traffic densities on the opposite bound to the monitoring station in veh km^{-1}	Continuous
$\bar{U}$	Hourly wind speed in m s^{-1}	Continuous
θ	Hourly wind direction in degree (°)	Continuous
RH	Hourly relative humidity in percentage (%)	Continuous
T	Hourly air temperature in degree Celcius (°C)	Continuous
JD	Day of the year	Discrete (1–365/366)
Day	Day of the week	Discrete (1–7)
H	Hour of the day	Discrete (1–24)

3. Results and discussion

Studying the relative influence of individual predictors on a specific response is amongst the most useful descriptions of the data under study. The measure is based on the number of times a variable is selected for splitting in the tree weighted by the squared improvement to the model as a result of each of the splits and averaged for all trees (Friedman and Meulman, 2003). The influence is scaled and the sum of all the predictors' relative influence adds to 100 where low values indicate low influence compared to higher values indicating stronger influence on the variation of NO_x. Table 5 shows the results of relative influence of predictors at each of the three sites with the urban site being represented by two BRT models, depending on whether monitored background NO_x or modelled background NO_x is used.

Table 5 highlights the difference in the relative influence of the monitored background concentrations of NO_x (37%) and the modelled ones (4%) at the urban site. Table 5 also highlights the high relative influence of both the near-bound traffic density (average of 18% across all sites) and the air temperature (13%). As a result of the differences observed in the relative influence of the monitored and modelled background NO_x of two urban BRT models and as a result of the relatively high influence (mean values above 10%) of traffic density and air temperature variables, these three parameters are studied in further detail using partial dependence plots. A partial dependence plot (Friedman, 2001) is a visualisation of the fitted function for each variable and is used to identify the dependence of roadside NO_x concentrations on each predictor after accounting for the average effect of the other remaining predictors. Also, two relatively high influencing variables (mean values above 10%) are the wind speed and wind direction. In order to study the dependence of roadside NO_x on the interaction of wind speed and wind direction, these two key predictors are further studied using 2-way partial dependence plots.

The first parameter to be studied is the background concentration of NO_x since studying the traffic contribution on roadside NO_x depends on the quality and resolution of background NO_x subtracted. Partial dependence plots are plotted for the monitored background NO_x at the urban site (Fig. 4a) and for the modelled background NO_x at each of the open motorway (Fig. 4c) and motorway tunnel (Fig. 4d) sites. A gradient approximately equal to 1 has been observed at the urban site where hourly local monitored background concentrations are used. Because only 0.1% (12 data points out of 11430 data points) of local background concentrations fall above 512 $\mu g\ m^{-3}$, a straight (zero gradient) line has been observed in the partial dependence plot of Fig. 4a which shows that the BRT was unable to predict NO_x concentrations in the range of 512–730 $\mu g\ m^{-3}$. For the open motorway and motorway tunnel sites, either very steep gradients or negligible gradients are observed at different background levels. Also, the relative influence of monitored background NO_x at the urban site is 37% while it is only 3% and 6% for the open motorway and motorway tunnel sites for which background maps are used.

In order to determine whether the reason behind this difference can be explained as a result of different road types or different background NO_x data source used, another BRT model is developed for the urban site using DEFRA's background maps of the urban area while keeping all other variables exactly the same. Monitored background NO_x on the urban site ranges between 0.3 and 730 $\mu g\ m^{-3}$ (Fig. 4a) while that of the modelled background NO_x ranges only between 15 and 50 $\mu g\ m^{-3}$ (Fig. 4b). Fig. 4b also shows that the relationship between roadside NO_x and modelled background NO_x becomes very similar to that of the open motorway site. In reference to Table 5, the relative influence of background NO_x at the urban site drops from 37% for the original model to 4% when using background maps. The lower range of modelled background NO_x compared to monitored background NO_x, the different relative influence obtained for monitored versus modelled

Table 5
Relative influence (%) of individual predictors on roadside NO_x concentrations [$\mu g\ m^{-3}$].

BRT model predictors	Urban		Open motorway	Motorway tunnel	Mean values[a]
	Monitored $[NO_x]_{bkgd}$	Modelled $[NO_x]_{bkgd}$			
$[NO_x]_{bkgd}$	37%	4%	3%	6%	4%
$K_{near\text{-}bound}$	19%	20%	10%	25%	18%
$K_{distant\text{-}bound}$	6%	10%	11%	–	11%
$\bar{U}$	12%	23%	9%	3%	12%
θ	9%	8%	31%	5%	15%
RH	4%	7%	8%	6%	7%
T	5%	10%	11%	18%	13%
JD	4%	12%	8%	17%	12%
Day	3%	4%	5%	10%	6%
H	1%	2%	4%	10%	5%

[a] Mean values exclude the urban site model with monitored $[NO_x]_{bkgd}$.

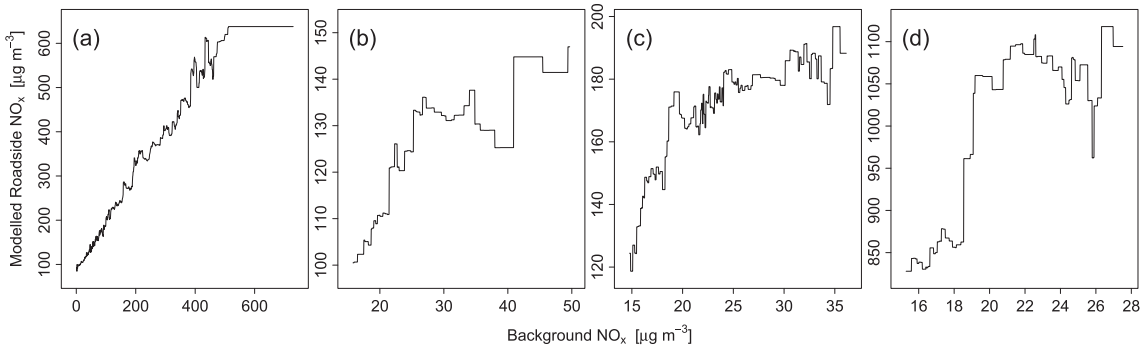

Fig. 4. Partial dependence of roadside concentrations of NO_x [µg m^{-3}] on (a) monitored, (b) modelled background concentrations of NO_x [µg m^{-3}] at the urban site, (c) modelled background concentrations of NO_x [µg m^{-3}] at the open motorway site, and (d) modelled background concentrations of NO_x [µg m^{-3}] at the motorway tunnel site.

background NO_x in the two BRT models for the same site, and the different partial dependence plots obtained for the exact same site when using modelled annual background maps highlight an issue associated with the use of the aggregate modelled background maps. Using these could produce high uncertainties and distort not only the results related to the background NO_x predictor but also the results of all other predictors as well; this shows that the obtained results (i.e. relationships between response and predictors) can be dependent on the quality and the temporal and spatial resolution of background data as a start.

This necessitates either setting up local background monitors at appropriate locations away from emission sources or otherwise enhancing the background maps used by DEFRA by improving their time/space resolution, improving the PCM model used for their calculation, and/or improving their validation with existing local background measurements. This would contribute to better results of quantifying the impact of other variables on air pollutants and produce more rigorous analysis results regarding the quantitative contribution of road traffic on NO_x concentrations.

The second predictor to be studied is the traffic density and particularly that of the 'near-bound'. The relative influence of the 'near-bound' traffic density is higher than that of the 'distant-bound' on the urban site while both bounds of the open motorway site show very similar results. This is mainly because the traffic density profiles of the open motorway are very similar at both bounds as shown in Fig. 2b. Fig. 5 shows the partial plots for the 'near-bound' traffic density at each site. These plots might not be a perfect representation of the effects of each variable on roadside NO_x, especially in the case of strong interactions. Hence, the relative difference in the slopes of the visualised fitted functions at the three different sites is considered a better approach to analysing these plots compared to the analysis of the absolute values. The calculated slopes (or rates) represent the increase of roadside NO_x in µg m^{-3} per unit increase of traffic density in veh km^{-1}.

Slopes of the fitted regression line to the data points between the 5th and 95th percentile of measured traffic density are calculated to highlight the differences of influence on roadside NO_x at the different road types. Rates of 5, 0.51, and 15 are calculated for the urban, open motorway, and motorway tunnel sites, respectively. The highest rate for the motorway tunnel which is approximately 3 times and 30 times that observed at the urban site and open motorway site, respectively, highlights the restricted dispersion and the limited impact of meteorological conditions on the variation of NO_x within the motorway tunnel. Also, the high rate observed at the urban site which is approximately 10 times that observed at the open motorway site can be a result of local buildings inducing air recirculation and restricting dispersion on the urban site, a criteria which is absent on open motorways resulting in the lowest rate of 0.51. It can also be a result of more frequent stop-start driving conditions on urban sites compared to high-speed motorways and/or a result of the closer distance from the traffic lane edge to the urban monitoring station (1.75 m) compared to that of the open motorway one (5.5 m).

The rates for the 'distant-bound' traffic density drop to 2 and 0.17 at each of the urban and open motorway sites, respectively. This is approximately 3 times lower compared to that observed at both the 'near-bound' traffic density of the urban and open motorway sites. This emphasises the stronger influence of traffic on the bound closer to the receptor at both road types. While the coefficient of determination, R^2, which indicates how well the regression line fits the data, remains respectively high (0.7) for the traffic density on the northbound of the urban site, it drops to 0.25 for the traffic density on the anti-clockwise direction of the open motorway. This suggests that at a 5-lane distance from the monitoring station, the influence of traffic emissions on roadside NO_x becomes much less explained in a linear regression context compared to two-lane distance despite the drop in its absolute value.

Fig. 5 highlights the influence of each traffic state on the variation of roadside NO_x. Two regression lines are fitted to the 'near-bound' traffic density data points corresponding to the free-flow traffic state and to the busy-flow traffic state, respectively. On all the three sites, the calculated rates of traffic density on roadside NO_x are higher when the increase in traffic density occurs within the free-flow conditions rather than busy-flow conditions. For each of the urban, open motorway, and motorway tunnel sites, the calculated rates within free-flow conditions are approximately 2.5 times, 1.2 times, and 2.8 times that observed for the busy-flow conditions. This observation can be a result of further dispersion of emissions induced by following vehicles when the space-headways are shorter during busy-flow compared to free-flow conditions or a result of the direct impact of different average vehicular emissions rates, i.e. higher grams per hour produced by road vehicles during free-flow compared with busy-flow conditions as a result of different driving conditions (traffic speeds, accelerations, etc …). While there were insufficient congested state and severely congested state data to unambiguously line-fit (both <5% available data and subject to significant variance), initial inspection, limited confidence acknowledged, suggests that the influence of traffic density most likely reduces, and maybe even levels out or breaks down completely in these two states. However, because of insufficient data above 44, 265, and 143 veh km^{-1} for each of the urban, open motorway, and motorway tunnel sites respectively (0.1% of traffic densities), the BRT model of each site is unable to predict NOx concentrations for traffic densities above these points resulting in a straight (zero gradient) lines.

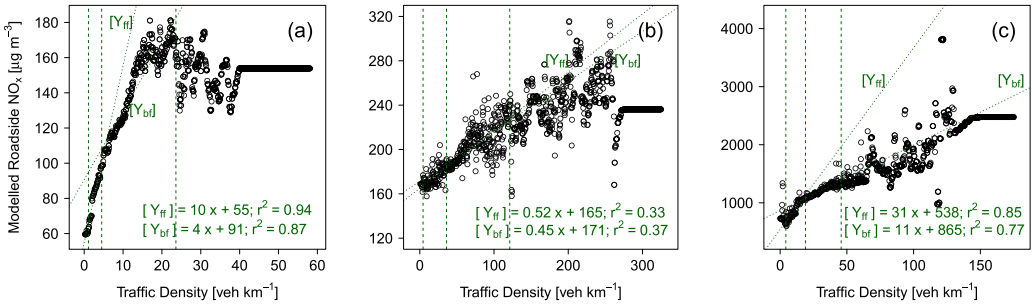

Fig. 5. Partial dependence of roadside concentrations of NO_x [µg m^{-3}] on 'near-bound' traffic densities [veh km^{-1}] at each of the (a) urban, (b) open motorway, and (c) motorway tunnel sites. The three vertical dotted lines represent the 5th percentile of the measured traffic density, traffic density point of split between free-flow and busy-flow, and 95th percentile of the measured traffic density. [Y_{ff}] fits the free-flow data points with traffic densities higher than the 5th percentile. [Y_{bf}] fits the busy-flow data points with traffic densities less than the 95th percentile. $Y = mx + c$, r^2 is the equation of the two fitted lines where m is the gradient, c is the y-intercept, and r^2 is the coefficient of determination.

The third predictor to be studied through the 1-way partial dependence plots is the ambient air temperature. Fig. 6 shows the partial plots of the air temperature for each site. Linear regression lines are fitted for the data points between the 5th and 95th percentile of the monitored air temperature to study the influence of air temperature on the variation of NO_x. Within these ranges of air temperature at each site, a sharp decrease in NO_x levels is noticed with the increase of air temperature with −1, −3, and −19 at each of the urban, open motorway, and motorway tunnel sites. Although at or just above the 95th percentile (so based on limited data), roadside NO_x concentrations typically appear to reach a minimum at around 22 °C at all sites.

The decrease in roadside NO_x with warmer air temperatures could be associated with Ozone formation. Several studies such as Bloomfield et al. (1995) and Pudasainee et al. (2006) highlight the influence of warm air temperatures on O_3 concentrations. Since sunlight intensity and solar radiation is largely associated with air temperature, lower air temperature can result in lower O_3 concentrations which eventually leads to lesser destruction of NO and NO_2. Higher NO_x concentrations would then lead to the formation of O_3 in the presence of sunlight which is associated with higher air temperatures. Carslaw and Taylor (2009) have also explained the decrease in NO_x with the increase in air temperature to be in keeping with plumes being increasingly diluted due to increasing atmospheric mixing.

High roadside NO_x concentrations at low air temperature noticed in Fig. 6 could also be associated with the emission source strength. Weilenmann et al. (2009) have studied the influence of air temperature (specifically at 23 °C, −7 °C, and −20 °C) on cold start CO, HC, and NO_x emissions for both petrol and diesel vehicles (up to Euro 4), all tested on a chassis dynamometer installed in an air-conditioned chamber. While no evident trend was observed on NO_x emissions for petrol vehicles, they have observed an evident increase in NO_x as the air temperature decreased for particular Euro categories of diesel vehicles (specifically Euro 4 diesel vehicles). Knowing that the test chamber for the New European Driving Cycle (NEDC) is required to have an air temperature condition between 20 and 30 °C (Martini et al., 2012), Dardiotis et al. (2012) have also observed higher NO_x emissions at -7 °C compared to emissions at 22 °C for most of their laboratory tested petrol vehicles with Euro 5a emission standard. Although the reason to this effect, which seems to depend on the vehicle fuel type and Euro standard, is not clear, Weilenmann et al. (2009) have suggested that higher loads due to higher friction at low air temperature might play a role. Dardiotis et al. (2012) have also pointed out the possibility that increased NO_x at low air temperatures for some vehicles might be due to a decrease in the Exhaust Gas Recirculation (EGR) rate to avoid water condensation in the EGR cooler and corrosion of the cooled EGR components.

While the analysis here shows the association between roadside NO_x and air temperature, the exact reason behind the observed relationship is outside the scope of this analysis and requires further specific experimental work that would help determine whether the observation is strictly a result of: restricted dispersion and O_3 formation; or a result of changes in vehicle emission characteristics which would require the air temperature conditions used in the NEDC to be further investigated.

The last two key parameters that have high relative influence on the variation of NO_x are wind speed/direction. Table 5 shows that, as expected, these two have higher influence on the urban and open motorway sites compared to the motorway tunnel site. The 2-way partial dependence plots showing the dependence of roadside

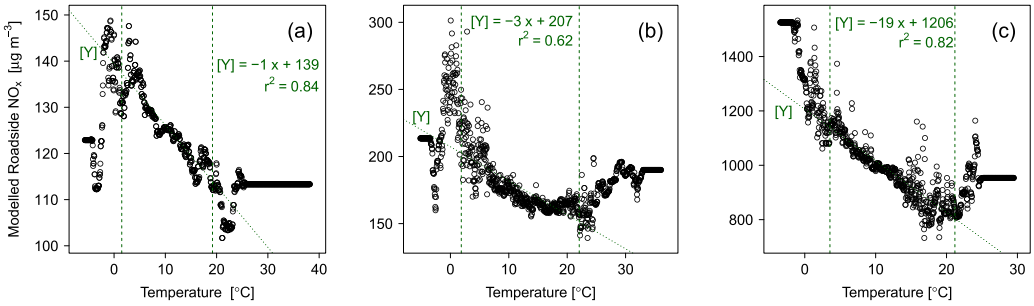

Fig. 6. Partial dependence of roadside concentrations of NO_x [µg m^{-3}] on ambient air temperatures [°C] at each of the (a) urban, (b) open motorway, and (c) motorway tunnel sites. The two vertical dotted lines show the 5th and 95th percentiles of the monitored ambient air temperature. '[Y]' line fits the points between the 5th and 95th percentiles. $Y = mx + c$, r^2 is the equation of the fitted line where m is the gradient, c is the y-intercept, and r^2 is the coefficient of determination.

NO_x on the interaction of wind speed/direction are shown in Fig. 7 for each site. Fig. 8 shows the variation on NO_x by wind speed and wind direction with the smooth continuous surface being calculated using the 'polarPlot' method in the 'openair' R package (R Core Team, 2012; Carslaw and Ropkins, 2012). In comparing Figs. 7 and 8, initial analysis shows that both modelling frameworks produce similar ranges of NO_x concentrations as well as identify key features of when the interaction between wind speed and wind direction result in high concentrations of NO_x.

At all three sites, high roadside NO_x are dependent on the road axis and the location of the monitoring station which are shown in Fig. 1. For the urban site for which the monitoring station is located on the southbound, north-eastern wind perpendicular to the street axis yields the highest concentrations. Being a canyon street with buildings imposing air recirculation (Wood et al., 2009), highest concentrations are found on its leeward side during perpendicular wind flow whereas on the open motorway for which the monitoring station is located on the clockwise direction, highest concentrations are found on its windward side since obstacles on either sides do not exist to induce air recirculation. Generally, an increase in wind speeds leads to a decrease in NO_x concentrations as expected since emissions from a ground level source take the form of a function that is inversely proportional to wind speed (Carslaw et al., 2006). The open motorway site, however, shows potential air traffic emission source at high wind speeds. Potential transport of emissions from the nearby Heathrow airport on the north-eastern region could be an explanation for this observation.

While the relative influence of wind speed/direction is low for the motorway tunnel site, Figs. 7c and 8c show complex interaction between the ambient wind conditions (outside the tunnel) and in-tunnel air quality. Extremely high concentrations of NO_x seem to be associated with the M25 motorway axis and surrounding minor roads. Whilst the dispersion of in-tunnel air pollutants is often restricted by the surrounding tunnel walls unlike open roadways, Figs. 7c and 8c suggest that at least for relatively short tunnels (this e.g. being 0.47 km) wind direction may still have a significant influence on air pollutant concentrations even within the tunnel. However, the complexity of the influence of ambient wind conditions (outside the tunnel) and in-tunnel wind conditions as well as the selection of a relatively far meteorological station for the analysis of in-tunnel air quality add up to the complexity of the observed wind-pollution polar plots and to the fact that both plots of the motorway tunnel (Figs. 7c and 8c) are the least similar compared to that of the urban and open motorway sites. Hence, identifying the contribution of ambient wind (outside the tunnel) to the variation of in-tunnel NO_x concentrations warrants further investigation.

4. Conclusion

This paper has explored and applied the Boosted Regression Trees (BRT) method to study the influence of different traffic-related variables and meteorological variables on the variation of roadside NO_x concentrations. The method allowed the examination of the relative importance of key independent predictors on the response variable. The method also facilitated the visualisation of the partial dependence of the response variable, NO_x ($\mu g\ m^{-3}$) on key predictors, mainly background NO_x, traffic density, and meteorological parameters.

The availability of both monitored background NO_x and modelled background NO_x on the urban site has allowed investigating their influence on roadside NO_x. In studying the dependence of roadside NO_x on both monitored background NO_x and modelled background NO_x, an issue associated with the use of modelled annual background maps has been found. Roadside NO_x was found to much more completely map onto measured than modelled background NO_x, suggesting the latter might not be adequately representing associated contributions. Despite this being investigated on the urban site only as a result of the local background monitoring station availability, initial recommendation to maximise roadside NO_x modelling capabilities might be to either set up local background monitors at appropriate locations away from emission sources or otherwise enhance the background maps used by DEFRA by improving their time/space resolution, improving the PCM model used for their calculation, and/or improving their validation with existing local background measurements. This would allow users to better account for background concentrations and consequently to better investigate the influence of a specific source on concentrations. This conclusion, however, can be further investigated using local background/roadside monitoring stations available on other sites.

The study of the traffic density partial plots has shown that the influence of traffic density on NO_x concentrations is larger when the increase occurs within free-flow conditions compared to busy-flow conditions. Due to the limited occurrence of congested and severely congested-flow conditions at the three sites, the study is limited to free and busy-flow conditions only. Future studies could make use of larger datasets or other sites where congested-flow data points are more abundant in order to test whether the influence would also be lower compared to less dense traffic states.

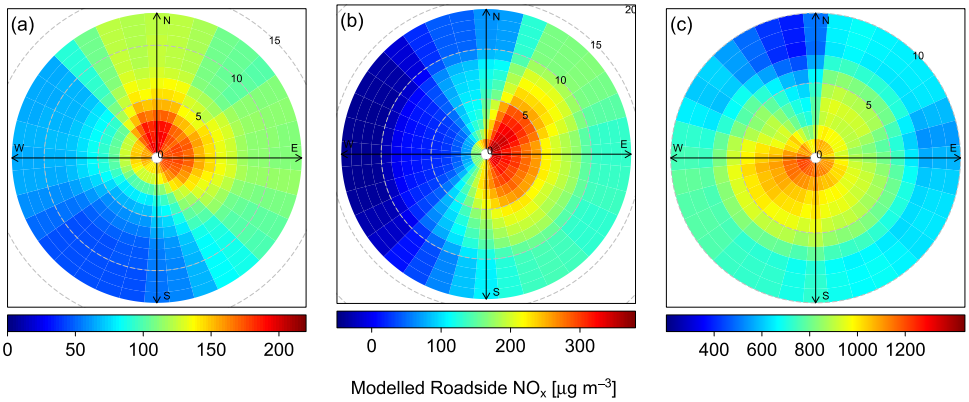

Fig. 7. Wind speed [m s^{-1}] and wind direction [°] variable interaction plots for the BRT modelled roadside NO_x [$\mu g\ m^{-3}$] at the (a) urban, (b) open motorway, and (c) motorway tunnel sites. The centre of each plot is at 0 m s^{-1} and the circumference is dependent on the wind speed of each site.

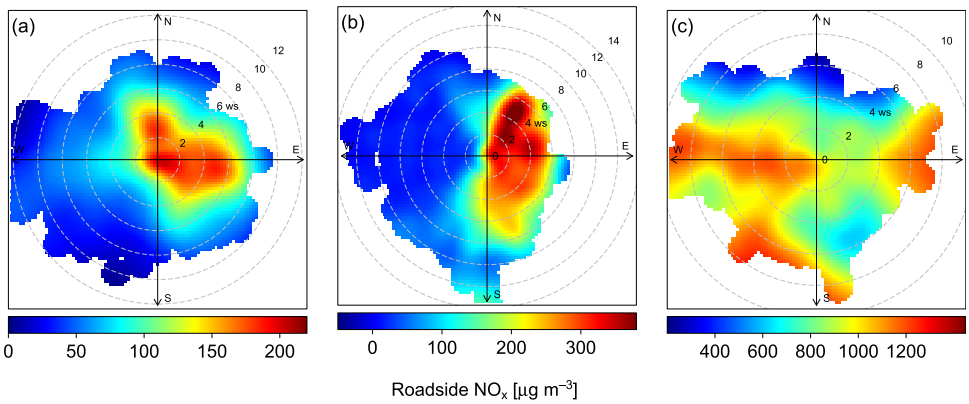

Fig. 8. Bivariate wind-pollution polar plots for roadside concentration of NO_x [µg m^{-3}] at the (a) urban, (b) open motorway, and (c) motorway tunnel.

Also, the influence varies for each of the three road types with the motorway tunnel showing the highest rate and the open motorway the lowest; this highlights the importance of the level of restricted dispersion on the contribution of traffic to ambient NO_x concentrations.

The BRT modelling technique can be used to study the influence of different predictors on response variation. Here, in particular, trends in the partial response relationship for roadside NO_x and traffic density indicates that there might be scope to develop this into a 'ventilation' metric for the 'pollution potential' of different traffic states and different roadside environments.

The study of ambient air temperature partial plots has shown that roadside NO_x concentrations increase with the decrease of ambient air temperatures with a minimum of roadside NO_x at around 22 °C; however, this is rather tentative since the minimum is at or just above the 95th percentile of the ambient air temperature data monitored at all sites. While this tentative observation could be associated to restricted dispersion at low ambient air temperatures or to poorer performance of emission control technologies in road transport vehicles at lower ambient air temperatures, the exact reason cannot be determined in this study and needs to be further examined through specific experimental work, although monitoring sites with different ranges of ambient air temperature would also be worth investigating using the methods described in this paper.

Meteorological data has been collected and used in the BRT model for each of the urban, open motorway, and motorway tunnel sites. Nearest meteorological stations to the sites are kilometres away from the air quality monitoring stations. Despite being within the same topographical areas, further investigation is needed to highlight the extent of the impact of distance between meteorological station and air quality monitoring station on air quality modelling results and to also highlight the impact of using meteorological parameters collected at air quality monitoring stations using low height, potentially obstructed, equipment versus meteorological parameters collected using distant unobstructed meteorological equipment.

While this study shows the influence of each variable on roadside concentrations, road traffic in a model like BRT represents the emission source and most other variables influence the variation of emissions and that is what leads to air quality concentration measurements. The question that arises for future research is how to best model the different types of contributions. For example, traffic flow is an indicator of emissions (a direct source) while wind speed is a secondary influence (modifying concentrations of emissions once produced). Further refinements to the approach, e.g. in the form of a hierarchic structure, might provide a more robust and real-world representative solution and a suitable framework for incorporating more complex interactions such as NO_x/O_3 source/sink behaviour.

Appendix A

This appendix explains how the four traffic states shown in Section 2.2. (Fig. 3) were identified using traffic flow, traffic speed, and traffic density data at the 'near-bound' (bound closest to the monitoring station) of the urban (A660 Headingley, Leeds), open motorway (M25 Staines), and motorway tunnel (M25 Bell Common) sites.

- Fig. A.1a, Fig. A.2a, and Fig. A.3a identify the approximate critical traffic density points which signal the start of the busy-flow traffic states at the three sites, respectively. This is identified by a sharp drop in traffic densities. Data points below the identified critical traffic density point represent the free-flow traffic state.
- Fig. A.1b, Fig. A.2b, and Fig. A.3b identify the approximate critical traffic density points which signal the start of the congested-flow traffic states at the three sites, respectively. For the urban site and the open motorway site (Fig. A.1b and Fig. A.2b, respectively), two critical traffic density points are identified depending on the capacity of the road. This is identified by a sharp drop in traffic densities. Data points between the previously identified critical traffic density point and the new identified critical traffic density point represent the busy-flow traffic state.
- Fig. A.1c, Fig. A.2c, and Fig. A.3c identify the approximate critical traffic speed points which signal the start of the severely congested-flow traffic states at the three sites, respectively. This is identified by a drop in traffic flows. Data points above the previously identified critical traffic density point and above the new critical traffic speed point represent the congested-flow traffic state. Data points above the previously identified critical traffic density point and below the new critical traffic speed point represent the severely congested-flow traffic state.

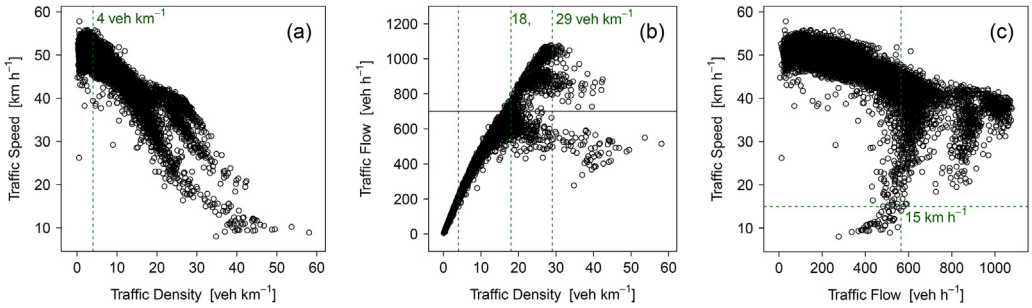

Fig. A.1. 'Near-bound' traffic state identification for the urban site with (a) speed-density diagram showing the split point of free-flow/busy-flow, (b) flow-density diagram showing the split points of busy-flow/congested-flow for flows lower than 700 veh h^{-1} and higher than 700 veh h^{-1}, and (c) speed-flow diagram showing the split point of congested-flow/severely congested-flow.

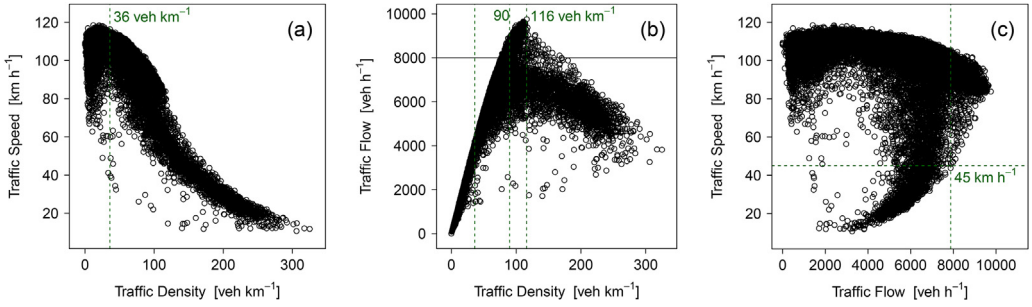

Fig. A.2. 'Near-bound' traffic state identification for the open motorway site with (a) speed-density diagram showing the split point of free-flow/busy-flow, (b) flow-density diagram showing the split points of busy-flow/congested-flow for flows lower than 8000 veh h^{-1} and higher than 8000 veh h^{-1}, and (c) speed-flow diagram showing the split point of congested-flow/severely congested-flow.

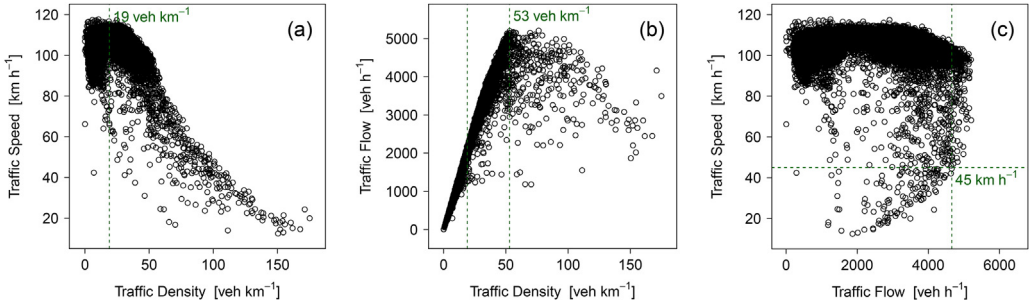

Fig. A.3. 'Near-bound' traffic state identification for the motorway tunnel site with (a) speed-density diagram showing the split point of free-flow/busy-flow, (b) flow-density diagram showing the split point of busy-flow/congested-flow, and (c) speed-flow diagram showing the split point of congested-flow/severely congested-flow.

References

AEAT, 2003. Automatic Urban and Rural Network Site Operator's Manual. AEAT/ENV/R1595. November 2003.

Aldrin, M., Haff, I.H., 2005. Generalised additive modelling of air pollution, traffic volume and meteorology. Atmos. Environ. 39 (11), 2145–2155. http://dx.doi.org/10.1016/j.atmosenv.2004.12.020.

Met Office, 2012. Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current). NCAS British Atmospheric Data Centre, 01 May 2012.

Bell, M., Chen, H., Hackman, M., Mccabe, K., Price, S., 2006. Using ITS to reduce environmental impacts. In: Proceedings of the 10th International Congress and Exhibition, Excel London, Paper 1271.

Bloomfield, P., Royle, A., Steinberg, L.J., Yang, Q., 1995. Accounting for meteorological effects in measuring urban ozone levels and trends. Atmos. Environ. 30 (17), 3067–3077. http://dx.doi.org/10.1016/1352-2310(95)00347-9.

Cappo, M., De'ath, G., Boyle, S., Aumend, J., Olbrich, R., Hoedt, F., Perna, C., Brunskill, G., 2005. Development of a robust classifier of freshwater residence in barramundi (Lates calcarifer) life histories using elemental ratios in scales and boosted regression trees. Mar. Freshw. Res. 56 (5), 713–723. http://dx.doi.org/10.1071/MF04218.

Carslaw, D.C., Beevers, S.D., Ropkins, K., Bell, M.C., 2006. Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. Atmos. Environ. 40 (28), 5424–5434. http://dx.doi.org/10.1016/j.atmosenv.2006.04.062.

Carslaw, D.C., Taylor, P.J., 2009. Analysis of air pollution data at a mixed source location using boosted regression trees. Atmos. Environ. 43 (22), 3563–3570. http://dx.doi.org/10.1016/j.atmosenv.2009.04.001.

Carslaw, D.C., Ropkins, K., 2012. Openair — an R package for air quality data analysis. Environ. Model. Softw. 27–28, 52–61. http://dx.doi.org/10.1016/j.envsoft.2011.09.008.

Dardiotis, C., Martini, G., Manfredi, U., 2012. Revision of Low Temperature Emission Standards for Petrol Vehicles. *European Commission, JRC Scientific and Policy Reports*.

De'ath, G., 2007. Boosted trees for ecological modelling and prediction. Ecology 88 (1), 243–251. http://dx.doi.org/10.1890/0012-9658(2007)88[243:BTFEMA]2.0.CO;2.

DEFRA, 2007. The Air Quality Strategy for England, Scotland, Wales and Northern Ireland. Volume 2.
DEFRA, 2012. Background Maps [online]. (Accessed 01.09.14.) Available from: http://laqm.defra.gov.uk/review-and-assessment/tools/background-maps.html.
DEFRA, 2014. Defra National Statistics Release: Emissions of Air Pollutants in the UK, 1970 to 2013. Statistical Release: 18 December 2014.
DfT, 2014. Traffic Counts – Transport Statistics – Department for Transport [online]. (Accessed 01.09.14.) Available from: http://www.dft.gov.uk/traffic-counts/.
EEA, 2011. Sector Share of Nitrogen Oxides Emissions (EEA Member Countries) [online]. (Accessed 01.09.14.). Available from: http://www.eea.europa.eu/data-and-maps/daviz/sector-share-of-nitrogen-oxides-emissions#tab-chart_1.
Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Animal Ecol. 77 (4), 802–813. http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stati. Data Anal. 38 (4), 367–378. http://dx.doi.org/10.1016/S0167-9473(01)00065-2.
Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22 (9), 1365–1381. http://dx.doi.org/10.1002/sim.1501.
Highways Agency, 2014. Where Are Smart Motorways Being Delivered? [online]. (Accessed 01.01.14.) Available from: http://www.highways.gov.uk/our-road-network/managing-our-roads/improving-our-network/smart-motorways/where-are-smart-motorways-being-delivered/.
HATRIS, 2007. HA Traffic Information System (HATRIS) [online]. (Accessed 01.01.14.) Available from: https://www.hatris.co.uk/.
Hitchins, J., Morawska, L., Wolff, R., Gilbert, D., 2000. Concentrations of submicrometre particles from vehicle emissions near a major road. Atmos. Environ. 34 (1), 51–59. http://dx.doi.org/10.1016/S1352-2310(99)00304-0.
Martini, G., Manfredi, U., Rocha, M., Marotta, A., 2012. Review of the European Test Procedure for Evaporative Emissions: Main Issues and Proposed Solutions. European Commission, JRC Scientific and Policy Reports.
Mayer, H., 1999. Air pollution in cities. Atmos. Environ. 33 (24), 4029–4037. http://dx.doi.org/10.1016/S1352-2310(99)00144-2.
Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards, T.C., 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecol. Model. 199 (2), 176–187. http://dx.doi.org/10.1016/j.ecolmodel.2006.05.021.
Mudakavi, J.R., 2010. Introduction to air pollution. In: Principles and Practices of Air Pollution Control and Analysis. I.K. International Publishing House, pp. 3–133.
NHMRC, 2008. Air Quality in and Around Traffic Tunnels. Final Report. Commonwealth of Australia [online]. (Accessed 01.01.14.) Available from: http://www.nhmrc.gov.au/guidelines/publications/eh42.
Pan, H., Bartolome, C., Gutierrez, E., Princevac, M., Edwards, R., Boarnet, M.G., Houston, D., 2013. Investigation of roadside fine particulate matter concentration surrounding major arterials in five Southern Californian cities. J. Air Waste Manag. Assoc. 63 (4), 482–498. http://dx.doi.org/10.1080/10962247.2013.763867.
Pudasainee, D., Sapkota, B., Shrestha, M., Kaga, A., Kondo, A., Inoue, Y., 2006. Ground level ozone concentrations and its association with NO_x and meteorological parameters in Kathmandu valley, Nepal. Atmos. Environ. 40 (40), 8081–8087. http://dx.doi.org/10.1016/j.atmosenv.2006.07.011.
R Core Team, 2012. R: a Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, *Vienna, Austria*, ISBN 3-900051-07-0. URL. http://www.R-project.org/.
Richmond-Bryant, J., Saganich, C., Bukiewicz, L., Kalin, R., 2009. Associations of PM 2.5 and black carbon concentrations with traffic, idling, background pollution, and meteorology during school dismissals. Sci. Total Environ. 407 (10), 3357–3364. http://dx.doi.org/10.1016/j.scitotenv.2009.01.046.
Ridgeway, G., 2007. Generalized Boosted Models: a Guide to the gbm Package. R package vignette, URL. http://CRAN.R-project.org/package=gbm.
Ridgeway, G., 2012. gbm: Generalized Boosted Regression Models. R package version 1.6–3.2. http://CRAN.R-project.org/package=gbm.
TRL, 2007. Primary NO_2 Emissions from Road Vehicles in the Hatfield and Bell Common Tunnels. Published Project Report PPR262.
TRL, 2011. The Highways Agency Roadside Air Pollution Monitoring Network Report – 2010 Volume I.
Uhrner, U., von Lowis, S., Vehkamaki, H., Wehner, B., Brasel, S., Hermann, M., Stratmann, F., Kulmal, M., Wiedensohler, M., 2007. Dilution and aerosol dynamics within a diesel car exhaust plume – CFD simulations of on-road measurement conditions. Atmos. Environ. 41 (35), 7440–7461. http://dx.doi.org/10.1016/j.atmosenv.2007.05.057.
Vardoulakis, S., Fisher, B.E., Pericleous, K., Gonzalez-Flesca, N., 2003. Modelling air quality in street canyons: a review. Atmos. Environ. 37 (2), 155–182. http://dx.doi.org/10.1016/S1352-2310(02)00857-9.
Weilenmann, M., Favez, J.Y., Alvarez, R., 2009. Cold-start emissions of modern passenger cars at different low ambient temperatures and their evolution over vehicle legislation categories. Atmos. Environ. 43 (15), 2419–2429. http://dx.doi.org/10.1016/j.atmosenv.2009.02.005.
Wood, C.R., Arnold, S.J., Balogun, A.A., Barlow, J.F., Belcher, S.E., Britter, R.E., Cheng, H., Dobre, A., Lingard, J.J.N., Martin, D., Neophytou, M.K., Petersson, F.K., Robins, A.G., Shallcross, D.E., Smalley, R.J., Tate, J.E., Tomlin, A.S., White, I.R., 2009. Dispersion experiments in central London: the 2007 DAPPLE project. Bull. Am. Meteorol. Soc. 90 (7), 955–969. http://dx.doi.org/10.1175/2009BAMS2638.1.