

RESEARCH

Open Access

Multivariate adaptive regression splines models for vehicular emission prediction

Seth Daniel Oduro^{1*}, Santanu Metia¹, Hiep Duc², Guang Hong¹ and Q.P. Ha¹

Abstract

Background: Rate models for predicting vehicular emissions of nitrogen oxides (NO_x) are insensitive to the vehicle modes of operation, such as cruise, acceleration, deceleration and idle, because these models are usually based on the average trip speed. This study demonstrates the feasibility of using other variables such as vehicle speed, acceleration, load, power and ambient temperature to predict (NO_x) emissions to ensure that the emission inventory is accurate and hence the air quality modelling and management plans are designed and implemented appropriately.

Methods: We propose to use the non-parametric Boosting-Multivariate Adaptive Regression Splines (B-MARS) algorithm to improve the accuracy of the Multivariate Adaptive Regression Splines (MARS) modelling to effectively predict NO_x emissions of vehicles in accordance with on-board measurements and the chassis dynamometer testing. The B-MARS methodology is then applied to the NO_x emission estimation.

Results: The model approach provides more reliable results of the estimation and offers better predictions of NO_x emissions.

Conclusion: The results therefore suggest that the B-MARS methodology is a useful and fairly accurate tool for predicting NO_x emissions and it may be adopted by regulatory agencies.

Keywords: Nitrogen oxide; On-board emission measurement system; Chassis dynamometer testing system; Emission

Background

Outdoor air pollution is reported as the main reason to cause 1.3 million annual deaths worldwide (World Health Organization 2015). Among air pollutants coming from natural effects (Duc et al. 2013), man-made emissions have been the main concern in air-quality modelling and control. Vehicular emissions, in this context, can bring serious impacts on air quality and thus, have received increasing research attention (Sharma et al. 2010). Road transport often appears as the single most important source of urban pollutant emissions in source apportionment studies (Maykut et al. 2007). In the coming decades, road transport is likely to remain a large contributor to air pollution, especially in urban areas.

For this reason, major efforts are being made for the reduction of polluting emissions from road transport. These include new powertrains and vehicle technology

improvements, fuel refinements, optimization of urban traffic management and the implementation of tighter emission standards (Querol et al. 2007). In recent decades, many emission models have been developed. Afotel et al. 2013 proposed regression models to estimate light-duty gasoline vehicle emissions of CO₂ based on vehicle velocity, acceleration, deceleration, power demand and time of the day. However, the model did not include NO_x emissions. Oduro et al. 2013 proposed multiple regression models with instantaneous speed and acceleration as a predictor variables to estimate vehicular emissions of CO₂ but not NO_x. Tóth-Nagy et al. 2006 proposed an artificial neural network-based model for predicting emissions of CO and NO_x from heavy-duty diesel conventional and hybrid vehicles. The methodology sounds promising, but applied to heavy-duty vehicles only, and the fit function contains many details which make the model difficult to understand. Emission model based on instantaneous vehicle power, which is computed on total resistance force, vehicle mass, acceleration, velocity, and drive-line efficiency, was developed by Rakha et al. 2011. However, the

*Correspondence: Daniel.SethOduro@studnet.uts.edu.au

¹ Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway, NSW 2007 Sydney, Australia
Full list of author information is available at the end of the article

model applies for fuel consumption and CO₂ emission factor and does not include the NO_x emission.

A key gap in our understanding of these emissions is the effect of changes in vehicle speed, power and load on average emission rates for the on-road vehicle fleet. Vehicle power, load and vehicle speed are closely linked to fuel consumption and pollutant emission rates (European Commission White Paper 2015). Improved understanding of the link between operating conditions and emissions could develop accurate models for prediction of vehicle emissions. The quality of the application of any road vehicle emission model largely depends on the representativeness of the emission factor such as carbon dioxide (CO₂), carbon monoxide (CO), nitrogen oxides (NO_x), volatile organic compounds (VOCs) and particulate matter (PM). This refers to the accuracy with which the emission factor can describe the actual emission level of a particular vehicle type and driving conditions applied to it.

This work focuses on using the MARS methodology to improve the prediction accuracy of chassis dynamometer and on-board measurement systems. The dynamometer testing is one of the three typical vehicle tailpipe emission measurement methods, where emissions from vehicles are measured under laboratory conditions during a driving cycle to simulate vehicle road operations (Frey et al. 2003). The real world on-board emissions measurement is widely recognized as a desirable approach for quantifying emissions from vehicles since data are collected under real-world conditions at any location travelled by the vehicle (Durbin et al. 2007). Variability in vehicle emissions as a result of changes in facility (roadway) characteristics, vehicle location, vehicle operation, driver, or other factors can be represented and analysed more reliably than with the other methods (Frey et al. 2002). This is because measurements are obtained during real world driving, eliminating the concern about non representativeness that is often an issue with dynamometer testing, and at any location, eliminating the setting restrictions inherent in remote sensing. Though this measuring technique seems to be more promising, the need to improve the prediction accuracy of emission factor especially with NO_x emissions by using effective statistical techniques is important in any emission inventory.

A number of the models discussed above either do not estimate NO_x emissions, or are so sophisticated as to require excessive data inputs. There needs to be a balance between the accuracy and detail of a model for its ease of application. Therefore, to enhance the prediction performance for the NO_x emissions, the boosting MARS (B-MARS) modelling approach is proposed in this paper. Here, we aim to estimate, with high accuracy, the NO_x emissions. The effectiveness of the model is then determined by grouping the data into two parts, one for building the model (learning) and the other for validating

the model (testing). The results are verified by comparing the experimental data, B-MARS and MARS predicted values. The remainder of this paper presents the data collection methods, namely via chassis dynamometer and on-board data collection, the MARS model and B-MARS methods.

Data collection methods

Chassis dynamometer

This study uses secondary data corrected by the New South Wales (NSW) Road and Maritime Service (RMS), Department of Vehicle Emission, Compliance Technology Operation. The data were collected on the second by second basis and four vehicles were used for the test. The test vehicles include Toyota, Ford, Holden and Nissan from 2007 and 2008 model year with an engine displacement ranging from 1.8L to 2.0L. A chassis dynamometer set-up in the laboratory simulates the resistive power imposed on the wheels of a vehicle, as shown in Fig. 1. It consists of a dynamometer that is coupled to drive lines that are directly connected to the wheel hubs of the vehicle, or to a set of rollers upon which the vehicle is placed, and which can be adjusted to simulate driving resistance. During testing, the vehicle is tied down so that it remains stationary as a driver operates it according to a predetermined time-speed profile and gear change pattern shown on a monitor. A driver operates the vehicle to match the speed required at the different stages of the driving cycle (Nine et al. 1999). Experienced drivers are able to closely match the established speed profile.

On-board data collection

Data from on-board instruments, can facilitate development of micro-scale emission models (Frey et al. 2003). When compared with conventional dynamometer tests under carefully controlled conditions, on-road data reflects real driving situations. Accordingly, second-by-second emissions data were collected using a Horiba On-Board Measurement System (OBS-2000), as shown in Fig. 2, with the same testing vehicles as with the dynamometer test cycle. The equipment is composed of two on-board gas analysers, a laptop computer equipped with data logger software, a power supply unit, a tailpipe attachment and other accessories. The OBS-2000 collects second-by-second measurements of nitrogen oxides NO_x, hydrocarbons (HC), carbon monoxide (CO), carbon dioxide (CO₂), exhaust temperature, exhaust pressure, and vehicle position (via a global positioning system, or GPS). Although the instrument measured other pollutants, the focus of this work was to build a model for NO_x emissions. For the measurement scale used, accuracy for the NO_x emission measurements, reported in percentage, was $\pm 0.3\%$. A two second lag in NO_x emission measurement was accounted for in the data spreadsheets.

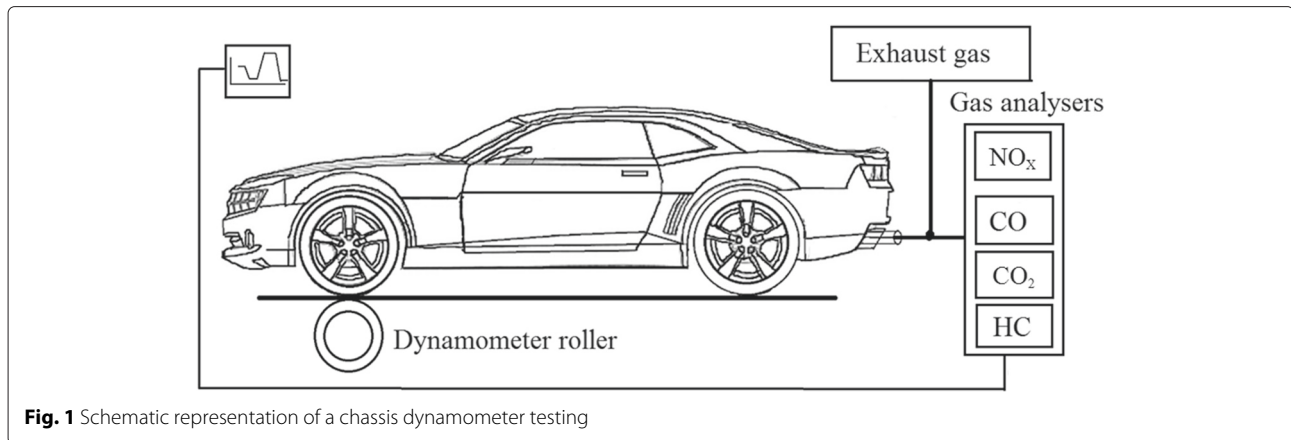


Fig. 1 Schematic representation of a chassis dynamometer testing

NO_x sensor calibration was carried out throughout the data collection period. To ensure consistently smooth and good data collection without frequent interruptions due to any possible unit malfunction, inability of batteries to stay charged and calibration issues throughout the period, proper maintenance and diagnostic procedures were strictly followed.

Methods

The Multivariate Adaptive Regression Splines (MARS) were introduced for fitting the relationship between a set of predictors and dependent variables (Friedman 1991). MARS is a multivariate, piecewise regression technique that can be used to model complex relationship. The space of predictors is divided into multiple knots in order to fit a spline function between these knots (Friedman 1991; MARS User Guide 2015). The basic problem in vehicular emission modelling is how best to determine the fundamental relationship between dependent variables, and vector of predictors, such as speed, acceleration, load, power, ambient temperature including other factors.

The MARS algorithm searches over all possible univariate hinge locations and across interactions among all variables. It does so through the use of combinations of variable called basis functions. The approach is analogous to the use of splines. This study aims at exploring the

potential of applying the MARS methodology to model NO_x emissions using the following set of input parameters: speed, acceleration, load, power and ambient temperature of chassis dynamometer and on-board emission measurements. The problem can be stated as a multivariate regression problem. Suppose that *N* pairs of input-output parameters are available: {*y_i, x_{1i}, …, x_{mi}*}_{1^N}, where the depend variable *y_i*, *i* = 1, 2 … , *N*, is the *i*th measure of NO_x and the predictor *x_{li}*, *i* = 1, 2 … , *N*, *l* = 1, 2, … , *m*, is the *i*th measure of the *l*th parameter. We assume that the data {*y_i, x_{1i}, …, x_{mi}*}_{1^N} are related through the following equation

$$y = f(x_1, \dots, x_m), \quad (x_1, \dots, x_m) \in D \subset R^m, \quad (1)$$

where *f*(·) is an unknown multivariate deterministic function and *D* is the domain of inputs. Since the true mapping in (1) is not known, it is desired to have a function $\hat{f}(x_1, \dots, x_m)$ that provides a “good” fit approximation of the output data. The good fit between $\hat{f}(x_1, \dots, x_m)$ and the output data is using the integrated mean square error (*MSE*) estimated.

$$MSE = \frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}(x_{1i}, \dots, x_{mi})]^2. \quad (2)$$

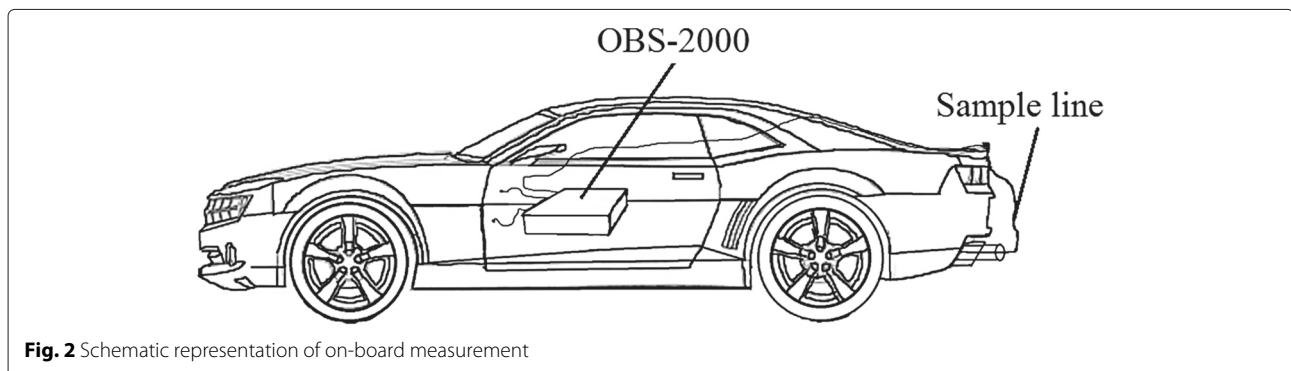


Fig. 2 Schematic representation of on-board measurement

To regularize the problem, that is, make it well-posed, a restriction is imposed for the solution $\hat{f}(x_1, \dots, x_m)$ as functions residing in the linear space:

$$f(\cdot) = \beta_0 + \sum_{m=1}^M \beta_m h_m(\cdot), \tag{3}$$

where $\{h_m(\cdot)\}_{m=1}^M$ is a set of basis functions and $\{\beta_m\}_{m=0}^M$ are coefficients of representation. In this paper, $h_m(\cdot)$ is the splines basis function defined as:

$$h_m(\cdot) = \prod_{k=1}^{K_m} [s_{k,m} \cdot (x_{\nu(k,m)} - t_{k,m})]_+, \tag{4}$$

where $s_{k,m}$ are variables that take values ± 1 , $\nu(k, m)$ labels the predictor variables and $t_{k,m}$ represents estimated values on the corresponding variables. The quantity K_m is the number of "splits" that give rise to each basis function β_m . Here the subscript "+" indicates a value of zero for negative values of the argument. The basis functions involved in (1) are known as "hockey sticks" basis function. MARS searches over the space of all inputs and predictors values (knots) as well as interactions between variables. Now, given the estimated coefficients $\{\beta_m^*\}_0^M$, basis functions $\{h_m^*(\cdot)\}_0^M$ and operation parameters describing a new measurement, the emission of the new measurement can be predicted by taking the following steps:

1. Segregate operation parameters including speed, acceleration, power, load and ambient temperature from the raw data.
2. Predict the emission NO_x by using the approximate function $\hat{f}(\cdot)$ with $\{\beta_m^*\}_0^M$ and $\{h_m^*(\cdot)\}_0^M$, that is

$$\hat{f}(x_{1i} \dots x_{mi}) = \beta_0^* + \sum_{m=1}^M \beta_m^* h_m^*(x_{1i} \dots x_{mi}),$$

$i = 1 \dots N$, where $\{x_{1i} \dots x_{mi}\}_1^N$ are from new measurements. The basis functions, together with the model parameters, are combined to produce the predictions given the inputs. The general MARS model equation is given as:

$$\hat{f}(X) = \beta_0 + \sum_{m=1}^M \beta_m h_m(X), \tag{5}$$

where $\{\beta\}_0^m$ are the coefficients of the model that are estimated to yield the best fit to the data, M is the number of sub-regions or the number of basis functions in the model, and $h_m(X)$ is the spline basis function given in (4).

This model searches over the space of all inputs and predictor values (referred to as "knots") as well as the interactions between variables. During this search, an increasingly larger number of basis functions are added to the model to minimize a lack-of-fit criterion. As a result of these operations, MARS automatically determines the most important independent variables as well as the most

significant interactions among them. It is noted that the search for the best predictor and knot location is performed in an iterative process. The predictors as well as the knot location, having the most contribution to the model, are selected first. Also, at the end of each iteration, the introduction of an interaction is checked for possible model improvements.

Model selection and pruning

In general, non-parametric models are adaptive and can exhibit a high degree of flexibility that may ultimately result in over fitting, if no measures are taken to counteract it. The second step is the pruning step, where a "one-at-a-time" backward deletion procedure is applied in which the basis functions with the least contribution to the model are eliminated. This pruning is based on a generalized cross-validation (GCV) criterion. The GCV criterion is used to find the overall best model from a sequence of fitted models, where a larger GCV value tends to produce a smaller model, and vice versa. The GCV criterion is estimated as the lack-of-fit criterion (Hastie and Tibshirani 2001).

$$\text{GCV} = \frac{1}{N} \frac{\sum_{i=1}^N (y_i - \hat{f}(X_i))^2}{\left[1 - \frac{\tilde{C}(M)}{N}\right]^2}, \tag{6}$$

where $\left[1 - \frac{\tilde{C}(M)}{N}\right]^2$ is a complexity function, and $\tilde{C}(M)$ is defined as $\tilde{C}(M) = C(M) + dM$, of which $C(M)$ is the number of parameters being fit and d represents a cost for each basis function optimization and is a smoothing parameter of the procedure. The higher the cost d is, the more basis functions will be eliminated (Put et al. 2004).

Boosting algorithm

To extend the results obtained in (Oduro et al. 2014), we propose to use a boosting algorithm for improving the performance of MARS model. The algorithm, introduced by Freund and Schapire 1997, has been successfully applied to several benchmark machine learning problems using supervised learning. Basically, boosting is an algorithm to form a strong learner by combining multiple weak learners whereby a new classifier is generated based on the result of the previously generated classifiers focusing on misclassified samples. The algorithm increases the weights of incorrect classified samples and decreases the weights of those classified correctly. The problem of applying least-square boosting (LS-Boosting) can be formulated as follows. Let x denote the feature vector and y the alignment accuracy. Given an input variable x , a response variable y and some samples $\{y_i, x_i\}_{i=1}^N$, the goal is to obtain an estimated or approximation $\hat{F}(x)$, of the function $F^*(x)$ mapping x to y , that minimises the expected

value of some specified loss function $L(y, F(x))$ over the joint distribution of all (y, x) values.

$$F^* = \arg \min_F L(y, F(x)). \tag{7}$$

In least squares (LS) boosting, the squared error loss is given by $L(y, F) = (y - F)^2/2$ and the pseudo-response is obtained as

$$\tilde{y} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} = y_i - F_{m-1}(x), \quad i = 1, 2, \dots, N. \tag{8}$$

Thus, for $i = 1, 2, \dots, N$ the minimisation of the data based estimate of the expected loss gives

$$(\rho_m, a_m) = \arg \min_{a, \rho} \sum_{i=1}^N [\tilde{y}_i - \rho h(x_i; a)]^2, \tag{9}$$

where $h(x; a)$ is the weak learner or base learner with basis functions $\{h(x, a_m)\}_{m=1}^M$ and ρ_m is the corresponding multiplier. The LS-boost algorithm is summarised below (Jerome 2001).

1. Initialize $F_0(x) = \bar{y}$.
2. For $m=1$ to M do:

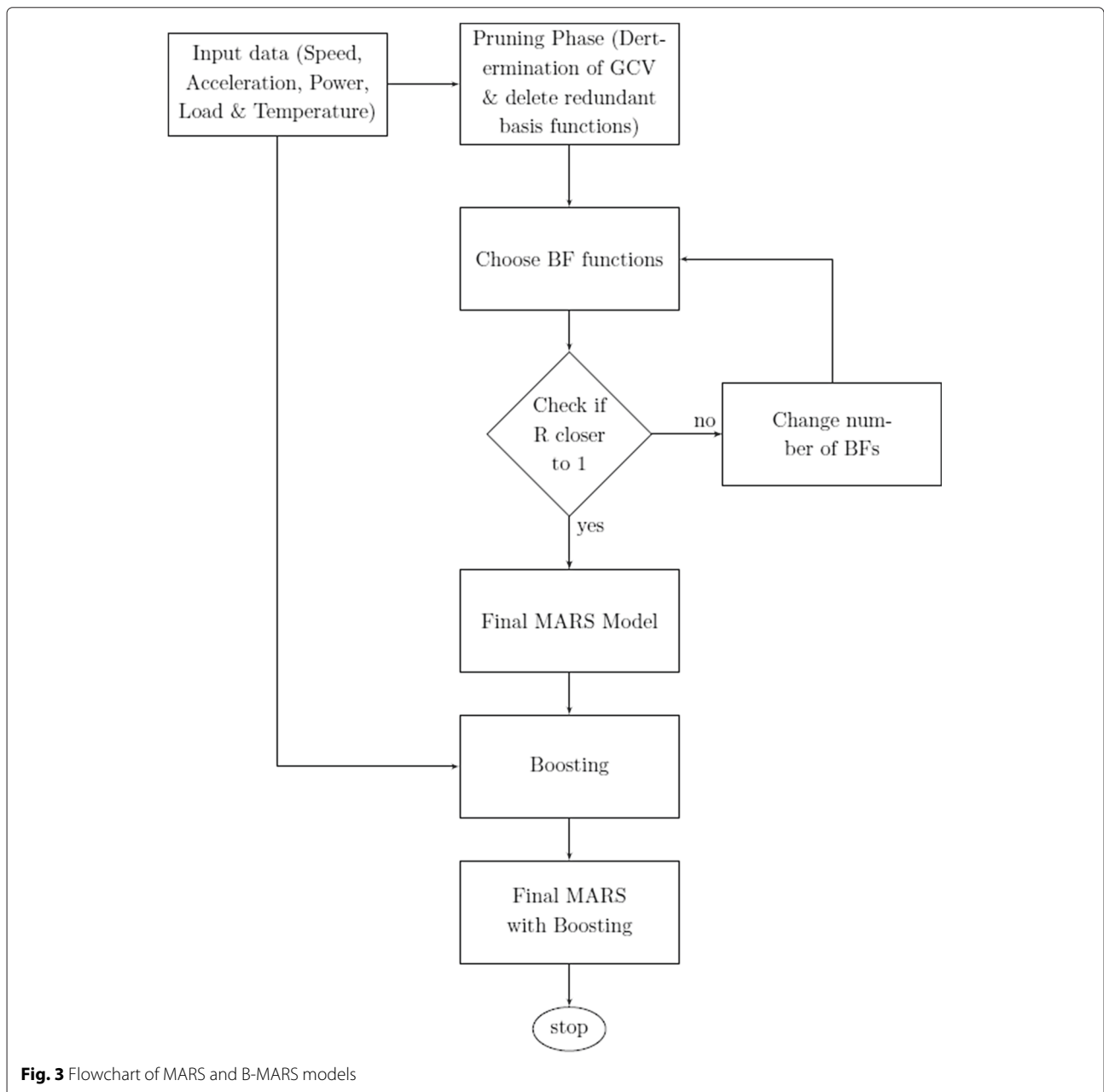


Fig. 3 Flowchart of MARS and B-MARS models

(a) Compute

$$\tilde{y}_i = y_i - F_{m-1}(x_i), \quad i = 1, N. \tag{10}$$

(b) Compute

$$(\rho_m, a_m) = \arg \min_{a, \rho} \sum_{i=1}^N [\tilde{y}_i - \rho h(x_i; a)]^2. \tag{11}$$

(c) The Update estimator at step m becomes

$$F_m(x) = F_{m-1}(x) + \rho_m h(x; a_m). \tag{12}$$

3. End for

4. Output the final regression function $F_m(x)$.

The flowchart in Fig. 3 shows the proposed models for MARS and B-MARS, whereby boosting in the latter is adopted to improve estimation performance by adjusting the weights of the classifiers.

Results and discussions

Five vehicular emission predictor variables, namely, speed (m/s), acceleration (m/s^2), power (W), temperature ($^{\circ} C$) and load (Nm) were used with the response variable of NO_x (g/s) in an attempt to identify the relationships that vehicular emission models developers wish to understand.

Table 1 List of basis functions of the MARS and their coefficients for on-board measurements

Beta factor	Basis function	Value
BF0		0.249827
BF1	Max(0, SPEED-8.11)	-0.000142
BF2	Max(0, SPEED-11.67)	0.000342
BF3	Max(0, SPEED-12.52)	0.000442
BF4	Max(0, SPEED-16.39)	0.0032363
BF5	Max(0, SPEED-23.89)	0.011587
BF6	Max(0, SPEED-24.17)	0.043903
BF7	Max(0.95-ACCEL, 0)	-0.001307
BF8	Max(0, ACCEL-1.25)	0.007307
BF9	Max(0, ACCEL-5.85)	0.011308
BF10	Max(0, ACCEL-7.21)	0.031102
BF11	Max(0, AMBT -22.12)	0.000231
BF12	Max(0, AMBT -23.47)	0.003130
BF13	Max(0, AMBT -24.76)	0.021131
BF14	Max(0, LOAD -10.53)	0.015618
BF15	Max(0, LOAD -52.34)	0.017966
BF16	Max(0, LOAD -60.16)	0.023225
BF17	Max(0, Power -8.98)	0.014877
BF18	Max(0, Power -21.32)	0.015679

Table 2 List of basis functions of the B-MARS and their coefficients for on-board measurements

Beta factor	Basis function	Value
BF0		0.218753
BF1	Max(0, SPEED-9.22)	-0.000112
BF2	Max(0, SPEED-12.54)	-0.000212
BF3	Max(0, SPEED-14.23)	0.000211
BF4	Max(0, SPEED-17.16)	0.002417
BF5	Max(0, SPEED-24.32)	0.011213
BF6	Max(0.98-ACCEL, 1.12)	-0.001126
BF7	Max(0, ACCEL-2.61)	-0.004267
BF8	Max(0, ACCEL-6.92)	0.011211
BF9	Max(0, ACCEL-7.56)	0.024123
BF10	Max(0, AMBT -22.51)	0.000212
BF11	Max(0, AMBT -23.78)	0.003451
BF12	Max(0, LOAD -13.15)	0.013617
BF13	Max(0, LOAD -54.56)	0.016541
BF14	Max(0, POWER -7.34)	0.012346
BF15	Max(0, POWER -9.76)	0.013145
BF16	Max(0, POWER -22.17)	0.012678

To explore factors affecting vehicular emission models, the present study provides results and some interpretations from the MARS and B-MARS models. Tables 1, 2, 3 and 4 summarize the variable selection results using MARS and B-MARS, whose beta factor coefficients β_m

Table 3 List of basis functions of the MARS and their coefficients for dynamometer testing

Beta factor	Basis function	Value
BF0		0.313578
BF1	Max(0, SPEED-6.43)	-0.000172
BF2	Max(0, SPEED-9.36)	0.000625
BF3	Max(0, SPEED-18.37)	0.005751
BF4	Max(0, SPEED-25.14)	0.063521
BF5	Max(0, ACCEL-1.12)	0.009433
BF6	Max(0, ACCEL-4.24)	0.056731
BF7	Max(0, ACCEL-6.24)	0.066312
BF8	Max(0, AMBT -21.54)	0.000321
BF9	Max(0, AMBT -23.15)	0.004433
BF10	Max(0, AMBT -24.62)	0.037215
BF11	Max(0, LOAD -15.67)	0.013211
BF12	Max(0, LOAD -45.67)	0.053412
BF13	Max(0, Power -13.76)	0.016813
BF14	Max(0, Power -20.64)	0.021213

Table 4 List of basis functions of the B-MARS and their coefficients for dynamometer testing

Beta factor	Basis function	Value
BF0		0.231567
BF1	Max(0, SPEED-7.11)	-0.000134
BF2	Max(0, SPEED-11.42)	0.000514
BF3	Max(0, SPEED-20.16)	0.004671
BF4	Max(0, SPEED-26.11)	0.051411
BF5	Max(0, ACCEL-1.23)	0.009671
BF6	Max(0, ACCEL-5.78)	0.032143
BF7	Max(0, AMBT-22.14)	0.000221
BF8	Max(0, AMBT-23.63)	0.003133
BF9	Max(0, AMBT-25.31)	0.028912
BF10	Max(0, LOAD-14.41)	0.012761
BF11	Max(0, LOAD-44.23)	0.041671
BF12	Max(0, Power -15.72)	0.015116
BF13	Max(0, Power -21.43)	0.021551

are denoted BF_m . In a MARS and B-MARS models, basis functions are used to predict the effects of independent variables on NO_X emission factor. The interpretation of B-MARS and MARS results is similar to but not as straight forward as that of classical linear regression models. A positive sign for the estimated beta factors for the basis function indicates increased NO_X emission, while a negative sign indicates the opposite. The value of beta factor

implies the magnitude of effect of the basis function (i.e., variable effect) on the NO_X emission. For the effect of each basis function, $\max(0, x - t)$ is equal to $(x - t)$ when x is greater than t ; otherwise the basis function is equal to zero. As shown in Tables 1, 2, 3 and 4, the MARS and B-MARS models contain 14, 15, 17 and 19 basis functions for on-board and dynamometer testing. The on-board measurements and dynamometer testing for MARS and B-MARS have similar interpretations. It can be observed that all the five predictor variables play crucial roles in determining NO_X vehicle emission. From Table 1, beta factors BF1, BF2, BF3, BF4, BF5 and BF6 account for the nonlinear effect of vehicle speed in the emission model.

The effect of speed on NO_X emissions can be explained as follows. By using the on-board measurements method for MARS, if the speed of the vehicle is lower than 8.11 m/s or 29.2 km/h for a short duration in traffic, it has a negligible impact on NO_X emission (indicated by BF0). However, for a longer queuing time, such as in large cities, the amount of NO_X emitted into the atmosphere can be significant as the NO_X emission increases with a corresponding increase in combustion temperature. The NO_X effect is increased as the speed increases from 11.67 m/s or 42 km/h (indicated by BF2-BF5) due to corresponding increase in combustion temperature. The emission rate can reach 0.043903 g/s when the speed is about 24.17 m/s or 82 km/h (indicated by BF6). This expected finding is consistent with previous findings in literature. From Carslaw et al. 2011, it is noted that NO_X emissions rise and fall in a reverse pattern to hydrocarbon emissions

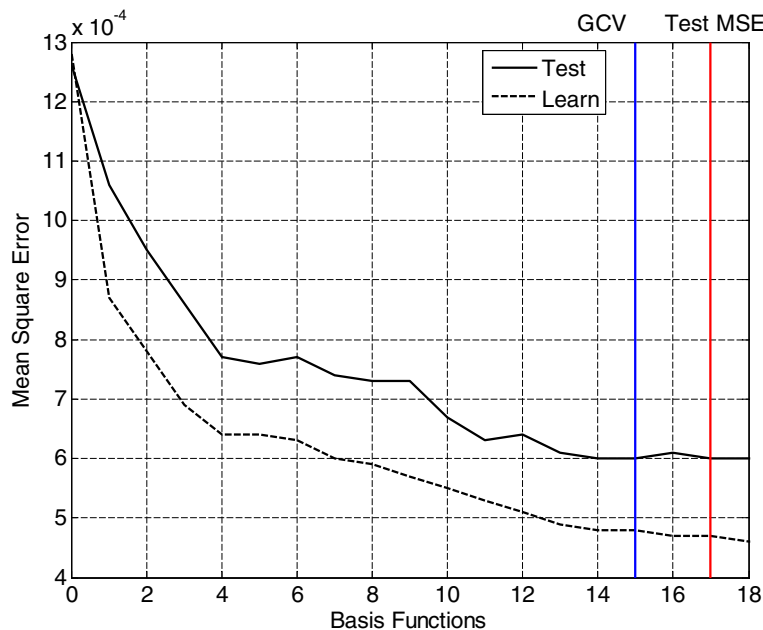


Fig. 4 Schematic representation of a on-board testing

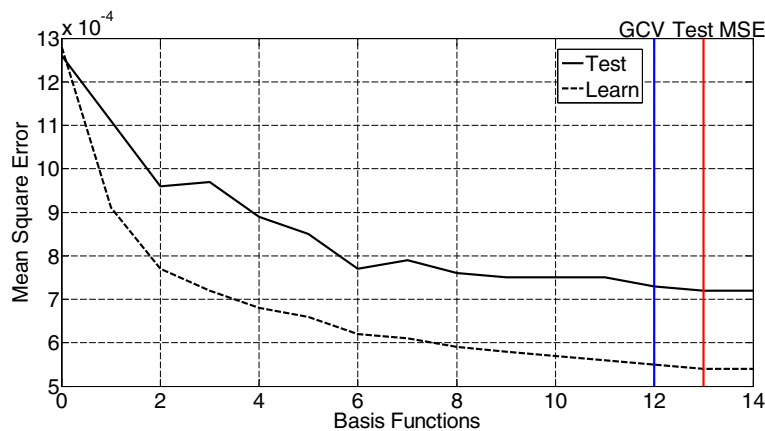


Fig. 5 Schematic representation of a chassis dynamometer testing

(HC). As the speed of the vehicle increase the mixture becomes leaner with more HC's at high temperatures in the combustion chamber, there appear excess oxygen molecules which combine with the nitrogen to form NO_x . From Table 1, as the speed increases (indicated by BF2-BF6) the total NO_x emission emitted from the tail pipe also increases. Beta factors (BF7-BF10) on Table 1 show the nonlinear effect of vehicle acceleration on the NO_x which can be described as follows. If the vehicle acceleration is less than 0.95 m/s^2 , NO_x emission will reduce by 0.0013075 g/s (indicated by BF7), but if the acceleration is increased from 1.25 m/s^2 , to 5.85 m/s^2 , the NO_x emission will increase by 0.0113075 g/s (indicated by BF8 and

BF9). The NO_x emission can reach more than 0.0311017 g/s when the acceleration exceeds 7.21 m/s^2 . This result is similar to that of the speed because of depressing the accelerator pedal increase acceleration as well as speed simultaneously.

The ambient temperature is also found to influence the NO_x emission as indicated by BF11, BF12 and BF13 of Table 1, the effects of ambient temperature on NO_x emission occurrence include: (1) if the ambient temperature is less than 22.12°C then it has no effect on vehicle NO_x emission (indicated by BF11); (2) if the ambient temperature is greater than 22.12°C but less 23.47°C , NO_x emission will increase by 0.00023075 g/s for 1°C increase

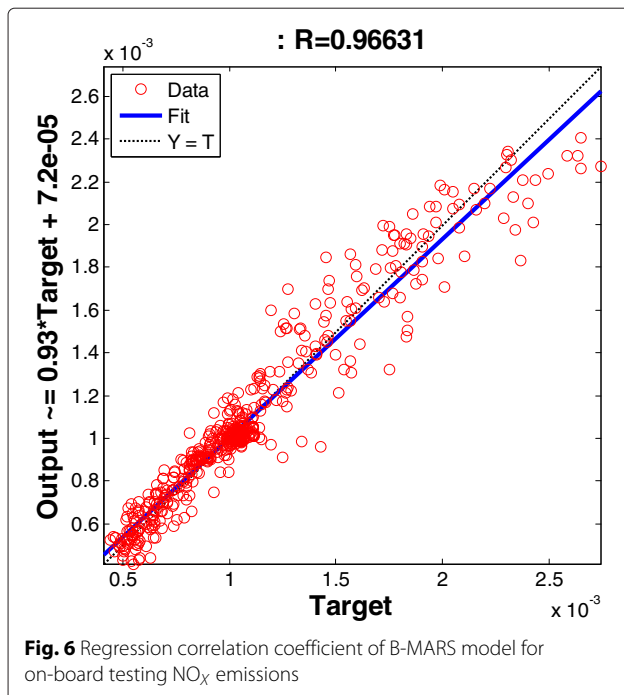


Fig. 6 Regression correlation coefficient of B-MARS model for on-board testing NO_x emissions

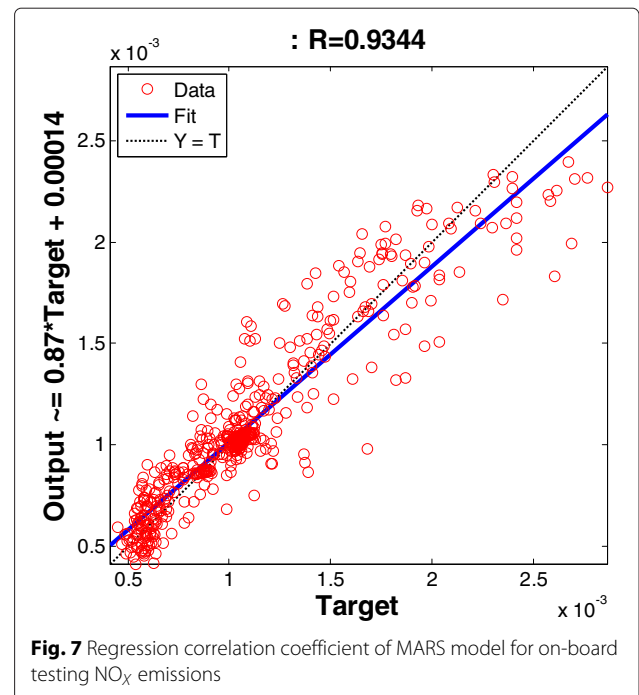
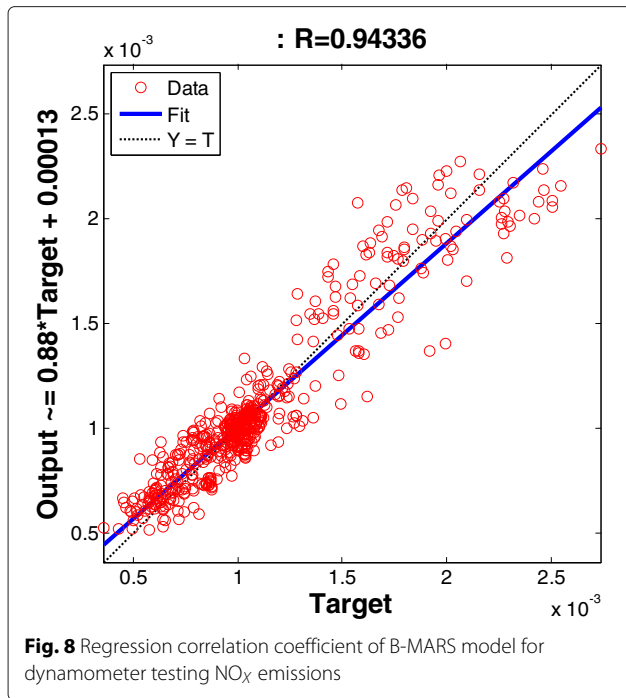


Fig. 7 Regression correlation coefficient of MARS model for on-board testing NO_x emissions



of ambient temperature (indicated by BF11 and BF12); (3) if the ambient temperature is greater than 23.47°C but less than 24.76°C , the vehicle NO_x emission will increase by 0.00313022 g/s for 1°C increase in ambient temperature (indicated by BF12 and BF13) and (4) if the ambient temperature is greater than 24.76°C the NO_x emission will increase by 0.02113075 g/s for 1°C increase in ambient

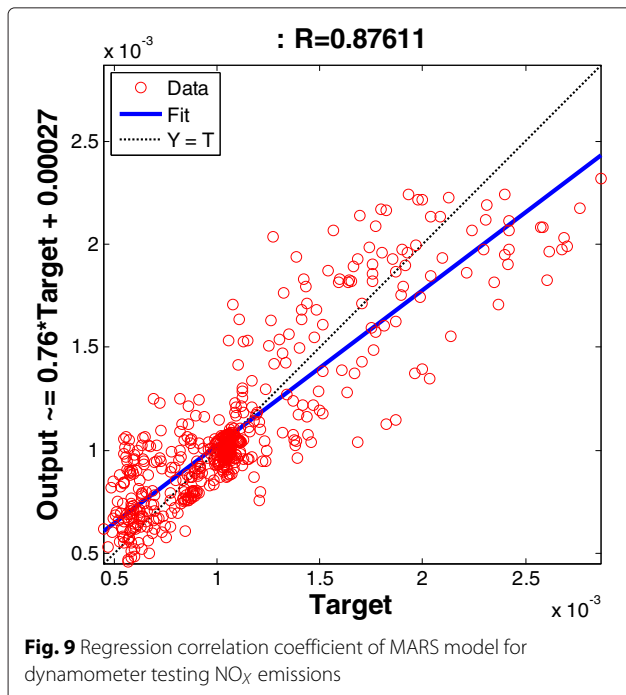
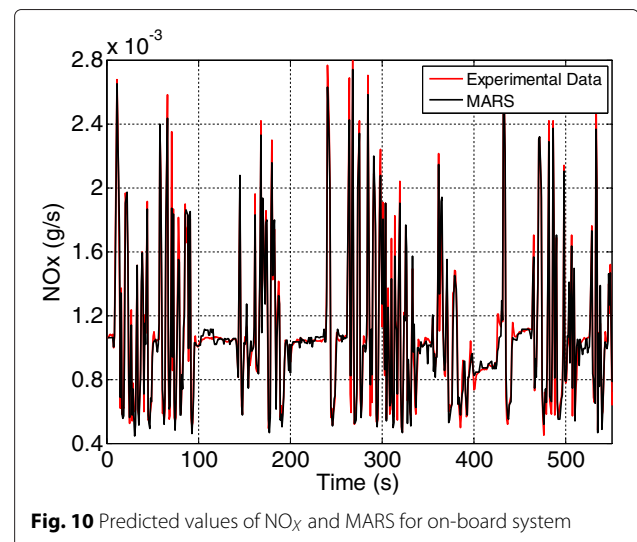
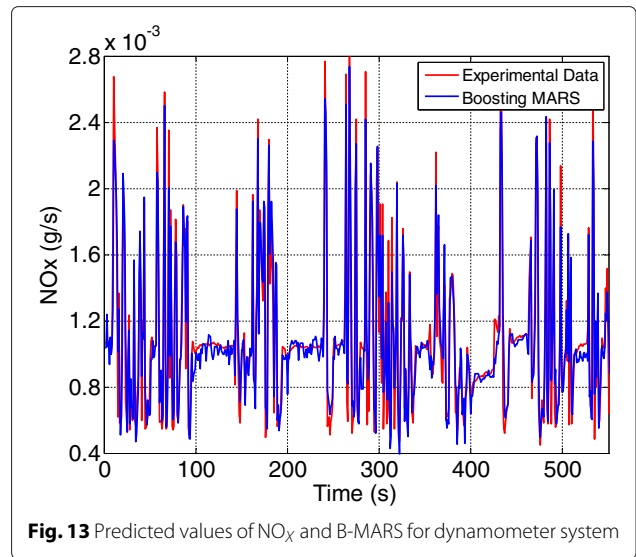
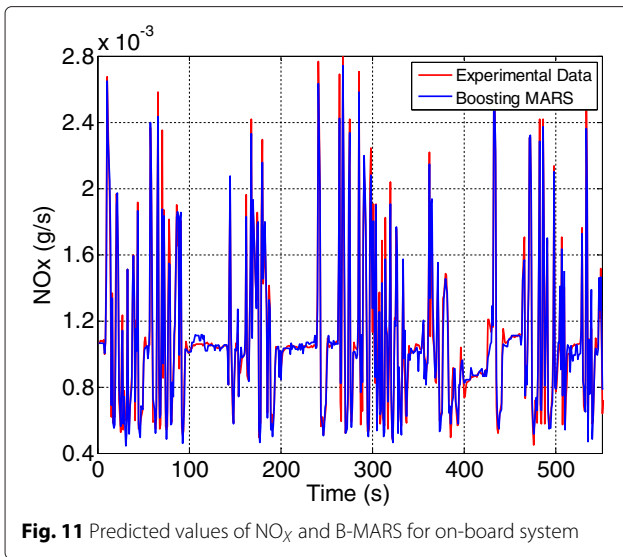


Table 5 Comparison of MARS, B-MARS and MLR model

Model	RMSE	MSE	R^2
MARS-OBS	0.00016	2.565×10^{-8}	0.87
MARS-DYN	0.00022	4.642×10^{-8}	0.77
B-MARS-OBS	0.00011	1.236×10^{-8}	0.93
B-MARS-DYN	0.00014	1.905×10^{-8}	0.89
MLR-OBS	0.00046	2.571×10^{-8}	0.51
MLR-DYN	0.00048	3.124×10^{-8}	0.50

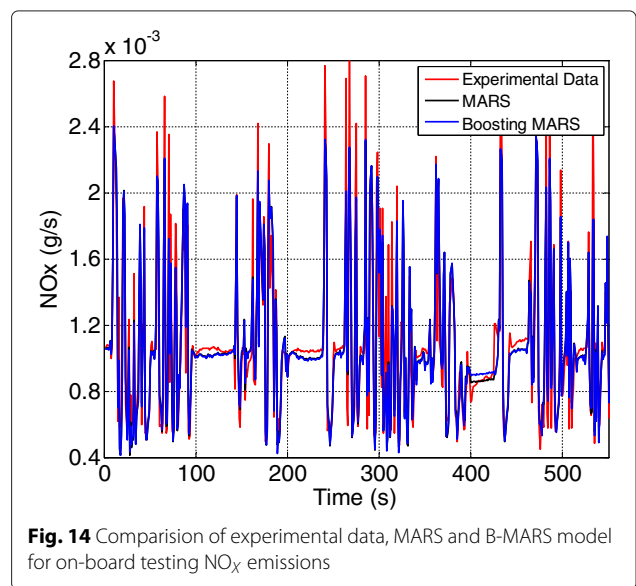
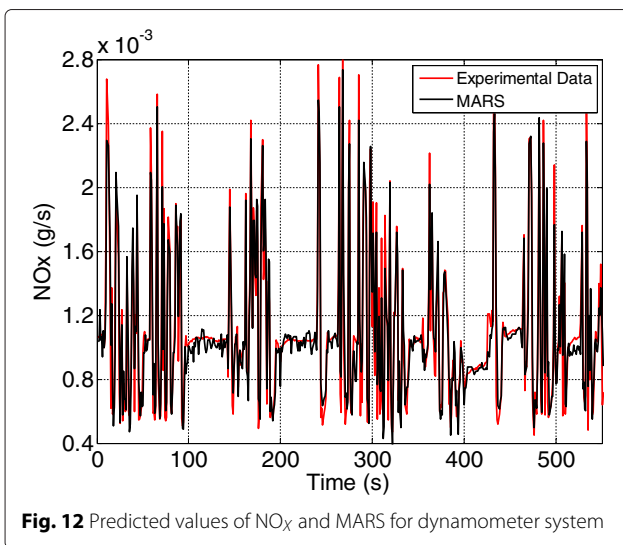
temperature (indicated by BF13). The higher ambient temperature resulting in more vehicle NO_x emission is expected, because NO_x is formed in a larger quantity in the cylinder as the combustion temperature exceeds the required limit. This finding is also consistent with previous explanation. In addition, temperatures greater than 24.76°C (B13) will significantly produce NO_x emissions. As indicated by BF14, BF15 and BF16, the MARS results show the effect of load: (1) if the load is less than 10.53 Nm , then it has no effect on NO_x emission (indicated by BF14); (2) if the load is greater than 10.53 Nm but less 52.34 Nm , the NO_x emission will increase by 0.01561811 g/s for 1 Nm increase of load (indicated by BF14 and BF15); (3) if the load is greater than 52.34 Nm but less than 60.15 Nm , the vehicle NO_x emission will increase by 0.0179656 g/s for 1 Nm increase in load (indicated by BF15 and BF16) and (4) if the load is greater than 60.15 Nm the NO_x emission will increase by 0.02324571 g/s for 1 Nm increase in load (indicated by BF16). As far as the effect of power on NO_x emission, BF17 and BF18 indicate that the occurrence can be described as: (1) if the power is less than 8.98 W , then it has no effect on vehicle NO_x emission (indicated by BF17); (2) if the power is greater than 8.98 W but less 21.32 W , NO_x emission will increase by 0.01567893

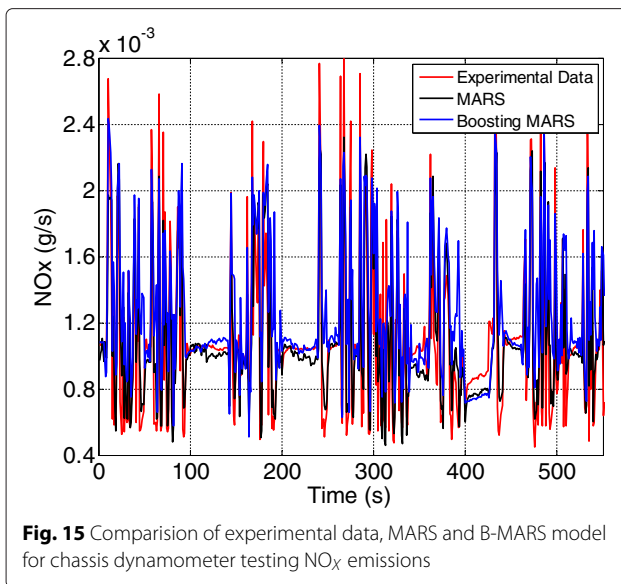




g/s for 1 W increase of power (indicated by BF17 and BF18); (3) if the power is greater than 21.23 W, the vehicle NO_x emission will increase by 0.01567893 g/s for 1 W increase in power (indicated by BF18). The NO_x emission as a result of the increasing load and power is expected, following the remark by Pierson et al. 1996 that driving a vehicle against a higher resistance will increase the engine load and power which will result in increases of the carbon dioxide (CO₂) and NO_x emissions. To illustrate the NO_x emission during real-world driving conditions and the dynamometer testing drive cycle, Figs. 4 and 5 show the MARS model that has the best performance basis on independent test samples. There were 556 data points used in the analysis, 65 % of which for building the model (Learn) and 35 % for validation (Test).

The on-board system models have seventeen and nineteen basis functions with the best model with the least mean square error occurring at 15th basis function for B-MARS and 17th for MARS, *R* values of 97 % and 93 % for B-MARS and MARS while the chassis dynamometer testing gave *R* of 94 % and 88 % for B-MARS and MARS models with the best model occurring at BF11 and 12. As shown in Figs. 6, 7, 8 and 9, the regression correlation coefficient *R* of the selected models clearly demonstrate strong positive correlation in the NO_x emissions model. In Table 5, we try to compare the MARS, B-MARS and Multiple Linear Regression (MLR) models. We learn that B-MARS obtained the best results and high accuracy with good *R*² values of 93 % and 89 % respectively for on-board and dynamometer testing while the MARS





model obtained R^2 of 87 % for the on-board and 77 % for the dynamometer testing. However, the MLR model shows the least performance with R^2 values of 51 % and 50 % for both the on-board and the dynamometer test. The improved performance of the B-MARS algorithm was clearly demonstrated with low root mean square error (RMSE) of 0.00011 and 0.00014 and mean square error of (MSE) 1.236×10^{-8} and 1.905×10^{-8} as compared with MARS, RMSE of 0.00016 and 0.00022, MSE 2.565×10^{-8} and 4.642×10^{-8} for the on-board and dynamometer testing. The contribution achieved by the boosting algorithm confirms its ability not only improving the prediction accuracy of the NO_x emission but also perform better in process. Figures 10, 11, 12, 13, 14 and 15 provide a detailed plot for comparing the experimental data with the predicted B-MARS and MARS techniques. Note that the predicted emissions follow the experimental data with sufficiently good precision with the B-MARS proving very strong in the NO_x emissions prediction. This suggests the robustness of the B-MARS algorithm and its capability of improving the accuracy of the MARS model in NO_x emissions prediction.

Conclusion

In this paper, we have proposed the use of Multivariate Adaptive Regression Splines (MARS) and Boosting Multivariate Adaptive Regression Splines (B-MARS) algorithms to effectively estimate vehicular NO_x emissions. The model approximates the nonlinear relationship between the NO_x emission, a function of speed, acceleration, temperature, power and load, considered as predictor variables. The B-MARS model is implemented with 14 and 17 piecewise-linear basis functions while the MARS model with 19 and 15 BFs. The model predicts the NO_x

emission by forming a weighted sum of the predictor variables; thus, the predicted emission changes in a smooth and regular fashion with respect to the input variations, offering some performance improvements. The results obtained indicate a promising application of the proposed algorithms to accurately estimate NO_x emissions with a reasonable accuracy. The method may usefully assist in a decision-making policy regarding urban air pollution.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed equally and significantly in writing this article. All authors read and approved the final manuscript.

Acknowledgements

The data used for this study were supplied by the Road and Maritime Service, Department of vehicle emission, Compliance Technology & Compliance Operations, NSW Office of Environment & Heritage, and Horiba Australia. Assistance provided by Paul Walker and Thomas Mahsling is gratefully acknowledged.

Author details

¹Faculty of Engineering and Information Technology, University of Technology Sydney, Broadway, NSW 2007 Sydney, Australia. ²Office of Environment and Heritage, Lidcombe, NSW 1825 Sydney, Australia.

Received: 21 January 2014 Accepted: 23 March 2015

Published online: 10 June 2015

References

- Afotey, B, Sattler, M, Mattingly, SP, Chen, VCP (2013). Statistical model for estimating carbon dioxide emissions from a light-duty gasoline vehicle. *Journal of Environmental Protection*, 4, 8–15.
- Carlsaw, D, Beevers, S, Westmoreland, E, Williams, M, Tate, J, Murrells, T, Stedman, J, Li, Y, Grice, S, Kent, A, Tsagatakis, I (2011). *Trends in NO_x and NO₂ Emissions and Ambient Measurements in the UK*. King's College London: University of Leeds Press.
- Duc, H, Azzi, M, Wahid, H, Ha, QP (2013). Background ozone level in the sydney basin: assessment and trend analysis. *International Journal of Climatology*, 33, 2298–2308.
- Durbin, T, Johnson, K, Cocker, ID, Miller, J, Maldonado, H, Shah, A, Ensfield, C, Weaver, C, Akard, M, Harvey, N (2007). Evaluation and comparison of portable emissions measurement systems and federal reference methods for emissions from a back-up generator and a diesel truck operated on a chassis dynamometer. *Environmental Science & Technology*, 41, 6199–6204.
- Freund, Y, & Schapire, R (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119–139.
- Frey, HC, Unal, A, Roupail, NM, Colyar, JD (2002). Use of on-board tailpipe emissions measurements for development of mobile source emission factors. In *Proceedings of US Environmental Protection Agency Emission Inventory Conference* (pp. 1-13). Atlanta, April.
- Frey, HC, Unal, A, Roupail, NM (2003). On-road measurement of vehicle tailpipe emissions using a portable instrument. *Journal of the Air and Waste Management Association*, 53, 992–1002.
- Friedman, JH (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19, 1–141.
- Hastie, T, Tibshirani, R, Friedman, J (2001). *The elements of statistical learning: Data mining, inference and prediction*, (pp. 337-343). Stanford, California: Springer-Verlag.
- Jerome, HF (2001). Greedy function approximation: A gradient boosting mac. In *The Annals of Statistics*, 29 (pp. 1189–1232). Chapman and Hall: Institute of Mathematical Statistics.
- Maykut, NN, Lewtas, J, Kim, E, Larson, TV (2007). Source apportionment of pm 2.5 at an urban improve site in seattle, washington. *Environmental Science and Technology*, 37, 5135–5142.

- Nine, RD, Clark, N, Daley, JJ, Atkinson, CM (1999). Development of a heavy-duty chassis dynamometer driving route. *Journal of Automobile Engineering*, 213, 561–574.
- Oduro, SD, Metia, S, Duc, H, Ha, QP (2013). CO₂ vehicular emission statistical analysis with instantaneous speed and acceleration as predictor variables, In *The 2nd International Conference on Control, Automation and Information Sciences* (pp. 158-163). Vietnam: Nha Trang.
- Oduro, SD, Metia, S, Duc, H, Hong, G, Ha, QP (2014). Prediction of no_x vehicular emissions using on-board measurement and chassis dynamometer testing, In *Proceedings of The 31st International Symposium on Automation and Robotics in Construction and Mining* (pp. 584-591). Sydney, Sydney Australia: University of Technology.
- Pierson, WR, Gertler, AW, Robinson, NF, Sagebiel, JC, Zielinska, B, Bishop, AW, Stedman, DH, Zweidinger, RB, Ray, WD (1996). Real world automotive emissions summary of studies in the fort mchenry and tuscarora mountain tunnels. *Atmospheric Environment*, 30, 2233–2256.
- Put, R, Xu, Q, Massart, D, Heyden, Y (2004). Multivariate adaptive regression splines (mars) in chromatographic quantitative structure-retention relationship studies. *Journal of Chromatography*, 1055, 11–19.
- Querol, X, Viana, M, Alastuey, A, Amato, F, Moreno, T, Castillo, S, Pey, J, de la Rosa, J, Sánchez de la Campa, A, Artíñano, B, Salvador, P, García Dos Santos, S, Fernández-Patier, R, Moreno-Grau, S, Negral, L, Minguillón, MC, Monfort, E, Gil, JI, Inza, A, Ortega, LA, Santamaría, JM, Zabalzah, J (2007). Source origin of trace elements in pm from regional background, urban and industrial sites of spain. *Atmospheric Environment*, 44, 7219–7231.
- Rakha, H, Ahn, K, Moran, K, Saerens, B, den Bulck E, V (2011). Simple Comprehensive Fuel Consumption and CO₂ Emissions Model Based on Instantaneous Vehicle Power, In *Transportation Research Board 90th Annual Meeting*. Washington DC, 23-27 January 2011, Paper No. 11-1009.
- Sharma, AR, Kharol, SK, Badarinath, KVS (2010). Influence of vehicular traffic on urban air quality – a case study of hyderabad, india. *Transportation Research Part D: Transport and Environment*, 15, 154–159.
- Tóth-Nagy, C, Conley, JJ, Jarrett, RP, Clark, NN (2006). Further validation of artificial neural network-based emissions simulation models for conventional and hybrid electric vehicles. *Journal of the Air & Waste Management Association*, 56, 898–910.
- European Commission White Paper. Road to a Single European Transport Area -Towards a Competitive and Resource Efficient Transport System, Brussels", COM, 144 Final. http://ec.europa.eu/transport/strategies/2011_white_paper_en.htm.
- MARS User Guide:San Diego, Salford System. <http://www.salford-systems.com>.
- World Health Organization Fact Sheet No 313 "Air Quality and Health Updated". <http://www.who.int/mediacentre/factsheets/fs313/en/>.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
