

## Research Article

# Genetic Algorithm Based Microscale Vehicle Emissions Modelling

Sicong Zhu,<sup>1</sup> LiSian Tey,<sup>2</sup> and Luis Ferreira<sup>3</sup>

<sup>1</sup>MOE Key Laboratory for Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, 3 Shangyuan Cun, Haidian District, Beijing 100044, China

<sup>2</sup>Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

<sup>3</sup>School of Civil Engineering, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Brisbane St Lucia, QLD 4072, Australia

Correspondence should be addressed to Sicong Zhu; [sczhu@bjtu.edu.cn](mailto:sczhu@bjtu.edu.cn)

Received 27 August 2015; Revised 24 November 2015; Accepted 25 November 2015

Academic Editor: Ivanka Stamova

Copyright © 2015 Sicong Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is a need to match emission estimations accuracy with the outputs of transport models. The overall error rate in long-term traffic forecasts resulting from strategic transport models is likely to be significant. Microsimulation models, whilst high-resolution in nature, may have similar measurement errors if they use the outputs of strategic models to obtain traffic demand predictions. At the microlevel, this paper discusses the limitations of existing emissions estimation approaches. Emission models for predicting emission pollutants other than CO<sub>2</sub> are proposed. A genetic algorithm approach is adopted to select the predicting variables for the black box model. The approach is capable of solving combinatorial optimization problems. Overall, the emission prediction results reveal that the proposed new models outperform conventional equations in terms of accuracy and robustness.

## 1. Introduction

Fuel consumption and emission estimation can be critical for comprehensive transportation planning. In light of the strong relationship between CO<sub>2</sub> emissions and fuel consumption [1] and the high volatility in crude oil prices, this requires more accurate and robust tools to quantify environmental impacts so that project evaluation can adequately address community expectations. Quantitative travel demand and emission models are necessary for the evaluation of future transport/land use options, as well as for the management of existing transport systems. The modelling of emissions is seen as an increasingly important tool in transportation planning and management.

Historically, car-following and traffic flow models have been developed using different theoretical basis. This has given rise to two main kinds of models of traffic dynamics, namely, microscopic representations, based on the description of the individual behavior of each vehicle, and macroscopic representations describing traffic as a continuous flow

obeying global rules [2]. Strategic travel demand models tend to be large and regional in nature whereas microsimulation models are used for detailed tactical or operational testing of options. Taking the highest macroscopic level as an example, the total vehicle flow and the average speed over an entire network may be all that is provided [3]. At the lowest level of the hierarchy, high-resolution microscopic transportation models typically produce second-by-second vehicle trajectories (location, speed, acceleration, etc.). Hence, the traffic modelling and emission modelling should match by the accuracy level and aggregation level. For instance, driving cycles used for vehicle emission testing are specified on a second-by-second speed-time profile. Microscopic traffic models should integrate real time emission prediction models, which are able to utilize high-resolution transportation modelling results, therefore generating potentially more precise emission estimations.

Taking the highest macroscopic level as an example, the total vehicle flow and the average speed over an entire network may be all that is provided [4]. Correspondingly,

a number of emission models deliver the predictions based on the aggregated traffic model outputs, for example, the ARTEMIS in Europe and Mobile 6 from US [5, 6]. The common macrolevel modelling approach used to produce a mobile source emission inventory is based on two processing steps. The first step consists of determining a set of emission factors that specifies the rate at which emissions are generated, and the second step is to produce an estimate of vehicle activity. The emission inventory is then calculated by multiplying the results of these two steps together. This methodology has two major shortcomings as follows.

*Inaccurate Characterization of Traffic Activities.* The overall rate of error in 20-year traffic forecasts resulting from strategic transport models is likely to be large due to the uncertainty attached to input forecasts. The predictive ability of current emission models depends on (a) the errors in the input data used to estimate the amount and characteristics of traffic flow—such data are usually the result of travel demand modelling analysis that is known to have a large degree of uncertainty associated with it—and (b) the adequacy and quality of the information on emission rates that is used in the models.

*Emissions Factors May Not Represent Actual Conditions Adequately.* The current methods used for determining emission factors are based on average driving characteristics embodied in a predetermined driving cycle. Emission measurements are used as base values to reconstruct statistically the relationship between emission rates and average vehicle speeds. These “averaged speeds” are at variance with the vehicle dynamic operation in microscope.

At the lowest level of the hierarchy, high-resolution microscopic transportation models typically produce second-by-second vehicle trajectories (location, speed, and acceleration). Driving cycles used for vehicle emission testing are also specified on a second-by-second speed-time profile. Microscopic models should be integrated with real time emission prediction models which are able to utilize high-resolution transportation modelling results, thereby generating potentially more precise emission estimates. Several commercial microsimulation traffic packages are widely used to estimate the emissions [7].

There have been a number of modelling approaches on microlevel proposed to estimate future vehicle emissions in conjunction with the outputs of transport models. One such approach is the use of engine power as the main predictive basis. Another is the use of vehicle speed and acceleration as predictive variables. There are three main types of modelling approaches, namely, power-based, speed-based, and hybrid models. Two models from each category, as shown in Table 1, have been analysed [13]. It is found that power-based models have good performance on CO<sub>2</sub> emission and fuel consumption. In contrast, the predicted HC, CO, and NO<sub>x</sub> do not match well with the measured results [8]. Moreover, it is difficult to collect some coefficients associated with instantaneous power, such as aerodynamic drag coefficient, vehicle frontal area, and gearing. The results from speed-based models highlight the need to model acceleration,

TABLE 1: Emission models assessed.

General approach	Model	Reference
Power-based	Commonwealth scientific and industrial research organisation (CRISCO)	[8]
	Comprehensive modal emissions model (CMEM)	[3]
Speed-based	Energy and emissions model (VT)	[9]
	Instantaneous traffic emissions	[10]
Hybrid	Microscale modelling	[11]
	Microscale modelling	[12]

deceleration, and cruising stages of the urban cycle separately. The instantaneous traffic emissions model, a speed-based approach which utilises the microtransportation simulation result as an input, was found to have merit based on the evaluation results. More complex models, whilst theoretically more desirable, may mean additional input measurement errors, such that the overall effect may not yield more accurate estimates [14].

The instantaneous traffic emissions model developed by Int Panis et al. [10] has been adopted by the AIMSUN traffic simulation model [7]. The latter integrates traffic simulation results with emission prediction equations. Emission functions for each vehicle were derived with instantaneous speed and acceleration as parameters using nonlinear multiple regression techniques. The model, shown in (1), was calibrated using data from twenty-five vehicles (six buses, two trucks, and seventeen cars) in Europe. The pollutants modelled are nitrogen oxide (NO<sub>x</sub>), hydrocarbon (HC), carbon dioxide (CO<sub>2</sub>), and particulate matter (PM):

$$E_n(t) = \max \left[ E_0, f_1 + f_2 v_n(t) + f_3 v_n(t)^2 + f_4 a_n(t) + f_5 a_n(t)^2 + f_6 v_n(t) a_n(t) \right], \quad (1)$$

where  $E_0$  is a lower limit of emission (g/s) specified for each vehicle and pollutant type;  $v_n(t)$  is instantaneous speed of vehicle  $n$  at time  $t$ ;  $a_n(t)$  is acceleration of vehicle  $n$  at time  $t$ ;  $f_1$  to  $f_6$  are emission constants specific for each vehicle and pollutant type determined by the regression analysis.

For certain pollutants, whenever visual inspection of the data plot reveals a clear distinction in the scatter for acceleration and deceleration, the approach suggests that different functional forms should be derived for different driving modes, such as acceleration (with  $a_n(t) \geq 0.5 \text{ m/s}^2$ ), cruising (with  $-0.5 \text{ m/s}^2 \leq a_n(t) < 0.5 \text{ m/s}^2$ ), and deceleration (with  $a_n(t) < -0.5 \text{ m/s}^2$ ). The modelling results for pollutant emissions such as HC are unsatisfactory, as shown in Section 4.

The current paper aims to develop vehicle emission models for predicting emission pollutants other than CO<sub>2</sub>. The models adopt genetic algorithm (GA) for selection of

TABLE 2: Selected average-aged vehicles for model development.

Vehicle <sup>(1)</sup>	Variant	Odometer (km)	Engine displacement (L)	Mass (kg)
323	Protégé	75,615	1.6	1215
Echo	—	53,859	1.5	921
Vectra	CD	81,666	2.0	1317
323	Protégé	62,229	1.8	1142
Camry	CSI	81,783	2.2	1395
Tarago	GLi	90,749	2.4	1615
Pulsar	LX	65,120	1.6	1067
Commodore	Executive	95,979	3.8	1654

<sup>(1)</sup> All vehicles are 4-cylinder except the Commodore (V6).

the predicting variables. This paper is structured as follows: Section 2 briefly describes the data sources used and the selection and validation prior to the model development; Section 3 proposes the model development approach; and Section 4 then discusses the findings. Finally, some overall conclusions and limitations are drawn from the results and future research avenues are recommended.

## 2. Data Sources, Selection, and Validation

The data used for analysis in the current paper were extracted from the Australian national in-service emissions study (NISE2) [15], which was developed using a Composite Urban Emissions Drive Cycle (petrol CUEDC). This cycle represents typical vehicle journeys in a given metropolitan study area, where data were collected in the Brisbane, Sydney, Melbourne, Adelaide, and Perth areas. The emission rates for CO<sub>2</sub>, CO, HC, and NO<sub>x</sub> of the test-bed vehicles from the NISE2 fleet, which travels on the composite urban driving cycle (CUEDC), were recorded second-by-second in addition to the instantaneous speed. Prior to analysis, the integrity of the emission measurements from NISE2 was reconfirmed and corrected to enhance the reliability of the instantaneous emissions using the method followed by Smit et al. [16]. In addition, cold start affected datasets were filtered prior to analysis by adopting an approach recommended by Favez et al. [17]. Eight average-aged passenger vehicles (which had travelled approximately 50,000–100,000 km each), as listed in Table 2, were shortlisted from the NISE2 database for model development. The vehicles selected are similar to Australian vehicle fleet characteristics (e.g., in mileage and age) [18]. The instantaneous emission observations were averaged over those vehicles.

The AIMSUN emission methodology was adopted for further evaluation using the NISE2 data subset described earlier. Each vehicle was “driven” through the speed profile of the CUEDC drive cycle which was simulated using a purpose-specific program. The averaged CO<sub>2</sub> emission observations from the selected eight vehicles were compared with the predicted second-by-second output from the AIMSUN model. The goodness of fit,  $R^2$ , for the two sets of data was 0.80, as shown in Figure 1.

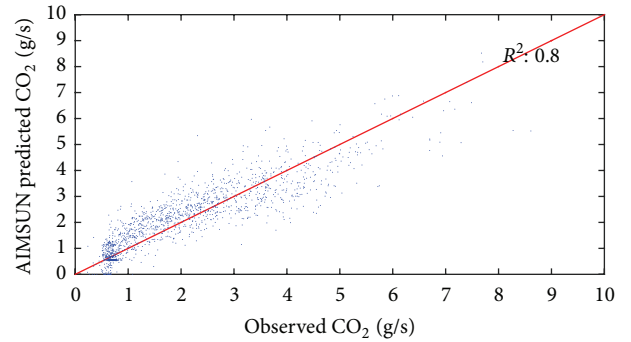


FIGURE 1: CO<sub>2</sub> by AIMSUN model predictions and averaged observed data.

## 3. Development of Emission Models

Microsimulation based modelling results for pollutants other than CO<sub>2</sub> are usually unsatisfactory [12]. The current paper proposes a methodology to improve those results. The concept is briefly described below.

Current microscopic models use a combination of instantaneous velocity and acceleration to predict various gaseous pollutants including HC and CO. Gaseous pollutants such as HC and CO are primarily formed during in-cylinder combustion processes depending on many factors such as air-fuel ratio, cylinder temperature and pressure, and engine speed [19]. The formation of HC and CO rises in a rich fuel environment. Thus, the fact of high correlation between CO<sub>2</sub> emission rate and fuel consumption was taken into account in the modelling of these gaseous pollutants. In addition, vehicle acceleration or deceleration leads to substantial change in fuel injection per combustion cycle. The change in air-fuel ratio forces the engine to adapt to a new equilibrium and tends to lead to a transient variation in pollutant formation [20]. This effect may be compounded by dynamic effects in the catalyst and exhaust system, such as catalyst malfunctioning, which can cause a sudden increase of the pollutant emissions. For these reasons, modelling of HC and CO, as the products of incomplete combustion fuel, should take the time-lag effect described above into consideration. Hence, several “historical” variables (i.e., variables at previous time steps of

$t-1$ ,  $t-2$ ,  $t-3$ , or  $t-4$  seconds) of the time-lag effect are introduced into the models as predicting variables.

A range of variables for the instantaneous and “historical” velocity, acceleration, and CO<sub>2</sub> emission rate were selected and tested:

$$[B_0, V_i, A_i, V_i^2, A_i^2, V_i A_i, V_i^3, A_i^3, V_i A_i^2, V_i^2 A_i, V_{i-1}, V_{i-2}, V_{i-3}, V_{i-4}, A_{i-1}, A_{i-2}, A_{i-3}, A_{i-4}, \text{Rate\_CO}_{2i}, \text{Rate\_CO}_{2i-1}, \text{Rate\_CO}_{2i-2}, \text{Rate\_CO}_{2i-3}, \text{Rate\_CO}_{2i-4}],$$

where  $B_0$  is constant;  $V_i$  is velocity at time  $i$ ;  $A_i$  is acceleration at time  $i$ ;  $\text{Rate\_CO}_{2i}$  is CO<sub>2</sub> emission rate at time  $i$ .

A total of 23 candidate predicting variables were identified. An exhaustive enumeration method is not a practical alternative, while stepwise and stagewise regression procedures produce only local optimum solutions [21]. Current highly divided emissions models that are reviewed in Section 1 cannot provide any prototype of variable combination. Compared with other artificial intelligent algorithms, unsupervised heuristic genetic algorithm is independent from any *a priori* knowledge, such as training dataset. In addition, the GA is a stochastic search process that mimics the natural process of “survival of the fittest” through the manipulation of a population of chromosomes [22]. With proper binary encoding, the GA can deliver a fair hybrid emission model which synthesizes variables of different models reviewed in Section 1. To be pointed out, the GA may not deliver an optimized result due to premature convergence and “Hamming cliff” problems [23]. However, it is more difficult, but not impossible, for a GA to become trapped in a local minimum unlike the more conventional gradient methods [24].

The elite individual solution is a balanced one with high accuracy and consisting of strong statistically related variables. The GA programme was compiled using Matlab 8.4. A selective weighted fitness for the GA, (2), was used to implement an automated variables selection procedure to build the calibration models based on least-square regression:

$$\text{Fitness}_i = \frac{1}{\sqrt{\sum_{i=1}^N (P_i - O_i)^2 / N}}, \quad (2)$$

where  $P_i = \sum \alpha_i \beta_i \text{Variable}_i$ ;  $P_i$  is estimated instantaneous emission rate at time  $i$ ;  $O_i$  is observed instantaneous emission rate at time  $i$ ;  $\alpha_i$  is binary logic control parameter after regression analysis;

$$\alpha_i = \begin{cases} 1 & \text{if } P \text{ value} < 0.05 \text{ this variable is included} \\ 0 & \text{otherwise;} \end{cases} \quad (3)$$

$\beta_i$  is coefficient derived from regression analysis;  $N$  is number of observations.

$P_i$  is the prediction of emission rate as a product of selected coefficients and the corresponding candidate variables whose  $P$  value of the  $t$ -statistical analysis is lower than 0.05. The reciprocal relationship of root-mean-square error enables the accurate prediction of solutions with high fitness values.

TABLE 3: Parameters tuning of genetic operators.

Population size ( $N$ )	600
Crossover rate ( $C$ )	1
Crossover type	Uniform crossover
Scaling method	Sigma
Sigma scaling coefficient	1
Mutation rate ( $M$ )	0.003
Mutation method	Bit string
Selection strategy ( $S$ )	Pure selection

Figure 2 shows the flowchart of the GA application. Firstly, chromosomes that represent feasible solutions of a nonlinear optimisation problem are randomly generated to form the initial population following the fitness test (see (2)). Each individual chromosome in binary format symbolises a potential solution to the emission modelling problem. Figure 3 demonstrates an example of an individual chromosome consisting of 23 bits. Each bit represents a corresponding variable in the list of all potential variables. The dichotomous data on each bit are the choice of “included in the model” valued 1 or “not included in the model” valued 0. The size of the solution is therefore  $2^{23}$ . In determining the fitness value, the chromosome tends to take more variables for better prediction accuracy. This may include irrelevant candidate variables and undermine the robustness of the model. In order to avoid this from happening, the calculation of the emission rate prediction,  $P_i$ , was limited to those statistically significant variables ( $P < 0.05$ ). In order to achieve acceptable solutions, the calculations of the fitness values and selection of chromosomes continued. This procedure evolves through many generations by a natural genetic process. The genetic process includes three types of operation, namely, crossover, selection, and mutation [25]. It repeats until the number of iterations exceeds a predefined limit. The number of chromosomes in a population and the number of iterations are set to 600 and 100, respectively. The tuning of genetic operator parameters follows the simple genetic algorithm proposed by Mitchell [25]; the details are presented in Table 3.

Figures 4(a) and 4(b) show the maximum and averaged fitness reached over the iterations. The averaged fitness for each generation increases substantially at the beginning stage and reaches stability towards the end. The elite fitness trends upward and fluctuates at the beginning. After 20 generations’ evolution, the fitness becomes mature. These results indicate that the chromosome evolvments improve the accuracy of the solution and reach a saturated condition eventually. After 30–40 different runs, the paper shortlists the results and selects the modelling equation based on statistical analysis. Table 4 shows three candidates for acceleration-phase HC emission modelling. Based on modelling fitness and variable significance, this paper selects candidate 1 to be the promising one. As the GA program can be configured to enable the mutation operator only, a random solution is generated by a  $600 \times 100$  random process. The chosen one is compared with the random solution. However, the randomly selected candidate is hardly a solution, due to variable significance

TABLE 4: HC acceleration phase modelling candidates.

(a) Candidate 1								
Modeling $R^2$	0.6474							
Variable name	$B_0$	$A$	$VA$	$V^3$	$V^2A$	$V_{i-1}$	$CO_2$	$CO_{2i-3}$
$P$ value	0.0151	0.0008	0.00172	$5.52E-05$	0.003202	$1.03E-06$	$7.94E-13$	0.016642
(b) Candidate 2								
Modeling $R^2$	0.6449							
Variable name	$B_0$	$V$	$VA$	$V^3$	$V^2A$	$V_{i-1}$	$A_{i-3}$	$CO_2$
$P$ value	0.0088	0.00264	$5.99E-03$	0.000214	$1.94E-02$	$5.54E-03$	0.035107	$1.19E-10$
(c) Candidate 3								
Modeling $R^2$	0.6493							
Variable name	$B_0$	$A$	$V^2$	$V^3$	$VA^2$	$A_{i-4}$	$CO_2$	
$P$ value	0.0492	0.01617	$5.2E-08$	$1.57E-07$	0.036635	0.022996	$3.23E-14$	

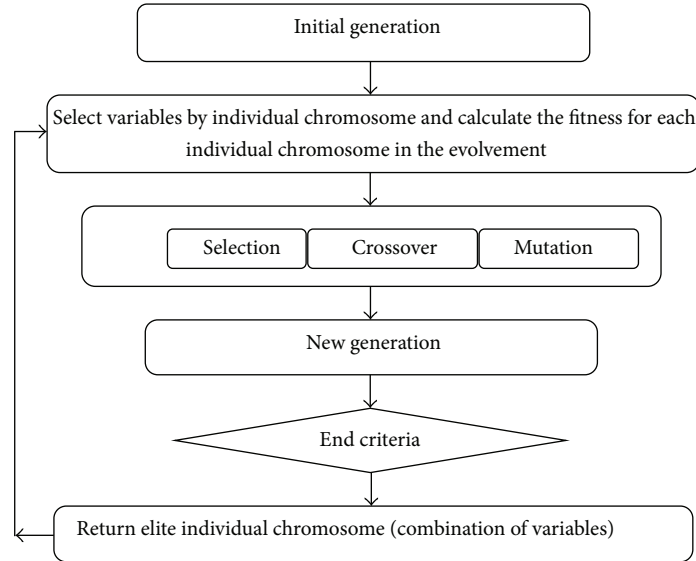


FIGURE 2: Flowchart of genetic algorithm.

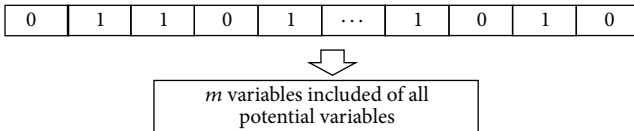


FIGURE 3: An individual chromosome.

constraint ( $P$  value). Hence the elite chromosome can be justified as the optimal (or subprime) solution.

## 4. Results and Main Findings

**4.1. General.** The proposed GA approach was applied to the HC pollutant for different driving modes, namely, acceleration, cruising, and deceleration. The models for the emission rate,  $Y$ , from the ultimate chromosome are presented in (4), (5), and (6), respectively. All the predicting variables

selected are statistically significant at  $P < 0.05$ . Both the instantaneous traffic emissions model (see (1)) and the newly modified equation parameters were calibrated by least-square regression on the same test-bed dataset that was used to develop the new model. To conduct meaningful comparisons of modelling-induced error, the goodness of fit,  $R^2$ , for the three driving modes, between the newly modified model and model (1) is shown in Table 5. The proposed new approach significantly improved the modelling results for HC, although the prediction accuracy for the deceleration component was relatively low. The deceleration component accounts for 42% of total driving cycle duration, but the summation of HC deceleration component only takes up 25% of total cycle measurements. As a result, the deceleration component does not play an important role as a percentage of total emissions. Due to the multidimensionality of the solution search space, the GA method is not certain to lead to a global optimum when formulated as an optimization



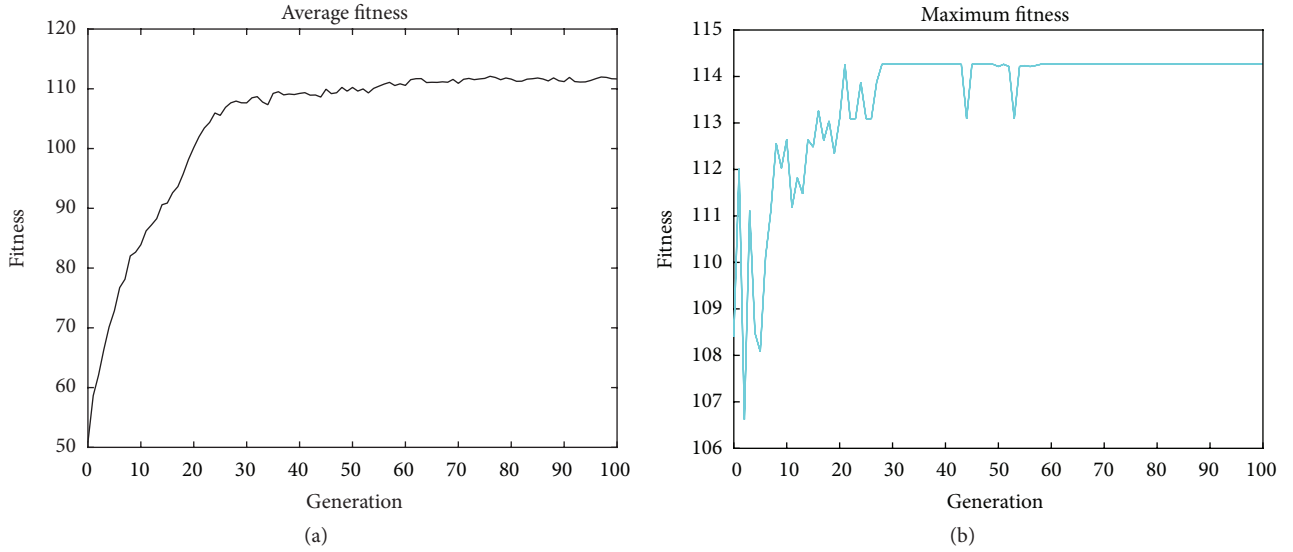


FIGURE 4: (a) Averaged fitness over generations. (b) Maximum averaged fitness over generations.

problem. GA has a tendency to converge towards local optima or even arbitrary points rather than the global optimum of the problem. This means that it is not wise to sacrifice short-term fitness to gain longer-term fitness. To tackle this issue, the proposed GA methodology relies on the fitness function, which enables quantification of individual solution appropriateness in terms of the statistical significance of both model accuracy and shortlisted variables. To test the fitness function, the size of the population was increased to 2,000 in order to enable more potential solutions to search in the multidimensional space, the results being similar to those previously presented. Moreover, taking the HC acceleration component as an example, the new proposed equation  $R^2$  is slightly lower than the theoretical maximum value by which an individual chromosome includes all candidature variables. Hence, the new equation selected by the new algorithm is a close approximation to the global optimum.

#### Acceleration Component

$$Y = B_0 + B_1 A_i + B_2 V_i A_i + B_3 V_i^3 + B_4 V_i^2 A_i + B_5 V_{i-1} + B_6 \text{Rate\_CO}_{2i} + B_7 \text{Rate\_CO}_{2i-3}. \quad (4)$$

#### Cruising Component

$$Y = B_0 + B_1 A_i + B_2 V_i^2 + B_3 V_i A_i + B_4 V_i^3 + B_5 V_i^2 A_i + B_6 V_{i-2} + B_7 V_{i-4} + B_8 \text{Rate\_CO}_{2i} + B_9 \text{Rate\_CO}_{2i-1} + B_{10} \text{Rate\_CO}_{2i-3}.$$

#### Deceleration Component

$$Y = B_0 + B_1 V_i + B_2 V_i^2 + B_3 V_i^3 + B_4 V_i A_i^2 + B_5 A_{i-3}, \quad (6)$$

where  $Y$  is emission rate (g/s) for each vehicle and pollutant type;  $V_i$  is velocity at time  $i$ ;  $A_i$  is acceleration at time  $i$ ;

TABLE 5: Results of model development: goodness of fit,  $R^2$ .

Driving mode	Overall	Acceleration	Cruising	Deceleration
		$R^2$		
GA-based model	0.80	0.65	0.75	0.15
Equation (1) model	0.69	0.47	0.60	0.08

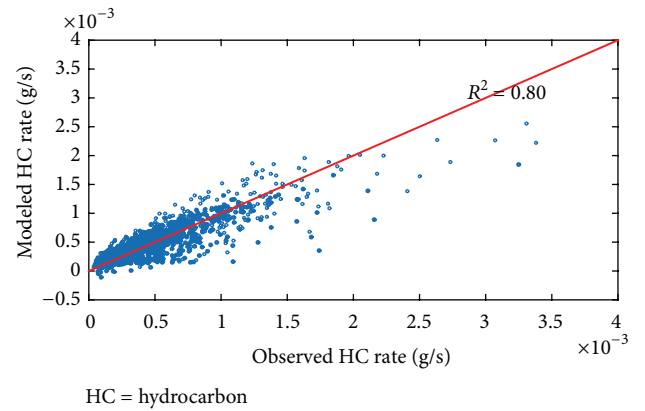


FIGURE 5: New model predicted versus measured HC.

$\text{Rate\_CO}_{2i}$  is  $\text{CO}_2$  emission rate at time  $i$ .  $B_1$  to  $B_n$  are emission coefficients for each vehicle and pollutant type.

The same experimental dataset was used to test the prediction accuracy of the model. Figure 5 shows the correlation between modelling predictions and HC measurements, and Figure 6 illustrates the corresponding residuals plots.

**4.2. Modelling Results and Validations.** The proposed new models were validated on different sets of data with various vehicle size and make for different age ranges, namely, new vehicles with mileages from 1,000 to 11,000 km; middle-aged

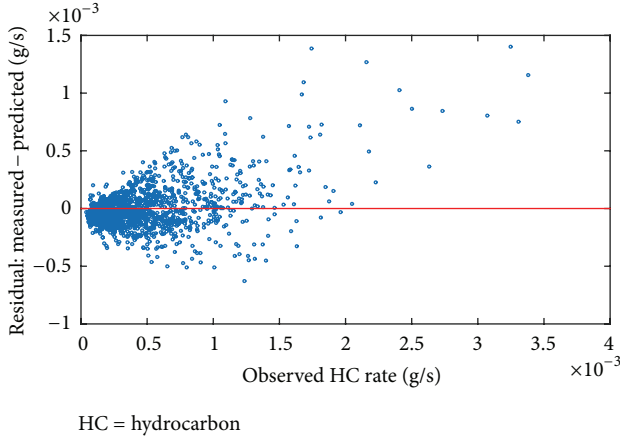


FIGURE 6: New model residuals.

vehicles with mileages from 75,000 to 95,000 km; and older vehicles with mileages from 130,000 to 140,000 km.

Comparisons of the goodness of fit,  $R^2$ , of the overall modelling results for the three vehicle age ranges, between the proposed new models (summation of (4) to (6)) and the instantaneous traffic emissions model (1) are shown in Tables 6(a) to 6(c), respectively. Overall, the emission prediction results prove that the proposed new models provide improved results. The new models are more robust and accurate for HC prediction.

Similarly, this methodology is applied to other pollutants including CO and  $\text{NO}_x$ . The new CO emission prediction equations for acceleration, cruising, and deceleration components are shown in (7), (8), and (9), respectively. The overall  $R^2$  for whole test-bed dataset is 0.70.

#### Acceleration

$$Y = B_0 + B_1 V_i^2 + B_2 V_i^3 + B_3 V_i A_i^2 + B_4 V_i^2 A_i + B_5 V_{i-2} + B_6 \text{Rate\_CO}_{2i} + B_7 \text{Rate\_CO}_{2i-4}. \quad (7)$$

#### Cruising

$$Y = B_0 + B_1 A_i + B_2 V_i^2 + B_3 V_i A_i + B_4 V_i^3 + B_5 V_i^2 A_i + B_6 V_{i-1} + B_7 A_{i-2} + B_8 \text{Rate\_CO}_{2i} + B_9 \text{Rate\_CO}_{2i-1} + B_{10} \text{Rate\_CO}_{2i-3}. \quad (8)$$

#### Deceleration

$$Y = B_0 + B_1 V_i^2 A_i + B_2 \text{Rate\_CO}_{2i-4}. \quad (9)$$

New  $\text{NO}_x$  emission prediction equations for acceleration, cruising, and deceleration components are shown in (10),

(11), and (12), respectively. The overall  $R^2$  for whole test-bed dataset is 0.82:

$$Y = B_0 + B_1 V_i^2 + B_2 V_i^3 + B_3 A_i^3 + B_4 V_i A_i^2 + B_5 V_i^2 A_i + B_6 A_{i-2} + B_7 \text{Rate\_CO}_{2i} + B_8 \text{Rate\_CO}_{2i-2} + B_9 \text{Rate\_CO}_{2i-3} \quad (10)$$

$$Y = B_0 + B_1 V_i + B_2 A_i + B_3 A_i^2 + B_4 V_i A_i + B_5 A_i^3 + B_6 V_i A_i^2 + B_7 V_i^2 A_i + B_8 V_{i-4} + B_9 A_{i-4} + B_{10} \text{Rate\_CO}_{2i} + B_{11} \text{Rate\_CO}_{2i-1} + B_{12} \text{Rate\_CO}_{2i-2} + B_{13} \text{Rate\_CO}_{2i-3} \quad (11)$$

$$Y = B_0 + B_1 V_i A_i + B_2 A_{i-4} + B_3 \text{Rate\_CO}_{2i-4}. \quad (12)$$

## 5. Paralleled Genetic Algorithm

Cantú-Paz [26] classified parallel GAs into three main types:

- (i) Global single-population master-slave GA.
- (ii) Single-population fine-grained GA.
- (iii) Multiple-population coarse grained GA.

Single-population fine-grained GA and multiple-population coarse grained GAs are suitable to tackle dynamic function optimization problems [27]. It has an important role in optimizing complex functions whose optima vary in time (learning-like process). In a master-slave GA there is a single population, but the evaluation of fitness is distributed among several processors. Matlab 8.4 enables the full functionality of the parallel language features by creating a special job on a pool of workers and connecting the pool to the Matlab client [28]. Distributed synchronous GA is based on distribution of workload among processors during the fitness function evaluation phase followed by single central population regeneration. Hence, the massive fitness computations are assigned to workers in order to improve the computation efficiency. When testing on a workstation with i7 CPU (3.6 GHz) and 32 G memory, one generation of genetic algorithm consumes 44.0 seconds without paralleled configuration and 11.8 seconds with 4-worker paralleled configuration.

## 6. Conclusions

Past research on modelling vehicle emissions other than  $\text{CO}_2$  reveals relatively weak predicting results. The current paper proposes a GA based methodology to determine the contributing variables for predicting vehicle emissions. This method provides a new approach to the selection of a combination of variables among a large potential set. The applications of the new models show enhanced results for modelling vehicle emissions, supporting the new variable selection methodology using GA. The modified fitness function for the proposed GA demonstrates the ability to establish a balanced multivariate model. In addition, the improved HC prediction results, obtained by introducing “historical”  $\text{CO}_2$

TABLE 6: (a) Model validation,  $R^2$  (new vehicles). (b) Model validation,  $R^2$  (middle-aged vehicles). (c) Model validation,  $R^2$  (older vehicles).

(a)								
	Vehicles							
	323	Echo	Vectra	323	Camry	Tarago	Pulsar	Commodore
GA-based model	0.68	0.59	0.84	0.51	0.82	0.80	0.41	0.84
Equation (1) model	0.61	0.55	0.79	0.44	0.80	0.72	0.38	0.80
(b)								
	Vehicles							
	323	Echo	Vectra	323	Camry	Tarago	Pulsar	Commodore
GA-based model	0.68	0.59	0.84	0.51	0.82	0.80	0.41	0.84
Equation (1) model	0.61	0.55	0.79	0.44	0.80	0.72	0.38	0.80
(c)								
	Vehicles							
	Festiva	Lancer	Astra	Civic	Astra	Lancer	Pulsar	
GA-based model	0.69	0.74	0.61	0.23	0.79	0.85	0.62	
Equation (1) model	0.56	0.69	0.53	0.16	0.67	0.82	0.27	

emission rates, support the time-lag effect hypothesis. The proposed GA methodology provides a solution for a combinatorial optimization problem, providing high modelling accuracy with statistically significant relationships between the selected predicting variables and the dependant variable.

Future research focus should extend the evaluation of models to include the full set of particles matter (PM). In addition, the accuracy of new models may be compromised when driving with loading or in hilly terrain. Moreover, the developed models produce relatively low accuracy in predicting emissions in deceleration mode. According to Heywood [19], HC, CO, and  $\text{NO}_x$  emissions depend on the fuel-air equivalence and injection timing. The sharp deceleration (with  $a_n(t) < -0.5 \text{ m/s}^2$ ) results in the misbalance of fuel-air equivalence and variance of injection timing.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors acknowledge the supports of this paper by the National Natural Science Foundation of China no. 51578052 and Beijing Municipal Project no. PXM2015-178215-000008.

## References

- [1] R. Akcelik and M. Besley, "Operating cost, fuel consumption and emission models in aaSIDRA and aaMOTION," in *Proceedings of the 25th Conference of Australian Institutes of Transport Research (CAITR '03)*, University of South Australia, Adelaide, Australia, December 2003.
- [2] E. Bourrel and J.-B. Lesort, "Mixing microscopic and macroscopic representations of traffic flow: hybrid model based on Lighthill-Whitham-Richards theory," *Transportation Research Record*, vol. 1852, pp. 193–200, 2003.
- [3] M. Barth and G. Scora, *Comprehensive Modal Emissions Model (CMEM), User's Guide, Version 3.01*, University of California, Riverside, Riverside, Calif, USA, 2006.
- [4] M. Barth, F. An, J. Norbeck, and M. Ross, "Modal emissions modeling: a physical approach," *Transportation Research Record*, vol. 1520, pp. 81–88, 1996.
- [5] M. André and M. Rapone, "Analysis and modelling of the pollutant emissions from European cars regarding the driving characteristics and test cycles," *Atmospheric Environment*, vol. 43, no. 5, pp. 986–995, 2009.
- [6] EPA, *User's Guide to MOBILE6.1 and MOBILE6.2; Mobile Source Emission Factor Model, EPA420-R-03-010*, US Environmental Protection Agency, Washington, DC, USA, 2003.
- [7] TSS, *Aimsun 6.1 Manual*, Transport Simulation System, Barcelona, Spain, 2010.
- [8] D. Y. C. Leung and D. J. Williams, "Modelling of motor vehicle fuel consumption and emissions using a power-based model," *Environmental Monitoring and Assessment*, vol. 65, no. 1-2, pp. 21–29, 2000.
- [9] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *Journal of Transportation Engineering*, vol. 128, no. 2, pp. 182–190, 2002.
- [10] L. Int Panis, S. Broekx, and R. Liu, "Modelling instantaneous traffic emission and the influence of traffic speed limits," *Science of the Total Environment*, vol. 371, no. 1–3, pp. 270–285, 2006.
- [11] Y. Qi, H. Teng, and L. Yu, "Microscale emission models incorporating acceleration and deceleration," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 348–359, 2004.
- [12] R. Smit and J. McBroom, "Use of microscopic simulation models to predict traffic emissions," *Road and Transport Research*, vol. 18, no. 2, pp. 49–54, 2009.
- [13] S. Zhu and L. Ferreira, "Evaluation of vehicle emissions models for micro-simulation modelling: using  $\text{CO}_2$  as a case study," *Road and Transport Research*, vol. 21, pp. 3–18, 2012.



- [14] R. Smit, H. Dia, and L. Morawska, "Road traffic emission and fuel consumption modelling: trends, new developments and future challenges," in *Traffic-Related Air Pollution*, pp. 29–68, Nova Science Publishers, New York, NY, USA, 2009.
- [15] Orbital Australia, *Second National In-Service Emissions Study (NISE2): Light Duty Petrol Vehicle Emissions Testing*, Road and Traffic Authority of New South Wales, 2009.
- [16] R. Smit, J. Steele, and M. Wilson, *Vehicle Emissions Factor Development—NISE2 Data Verification*, PaeHolmes, Brisbane, Australia, 2010.
- [17] J.-Y. Favez, M. Weilenmann, and J. Stilli, "Cold start extra emissions as a function of engine stop time: evolution over the last 10 years," *Atmospheric Environment*, vol. 43, no. 5, pp. 996–1007, 2009.
- [18] ABS, *Motor Vehicle Census*, 9309.0, Australian Bureau of Statistics, Canberra, Australia, 2009.
- [19] J. B. Heywood, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York, NY, USA, 1988.
- [20] R. J. North, R. B. Noland, W. Y. Ochieng, and J. W. Polak, "Modelling of particulate matter mass emissions from a light-duty diesel vehicle," *Transportation Research Part D: Transport and Environment*, vol. 11, no. 5, pp. 344–357, 2006.
- [21] S. Chatterjee, M. Laudato, and L. A. Lynch, "Genetic algorithms and their statistical applications: an introduction," *Computational Statistics and Data Analysis*, vol. 22, no. 6, pp. 633–654, 1996.
- [22] I. Vladislavjevic, P. T. Martin, and A. Stevanovic, "Integration of mathematical and physical simulation to calibrate car-following behaviour of unimpaired and impaired drivers," *World Review of Intermodal Transportation Research*, vol. 1, no. 4, pp. 403–418, 2007.
- [23] J. J. Grefenstette, "Optimization of control parameters for genetic algorithms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 16, no. 1, pp. 122–128, 1986.
- [24] R. Q. Sardiñas, M. R. Santana, and E. A. Brindis, "Genetic algorithm-based multi-objective optimization of cutting parameters in turning processes," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 2, pp. 127–133, 2006.
- [25] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, Mass, USA, 1998.
- [26] E. Cantú-Paz, "A survey of parallel genetic algorithms," *Calculateurs Paralleles, Reseaux et Systems Repartis*, vol. 10, pp. 141–171, 1998.
- [27] G. Luque and E. Alba, *Parallel Genetic Algorithms Theory and Real World Applications*, Springer, New York, NY, USA, 2011.
- [28] MathWorks, *MATLAB Documentation*, MathWorks, Natick, Mass, USA, 2012.

