Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268 | Vol. 13 | No. 4 | pp. 781-786 | 2014

Original Research Paper

Model Selection for Emission Models Based on Emission Factors

Jing M. A., Shuping Huang and Kasjewruy Pateuro*
School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
*University of Eastern Finland, 80101 Joensuu, Finland
Corresponding author: Shuping Huang

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 31-9-2014
Accepted: 30-10-2014

Key Words:
Emission factor
On-road test
MOBILE model
MOVES model

ABSTRACT

Transportation industry accounts for the majority of air pollutant emission, which is one of the leading factors of climate change. Many vehicle emission models have been proposed focusing on emission factors to study the pollutant emission issue. However, little research has been done in China on emission models, though China is a major air pollutant-emitting country. This paper first introduces two emission models-MOBILE model and MOVES model, then proposes model selection methods including AIC (Akaike information criterion) and BIC (Bayesian information criterion) to compare and select the better model which could be applied in China. Experimental results show that the value of AIC and BIC for MOVES model is significantly lower than MOBILE model, which implies that MOVES model has better performance on real data fitness and prediction. Our experimental findings may be useful for future research on air pollutant emission modelling in China.

INTRODUCTION

Gaseous pollutants including CO, HC and NO_x are complained as the major causes of climate change, among which, transportation industry is the major pollutant source. According to the estimation of IEA (International Energy Agency), the gaseous pollutants from the road transportation accounts for 72% of the total emission from the transportation industry (Julio et al. 2014). It is critical for researchers and policy makers to set rules on vehicle fuel consumption and pollutant emission that could help to control the air pollution. The reasonable of rules is based on accurate measurements, though needing huge effort in on-road test. Emission models were established to decrease on-road test as well as guarantee the accuracy.

Large bodies of emission models have been developed to estimate the emission factor, which is the most important indicator to evaluate the emission behaviours (Wang et al. 2005). Most of the existing emission models such as MOBILE, EMFAC and COPERT are based on data collected from in-laboratory emission test, and many researchers argue that the predefined laboratory models cannot represent real-world emissions (Barth et al. 1996, Noland et al. 2004). The portable emission measurement systems (PEMS) designed for measuring on-road emissions on a second by second basis provide a better method for developing vehicle emission models, which are closer to the real-world emissions (Noland et al. 2004). The PEMS equipment is widely used to measure the on-road emissions of cars, trucks and mini-buses (Qiao et al. 2005, Guo et al. 2007).

Since 2003, a variety of emission models have been proposed to simulate the on-road emissions, such as MOBIEL, VT-MICRO, CMEM, COPERT, EMFAC, among which the MOBILE model obtained the relatively best performance on prediction accuracy for mini-bus, heavy-duty diesel vehicle and light-duty gasoline vehicles (Rakha et al. 2003, Bai et al. 2009, Marmur & Mamane 2003, Guo et al. 2007, Wang et al. 2005, Wan et al. 2005).

In 2009, US-EPA (United States Environment Protection Agency) has set up the MOVES model to replace the MOBILE model. The NO_x emission factors estimated by MOVES were higher than MOBILE prediction while the CO predictions from MOVES were much lower than MOBIEL (Kota et al. 2012). Another study established in a traffic tunnel in California also certificates the results. The NO_x predictions of MOVES were higher than those of MOBILE by approximately 10%. It also reported that CO concentrations predicted by MOVES were 30% lower than those predicted by MOBILE 6.2 (Fujita et al. 2012). Though the two models appear differently in CO and NO_x.

As an emerging economy giant, China contributes approximately 25% of global air pollutant emission. However, China has not established some useful vehicle emission models to support the accurate emission prediction as it has set a goal to reduce the transportation pollutant emission. Thus, it is critical and reasonable to compare and select an appropriate model from the existing ones that could be quickly applied in China.

Model selection methods like AIC and BIC are widely

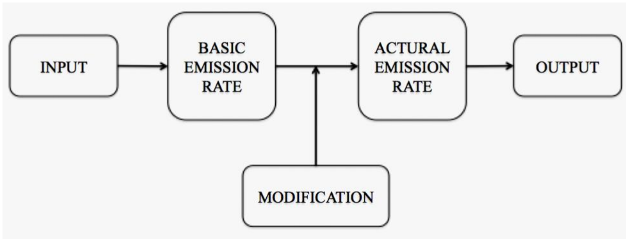

Fig. 1: Structure of MOBILE Model.

used in behavioural ecology (Burnham et al. 2011), civil engineering (Huang et al. 2007) and other areas. Accuracy of both fitting and predicting can be evaluated simultaneously using AIC and BIC methods. The model selection studies in the transportation industry are restricted to the category of road safety (Girma 2004), while the application in the area of roadway emission is still a new concept yet and needs further investigation.

This paper tries to propose a method to address the issue of selecting appropriate emission models and aims to improve the pollutant emission prediction accuracy. The proposed method will be based on the metrical data and the predicted data calculated by MOBILE model and MOVES model. Furthermore, the applicability and predictability of the selected model for China will be validated.

EMISSION MODEL DESCRIPTION

The emission factor means the amount of gaseous pollutants per kilometre, which reflects the emission behaviours. The formula of emission factor is as follows:

$$\text{Emission Factor (g/kg)} = \frac{\text{Total amount of pollutant (g)}}{\text{Mileage (km)}} \quad \text{...(1)}$$

Where, Total amount of pollutant means the total mass of CO, HC or NOx (g); Mileage means the total distance the vehicle has travelled (km).

MOBILE is an emission factor model that produces an estimation of the average emission rate in g/mile for a set of vehicles under a particular set of circumstances, which contains three steps: calculation of base emission rates, aggregation of base emission rates, and application of correction factors (Harrington et al. 1998). Distribution of vehicles, mileage, speed, I/M system, environment, pavement and other factors are all taken into account in the MOBILE model. The main structure of the MOBILE model is shown in Fig. 1.

MOVES model contains a set of model functions including an activity generator, a source bin distribution generator, an operating mode distribution generator, and an emission calculator, which is shown in Fig. 2 (Bai et al. 2009).

The operating mode distribution generator classifies the vehicle operating modes into different bins, which are associated with vehicle specific power (VSP) and speed (Jimenez 1999). VSP proposed by Jimenez-Palacios, reflects the power demand placed on a vehicle when the vehicle operates in various modes. The calculation of VSP for light duty gasoline vehicle (LDGV) is shown as follows:

$$VSP = v \times [1.1a+0.132] + 0.000302 \times v^3 \quad \text{...(2)}$$

Where, v is the vehicle speed (km/h); a means the acceleration (m/s^2).

By classifying the vehicle operating modes through VSP and speed, emission rates can be interpreted to the BIN modes, which are given in Table 1 (Ko 2011).

By inputting the information of vehicles, roadway,

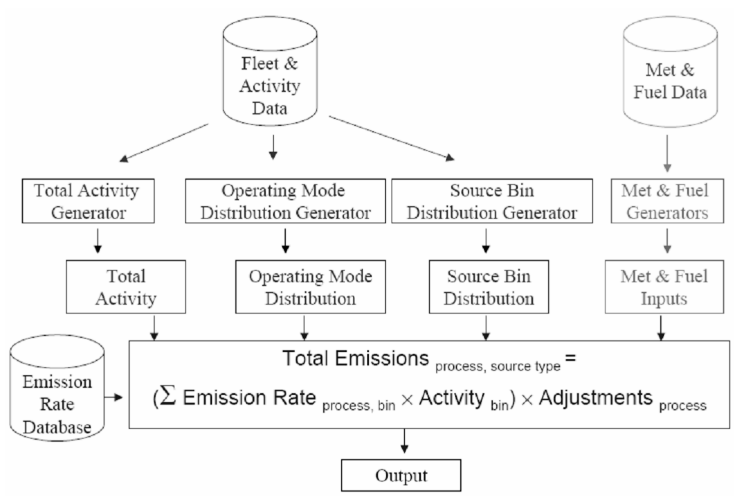

Fig. 2: Structure of MOVES model.

weather and environment, macroscopic emission factors can be put out based on the MOBILE model. The emission factors calculated by the MOVES model can be corresponded to the instantaneous speed and acceleration. These two models are based on different databases, so that we can have a comparison to decide which model has the better fitness and prediction accuracy to our observed data.

DATA ANALYSIS OF EMISSION FACTOR

The vehicle emission data are detected and captured by the PEMS equipment OBEAS-3000, with the sample of 1769 sets of CO, HC, NO_x pollutant emission amounts and mileage measured by GPS.

According to the test data, by removing the data of starting period, 1611 sets of emission data were used to calculate the emission factor. The emission factors of CO, HC and NO_x for three road sections were calculated and given in Table 2.

When predicting the emission factors by MOBILE model, several parameters should be taken into consideration, which will have great impacts on emission factors. In this case, the average speed, fuel Reye vapour pressure (RVP), temperature and other parameters were reset in the MOBILE model. The inputs of the parameters are listed in Table 3. The outputs of the MOBILE model for the three road sections are listed in Table 4.

In the MOVES model, instantaneous VSP can be calculated with formula 2. The instantaneous operating mode of the vehicle can be distinguished with instantaneous VSP and speed according to Table 1. Each Bin mode corresponds to a set of emission conditions of CO, HC and NOx. Emission factors of those three pollutants are computed by 1561 instantaneous test data. The results of the emission factors for MOVES model are listed in Table 5.

MODEL SELECTION METHOD

Suitable models can interpret the observed data, and predict the future simultaneously. When choosing from a set of candidate models, how could we know the selected model is better than other choices? Several methods have been set up to evaluate the fitness between model generated results and observations. But these methods, such as SSE, R2 and MSE, do have limitations that they could not guarantee the accuracy of the model-predicted data. So the model selection methods such as AIC and BIC are required in the field of emission models.

The R2 reflects the goodness of fit, which can be calculated as:

$$R^2 = 1 - \frac{SSE}{TSS} \qquad ...(3)$$

Table 1: Instantaneous VSP and speed to BIN emission inventory.

	Instantaneous speed (mph)		
Instantaneous VSP (kw/tonne)	0-25	25-50	>50
<0	Bin 11	Bin 21	
0-3	Bin 12	Bin 22	
3-6	Bin 13	Bin 23	
6-9	Bin 14	Bin 24	
9-12	Bin 15	Bin 25	
12 and greater	Bin 16	Bin 26	
12-18		Bin 27	Bin 37
18-24		Bin 28	Bin 38
24-30		Bin 29	Bin 39
30 and greater		Bin 30	Bin 40
6-12			Bin 35
< 6			Bin 33

Table 2: Emission factors of test data.

Section	CO Emission Factor(g/km)	HC Emission Factor(g/km)	NO_x Emission Factor(g/km)
1	13.710	0.323	3.919
2	12.576	0.259	2.952
3	16.568	0.323	3.226

Table 3: Parameters input in MOBILE model.

Vehicle Type		LDGV
Vehicle of the Year		5
Altitude		low
Temperature (F)	Minimum	34.7
	Maximum	48.2
Nominal Fuel RVP (psi)		12.7
Fuel Sulphur Content (ppm)		150
Average Speed of Section 1(km/h)		14.330
Average Speed of Section 2 (km/h)		21.000
Average Speed of Section 3 (km/h)		20.876

Table 4: Emission factors of MOBILE model.

Section	CO Emission Factor(g/km)	HC Emission Factor(g/km)	NO_x Emission Factor(g/km)
1	14.330	0.275	0.219
2	13.570	0.220	0.192
3	13.580	0.221	0.193

Where, SSE is the residual sum of square; TSS is the sum of squares of deviation.

The value of R^2 is closer to 1, the model has higher imitative effect.

The Akaike information criterion (AIC) is a measure of the relative quality of a statistical model, for a given set of data (Akaike 1974). The complexity of the model can also be evaluated at the same time, which is defined as:

$$AIC = -2\log(L(\hat{\theta}|y)) + 2p \qquad ...(4)$$

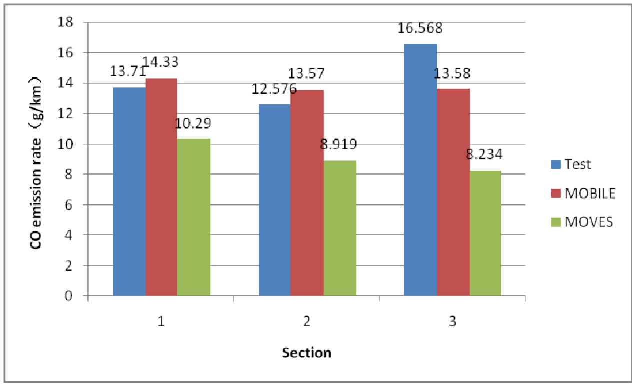

Fig. 3: Comparisons of CO emission factors within the two models.

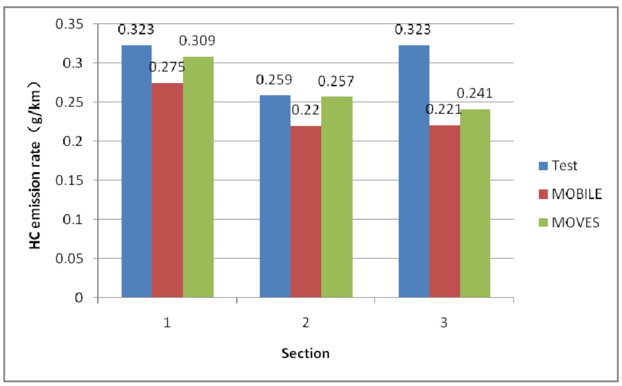

Fig. 4: Comparisons of HC emission factors within the two models.

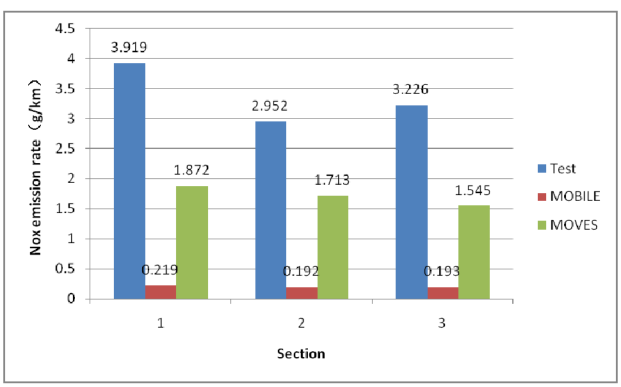

Fig. 5: Comparisons of NO_x emission factors within the two models.

Where, $L(\hat{\theta}|y)$ is the maximum likelihood function; $\hat{\theta}$ means the parameter vector estimated by y; p is the number of parameters in the model.

When facing the condition of the least square, residual sum of squares can be used in calculating the AIC value, so the formula can be adapted as:

$$AIC = n \times \ln(SSE) + 2k - n \times \ln(n) \quad ...(5)$$

Where, SSE is the residual sum of squares; k equals to the number of parameters in the model plus 1; n equals to the number of the observed data.

Highly fitted models are warmly welcomed by AIC, the model who has the minimum AIC value has the priority to take into consideration. However, the AIC tries to find a model that can well fit the observed data with the least parameters, so the situation of over-fitting should be avoid.

Bayesian Information Criterion (BIC) (Akaike 1974) is the integration of likelihood function based on the Bayesian theory, which can be defined as:

$$BIC = -2\log(L(\hat{\theta}|y)) + \log(n)p \quad ...(6)$$

Where, n stands for the number of the observed data.

Similarly to the AIC, the formula of BIC can also be adapted as:

$$BIC = n \times \ln(SSE) + k \times \ln(n) - n \times \ln(n) \quad ...(7)$$

To improve the maximum likelihood rate, some parameters are artificially added into the model, which causes over-fitting. The AIC and BIC try to reduce the impact of over-fitting, however, there is no rule that which method should be used in model selection. Useful information for model selection can be obtained from using AIC and BIC together (Kuha 2004). Generally, AIC based on the K-L distance model, while BIC based on the Bayesian theory, AIC can be derived as BIC when facing the Bayesian framework (Burnham et al. 2002, 2004). The AIC penalizes the number of parameters less strongly than BIC. BIC tends to choose the "true model" when facing a large number of data, while AIC tends to choose the complex one. AIC is asymptotically optimal in selecting the model with the least mean squared error when the exact "true model" is not in the candidate sets (Yang 2005). However, when facing the limited number of data, BIC tends to choose simple models due to the high coefficient for punishing the complexity of the models.

DATA ANALYSIS OF MODEL SELECTION

The values to evaluate goodness of fit are listed in Table 6 according to the formula 3. The values of R^2 for the two models are close to 1, which means the models that fit the test data are reasonable. So the model selection methods are used to determine that which model has higher accuracy on prediction.

The real-world data and the predictive data calculated by MOBILE and MOVES models of the CO, HC and NO_x emission factors are used to figure out the AIC and BIC values with formulae 5 and 7. The results of section 1 are given in Table 7

Table 5: Emission factors of MOVES model.

Section	CO Emission Factor(g/km)	HC Emission Factor(g/km)	NO_x Emission Factor(g/km)
1	10.290	0.309	1.872
2	8.919	0.257	1.713
3	8.234	0.241	1.545

Table 6: R^2 value for models.

	R^2 value
MOBILE	0.9289
MOVES	0.8069

Table 7: Model selection for Section 1.

	Real-World	MOBILE	MOVES
CO Emission Factor (g/km)	13.71	14.33	10.290
HC Emission Factor (g/km)	0.323	0.275	0.309
NO_x Emission Factor (g/km)	3.919	0.219	1.872
SSE		14.077	15.889
AIC		22.638	11.001
BIC		14.525	8.297

Table 8: Comparison of AIC value.

Section	MOBILE	MOVES
1	22.638	11.001
2	21.162	10.810
3	53.136	27.847

Table 9: Comparison of BIC value.

Section	MOBILE	MOVES
1	14.525	8.297
2	13.049	8.106
3	15.286	12.842

Table 10: Verification Results of MOVES Model.

	Real-World	MOVES	Rate of Deviation
CO Emission Factor (g/km)	8.681	8.499	2.09%
HC Emission Factor (g/km)	0.218	0.246	13.05%
NO_x Emission Factor (g/km)	2.088	1.622	22.35%

According to the computation of AIC and BIC from the Tables 8 and 9, the values for the MOVES model are correspondingly less than that for the MOBILE model, and it can be inferred that the MOVES model has superiority over the MOBILE model on predicting the emission factors in this case.

DISCUSSION

Different fitting results will emerge when taking the three emission factors separately. When considering the CO emission factor, the fitting values of the MOBILE model have better fitting precision to actual values, which have a margin of error limited to 20%, while the deviations for MOVES model are 24.95%, 29.08% and 50%. However, the fitting values of the MOBILE model are wildly inaccurate when fitting the HC and NO_x emission factors. The deviations between MOVES and test data are 4.39%, 0.65%, 25.44%, 52.23%, 41.96% and 52.11% compared with 14.86%, 15.06%, 31.58%, 94.41%, 93.50% and 94.02% which are calculated between MOBILE and test data. The high rates of deviation imply that it is unreasonable to compare the emission models by numerical comparison. Although the MOBILE model has higher imitative effect on CO emission factor and the MOVES model has higher imitative effect on HC and NO_x emission factor, all the emission factors should be taken into consideration synthetically when choosing the emission model. The results conducted by the model selection methods from Tables 7 and 8 reflect the comparison of model selection.

The results of the AIC and BIC show that the data predicted by the MOVES model are smaller than that by the MOBILE model. Such phenomenon indicates that the MOVES model has better performance on fitting and predicting the real-world data. Apart from selecting suitable models, other issues concerning the emission can be suggested from the results of AIC and BIC as follows:

1. MOBILE model output is closer to the real-world data in CO emission factors, but the overall effect of the model falls behind the MOVES model. Such situation can be imputed to the less attention to the HC emission factors in MOBILE model.
2. Speed and acceleration of the vehicles are the main concerns in MOVES model, while MOBILE model mainly focuses on the vehicle condition rather than driving condition. The results of AIC and BIC may indicate that the speed and acceleration are the key factors of the vehicle emission behavior.
3. The BIC result tends to choose the simpler model when the data are limited. In this case, however, the BIC result suggests MOVES model is much simple. Furthermore, the data predicted by MOVES model can be corresponded to the instantaneous condition of the vehicle, while MOBILE model is based on the average speed, which may be another reason why the MOVES model is more accurate than the MOBILE model.

To verify that the MOVES model is more accurate to the reality, the verification result is listed as in Table 10. The deviation between the MOVES and the real-world data indicates that the MOVES model can be a reflection of the reality, while the high rate of deviation on HC and NO_x emission factors may have some reasons.

As the MOVES model is based on the US emission database, the sulphur and nitrogen content in US benzene standard is 10 ppm, which is less than the 50 ppm in China, the NO_x emission factor is much lower in MOVES model than the case in Shanghai. On the other hand, hydrogenation equipment should be used to reduce the sulphur and nitrogen content in the gasoline, which can be the explanation of the high HC emission factor in the MOVES model.

Zero calib has been done in the experiment, though, the real-world data may have some deviation to reflect the reality inevitably. The model calculate the data considering factors concerning emission, which in some sense is still in ideal condition.

CONCLUSION

Based on the on-road test data of PEMS equipment, the AIC values for MOVES model (11.001, 10.810, 27.847) are less than the values for MOBILE model (22.638, 21.162, 53.136). Similarly, the BIC values for MOVES model (8.297, 8.106, 12.842) are also less than that for MOBILE model (14.525, 13.049, 15.286). Conclusions can be drawn that MOVES model has better reflection on fitness and prediction than the MOBILE model based on the experimental data of the light-duty gasoline vehicle.

Though the predictive values of CO emission factors calculated by the MOBILE model are more fitted to the real-world data, the other emission factors have large gap between the real-world data and the MOBILE model data. Moreover, the MOBILE model is a macroscopic model, the deviation between the predictive data and the real-world data can be occurred when predicting a specific vehicle.

This paper provided a useful method to compare and select a more accurate emission model in China to predict the real air pollutant emission. The results of model selection can be applied for further study on emission factors.

REFERENCES

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19: 716-723.

Bai, S., Eisinger, D. and Niemeier, D. 2009. MOVES vs. EMFAC: A comparison of greenhouse gas emissions using Los Angeles County. Transportation Research Board Annual Meeting, Washington DC.

Barth, M., An, F. and Norbeck, J. 1996. Modal emissions modeling: A physical approach. Transportation research record: Journal of the Transportation Research Board, 1520: 81-88.

Burnham, K.P. and Anderson, D.R. 2002. Model selection and multimodel inference: A practical information-theoretic approach, Springer, 336-340.

Burnham, K.P. and Anderson, D.R. 2004. Multimodel inference understanding AIC and BIC in model selection. Sociological Methods and Research, 33(2): 261-304.

Burnham, K.P., Anderson, D.R. and Huyvaert, K.P. 2011. AIC model selection and multimodel inference in behavioral ecology: Some background, observations and comparisons. Behavioral Ecology and Sociobiology, 65(1): 23-35.

Fujita, E.M., Campbell, D.E. and Zielinska, B. 2012. Comparison of the MOVES2010a, MOBILE6.2 and EMFAC2007 mobile source emission models with on-road traffic tunnel and remote sensing measurements. Journal of the Air & Waste Management Association, 62: 1134-1149.

Girma, B. 2004. Models relating traffic safety with road environment and traffic flows on arterial roads in Addis Ababa. Accident Analysis and Prevention, 36(5): 697-704.

Guo, S.X., Yu, L. and Song, G.H. 2007. A comparative analysis of real-world emission factors and MOBILE6 predictive value for a heavy-duty diesel vehicle. Journal of Safety and Environmental Engineering, 6(2): 14-16.

Harrington, W., McConnell, V.D. and Cannon M.A. 1998. Behavioral Analysis of EPA's MOBILE Emission Factor Model. US Department of Transportation, Federal Highway Administration.

Huang, S.P., Mahadevan, S. and Rebba, R. 2007. Collocation-based stochastic finite element analysis for random field problems. Probabilistic Engineering Mechanics, 22(2): 194-205.

Jimenez, J. 1999. Understanding and Quantifying Motor Vehicle Emissions with Vehicle Specific Power and TILDAS Remote Sensing. Massachusetts Institute of Technology, Cambridge.

Julio, L., Rafael, B. and Alberto, G. 2014. A methodology to compute emission projections from road transport (EmiTRANS). Technological Forecasting and Social Change, 81: 165-176.

Ko, M.H. 2011. Incorporating vehicle emission models into the highway design process. Texas A&M University.

Kota, S.H., Ying, Q. and Schade, G.W. 2012. MOVES vs. MOBILE6.2: differences in emission factors and regional air quality predictions. Transportation Research Board Annual Meeting, Washington DC.

Kuha, J. 2004. AIC and BIC comparisons of assumptions and performance. Sociological Methods & Research, 33(2): 188-229.

Marmur, A. and Mamane, Y. 2003. Comparison and evaluation of several mobile-source and line-source models in Israel. Transportation Research Part D: Transport and Environment, 8(4): 249-265.

Noland, R.B., Ochieng, W.Y. and Quddus, M.A. 2004. Vehicle emissions and performance monitoring system: Initial analysis of tailpipe emissions and vehicle performance. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, DC.

Qiao, F.X., Yu, L. and Michal, V.L. 2005. On-road vehicle emission and activity data collection and evaluation in Houston, Texas. Transportation Research Record. Journal of the Transportation Research Board, 1941: 60-71.

Rakha, H., Ahn, K. and Trani, A. 2003. Comparison of MOBILE5a, MOBILE6, VT-MICRO and CMEM models for estimating hot-stabilized light-duty gasoline vehicle emissions. Canadian Journal of Civil Engineering, 30(6): 1010-1021.

Wan, T., Yu, L. and Pei, W. W. 2005. A comparative analysis of real-world emission factors and MOBILE6 for Jieda. Journal of Environmental Protection in Transportation, 26(2) : 11-14.

Wang, W., Yu, L. and Liu, J. 2005. Comparison and analysis of mini-bus real-world emission factors and forecasting results of MOBILE6. Journal of Environmental Protection in Transportation, 26(3): 40-43.

Yang, Y. 2005. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. Biometrika, 92(4): 937-950.