Construction Materials and Structures
S.O. Ekolu et al. (Eds.)
IOS Press, 2014

doi:10.3233/978-1-61499-466-4-1445

Comparison of simple linear regression and multiple linear regression for estimating fuel use and emission rates for excavators

Heni FITRIANI[a,1] and Phil LEWIS[b]
[a]*Civil Engineering Department, University of Sriwijaya, Indonesia*
[b]*School of Civil and Environmental Engineering, Stillwater, Oklahoma, USA*

Abstract. Heavy-duty diesel (HDD) construction equipment consumes a significant amount of fuel and subsequently emits a substantial amount of pollutants into the environment. In most construction activities, HDD construction equipment is the primary source of emissions. The purpose of this paper is to demonstrate the comparative models for estimating fuel use and emission rates for HDD construction equipment specifically excavators. Second by second data were collected from portable emission measurement system (PEMS), containing fuel use and emission rates datasets along with engine performance data from three excavators. Emission pollutants include nitrogen oxides (NO_x), hydrocarbons (HC), carbon monoxide (CO), carbon dioxide (CO_2), and particulate matter (PM). For each excavator, predictive models were developed using simple linear regression (SLR) and multiple linear regression (MLR). Results yielded that the MLR accounted for the highest percentage of variability in the data compared to SLR based on the values of coefficient of determination (R^2) for each model. In order to exhibit the significant impact of which engine data that may affect the emission rates, the variable impact analysis was also conducted.

Keywords. Heavy-duty diesel (HDD) construction equipment, fuel use, emission, portable emission measurement system (PEMS)

Introduction

The construction sector plays an essential role in improving climate change due to the impact of greenhouse gas (GHG) emissions primarily caused by its major activities. Construction activities consume a significant amount of fuel and consequently emit a substantial amount of pollutants into the environment. According to the United States Environmental Protection Agency [2], there are approximately two million items of construction and mining equipment in the United States that spend about six billion gallons of diesel fuel annually. Furthermore, in most construction activities, heavy-duty diesel (HDD) construction equipment is the primary source of emissions. The EPA also estimates that in 2005, HDD construction vehicles produced U.S. national annual totals of 657,000 tons of NO_x, 1,100,000 tons of CO, 63,000 tons of PM_{10} and 94,000 tons of SO_2 [2].

[1]Corresponding author: henifitriani79@yahoo.com

Of these pollutants, NO_x and PM are the most prominent among HDD equipment [3]. Other pollutants found in diesel exhaust (DE) include hydrocarbons (HC) and carbon dioxide (CO_2). In order to quantify and characterize the HDD emissions problem, reliable prediction models are needed; however, most emission prediction tools are based on engine dynamometer data and not real-world data [1]. The objective of this paper is to demonstrate two different predictive modeling methodologies for estimating fuel use and emission rates for HDD construction equipment, specifically excavators, based on real-world data.

1. Previous work

As the need of conforming to emission standards has been largely increasing, numerous studies have been extensively piloted to quantify and characterize emissions and energy consumption of HDD construction equipment. Many studies have been completed using experimental designs such as dynamometer tests and real-world in-use measurements. Dynamometer tests are commonly used in quantifying emissions at steady-state conditions in the laboratory. Other studies conducted emission quantification by engaging Portable Emission Measurement Systems (PEMS), models, and simulations. The Environmental Protection Agency (EPA) and other government agency also develop other models such as the Nonroad model, the Offroad model, and the Urbemis model.

PEMS is generally used to gather fuel use and emissions field data of vehicles based upon real-world measurement. In-use emissions quantification enables data collection by capturing the actual duty cycle on second by second basis measurement. Commercial PEMS are obtainable for any kinds of applications as well as for different types of fuel use. Some of the most prominent real-world emissions measurements from HDD construction equipment were completed by the researchers at North Carolina State University [1, 5, 6, 7, 8]. Other researchers from West Virginia University and the University of California – Riverside also directed their studies on the use of on-board emission measurement for particular construction equipment.

2. Methodology

This paper presents two different predictive modeling methodologies for estimating fuel use and emissions rates based on the real-world dataset from the research team at North Carolina State University. Simple linear regression (SLR) and multiple linear regression (MLR) models were developed and compared for three excavators. Engine performance data from the excavators, including manifold absolute pressure (MAP), revolutions per minute (RPM), and intake air temperature (IAT), were used to develop prediction models for fuel use and emission rates of NO_x, HC, CO, CO_2, and PM.

Table 1 displays the summary of engine attribute data for each excavator, including engine size (HP), displacement, model year, and EPA engine tier. The rated engine horsepower (HP) ranged from 93 HP to 254 HP and the model year ranged from 1998 to 2003; thus, all three of the excavators were either EPA engine tier 1 or 2.

Table 1. Summary of engine attribute data

Equipment	Horsepower (HP)	Displacement (Liters)	Model Year	Engine Tier
Excavator 1	254	8.3	2001	1
Excavator 2	138	6.4	2003	2
Excavator 3	93	3.9	1998	1

3. Simple linear regression

Simple linear regression models were developed to determine the relationship between a single response variable and a single predictor variable. Since it has been shown by others that MAP is highly correlated to fuel use and emission rates [4, 5, 7], simple linear regression models were formulated based on the relationship between MAP as a predictor variable and fuel use as a response variable, as well as MAP and mass per time (grams per second) emission rates of NO_x, HC, CO, CO_2, and PM. These SLR models take the form of:

$$Y_{1\text{-}6} = ax + c \quad (1)$$

where:

$Y_{1\text{-}6}$ = Fuel use or emission rate of NO_x, HC, CO, CO_2, or PM (g/s)
a = slope of the regression line
x = MAP (kilopascal)
c = y-intercept of regression line

4. Multiple linear regression

Multiple linear regression was used to model the relationship between three predictor variables based on engine performance data (MAP, RPM, IAT) and one response variable (either fuel use or emission rate of NO_x, HC, CO, CO_2, and PM).

The MLR equations for fuel use and emission rates for each pollutant take the form of:

$$Y_{1\text{-}6} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \quad (2)$$

where:

$Y_{1\text{-}6}$ = Fuel use or emission rate of NO_x, HC, CO, CO_2, or PM (g/s)
X_1 = MAP (kilopascal)
X_2 = Revolutions Per Minute (RPM)
X_3 = Intake Air Temperature (Celsius degrees)
β_0, β_1, β_2, β_3 = Coefficients of linear relationship

5. Results

This section presents the results for two predictive modeling methodologies – SLR and MLR- for excavators, as well as variable correlations and model comparisons for those models. The validation results for the models for all three excavators are presented in Table 5.

Table 2 shows the summary of the Pearson correlation coefficients for all three excavators, indicating the relationship between engine data, fuel use, and emission rates. MAP has a strong positive relationship with fuel use and emission rates of NO_x, CO_2, and PM, but a moderate positive relationship with HC and CO. RPM has the second strongest relationship with fuel use and emission rates. Meanwhile, IAT has the weakest relationship with fuel use and emission rates as indicated by the lower (and sometimes negative) values of correlation to the specified response variables.

Table 2. Summary of Pearson correlations coefficients

Equipment	Engine Data	Fuel Use	NO_x	HC	CO	CO_2	PM
EX 1	MAP	0.99	0.97	0.59	0.74	0.99	0.94
	RPM	0.80	0.74	0.63	0.85	0.79	0.74
	IAT	0.56	0.59	0.07	0.37	0.57	0.51
EX 2	MAP	0.98	0.92	0.62	0.47	0.98	0.94
	RPM	0.85	0.85	0.62	0.57	0.85	0.69
	IAT	0.55	0.56	0.33	0.30	0.55	0.44
EX 3	MAP	0.96	0.94	0.44	0.14	0.96	0.57
	RPM	0.84	0.79	0.42	0.23	0.84	0.47
	IAT	0.32	0.40	0.36	-0.12	0.32	0.44

5.1 Simple linear regression models

Based on their high correlation values, SLR models were developed using MAP as a predictor variable to predict fuel use and emission rates of each pollutant. Table 3 presents the results of the SLR models for all three excavators. These models are based on more than 19,000 observations of second-by-second, real-world fuel use and emissions data for excavator 2 and 3, and around 7,000 observations for excavator 1. Based on the coefficient of determination (R^2), these models accounted for a high percentage of the variability in the data for fuel use, NO_x, CO_2 and PM. HC and CO had the lowest R^2 value, indicating much variability in the data, and therefore were more difficult to predict.

Table 3. Summary of SLR models for all excavators

Equipment	Response	Equations	R^2
Excavator 1	Fuel Use	$Y_1 = 9.9429\ X_1 + 0.4704$	0.982
	NO_x	$Y_2 = 0.3545\ X_1 + 0.0242$	0.948
	HC	$Y_3 = 0.0054\ X_1 + 0.0024$	0.351
	CO	$Y_4 = 0.0175\ X_1 + 0.0066$	0.543
	CO_2	$Y_5 = 31.431\ X_1 + 1.4720$	0.982
	PM	$Y_6 = 3.8619\ X_1 + 0.1076$	0.881
Excavator 2	Fuel Use	$Y_1 = 6.4485X_1 + 0.5302$	0.963
	NO_x	$Y_2 = 0.1202\ X_1 + 0.0209$	0.850
	HC	$Y_3 = 0.0083\ X_1 + 0.0031$	0.390
	CO	$Y_4 = 0.0239X_1 + 0.0142$	0.219
	CO_2	$Y_5 = 20.358X_1 + 1.6475$	0.963
	PM	$Y_6 = 1.8463X_1 + 0.0354$	0.888
Excavator 3	Fuel Use	$Y_1 = 3.9492\ X_1 + 0.1231$	0.930
	NO_x	$Y_2 = 0.1231\ X_1 + 0.0098$	0.876
	HC	$Y_3 = 0.0084X_1 + 0.0021$	0.194
	CO	$Y_4 = 0.0051X_1 + 0.0055$	0.018
	CO_2	$Y_5 = 12.468\ X_1 + 0.3748$	0.929
	PM	$Y_6 = 1.0842\ X_1 - 0.0099$	0.333

X_1 = MAP

5.2 Multiple linear regression models

Based on the correlation matrix in Table 2, MAP and RPM are highly correlated to fuel use and emissions rate for most of pollutants. Even though IAT has a lower correlation to fuel use and emissions rate, IAT was still used as an input variable for the MLR models because it may still have some predictive power.

Table 4 summarizes the models for fuel use and emissions rates for all three excavators. Overall, the MLR models yielded higher R^2 values than the SLR models for their respective response variables. The MLR R^2 values for fuel use and emission rates for NO_x, HC, CO_2 and PM indicate that the models perform well. The model for CO, however, accounted for less than 50% of the variability in the data; thus, the MLR models also indicate that emission rates of CO are more difficult to predict compared to fuel use and the other pollutants.

Table 4. Summary of MLR models for all excavators

Equipment	Response	Equations	R^2
Excavator 1	Fuel Use	$Y_1 = -5.748 + 0.0728 X_1 + 0.000301X_2 - 0.0296X_3$	0.985
	NO_x	$Y_2 = -0.2093 + 0.00247X_1 - 0.00002 X_2 + 0.000176X_3$	0.954
	HC	$Y_3 = 0.0056 + 0.000034 X_1 + 2.64E\text{-}6 X_2 - 0.00021X_3$	0.582
	CO	$Y_4 = -0.00003 + 0.000041 X_1 + 0.000011X_2 - 0.00018X_3$	0.801
	CO_2	$Y_5 = -18.21 + 0.230X_1 + 0.00093 X_2 - 0.093X_3$	0.985
	PM	$Y_6 = -2.21 + 0.0293X_1 - 0.0136X_3$	0.880
Excavator 2	Fuel Use	$Y_1 = -5.07 + 0.0524 X_1 + 0.00069 X_2 - 0.0085 X_3$	0.972
	NO_x	$Y_2 = - 0.089 + 0.00082 X_1 + 0.000024 X_2 + 0.000134X_3$	0.884
	HC	$Y_3 = -0.0024 + 0.000048X_1 + 3.14E\text{-}6X_2 - 0.00008X_3$	0.402
	CO	$Y_4 = -0.0004 + 0.000013 X_1 + 0.000019 X_2 - 0.00024 X_3$	0.340
	CO_2	$Y_5 = -16.05 + 0.166X_1 + 0.00213 X_2 - 0.0262 X_3$	0.972
	PM	$Y_6 = -1.53 + 0.021X_1 - 0.00026X_2 - 0.0064X_3$	0.913
Excavator 3	Fuel Use	$Y_1 = -2.343 + 0.0295X_1 + 0.00006X_2 - 0.007X_3$	0.935
	NO_x	$Y_2 = -0.079 + 0.00096X_1 - 5.33E\text{-}6X_2 + 0.000096X_3$	0.880
	HC	$Y_3 = -0.0071 + 0.000034X_1 + 1.57E\text{-}6X_2 + 0.000094X_3$	0.250
	CO	$Y_4 = 0.0094 - 0.00005X_1 + 9.92E\text{-}6X_2 - 0.00018X_3$	0.096
	CO_2	$Y_5 = -7.409 + 0.0932X_1 + 0.00017X_2 - 0.022X_3$	0.934
	PM	$Y_6 = -1.142 + 0.0081X_1 - 0.00013X_2 + 0.0104X_3$	0.390

Table 5. Comparison of validation results for SLR and MLR

Equipment	Pollutants	SLR			MLR		
		m	b	R^2	m	b	R^2
Excavator 1	Fuel Use	0.982	0.045	0.982	0.983	0.044	0.985
	NO_x	0.948	0.005	0.948	0.944	0.004	0.951
	HC	0.352	0.002	0.351	0.573	0.002	0.575
	CO	0.542	0.005	0.543	0.773	0.003	0.759
	CO_2	0.982	0.143	0.982	0.981	0.107	0.985
	PM	0.881	0.107	0.881	0.873	0.099	0.886
Excavator 2	Fuel Use	0.963	0.074	0.963	0.974	0.063	0.971
	NO_x	0.850	0.007	0.850	0.887	0.006	0.879
	HC	0.392	0.003	0.390	0.441	0.003	0.434
	CO	0.220	0.015	0.219	0.322	0.013	0.327
	CO_2	0.963	0.234	0.963	0.974	0.206	0.971
	PM	0.889	0.052	0.888	0.917	0.053	0.909
Excavator 3	Fuel Use	0.930	0.120	0.930	0.936	0.113	0.935
	NO_x	0.875	0.007	0.876	0.878	0.007	0.878
	HC	0.193	0.004	0.194	0.243	0.004	0.239
	CO	0.018	0.008	0.018	0.105	0.007	0.100
	CO_2	0.930	0.381	0.929	0.933	0.354	0.934
	PM	0.333	0.284	0.333	0.384	0.252	0.387

5.3 Model comparison

Model validations for the three excavators were developed in order to compare and evaluate the performance of SLR and MLR methodologies. The models were validated by plotting the predicted versus actual results for each model and fitting a trend line to the data. For each trend line, the values of accuracy (m), bias (b), and precision (R^2) were determined. As shown in Table 5, MLR produces higher R^2 values compared to SLR for fuel use and all emissions rates. SLR has the lowest R^2 value for fuel use and emissions rates. Overall, MLR outperformed SLR with respect to precision, accuracy, and bias. In most cases, the MLR approach produced highly precise models for NO_x, CO_2, and PM; while the models for HC and CO were likely to be moderately precise with R^2 values ranging from 0.50 – 0.87.

5.4 Variable impact analysis

Using the MLR models, a variable impact analysis was conducted to determine the percentage of contribution of the input variables (MAP, RPM, and IAT) to the prediction of fuel use and emission rates of each pollutant. Table 6 presents the summary of the variable impact analysis for all three excavators. MAP is the most significant variable for fuel use, CO, CO_2, and PM which are 63.96%, 53.71%, 70.48% and 59.50%, respectively. RPM, however, has the most contribution for NO_x. IAT had the highest impact for HC.

Table 6. Variable impact analysis for average excavators

Engine Data	Fuel Use	NO_x	HC	CO	CO_2	PM
MAP	63.96%	40.13%	24.39%	53.72%	70.48%	59.50%
RPM	27.21%	43.20%	25.91%	35.02%	23.56%	19.67%
IAT	8.83%	16.57%	49.70%	11.26%	5.96%	20.83%

6. Conclusions and recommendations

The purpose of this paper was to demonstrate two different predictive modeling methodologies for estimating fuel use and emission rates of pollutants using real-world data. Based on the summary of Pearson correlation coefficients, MAP had a high positive correlation to fuel use and emission rates of NO_x, CO_2, and PM, but had a moderate positive relationship with HC and CO. Although not as highly correlated, RPM had a strong positive relationship with fuel use and emissions. IAT was shown to have the least impact of the three engine performance variables on predicting fuel use and emission rates. It is recommended that other engine performance data, such as engine load or throttle position, be considered for future studies. For two modeling approaches, CO proved to be the most difficult pollutant emission rate to predict, as evidenced by its low R^2 values. Typically, there is high variability in CO data which confounds the prediction effort, as well as the fact that CO did not have a strong correlation with any of the engine data predictor variables. It is recommended that strong relationships between CO and other variables be considered. For example, it there exists a strong relationship between CO and fuel use (which is accurately and precisely predicted by each of the three modeling approaches), then fuel use may be used as a predictor variable for CO.

With regard to variable impact analysis, it can be concluded that MAP has the highest percentage of contribution in the prediction of fuel use and emission rates, accounting for approximately 60% of total impact, although for HC and CO it had the second highest impact. For these two pollutants, RPM had the highest impact but it was second for fuel use, CO, CO_2, and PM. Although IAT had the lowest ranking impact among the three engine performance variables, it still may have some predictive power, especially for CO. For strictly prediction purposes, it is recommended that all three engine performance variables be used to estimate fuel use and emission rates. Based on the model comparisons, MLR models generally performed better with respect to precision, accuracy, and bias. In most cases, the MLR approach produced highly precise models for NO_x, CO_2, and PM; while the models for HC and CO were moderately precise. These models are still useful. The simplicity of the one variable SLR models may be appealing to some users, such as fleet managers, that want to estimate the fuel use and emissions footprints of their equipment. Other users, such as engine manufacturers, may like the MLR approach because they would be able to reasonably estimate each of the engine performance variables.

Overall, the results of this study help to quantify and characterize the air pollution problem from HDD equipment used in construction. Although only excavators were addressed in this paper, the methodologies presented may certainly be used to develop fuel use and emissions models for other types of equipment. In order to further characterize this emissions problem, it is recommended that other types of equipment, such as backhoes, bulldozers, motor graders, track loaders, wheel loaders, and off road trucks, be targeted for future modeling efforts.

Acknowledgement

The authors acknowledge the use of the real-world non-road equipment and emission database that was developed at North Carolina State University by Dr. H. Christopher Frey and Dr. William Rasdorf.

References

[1] Abolhasani, S., Frey, H. C., Kim, K., Rasdorf, W., Lewis, P., & Pang, S.-H, Real-World In-Use Activity, Fuel Use, and Emissions for Nonroad Construction Vehicles: A Case Study for Excavators. *Journal of the Air & Waste Management Association*, Vol 58(8), pp. 1033-1046, 2008.

[2] EPA, *Users guide for the final NONROAD2005 model*, EPA-420-R-05-013, Ann Arbor, MI, 2005.

[3] *Environmental Protection Agency Clean Air Act Advisory* Committee (EPA CAAAC), Recommendations for reducing emissions from the Legacy Diesel Fleet. U.S. Environmental Protection Agency, Washington D.C, 2006.

[4] Fitriani, H, Development of Predictive Modeling Tools for Estimating Fuel Use and Emissions Rates for Heavy-Duty Diesel Construction Equipment, *Doctor of Philosophy, Oklahoma State University*, Stillwater, OK, 2014.

[5] Frey, H. C., Rasdorf, W., Kim, K., Pang, S.-H., Lewis, P., & Abolhassani, S, Real-World Duty Cycles and Utilization for Construction Equipment in North Carolina, Dept. of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC, 2008.

[6] Kim, K., Operational Evaluation of In-UseEmissions and Fuel Consumption of B20 Biodiesel versus Petroleum Diesel-Fueled Onroad Heavy-Duty Diesel Dump Trucks and Nonroad Construction Vehicles. *Doctor of Philosophy, North Carolina State University*, Raleigh, NC, 2007.

[7] Lewis, P. Estimating Fuel Use and Emission Rates of Nonroad Diesel Construction Equipment Performing Representative Duty Cycles. *Doctor of Philosophy, North Carolina State University*, Raleigh, NC, 2009.

[8] Rasdorf, W., Frey, C., Lewis, P., Kim, K., Pang, S.-H., & Abolhassani, Field Procedures for Real-World Measurements of Emissions from Diesel Construction Vehicles, *Journal of Infrastructure Systems*, Vol 16(3), pp. 216-225, 2010.