2009

Comparative Models of Hydrocarbon Emissions for a Diesel Engine Operating at Constant Loads and Speeds

Dennis G. Watson
Southern Illinois University Carbondale, dwatson@siu.edu

David R. Bostic
Southern Illinois University Carbondale

Tony V. Harrison
JPT Integrated Solutions, Inc

Recommended Citation

Watson, Dennis G., Bostic, David R. and Harrison, Tony V. "Comparative Models of Hydrocarbon Emissions for a Diesel Engine Operating at Constant Loads and Speeds." *Transactions of the ASABE* 52, No. 4 (Jan 2009): 1079-1087. doi:10.13031/2013.27778.

Comparative Models of Hydrocarbon Emissions for a Diesel Engine Operating at Constant Loads and Speeds

D. G. Watson, D. R. Bostic, T. V. Harrison

Abstract. *Linear multiple regression (LMR) and nonlinear polynomial network (NPN) models were developed from data collected from ISO 8178-4 (1996) test cycle B-type tests (ISO) and an expanded set of tests (EXP) to predict hydrocarbon (HC) emissions from a diesel engine. LMR using the ISO training data (R^2 = 0.94) resulted in overfitting of the model as applied to the evaluation data (R^2 = 0.49). LMR based on the expanded data (R^2 = 0.68) was a better LMR model when applied to the evaluation data (R^2 = 0.64). NPN using the expanded training data (R^2 = 0.99) resulted in the best model when applied to the evaluation data (R^2 = 0.98) and is preferred for predicting HC when the larger set of test mode data are available. NPN using the ISO training data (R^2 = 0.99) resulted in a satisfactory fit for the evaluation data (R^2 = 0.91), although with a higher average absolute error (0.52 vs. 0.42 g/kWh) than NPN using the EXP training data. This model was also considered suitable for predicting HC. Results of this initial study suggest that data could be collected during ISO 8178-4 emission tests and modeled with NPN to predict HC emissions for a diesel engine operating at various constant speeds and loads.*

Keywords. *Diesel engines, Emissions, Hydrocarbon, Modeling, Polynomial network, Regression.*

Even with increased emissions regulations, the availability and flexibility of fossil-fueled, internal combustion engines results in new applications on a regular basis. This is especially true in agriculture, where remote power requirements are common. In the past, emissions regulations were only for on-road engines, but by 2001, emissions of non-road power units were being regulated (CCR, 1999). Engine manufacturers are required to have engine emissions certified per acceptable standards, such as ISO 8178-4 (ISO, 1996).

The ISO 8178-4 standard for emissions measurement (ISO, 1996) includes universal test cycle B, which includes the engine speed and load combinations of the other test cycles. Test cycle B specifies 11 test modes of engine load and speed combinations for emissions measurement, specifically 10%, 25%, 50%, 75%, and 100% torque at rated speed and an intermediate speed and no load at low idle. Overall emission values are determined by averaging (other test cycles require weighting) the emissions of the test modes. The goal of the ISO standard was to minimize test modes while ensuring that test cycles were representative of actual engine operation (ISO, 1996). Hansson et al. (2001) found hydrocarbon (HC) emissions of representative average tractor use to be as much as 50% higher than values determined according to ISO 8178-4. They concluded it was not possible to design one set of emissions factors that produced representative results for all types of tractors and work operations (Hansson et al., 2001). Although the results of the ISO 8178-4 emissions tests meet regulatory requirements, the certified emission values are limited for representing actual emissions of an operating engine.

Besides using the ISO 8178-4 test modes to compute a set of overall emissions values, additional data from tests may be useful for developing a mathematical model for predicting emissions at various load and speed combinations. Predicting emissions of a stationary or portable engine used to power a relatively constant load, such as an irrigation pump, would be an initial test of this hypothesis. In some cases, engines have been sized to operate at rated continuous power, but in other cases, engines operate at less than rated power and may be considerably overpowered for an application. Data from ISO 8178-4 tests may be insufficient to model emissions of engines operating at speeds different than the two tested speeds or at loads between the tested loads. Emission data from additional loads and speeds may produce a better model of the range of potential operating conditions.

Models have been developed for diesel-powered, heavy-duty, on-road vehicles. Ramamurthy et al. (1998) fit a polynomial curve to emissions based on axle power of a heavy-duty diesel vehicle. Krijnsen et al. (2000) successfully modeled NO_x emissions from a diesel engine using an artificial neural net, a split and fit algorithm, and a nonlinear polynomial model. Yanowitz et al. (2002) used test data from a heavy-duty transient test to predict diesel emissions based on engine power and found a good linear correlation between rates of horsepower increase and particulate matter emissions.

Submitted for review in October 2008 as manuscript number PM 7786; approved for publication by the Power & Machinery Division of ASABE in July 2009.

The authors are **Dennis G. Watson, ASABE Member,** Associate Professor, and **David R. Bostic,** former Graduate Assistant, Department of Plant, Soil, and Agricultural Systems, Southern Illinois University, Carbondale, Illinois; and **Tony V. Harrison,** Chief Operating Officer, JPT Integrated Solutions, Inc., Gainesville, Florida. **Corresponding author:** Dennis G. Watson, 1205 Lincoln Dr., Rm 176, MC 4415, Southern Illinois University, Carbondale, IL 62901; phone: 618-453-6979; e-mail: dwatson@siu.edu.

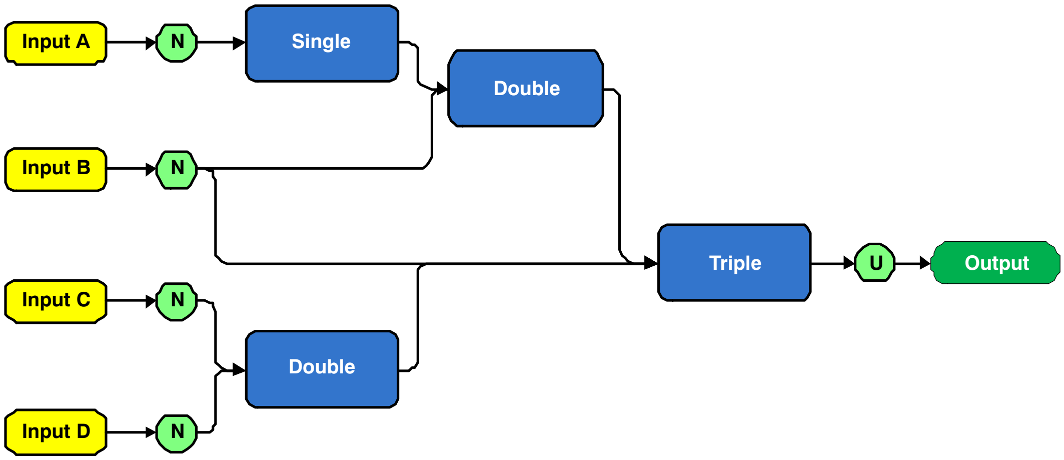

Figure 1. Example of polynomial network with N symbolizing normalizing function, U symbolizing unitizing function, and single, double, and triple indicating the number of inputs in a network node.

Objective

A study was conducted to compare models derived from two data sets and two modeling methods for predicting HC emissions of a diesel engine operating at constant loads and speeds. For modeling purposes, the target range of engine operation was 1500 to 2500 rpm (rated speed) with 10% increments of torque starting at 40% up to 100%. The data sets consisted of data obtained from tests similar to the ISO 8178-4 B test cycle and an expanded set of tests with additional loads and speeds. The data included engine operating conditions from the engine's controller area network (CAN) and torque, emissions, and ambient condition sensors. The modeling methods used were linear multiple regression (LMR) and nonlinear polynomial network (NPN).

The results of the study provide comparative data on the relative suitability of the ISO 8178-4 test cycle B-type data and expanded data for predicting emissions of a diesel engine running at a constant load and speed, as modeled with LMR and NPN.

Nonlinear Polynomial Network Modeling

NPN modeling is a non-parametric, self-organization approach in which underlying relationships of variables are automatically discovered by the NPN algorithm. In this context, a network is a function represented by the composition of many functions (Barron and Barron, 1988) (see fig. 1 for example network). NPN is closely related to the group method of data handling (GMDH) algorithm developed in Kiev, Ukraine, and published by Ivakhnenko (1968). Barron et al. (1984) described early polynomial network software development in the U.S. as based on the GMDH described by Ivakhnenko (1971). According to Farlow (1984), Ivakhnenko's work was prompted by the requirement of many mathematical models to know details about a system that are generally unknown. A method was needed that relied on objective methods rather than biases of the researchers (Farlow, 1984).

NPN software programs based on GMDH-type algorithms have been described using various terms, including polynomial network (Barron et al., 1984; Drake et al., 1994; Griffin et al., 1994; Kleinsteuber and Sepehri, 1996), abductory induction (Montgomery, 1989), abductive reasoning network (Montgomery and Drake, 1991), and abductive polynomial network (Drake and Kim, 1997). More recently, polynomial network software programs have been classed as data mining tools (Agarwal, 1999; Kim, 2002; King et al., 1998; and Pyo et al., 2002). Polynomial networks have been used for a wide range of modeling applications, including defense (Montgomery et al., 1990), financial (Stepanov, 1974; Kim, 2002), medical (Abdel-Aal and Mangoud, 1997; Griffin et al., 1994), process control (Silis and Rozenblit, 1976), and agriculture (Duffy and Franklin, 1975; Ivakhnenko et al., 1977; Lebow et al., 1984; Pachepsky and Rawls, 1999; Reddy and Pachepsky, 2000).

Equipment and Procedures

A 2003 John Deere 4045T, 4.5 L, inline four-cylinder, EPA Tier 2, turbocharged diesel engine was used for this study. Peak torque was 394 N·m at 1400 rpm, and rated power was 86 kW (77 kW for continuous operation) at 2500 rpm. This engine was equipped with an SAE J1939 CAN (SAE, 2002). The equipment used for data collection and storage included a CAN protocol adapter, programmable automation controller (PAC), ambient condition sensors, and computer with LabVIEW (National Instruments, 2003). This equipment was previously described by Hogan et al. (2007) and Watson et al. (2008).

A P-400B hydraulic dynamometer (M&W Gear, Gibson City, Ill.) was used to provide an engine load. A TMS 9000 torque measurement system (Honeywell International, Morristown, N.J.) was used to measure torque. The system consisted of a rotating torque sensor, mounted between the engine flywheel and the dynamometer driveshaft, and a signal processing module. The two components had a typical accuracy of 0.50% and 0.002%, respectively. The output of the signal processing module was connected to the PAC.

Exhaust emissions were collected and analyzed by an FGA4000XD gas analyzer (GA) (Infrared Industries, Hayward, Cal.). The GA used non-dispersed infrared light to measure HC. The GA also measured exhaust temperature, pressure, and air to fuel ratio. Exhaust gases were collected by connecting a tube to the exhaust system upstream from the muffler. GA output was connected to the PAC. The GA measured HC in parts per million (ppm), and units of g/kWh were calculated as specified by Infrared Industries.

Table 1. Variables measured and calculated during engine emissions tests and used for modeling.

Variable	Source
Engine speed (rpm)	CAN
Percent torque	CAN
Percent load	CAN
Percent friction[a]	CAN
Fuel flow rate (L/h)	CAN
Engine fuel temperature (°C)	CAN
Coolant temperature (°C)	CAN
Intake manifold temperature (°C)	CAN
Torque (N·m)	TMS 9000
Flywheel power (kW)	Calculated
Ambient temperature (°C)	ICTD sensor[b]
Relative humidity (%)	HIH3610[c]
Atmospheric pressure (mbar)	WS16BP[d]
Exhaust temperature (°C)	GA
Exhaust pressure (kPa)	GA
Air to fuel ratio	GA
HC (ppm)	GA
HC (g/kWh)	Calculated

[a] Percent friction is the percentage of engine torque required by the engine and includes frictional and thermodynamic losses of the engine and the losses of fuel, oil, cooling pumps, and accessories (SAE, 2002, suspect parameter number (spn) 514).
[b] Source: Opto 22 (Temecula, Cal.).
[c] Source: Honeywell International (Morristown, N.J.).
[d] Source: NovaLynx (Grass Valley, Cal.).

SAE standard J1939-71 defined variables potentially available on the CAN (SAE, 2002). Eight variables were available that were related to engine performance. These were included in the 18 variables measured or calculated for the emissions tests (table 1).

Two data sets, each from different experiments, were compared to determine which produced the best prediction model. The first data set was called ISO and was based on ISO 8178-4 test cycle B (ISO, 1996). The rated speed was 2500 rpm, and an intermediate speed of 1500 rpm was selected. No-load tests were substituted for the 10% torque tests, since testing equipment would not support the low 10% of maximum torque for the rated or intermediate speeds. The 11 torque and engine speed combinations were replicated four times for a total of 44 tests. Although ISO 8178-4 does not require replications, they were added to provide additional data for modeling and model evaluation.

The second data set expanded on the ISO test modes and was called EXP (expanded). It was designed to provide more data points by testing loads at 10% intervals between 40% and 100% of maximum torque for each engine speed and using additional speed settings. Emissions data were collected while the diesel engine was operated at 0%, 25%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% of maximum torque for each engine speed of 1500, 1750, 2000, 2250, and 2500 (rated power) rpm. For each test mode combination of percent of maximum torque and engine speed, the actual torque level used for the test was determined by measuring the maximum observed torque at the engine speed and multiplying by the percent torque. The 45 torque and engine speed combinations were replicated four times for a total of 180 tests. Other than the torque and speed combinations, there were no other differences in equipment and procedures in collecting data for the ISO and EXP data sets.

ISO 8178-4 (ISO, 1996) specifies that the test time for each mode be no less than ten minutes. This includes seven minutes for engine adjustment and stabilization, and a minimum of three minutes for data collection. For this study, a minimum of five minutes was used to adjust the engine speed and load, and allow the engine to stabilize. Once the engine stabilized at the desired settings, two additional minutes passed before data collection started. Then the LabVIEW program automatically recorded CAN, torque, ambient condition, and GA data at 30-second intervals during an eight minute test run. After each test mode ended, the data points were averaged. The summary data for each test mode was stored in a comma separated values file.

The data were combined into three files for modeling. All data from replications 1, 2, and 3 of the ISO tests were combined into one file for the ISO training data. Likewise, all data from replications 1, 2, and 3 of the EXP tests were combined into one file for the EXP training data. Data from replication 4 of both the ISO and EXP tests were combined into one file for the evaluation data set. The first 16 variables from table 1 were used as independent variables (inputs), and HC in g/kWh was the dependent (output) variable. Although the resulting training sample sizes (n = 33 for ISO and n = 135 for EXP) were relatively small for LMR and were expected to result in overfitting, the LMR models were included as a comparison to the NPN models, which have been found to be more efficient than LMR with small sample sizes (Stepanov, 1974).

SAS 9.1 (SAS, 2007) was used to compute correlation and regression coefficients for the 16 inputs to HC. Two LMR models were developed, one each for the ISO and EXP training data. The form of the regression equation was:

$$Y' = a + b_1 X_1 + b_2 X_2 + ... + b_k X_k \quad (1)$$

where Y' is the predicted HC (g/kWh), a is the intercept constant, b_k is the regression coefficient for the kth predictor variable, and X_k is the kth predictor variable. Both of the LMR models were evaluated by using the resulting equations to predict HC with inputs from the evaluation data set.

Two NPN models were developed, one each for the ISO training and EXP training data. ModelQuest (MarketMiner, 2004) software was used to complete the steps to derive the NPN model. ModelQuest software has been used by other researchers, including Abdel-Aal and Mangoud (1997), Agarwal (1999), Cerullo and Cerullo (2006), Kim (2002), and Reddy and Pachepsky (2000).

The NPN was calculated one layer at a time. The initial (or input) layer consisted of normalizing the 16 inputs to a mean of zero and standard deviation of one. For each subsequent layer of the network, each possible combination of inputs from the prior layer were combined into third-order polynomial equations with each combination of single, double, and triple inputs using the following equations (Montgomery, 1989):

$$\text{Single} = w_0 + w_1 x_1 + w_2 x_1^2 + w_3 x_1^3 \quad (2)$$

$$\text{Double} = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + w_6 x_1^3 + w_7 x_2^3 \quad (3)$$

$$\text{Triple} = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_1^2 + w_5 x_2^2 + w_6 x_3^2 + w_7 x_1 x_2 + w_8 x_1 x_3 + w_9 x_2 x_3 + w_{10} x_1 x_2 x_3 + w_{11} x_1^3 + w_{12} x_2^3 + w_{13} x_3^3 \quad (4)$$

where w_i are the coefficients, x_i are the input variables, Single indicates an equation with one input variable, Double indicates an equation with two input variables, and Triple indicates an equation with three input variables.

A selection criterion was applied to each of the single, double, and triple equations, along with inputs from the prior layer to select the best predictors for input to the next level. The selection criterion was also applied to the resulting network. Until the selection criterion for the network was met, additional layers were added to the network using inputs calculated in the prior layer. Predicted squared error (PSE) was used as the selection criterion. Barron (1984) defined PSE as consisting of a squared error term based on the training data and an overfit penalty term as follows:

$$\text{PSE} = \text{TSE} + 2\sigma_p^2(K/N) \quad (5)$$

where TSE is the training squared error, $2\sigma_p^2$ is the prior estimate of true error variance, K is the number of coefficients estimated to minimize TSE, and N is the number of training observations.

As each coefficient was added to reduce the error of the NPN, the overfit penalty increased. The overfit penalty is designed to keep a model from overfitting the training data to the extent that it performs poorly on future observations. Once PSE for a layer increased from the prior layer, the NPN from the prior layer was selected as the best model. The resulting value (normalized HC) was converted to units of g/kWh.

Each of the NPN models was evaluated with the same evaluation data set as the LMR models. The models were compared based on coefficient of determination (R^2), average error, and maximum error.

Results

Data from each of the ISO and EXP training data sets were analyzed using correlation. Table 2 lists the input variable names and correlation coefficients for each set of training data. For the ISO training data, correlations for 10 of the 16 input variables with HC were significant at $p < 0.0001$, with three additional inputs significant at $p < 0.05$. Only percent friction, atmospheric pressure, and exhaust pressure were not significant. Correlations of the EXP training data found 9 of the 16 input variables significant at $p < 0.0001$, with one additional input significant at $p < 0.05$. The six EXP training data inputs not significantly correlated to HC included the three from the ISO training data, plus engine speed, engine fuel temperature, and exhaust temperature.

Correlations of each input variable to HC were stronger for the smaller ISO training data with the exception of atmospheric pressure. For the ISO data, the strongest correlations were ambient temperature, percent load, torque, percent torque, flywheel power, and fuel flow rate, with correlations in the range of −0.76 to −0.71. For the EXP data, the five strongest correlations of −0.60 to −0.55 were for percent torque, torque, percent load, flywheel power, and fuel flow rate.

Table 2. Pearson correlation coefficient of each independent variable to HC (g/kWh) for each of the ISO and EXP training data sets.

Variable	Correlation Coefficient (r)[a] ISO	EXP
Engine speed (rpm)	-0.38 *	-0.04
Percent torque	-0.74 **	-0.60 **
Percent load	-0.75 **	-0.59 **
Percent friction	-0.26	-0.04
Fuel flow rate (L/h)	-0.71 **	-0.55 **
Engine fuel temperature (°C)	-0.62 **	-0.14
Coolant temperature (°C)	-0.53 *	-0.40 **
Intake manifold temperature (°C)	-0.69 **	-0.50 **
Torque (N·m)	-0.75 **	-0.60 **
Flywheel power (kW)	-0.73 **	-0.59 **
Ambient temperature (°C)	-0.76 **	-0.40 **
Relative humidity (%)	0.66 **	0.23 *
Atmospheric pressure (mbar)	0.03	-0.05
Exhaust temperature (°C)	-0.44 *	-0.16
Exhaust pressure (kPa)	-0.07	-0.05
Air to fuel ratio	0.64 **	0.37 **

[a] * = correlation coefficient significant with $p < 0.05$.
** = correlation coefficient significant with $p < 0.0001$.

Linear Multiple Regression (LMR) Models

LMR was used to fit an equation to the combination of the 16 input variables from the ISO training data to predict HC. The resulting equation accounted for approximately 94% of observed variance in HC in the training data ($F_{16,16} = 15.08$, $p < 0.0001$, adjusted $R^2 = 0.88$). Table 3 lists the regression coefficients and standardized coefficients for each input. Inputs of flywheel power, fuel flow rate, and percent torque had the highest weights, but none of the weights was significant at $p < 0.05$.

LMR was also used with the EXP training data to fit an equation to HC. The resulting equation accounted for

Table 3. Regression coefficients and standardized coefficients for each of the ISO and EXP training data sets.

Variable	Regression Coefficients[a] ISO	EXP	Standardized Coefficients ISO	EXP
y-intercept	424.606	84.118	0	0
Engine speed (rpm)	-0.001	0.013*	-0.389	1.342
Percent torque	-0.220	-0.230	-3.107	-1.507
Percent load	-0.039	0.117	-0.646	0.880
Percent friction	0.093	-3.494**	0.146	-2.161
Fuel flow rate (L/h)	1.2434	-0.457	4.846	-0.810
Engine fuel temp. (°C)	-0.307	-0.165*	-1.041	-0.221
Coolant temperature (°C)	-0.065	-0.278	-0.088	-0.140
Intake manifold temp. (°C)	0.099	0.420**	2.031	4.086
Torque (N·m)	0.031	-0.032	2.204	-1.093
Flywheel power (kW)	-0.409	-0.304	-6.070	-2.228
Ambient temperature (°C)	0.189	-0.045	0.346	-0.042
Relative humidity (%)	0.103	-0.045	0.104	-0.062
Atmos. pressure (mbar)	-0.508	0.139	-0.325	0.106
Exhaust temperature (°C)	-0.320	-0.072	-0.400	-0.066
Exhaust pressure (kPa)	0.156	-0.216	0.079	-0.161
Air to fuel ratio	-0.147	-0.164**	-0.897	-0.506

[a] * = regression coefficient significant with $p < 0.05$.
** = regression coefficient significant with $p < 0.0001$.

Table 4. Comparative performance of LMR and NPN models derived from each of ISO and EXP data sets and the evaluation data.

Data Set	Model Strategy	Training Data[a]			Evaluation Data[b]		
			Absolute Error			Absolute Error	
		$R^{2[c]}$	Mean	Max.	$R^{2[c]}$	Mean	Max.
ISO	LMR	0.94	0.37	1.58	0.49	1.56	7.02
	NPN	0.99	0.20	0.60	0.91	0.52	6.31
EXP	LMR	0.68	1.37	10.71	0.64	1.27	8.57
	NPN	0.99	0.20	2.74	0.98	0.42	5.06

[a] Actual HC (g/kWh) values in the ISO data set ranged from 0.05 to 5.79, with a mean of 1.37 and standard deviation of 2.00.
Actual HC (g/kWh) values in the EXP data set ranged from 0.01 to 23.57, with a mean of 1.24 and a standard deviation of 3.50.

[b] Actual HC (g/kWh) values in the evaluation data set ranged from 0.01 to 11.89, with a mean of 1.24 and a standard deviation of 2.80.

[c] Coefficient of determination between actual and predicted HC for the respective data set.

approximately 68% of the observed variance in HC ($F_{16,118}$ = 15.69, $p < 0.0001$, adjusted $R^2 = 0.64$). Table 3 lists the regression coefficients and standardized coefficients for each input. Inputs of intake manifold temperature, flywheel power, percent friction, and percent torque had the four highest weights. Percent friction, intake manifold temperature, and air to fuel ratio had coefficients significant at $p < 0.0001$. Regression coefficients for engine speed and engine fuel temperature were significant at $p < 0.05$.

The respective regression equations of the ISO and EXP training data were applied to the evaluation data to predict HC. Table 4 summarizes the R^2, mean absolute error, and maximum absolute error of each model applied to the training data and evaluation data. When applied to the evaluation data, the R^2 for the LMR model based on the ISO training data dropped from 0.94 to 0.49. The R^2 for the EXP data dropped from 0.68 to 0.64. When the ISO-based model was applied to the evaluation data, the mean absolute error (1.56) and maximum absolute error (7.02) were higher than the respective error of the training data (0.37 and 1.58) by ratios of 4.2 and 4.4, respectively. When the EXP-based model was applied to the evaluation data, the mean and maximum absolute errors actually decreased compared to the EXP training data.

Nonlinear Polynomial Network (NPN) Models

The ISO training data were used to develop a NPN model to fit an equation to HC based on the 16 input variables. The resulting NPN is depicted in figure 2. Of the 16 input variables, only torque and exhaust temperature were used by the resulting polynomial network. The predicting network accounted for approximately 99% of the observed variance in HC and consists of the following network of equations:

$$T_n = -1.2661 + 0.0071T \tag{6}$$

$$EX_n = -17.6296 + 0.4001EX \tag{7}$$

$$DB = -0.7365 + 0.7404T_n^2 - 0.4231T_n^3 + 0.2337T_n\,EX_n - 0.1819T_n^2\,EX_n - 0.1522T_n\,EX_n^2 \tag{8}$$

$$HC = 1.3691 + 1.9983DB \tag{9}$$

where T_n is the normalized torque (N·m), T is the observed torque (N·m), EX_n is the normalized exhaust temperature (°C), EX is the observed exhaust temperature (°C), DB indicates double, i.e., a network node with two inputs (eq. 3), and HC is hydrocarbon emissions (g/kWh).

The EXP training data were likewise used to develop an NPN model to fit an equation to HC based on the 16 input variables. The resulting polynomial network is depicted in figure 3. The only input variables used in the model were torque and air to fuel ratio. The predicting network accounted for approximately 99% of the observed variance in HC and consists of the following network of equations:

$$T_n = -1.9012 + 0.0083T \tag{10}$$

$$AFR_n = -3.1372 + 0.0925AFR \tag{11}$$

$$DB_1 = -0.4145 - 0.0729T_n + 0.7182T_n^2 - 0.4281T_n^3 - 0.1578AFR_n + 0.4491T_n\,AFR_n - 0.219T_n^2\,AFR_n \tag{12}$$

$$DB_2 = -0.5987 - 2.3356DB_1 - 4.0739DB_1^2 + 0.9741DB_1^3 + 0.3609AFR_n + 1.7597DB_1\,AFR_n + 1.8434DB_1^2\,AFR_n - 0.1562AFR_n^2 - 0.4836DB_1\,AFR_n^2 + 0.0209AFR_n^3 \tag{13}$$

$$HC = 1.2416 + 3.4954DB_2 \tag{14}$$

where AFR_n is the normalized air to fuel ratio, AFR is the observed air to fuel ratio, and DB_1 and DB_2 indicate double, i.e., network nodes with two inputs (eq. 3).

The EXP-based NPN model was more complex than the ISO-based NPN model, with an additional layer in the polynomial network. Each of the NPN models used two input variables, and both models had torque as a common input. The other inputs of exhaust temperature and air-to-fuel ratio had a very low correlation (r = –0.078).

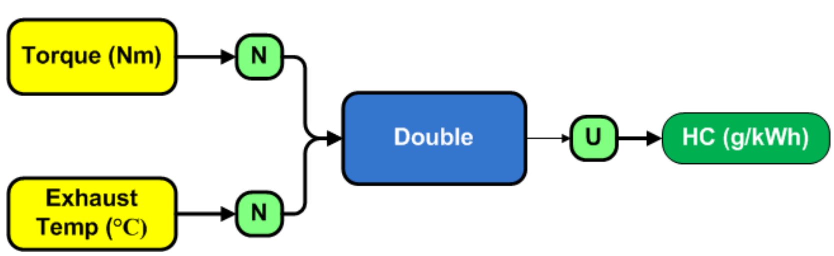

Figure 2. Polynomial network generated from ISO training data.

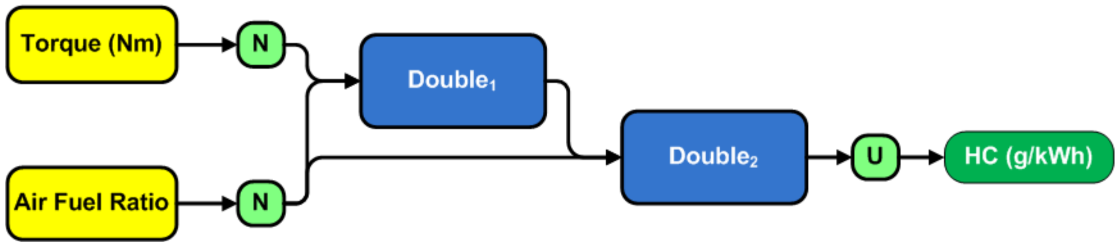

Figure 3. Polynomial network generated from EXP training data.

The NPNs of the ISO and EXP training data were applied to the evaluation data to predict HC. Table 4 summarizes the R^2, mean absolute error, and maximum absolute error of each model applied to the training data and evaluation data. When applied to the evaluation data, the R^2 for the NPN model based on the ISO training data dropped from 0.99 to 0.91. The R^2 for the EXP data dropped from 0.99 to 0.98. For the ISO-based model applied to the evaluation data, the ratio of the mean absolute error for the evaluation data (0.52) to that of the training data (0.20) was 2.6 ,and the ratio of the maximum absolute error for the evaluation data (6.31) to that of the training data (0.60) was 10.5. Mean (0.42) and maximum (5.06) absolute error in predicting HC for evaluation data based on the EXP model were ratios of 2.1 and 1.8, respectively, of the mean (0.2) and maximum (2.74) absolute error for the training data.

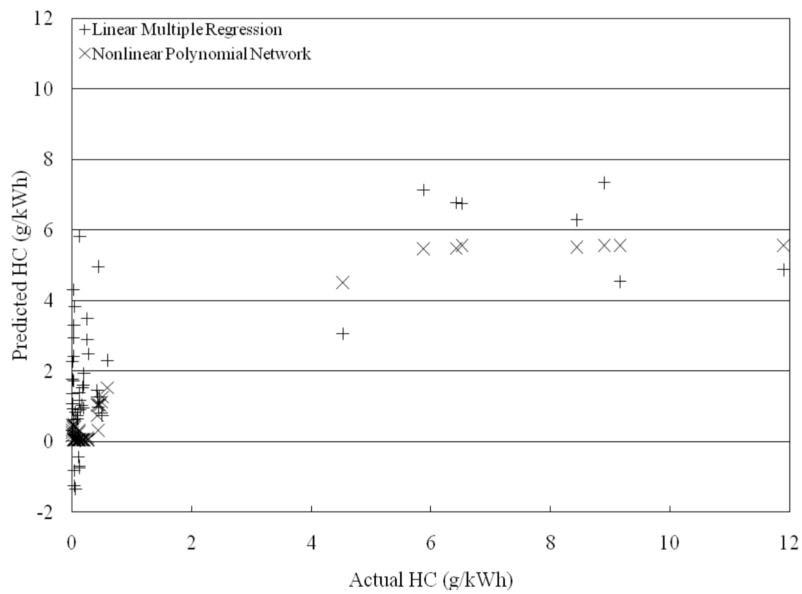

Figure 4. Predicted HC values for evaluation data, from LMR and NPN models, derived from ISO training data.

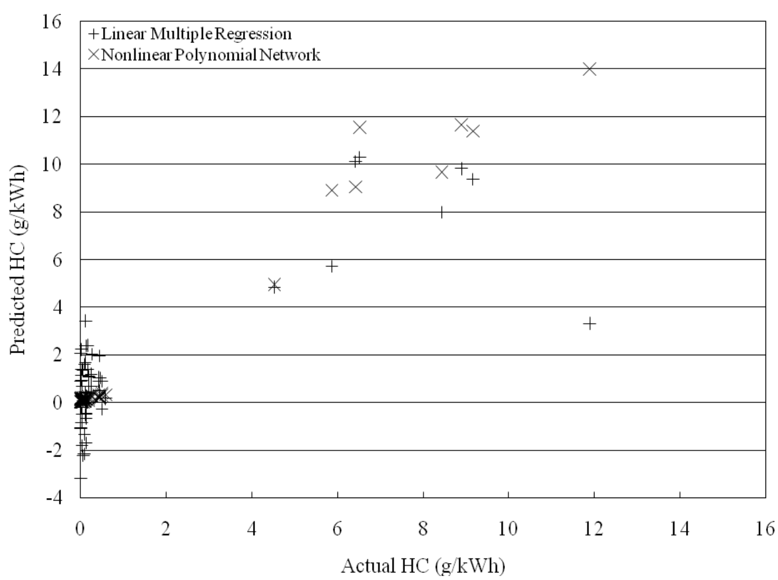

Figure 5. Predicted HC values for evaluation data, from LMR and NPN models, derived from EXP training data.

The relative accuracies of the ISO-based LMR and NPN models in predicting the evaluation data are illustrated in figure 4. Actual HC values were 4.5 g/kWh or higher when the engine was operated at no load (0% torque). For all the other test modes, actual HC values were below 0.6 g/kWh. When measured HC was below 0.6 g/kWh, the LMR model predicted HC values ranging from −1.3 to 5.8 g/kWh. In contrast, the NPN model predicted values ranged from 0.05 to 1.5 g/kWh. Both models underpredicted when actual HC was above 8.0 g/kWh, and the highest actual output of nearly 12 g/kWh is where each model had its maximum absolute error.

The relative accuracies of the EXP-based LMR and NPN models in predicting the evaluation data are depicted in figure 5. When actual HC was below 0.6 g/kWh, the LMR model predicted HC values ranging from −3.2 to 3.3 g/kWh. In contrast, the NPN model predicted values ranged from 0.01 to 0.39 g/kWh. The LMR model had its maximum absolute error when actual HC was nearly 12 g/kWh. The NPN model had its maximum absolute error when actual HC was about 6.5 g/kWh.

Discussion

ISO and EXP data sets from diesel engine emissions tests were compared for generating models to predict HC emissions of an engine operating at a range of constant loads and speeds. Modeling methods of LMR and NPN were used with both data sets for comparison. Models were evaluated with data consisting of a fourth replication of data from the same tests used for the ISO and EXP training data. The target operating range to model was 1500 to 2500 rpm with loads of 40% to 100% of the maximum torque for each engine speed.

The LMR model developed with the ISO training data ($R^2 = 0.94$) was able to explain nearly half the variation in the evaluation data ($R^2 = 0.49$). Although the ISO training data alone resulted in a strong relationship, the resulting model was overfitted (i.e., fitted to both the signal and noise and thus fit the training data better, with higher R^2, than it could predict new values) and was not as effective at predicting the evaluation data with the additional engine operating conditions. Results of the ISO-based LMR model indicated that ISO 8178-4 test cycle B-type test modes alone would not be sufficient to model the target range of engine operation.

A second LMR model was developed using the EXP training data ($R^2 = 0.68$). This data included engine speeds from 1500 to 2500 rpm in 250 rpm increments and 40%, 50%, 60%, 70%, 80%, 90%, and 100% of maximum torque for each speed. When applied to the evaluation data ($R^2 = 0.64$), the model outperformed the ISO-based LMR model. The additional test modes in the EXP training data improved the LMR model evaluation but left 36% of the variation of HC unexplained.

An NPN model was developed with the ISO training data ($R^2 = 0.99$). When applied to the evaluation data ($R^2 = 0.91$), it explained 91% of the variance of HC, compared to 49% of the variance with the LMR model. The second NPN model developed with the EXP training data ($R^2 = 0.99$) maintained nearly the same performance with the evaluation data ($R^2 = 0.98$). Average absolute error for the two NPN models was similar, with 0.52 and 0.42 g/kWh, respectively, for the ISO and EXP models.

Actual HC emissions from test modes with a load applied to the engine, ranged from 0.01 to 0.6 g/kWh. The LMR method predicted HC ranging from −1.3 to 5.8 g/kWh for the ISO-based model and −3.2 to 3.3 g/kWh for the EXP-based model. In comparison, the NPN method predicted values ranged from 0.05 to 1.5 g/kWh for the ISO-based model and 0.01 to 0.39 g/kWh for the EXP-based model. The NPN method was more effective at modeling the emission rates due to a combination of a nonlinear relationship and the polynomial network method.

Both of the NPN models used torque as an input. When comparing the standardized coefficients for the LMR models (table 3), based on absolute value, torque had the fourth and sixth highest values for the ISO and EXP models, respectively. The other NPN inputs, exhaust temperature for the ISO-based model and air to fuel ratio for the EXP-based model, were the ninth highest absolute standardized coefficients for the respective LMR models. For both models, the NPN method rejected inputs with higher absolute standardized coefficients. For our data, the NPN method of testing network nodes of third-order polynomials with all combinations of single, double, or triple inputs (eqs. 2, 3, and 4) to find the best polynomial relationship resulted in better fitting (R^2) models than LMR.

The expanded data from the EXP test modes would be preferable for modeling emissions of an engine running at a constant load and speed as described for this study. In the absence of EXP data, data collected from ISO 8178-4 test cycle B would be sufficient to model HC emissions of a diesel engine with NPN. The NPN method was superior to LMR in predicting HC of the evaluation data, even when trained with the smaller sample size ($n = 33$) of the ISO training data.

We do not dispute the conclusion of Hansson et al. (2001) that it was not possible to design one set of emissions factors that produced representative results for all types of tractors and work operations. Although our EXP test modes resulted in better models for our target range of operation, these test modes would not describe every engine application. However, our results do indicate that if a broader range of data (torque, air to fuel ratio, exhaust temperature, etc.) were made available from ISO 8178-4 emission tests, it may be possible to predict HC emissions of an engine operating at a constant load and speed using NPN. Inputs used for the NPN models consisted of torque, exhaust temperature, and air to fuel ratio. Instruments are readily available to measure exhaust temperature or air to fuel ratio. Torque data for an operating engine could be derived from the available percent torque on the CAN. Assuming the percent torque value is an integer, the reduced resolution may adversely affect a model by increasing the error in HC predictions. Another option would be to install strain gauge transducers to measure torque. Hansson et al. (2003) used this method to measure torque at the transmission input shaft of a tractor.

Conclusions

This study of using ISO and EXP data sets with LMR and NPN modeling to predict HC produced the following conclusions:

- LMR using the ISO training data ($R^2 = 0.94$) resulted in overfitting of the model, as applied to the evaluation data ($R^2 = 0.49$).

- LMR using the EXP training data ($R^2 = 0.68$) resulted in a better fit for the evaluation data ($R^2 = 0.64$) than the ISO training data, but the model underperformed compared to the NPN models.
- NPN using the EXP training data ($R^2 = 0.99$) resulted in the best model when applied to the evaluation data ($R^2 = 0.98$) and is recommended for predicting HC when the larger set of test mode data is available.
- NPN using the ISO training data ($R^2 = 0.99$) resulted in a satisfactory fit for the evaluation data ($R^2 = 0.91$), although with a higher average absolute error (0.52 vs. 0.42 g/kWh) than NPN using the EXP training data. This model was also considered suitable for predicting HC.
- Considering the time and resource costs of the additional 34 test modes for the EXP training data, the ISO training data was sufficient to satisfactorily model HC when NPN was used.
- This study of one engine suggests that data could be collected during ISO 8178-4 emission tests and modeled with NPN to predict HC emissions for a diesel engine operating at various constant speeds and loads. Similar studies with other engines are needed to determine the applicability of these results.

References

Abdel-Aal, R. E., and A. M. Mangoud. 1997. Modeling obesity using abductive networks. *Computers and Biomed. Res.* 30(6): 451-471.

Agarwal, A. 1999. Abductive networks for two-group classification: A comparison with neural networks. *J. Applied Business Res.* 15(2): 1-12.

Barron, A. R. 1984. Predicted squared error: A criterion for automatic model selection. In *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, 87-103. S. J. Farlow, ed. New York, N.Y.: Marcel Dekker.

Barron, A. R., and R. L. Barron. 1988. Statistical learning networks: A unifying view. In *Computing Science and Statistics: Proc. of the 20th Symposium on the Interface*, 192-203. E. J. Wegman, D. T. Gantz, and J. J. Miller, eds. Alexandria, Va.: American Statistical Association.

Barron, R. L., A. N. Mucciardi, F. J. Cook, J. N. Craig, and A. R. Barron. 1984. Adaptive learning networks: Development and applications in the United States of algorithms related to GMDH. In *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, 25-65. S. J. Farlow, ed. New York, N.Y.: Marcel Dekker.

CCR. 1999. California exhaust emission standards and test procedures for new 2001 and later off-road large spark-ignition engines. California Code of Regulations, Title 13, Chapter 9, Article 4.5. Sacramento, Cal.: State of California.

Cerullo, M. J., and M. V. Cerullo. 2006. Using neural network software as a forensic accounting tool. *Information Systems Control Journal* 2006, vol. 2. Available at: www.isaca.org/Template.cfm?Section=Archives&Template=/ContentManagement/ContentDisplay.cfm&ContentID=30759.

Drake, K. C., and R. Y. Kim. 1997. Abductive information modeling applied to financial time series forecasting. In *Nonlinear Financial Forecasting, Proc. 1st INFFC*, 95-108. Haymarket, Va.: Finance and Technology Publishing.

Drake, K. C., R. Y. Kim, T. Y. Kim, and O. D. Johnson. 1994. Comparison of polynomial network and model-based target recognition. In *Proc. Sensor Fusion and Aerospace Applications II*, 2-11. Bellingham, Wash.: SPIE.

Duffy, J. J., and M. A. Franklin. 1975. A learning identification algorithm and its application to an environmental system. *IEEE Trans. Systems, Man, and Cybernetics* 5(2): 226-240.

Farlow, S. J. 1984. The GMDH algorithm. In *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, 1-24. S. J. Farlow, ed. New York, N.Y.: Marcel Dekker.

Griffin, M. P., D. F. Scollan, and J. R. Moorman. 1994. The dynamic range of neonatal heart rate variability. *J. Cardiovascular Electrophysiology* 5(2): 112-124.

Hansson, P.-A., M. Lindgren, and O. Noren. 2001. A comparison between different methods of calculating average engine emissions for agricultural tractors. *J. Agric. Eng. Res.* 80(1): 37-43.

Hansson, P.-A., M. Lindgren, M. Nordin, and O. Pettersson. 2003. A methodology for measuring the effects of transient loads on the fuel efficiency of agricultural tractors. *Applied Eng. in Agric.* 19(3): 251-257.

Hogan, J. A., D. G. Watson, and T. V. Harrison. 2007. Data points and duration for estimating fuel consumption of a LPG engine. *Agric. Eng. Intl.: CIGR EJournal* IX: 1-10.

ISO. 1996. ISO 8178-4: Reciprocating internal combustion engines, exhaust emission measurements: Part 4. Test cycles for different engine applications. Geneva, Switzerland: International Organization of Standardization.

Ivakhnenko, A. G. 1968. The group method of data handling: A rival of the method of stochastic approximation. *Soviet Automatic Control* 13(3): 43-55.

Ivakhnenko, A. G. 1971. Polynomial theory of complex systems. *IEEE Trans. Systems, Man, and Cybernetics* 1(4): 364-378.

Ivakhnenko, A. G., V. S. Stepashko, M. G. Khomovnenko, and Y. P. Galyamin. 1977. Self-organization of dynamic models of growth of agricultural crops for control of irrigated crop rotation. *Soviet Automatic Control* 10: 23-33.

Kim, K. S. 2002. Value management and common accounting performance measures for corporations. *Expert Systems with Applications* 22(4): 331-336.

King, M. A., J. F. Elder, B. Gomolka, E. Schmidt, M. Summers, and K. Toop. 1998. Evaluation of fourteen desktop data mining tools. *IEEE Intl. Conf. Systems, Man, and Cybernetics.* Piscataway, N.J.: IEEE.

Kleinsteuber, S., and N. Sepehri. 1996. A polynomial network modeling approach to a class of large-scale hydraulic systems. *Computers and Electrical Eng.* 22(2): 151-168.

Krijnsen, H. C., R. Bakker, W. E. van Kooten, H. P. Calis, R. P. Verbeek, and C. M. van den Bleek. 2000. Evaluation of fit algorithms for NOx emission prediction for efficient DeNOx control of transient diesel engine exhaust gas. *Ind. and Eng. Chem. Res.* 39(8): 2992-2997.

Lebow, W. M., R. K. Mehra, and P. M. Toldalagi. 1984. Forecasting applications of GMDH in agricultural and meteorological time series. In *Self-Organizing Methods in Modeling: GMDH Type Algorithms*, 121-148. S. J. Farlow, ed. New York, N.Y.: Marcel Dekker.

MarketMiner. 2004. *MarketMiner ModelQuest Analyst Version 6.0 User's Guide.* Charlottesville, Va.: MarketMiner, Inc.

Montgomery, G. J. 1989. Abductive diagnostics. In *Proc. 7th AIAA Computers in Aerospace Conf.*, 267-275. Washington, D.C.: American Institute of Aeronautics and Astronautics.

Montgomery, G. J., and K. C. Drake. 1991. Abductive reasoning networks. *Neurocomputing* 2(3): 97-104.

Montgomery, G. J., P. Hess, and J. S. Hwang. 1990. Abductive networks applied to electronic combat. In *Proc. SPIE 1294: Applications of Artificial Neural Networks*, 454-465. Bellingham, Wash.: SPIE.

National Instruments. 2003. *LabVIEW 7 Express User Manual.* Austin, Tex.: National Instruments.

Pachepsky, Y. A., and W. J. Rawls. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. *SSSA J.* 63(6): 1748-1757.

Pyo, S., M. Uysal, and H. Chang. 2002. Knowledge discovery in database for tourist destinations. *J. Travel Res.* 40(4): 396-403.
Ramamurthy, R., N. N. Clark, C. M. Atkinson, and D. W. Lyons. 1998. Models for predicting transient heavy-duty vehicle emissions. SAE Paper 982652. Warrendale, Pa.: SAE.
Reddy, V. R., and Y. A. Pachepsky. 2000. Predicting crop yields under climate change conditions from monthly GCM weather projections. *Environ. Modeling and Software* 15(1): 79-86.
SAE. 2002. SAE J1939-71 AUG2002 Recommended practice for truck and bus control and communications network: Vehicle application layer. Warrendale, Pa.: SAE.
SAS. 2007. SAS 9.1. Cary, N.C.: SAS Institute, Inc.
Silis, Y. Y., and A. B. Rozenblit. 1976. Algorithm for construction of decision function in the form of a complex logic proposition. *Soviet Automatic Control* 9(2): 1-5.
Stepanov, V. A. 1974. Results of mass checking of GMDH efficiency in long-term prediction of demand for consumer goods. *Soviet Automatic Control* 7: 46-51.
Watson, D. G., T. V. Harrison, and R. W. Steffen. 2008. Data points and duration for estimating fuel consumption of a diesel engine. *Agric. Eng. Intl.: CIGR EJournal* X: 1-9.
Yanowitz, J., M. S. Graboski, and R. I. McCormick. 2002. Prediction of in-use emissions of heavy-duty diesel vehicles from engine testing. *Environ. Sci. and Tech.* 36(2): 270-275.