REAL-TIME TRANSIENT SOOT AND NO_X VIRTUAL SENSORS FOR DIESEL ENGINE USING NEURO-FUZZY MODEL TREE AND ORTHOGONAL LEAST SQUARES

Rajit Johri
Mechanical Engineering,
University of Michigan
Ann Arbor, Michigan, USA
48109
rajit@umich.edu

Ashwin Salvi
Mechanical Engineering,
University of Michigan
Ann Arbor, Michigan, USA
48109
asalvi@umich.edu

Zoran Filipi*
Mechanical Engineering,
University of Michigan
Ann Arbor, Michigan, USA
48109
filipi@umich.edu
* Corresponding Author

ABSTRACT

Diesel engine combustion and emission formation is highly nonlinear and thus creates a challenge related to engine diagnostics and engine control with emission feedback. This paper presents a novel methodology to address the challenge and develop virtual sensing models for engine exhaust emission. These models are capable of predicting transient emissions accurately and are computationally efficient for control and optimization studies.

The emission models developed in this paper belong to the family of hierarchical models, namely "neuro-fuzzy model tree". The approach is based on divide-and-conquer strategy i.e. to divide a complex problem into multiple simpler subproblems, which can then be identified using simpler class of models. Advanced experimental setup incorporating a medium duty diesel engine is used to generate training data. Fast emission analyzers for soot and NO_X provide instantaneous engine-out emissions. Finally, the Engine-In-the-Loop is used to validate the models for predicting transient particulate mass and NO_X.

Keywords: transient diesel emissions, soot model, neuro-fuzzy model tree, hierarchical models, orthogonal least squares (OLS), multi-level pseudo random signal (MPRS).

INTRODUCTION

Growing environmental concerns, stringent emission regulations and demand for increased fuel efficiency will require advanced engines and control strategies. The problem of meeting emission regulations is particularly tough for diesel engines compared to their gasoline counterparts. Transient diesel engine particulates and NO_X emissions are very complex phenomena owing to the nature of diesel combustion. To meet EPA standards, modern diesel engines employ a large number of actuators like multiple fuel injections, variable geometry turbochargers (VGT), exhaust gas recirculation (EGR) etc. which adds to complexity. Aftertreatment is necessary to bring the tailpipe emissions down to compliance levels. However, the burden is equally shared by the in-cylinder clean combustion strategies, advanced catalyst and diesel particulate filters since size and cost of aftertreatment is an issue.

Advanced approaches such as model based predictive control, closed loop combustion control and development of advanced supervisory strategies for hybrid propulsion systems will be essential for coping with new regulations and complex hardware. In all cases, model-based soot and NO_X virtual sensors can provide real-time predictions and enable strategies that require feedback of emissions under transient operating conditions. This paper pursues development of such virtual sensors for onboard vehicle application or powertrain system optimization.

Previous work done by Hagena et al. [1] and Kirchen et al. [2] showed that transient soot emissions account for almost half of total soot emission over a driving schedule. Steady state map based models fail to capture the transient nature of emission when engine is operated transiently and underestimate soot production. This can be seen from Figure 1, as the integrated area under the transient trace is much larger compared to the quasi-steady state curve. The transient spike is higher and precedes the quasi-steady state prediction. The quasi-steady state estimates deviate considerably at the initiation of transient when conditions are irregular. Transient conditions easily dominate the emission trends for a heavy-duty vehicle, particularly over an aggressive driving schedule like FTP75.

Dealing with transients needs to be part of overall low-emissions strategy, as more than half the soot particulates can be attributed to rapid increase in load [1].

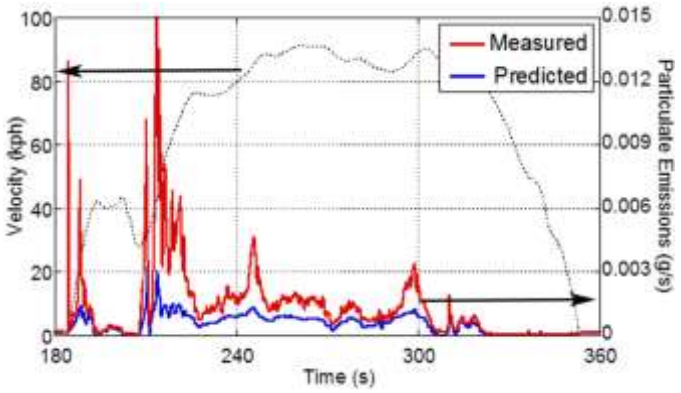

Figure 1: Quasi-steady state model prediction compared with measured soot emissions [12]

Present day emission models generally fall into two categories, at the extreme ends of the spectrum when it comes to complexity and computational speed: (i) lookup table based steady state models, and (ii) computational fluid dynamics (CFD) with chemical kinetics based models. CFD and chemical kinetics based models can capture transient effects but are very complex and computationally slow. This makes them impractical to be used with optimization routines and as virtual sensors with real engine. Therein lies the impetus for transient emission models, fast enough to be employed for control-oriented problems or as virtual real-time sensors, and yet detailed enough to capture system dynamics accurately.

Recently, researchers have proposed empirical models that capture emission dynamics based on certain engine parameters. Kirchen [2] developed a mean value model for soot and showed the effectiveness with tip-in operations. The model included empirical correlations relating engine out emission with engine operating conditions. Bayer and Foster [3] developed phenomenological model for predicting soot. Brahma et al. [4] linked these phenomenological models with neural network and trained to predict soot. A major drawback of the above proposed models is the required knowledge of in-cylinder pressures and available fuel mass as model inputs.

This paper presents a novel neuro-fuzzy model tree for predicting transient soot and NO_X emissions for diesel engine. The model is intended to run on a microprocessor in real-time and predict engine-out soot and NO_X emissions using signals from the ECU and low-cost physical sensors. The neuro-fuzzy model tree based emission sensor is capable of learning complex, nonlinear and multidimensional association between inputs and outputs. The neuro-fuzzy model has parallel structure with respect to local models and thus can be efficiently implemented in hardware. Emission models developed in this study are driven by experimental data and are specific to a particular diesel engine but the methodology developed is universal and can be applied to any other engine.

The paper begins with a brief overview of combustion and emission formation in diesel engine. The next section provides a detailed description of Engine-in-the-Loop facility at the University of Michigan. The EIL setup is used to characterize the engine transients and to subsequently validate the emissions model. Next, a perturbation signal is designed specifically for characterizing the dynamic engine operation and resulting emissions, followed by the description of the neuro-fuzzy modeling approach and the training algorithm. The algorithm is augmented with Orthogonal Least Squares (OLS) for input regressor selection. Finally, the model validation results are presented. The emission model predictions are compared with actual measurements from fast analyzers. The paper ends with conclusions.

BACKGROUND - DIESEL COMBUSTION AND EMISSION

Diesel combustion is a very complex process. Optical studies combined with analyses of engine cylinder pressure data have led to the widely accepted phenomenological understanding proposed by Dec [5]. After injection, the fuel evaporates and mixes with air, and owing to very high temperatures, autoignites after a delay (*ignition delay*). The fuel/air mixture prepared during the ignition delay period burns rapidly and this is referred to as *premixed phase* of burning. Following the premixed combustion, the fuel injection continues through the *mixing-controlled burn phase*. The liquid core of injected fuel persists and the fuel droplets downstream of the liquid core are evaporated, facilitated by turbulent air entrainment. This results in formation of relatively uniform, high equivalence ratio (fuel/air of 2-4) zone, extending ahead and around the liquid fuel core. A standing premixed flame forms at the boundary of this gaseous fuel/air zone, and owing to excessive rich conditions, produces polycyclic aromatic hydrocarbons (PAH - soot precursor) and solid particles. The soot particles are initially small but grow in size and concentration as they move towards the head vortex. The particle accumulation process continues in the head-vortex zone surrounded by a thin diffusion flame. The outer edge of diffusion flame is surrounded by OH radicals and oxygen molecules, which oxidize particles that reach outer boundary. The high temperature and presence of oxygen is highly conducive for NO_X production. The NO_X production continues even after end of injection due to latter part of diffusion burning.

Even though the fundamentals of the emission formation process remain same with transient operation of the engine, it makes modeling them even more complex due to constantly changing combustion environment and engine subsystem interactions. Fluctuations in charge composition, stochastic nature of turbulence, mixing and combustion makes the interaction of species in engine cylinder inconsistent. Instantaneous composition in the cylinder and flow field goes through dramatic excursions from the steady state values after a rapid change of engine command. Thus, steady state emission models cannot capture these effects and are incapable of

predicting transient emissions [6]. In summary, multiple aspects of the very complex phenomena in the combustion chamber have to be understood, and experimental insights are necessary to support virtual sensor development.

EXPERIMENTAL SETUP

This section gives a brief overview of the Engine-In-the-Loop (EIL) setup at the University of Michigan, Figure 2.

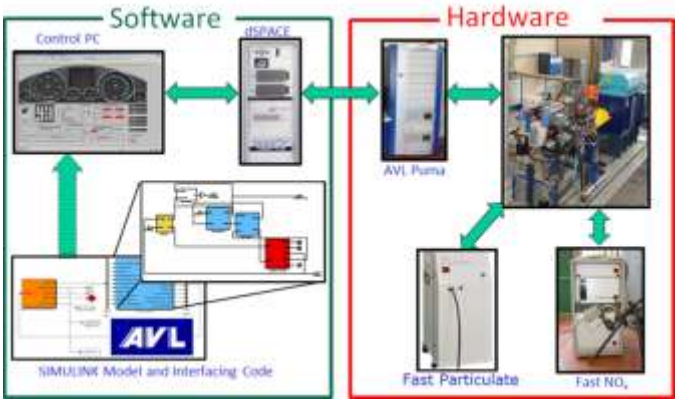

Figure 2: Engine-In-the-Loop test cell configuration

ENGINE SPECIFICATIONS

A 6.4 L V-8 direct-injection diesel engine manufactured by the Navistar Ltd. is used for this work. Engine specifications are given in Table 1. The engine incorporates advanced technologies to provide high power density while meeting 2007 emissions standards. A common rail direct injection (CRDI) system permits precise control of fuel injection timing, pressure, quantity, and number of injections. An EGR circuit allows for introducing cooled exhaust gases into intake manifold and reduces NO_X emissions. The dual stage VGT is used to enhance engine performance and EGR control.

Table 1: Diesel engine specifications

Engine Type	DI 4-Stroke Diesel Engine
Configuration	V-8, Cam-in-Crankcase, 90° V-8
Bore x Stroke	98mm x 105mm
Displacement	6.4L
Rated Power	261kW @ 3000RPM
Rated Torque	881Nm@ 2000 RPM
Compression Ratio	16.7 : 1
Valve Lifters	Push Rod-Activated Rocker Arm
Aspiration	Variable Geometry Dual stage Turbocharger / Intercooler
Fuel Delivery System	Common Rail Direct Injection (CRDI)

TEST CELL SYSTEMS

The engine is coupled to a 330 kW AVL ELIN series 100 APA Asynchronous Dynamometer. This dynamometer is especially well suited for transient testing with a 5 ms torque response time and a -100% to +100% torque reversal time of 10 ms. AVL PUMA Open system orchestrates engine operation in the test cell, and provides monitoring and control of test cell functions. The AVL PUMA interfaces and communicates with dSPACE real-time system. This facilitates concurrent running of engine with virtual driveline and vehicle [7]. The engine is fully instrumented with time-based measurements like temperature, manifold pressure and flow rates, as well as crank-angle based measurements such as in-cylinder and fuel injection pressures.

EMISSIONS MEASUREMENT

Fast NO_X

CLD 500 Fast NO_X analyzer is used for accurate temporal measurement of NO_X. It consists of a chemiluminescent detector with a 90%→10% response time of less than 3 ms for NO, and less than 10ms for NO_X. To achieve very fast response, the detectors in remote sample heads are positioned very close to the sample point and use vacuum to convey the sample gas through narrow heated capillaries. The Fast NO_X analyzer provides NO_X concentration in parts per million (ppm).

Fast Particulate Sizer

Temporally resolved particulate concentrations are obtained using a differential mobility spectrometer (DMS) 500. This instrument measures the number of particles and their spectral weighting in 5 nm to 1000 nm size range with a time response of 200 ms. The DMS uses a corona discharge to place a prescribed charge on each particle. The charged particles are then carried along a classifier column by a sheath of clean air and attracted to one of the electrodes in the array depending on their size and aerodynamic drag. As the particles land on the grounded rings, they give up their charge and the outputs from the electrometers are processed in real time to provide spectral distribution. More details about the experimental setup are given in [7].

APPROACH

The emission (soot and NO_X) formation in diesel engine is highly nonlinear and displays complex dynamic behavior with change in operating condition. In addition, emission formation is not only a function of present operating condition but previous time history as well. A single model capable of capturing all the nonlinearities, particularly under highly dynamic operating conditions will invariably have complex structure and very high order. In addition, training such a model will pose numerical challenges. This paper proposes an approach with multiple local models to circumvent this difficulty. The engine operating space is partitioned into multiple smaller subspaces with each subspace having its own model.

The model is based on experimental data. A specially designed perturbation signal is used for generating training data and the models are later validated using a completely different data set. The flow chart in Figure 3 gives an overview of the algorithm presented in this paper with details in subsequent sections.

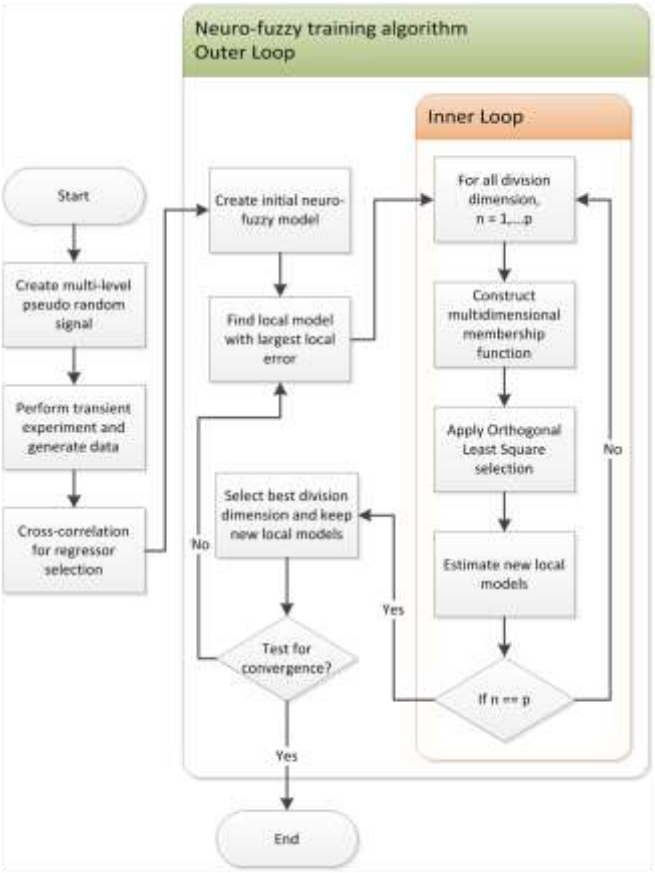

Figure 3: Algorithm for creating virtual emission sensors

PERTURBATION SIGNAL DESIGN

System identification (SYSID) of black box system requires application of perturbation signal for exciting system and gathering data. Choice of perturbation signal provides an upper bound on the accuracy of the black box model independent of the model structure and architecture. For nonlinear system identification, two aspects of perturbation signal are important. First, the signal must be persistently exciting and second, the signal should be rich enough to excite all frequencies and nonlinearities in the system. For a linear model, a binary random signal suffices, but it is not suitable for nonlinear system identification because it is not persistently exciting in amplitude. The selection of perturbation signal for nonlinear system identification requires more careful consideration. A multi-level Pseudo Random Signal (m-PRS) or Amplitude Modulated PRBS (APRBS) is applied in this paper for nonlinear system identification. m-PRS signals are periodic, deterministic, persistently exciting and have autocorrelation function similar to white noise. These characteristics make m-PRS signal well suited for this type of work. The theory behind generation of m-PRS is well developed [8], [9]. The m-PRS signal is generated using q-level shift registers (Figure 4) where q is prime or power of primes. Details about m-PRS signal generation for specific problem is given in [8].

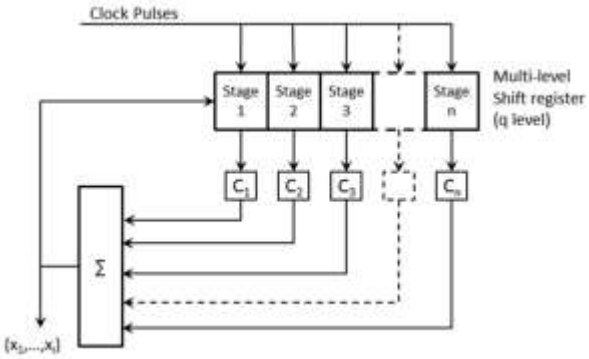

Figure 4: Schematic for generating m-level pseudo random signal

Modern diesel engine is highly complex system with nonlinear responses to action of multiple actuators. Furthermore, the diesel engine emission formation is a highly nonlinear process owing to complex mixing and chemistry. To create a transient diesel engine emission model valid over entire operating region, the test signal should excite all the engine operating frequencies. This will ensure that the training data are "rich". Using a priori information about the engine and preliminary tests like step and stair case excitation, information about bandwidth of system dynamics, dominant settling time, etc. is obtained [10], [11]. Figure 5 gives the result of one such staircase test performed on the engine. This information is used to create perturbation signal i.e. the signal with appropriate frequency range, switching time and amplitude level. The switching time for signal is short enough to prevent capturing predominantly steady state data but long enough to allow engine transients to fully develop before the next instance of signal is sent.

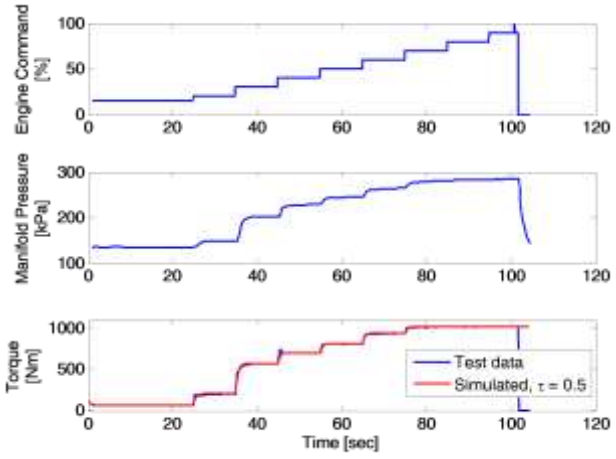

Figure 5: Staircase test @ 2000 RPM

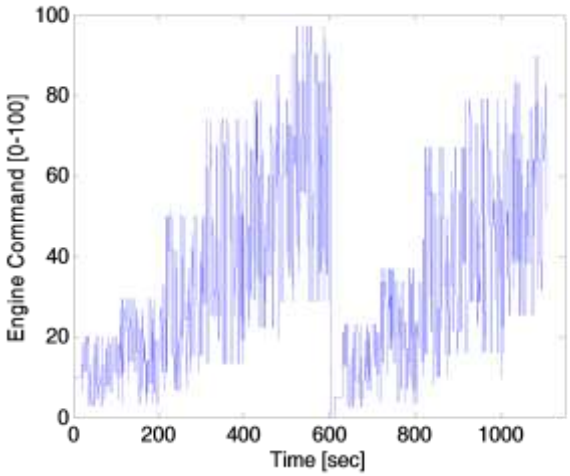

Figure 6: Throttle signal to engine for SYSID

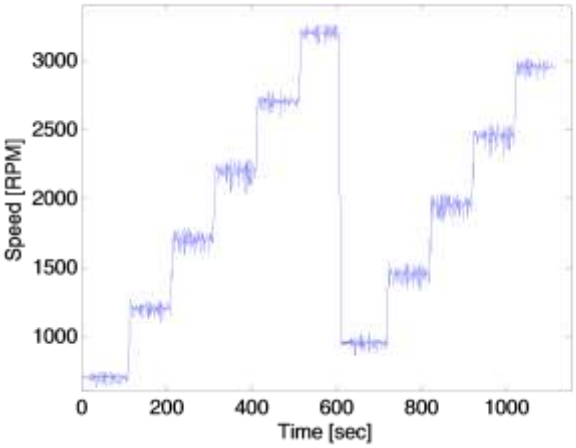

Figure 7: Speed signal to engine for SYSID

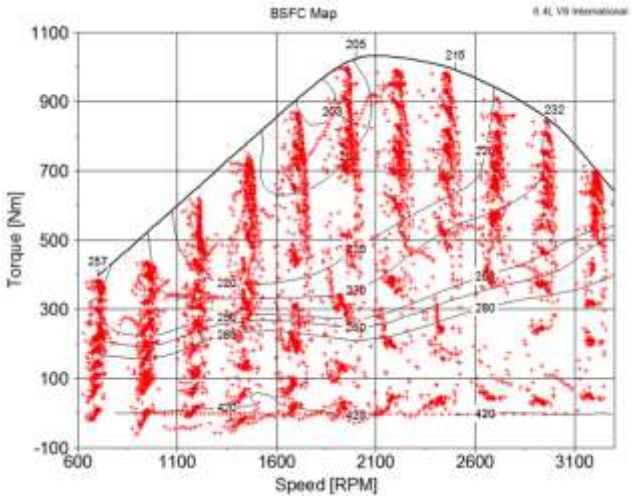

Figure 8: Engine visitation points during SYSID test

For this paper, an 11-step m-PRS signal is created and applied to engine as throttle (fueling) command. Figure 6 and Figure 7 shows the engine throttle and engine speed signal applied to engine. Figure 8 shows the engine operating points on 2-D engine operating space spanned by speed and torque. Similar plots can be generated with different axis, e.g. engine boost and EGR valve opening angle, with respect to other variables. Figure 8 shows that the whole engine operating space is completely covered and the obtained training data are "rich".

SELECTION OF INPUT REGRESSORS

A three-step process is employed in this paper for selection of input regressor set. First, a set of potential input signals is generated based on the knowledge of diesel engine combustion and emission formation like engine speed and fuel injection. Hagena's [12], [13] work gives an insight into potentially relevant signals for predicting transient emissions. Hagena et al. [13] showed that transient NO_X spikes are primarily dependent on lag between increased fueling and boost combined with EGR starvation, while the particulate transients are initiated by step change in fueling commands. The latter are highly dependent on rate of change in fueling and transient excursions of residual concentration. Important inclusions based on Hagena's work are the rate of change of fuel injection and post injection.

The above set of input variables are down selected based on ease of availability of signal. The ease of availability means the signal should be readily available either through onboard engine sensor or from ECU. Residual content in the fresh charge cannot be easily measured onboard engine and hence is not included in the set of input signals. However, EGR valve is actuated by ECU and the percentage opening of that valve is available. Hence, this signal is considered instead.

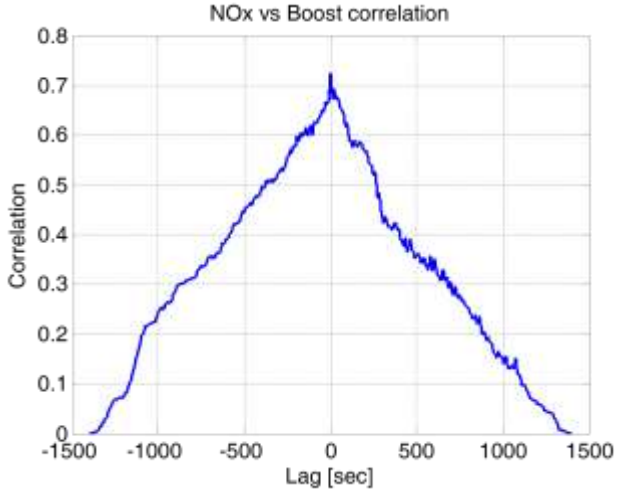

Figure 9: Cross-correlation of fast NO_X and boost

Finally, the available input signals are cross-correlated with soot and NO_X emissions. Cross-correlation is a sliding inner product of two signals and gives the measure of similarity of two signals with one signal time-shifted.

$$corr_{xy}(\kappa) = \frac{1}{N}\sum_{i=1}^{N-|\kappa|} x(i)y(i-\kappa) \qquad (1)$$

where *x(i)* is one of the selected input variable, *y(i)* is the emission with time instant *i* and κ is the time shift or "lag" between two signals.

The signals with very high correlation are included in the input data set. Figure 9 gives one such example of high cross-correlation of NO_X with input variable, boost. Cross-correlation also shows that emission formation is not only a function of present values, but depends on the time history of the input variable as well. Therefore, time shifted input signals are also considered as separate inputs.

NEURO-FUZZY MODEL TREE

The neuro-fuzzy model tree belongs to hierarchical class of models and is also known as Takagi-Sugeno fuzzy models. The underlying principle is a divide-and-conquer strategy, whereby, the operating space is subdivided into multiple smaller subspaces and individual submodels are used for identification. Hence, the complex problem is subdivided into multiple simpler problems, which are then identified using simpler models. The local models can be linear. The challenge lies in devising correct division and training strategy for local models.

The concept of multiple models to represent physical system has been independently developed in many fields like artificial intelligence and statistics with different names like operating regime based models [14], [15], [16] and piecewise local models [17]. The two level nested structure of the model combined with orthogonal least square routine makes the model developed in this paper different from previous approaches. The model has a soft partitioning with Gaussian validity function.

The models can be considered as an extension of radial basis networks with output neuron replaced with local functions. Hence, the validity function is weighted with their corresponding local functions [18]. The neuro-fuzzy model tree can be expressed as

$$y = \sum_{i=1}^{M} f\left(\bar{w}, \bar{u}\right) \cdot \Phi_i(\bar{u}) \tag{2}$$

i.e. the output of model is the weighted sum of all local sub models *f(.)*, and validity functions determine the regions of input space where that particular model is active.

$\Phi_i(\bar{u}) = \dfrac{\mu_i}{\sum_{j=1}^{M} \mu_j(\bar{u})}$ is the validity function

where,

$$\mu_j(\bar{u}) = \exp\left(-\frac{1}{2} \cdot \left(\frac{\left(u_1 - c_{i1}\right)^2}{\sigma_{i1}^2} \right) + \cdots + \left(\frac{\left(u_n - c_{in}\right)^2}{\sigma_{in}^2} \right) \right) \tag{3}$$

with center coordinates c_{ij} and dimension individual standard deviation σ_{ij} [18].

In this work, all the validity functions are Gaussian. The validity functions are normalized and add up to 1.

$$\sum_{i=1}^{M} \Phi_i(\bar{u}) = 1 \tag{4}$$

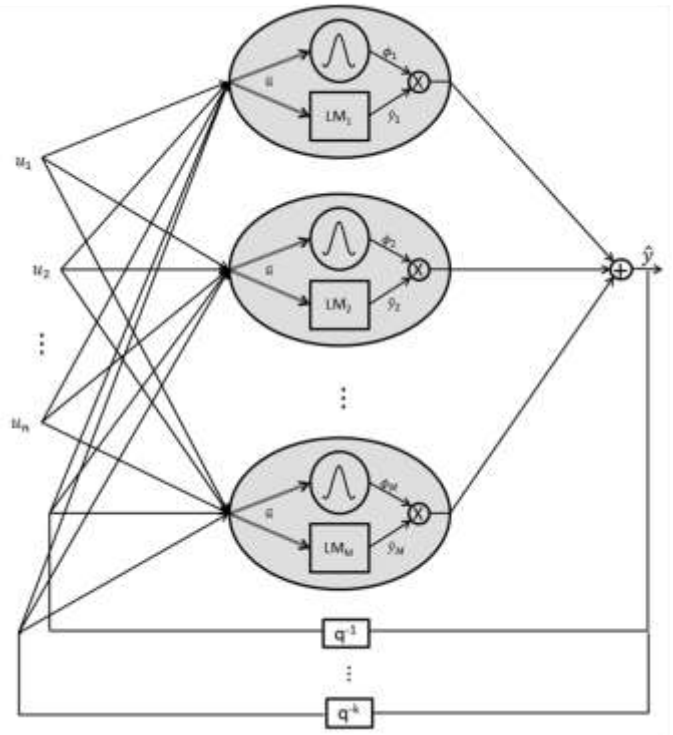

Figure 10: Network structure of a dynamic neuro-fuzzy model with *M* local models and *n* inputs with *k* tapped-delay output feedback for transient systems

The neuro-fuzzy model tree can be used to model transient systems by using external dynamics approach, Figure 10. External dynamics model consists of two parts [18]: a nonlinear static approximator and an external dynamic filter. The neuro-fuzzy model tree serves as a nonlinear static approximator, while a time delay feedback of output behaves as a dynamic filter. In other words, recurrent neuro-fuzzy model tree architecture is adopted to predict transient systems. The neuro-fuzzy model tree in Figure 10 has *M* local models, each with an associated validity function that determines the region of validity of local model.

The training algorithm is based on local linear model tree algorithm proposed by Nelles [18]. The algorithm has two loops. The outer loop optimizes the input space partitioning, given by the center and standard deviation of validity function while the inner loop calculates the optimal weights for the local model in least square sense.

The algorithm proceeds by partitioning the input space into hyperrectangles and then training a local model for every hyperrectangle space. The center of each hyperractangle holds the validity function with standard deviation based on the boundary of this space. A validity function with large standard deviation will have large influence area. Similarly, a validity function with small standard deviation will have extremely localized area of influence. The model tree grows by

partitioning the hyperrectangle space, which has the worst performing local model i.e. model with maximum local error, Eq. (5). The hyperrectangle space is then divided further into two smaller regions with individual local models. All the possible directions are evaluated for new division and the one that gives the maximum improvement is chosen for the new division dimension for hyperrectangle.

$$J_i = \sum_{j=1}^{N} \Phi_i(\bar{u}(j))(y(j) - \hat{y}(j))^2 \quad (5)$$

i.e. the local error is the sum squared error for all data, N weighted by the validity function.

Any optimization algorithm can be employed to calculate the optimal weights for the local models. A weighted least square is employed in this paper. The estimation can be carried out either globally or locally. Global estimation requires calculating weights for all the local models together and hence takes into consideration influence of overlapping validity function. It is more computationally intensive as the weights are refreshed for every local model with addition of newer local models. In contrast, local estimation optimizes the weights separately, neglecting the interaction with other local models. The local model estimation approaches global estimation with validity function standard deviation, $\sigma \rightarrow 0$, i.e. no interaction between local models. The model error increases with higher σ i.e. larger interaction between local models. Local estimation approach also lends itself to having different local model architectures and is preferred in this application.

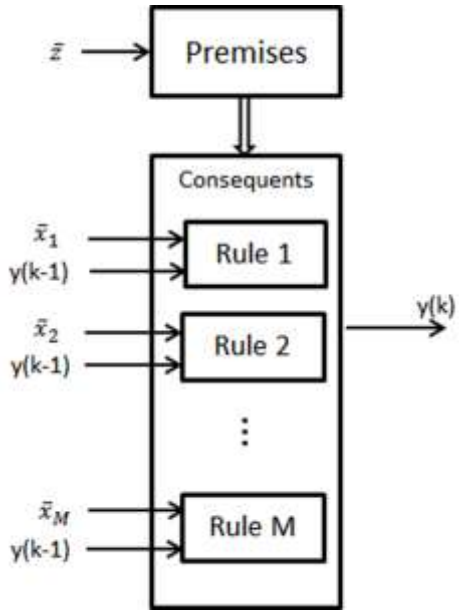

Figure 11: Construction of a neuro-fuzzy model

The training algorithm for neuro-fuzzy model tree is augmented with the OLS algorithm for automatic selection of input structure for local models from the available input vector. A brief description of OLS technique is given in the next section. To keep computational cost low, both the weight estimation for models and input selection using OLS are done locally. The OLS algorithm is nested in the inner loop of neuro-fuzzy model tree algorithm and executed before optimization of local model weights.

Figure 11 summarizes our neuro-fuzzy approach with OLS modeling strategy. The neuro-fuzzy model tree algorithm divides the input space based on rule premise, ***z***. The algorithm is able to extract variables in a rule premise vector ***z,*** which have a significant nonlinear influence on the process output and create partition rules. It subsequently trains the local model valid for a specific input region. The OLS algorithm helps in structure selection of local models by pruning the input variables. Each local model can have different set of input variables. The premise vector, ***z*** need not be similar to consequent vector, ***x***. The validity function (premises) depends on ***z*** and local models (consequents) depend on ***x***.

ORTHOGONAL LEAST SQUARE

OLS is a linear subset selection technique. Chen et al. [19] gave a detailed overview of algorithm and its application for nonlinear system identification. The standard OLS routines for local subset selection are modified in this paper, to include the weighting of data in local neuro-fuzzy models.

The OLS method involves the transformation of the set of input regressors x_i into set of orthogonal basis vectors and then calculating the individual contribution to the desired output variance from each basis vector.

$$\bar{y} = \bar{X}\bar{\theta} + e \quad (6)$$

where $\bar{X}$ is the regression matrix, $\bar{\theta}$ is the parameter vector and e is the error. The regression matrix is decomposed into $\bar{X} = \bar{V}\bar{W}$ with $\bar{W}$ being the triangular matrix and $\bar{V}$ being the matrix with orthogonal columns. The space spanned by set of orthogonal basis vectors v_i is the same as space spanned by the set of x_i. The model then can be rewritten as

$$\bar{y} = \bar{V}g + e \quad (7)$$

with $g = \bar{W}\bar{\theta}$ which can be derived by any orthogonalization method like Gram-Schmidt or Householder transformations. The output variance is then be given by

$$\bar{y}^T\bar{y} = \sum_{i=1}^{n} g_i^2 v_i^T v_i + e^T e \quad (8)$$

The output variance due to regressor v_i is given by term $g_i^2 v_i^T v_i$ in the Eq. (8). A regressor is important if this term is large or in other words, the error reduction ratio provides a criterion for subset selection.

$$[err]_i = \frac{g_i^2 v_i^T v_i}{\bar{y}^T\bar{y}} \quad (9)$$

The algorithm proceeds with transformation of original regressors into orthogonalized basis vectors. Then the regressor with maximum error reduction ratio is chosen and subtracted from original regressors. The remaining regressors are re-orthogonalized and the whole algorithm is repeated until desired number of regressors has been obtained or the unexplained error falls below a given threshold. The standard OLS algorithm computational demand grows with increase in potential regressors and size of data set. The paper employs a modified fast OLS algorithm [20] to greatly reduce computational effort.

SOOT AND NO_X VIRTUAL SENSOR

The soot and NO_X transient emission model in this paper is a two level neuro-fuzzy model tree. Figure 12 depicts the structure of the model. The local models for the top-level neuro-fuzzy model are also a neuro-fuzzy model tree. The second tier neuro-fuzzy model trees in turn have linear submodels.

The root node denotes the whole input space. The root node is partitioned by engine speed. The choice of this partition is based on the signal used to generate training data shown in Figure 7. The Figure 13 shows the soft partition of first level models based on engine speed with Gaussian validity function. The first level models are further partitioned into multiple local models (leaf nodes) using the neuro-fuzzy algorithm described earlier.

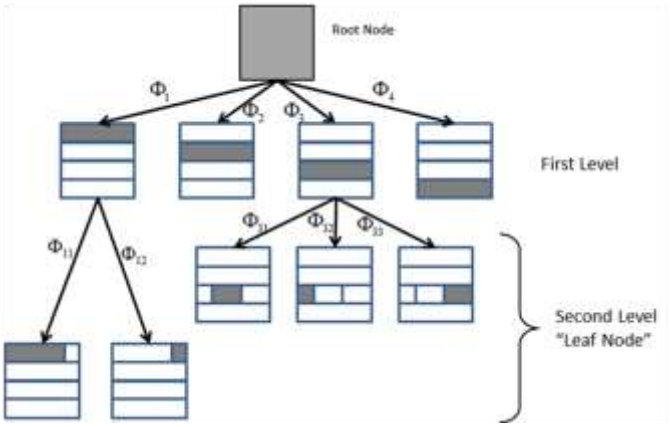

Figure 12: Hierarchical model structure of neuro-fuzzy model tree based emission sensors

The leaf nodes are partitioned in a 4-D hyperspace. The rule premise space is spanned by mass of fuel injected, boost, EGR valve angle and post fuel injection, whereas the rule consequent space includes engine speed, fuel injected per cylinder, boost pressure, EGR valve angle, rate of change of fuel injection, post fuel injection, their previous time histories and previous emission output. Table 2 and Table 3 give the list of included time histories for each signal. In order to introduce nonlinearity in input space, the second order multiplication of input data set is also used. Though the formulation now includes nonlinearities, the structure of submodel is still linear. Each input is preprocessed and normalized between 0 and 1. The normalization has the benefit of making the learning more numerically stable. The output is then anti-normalized to recover values in original range.

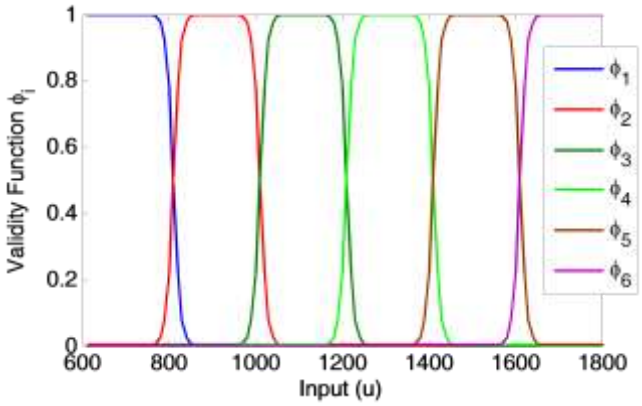

Figure 13: Soft model partition based on engine speed with Gaussian validity function

Table 2: Time delay in input signals considered for NO_X model

Input Signal	Time History (sec)
Speed (RPM)	0
Fuel Injected (mg/str)	0, 0.1, 0.2, 0.3, 0.4
Boost (bar)	0
Angle of EGR Valve (θ)	0, 0.1
Post Fuel Injection (mg/str)	0
Rate of Fuel Injection	0

Table 3: Time delay in input signals considered for soot model

Input Signal	Time History (sec)
Speed (RPM)	0
Fuel Injected (mg/str)	0, 0.1, 0.2, 0.3, 0.4, 0.5
Boost (bar)	0
Angle of EGR Valve (θ)	0, 0.1, 0.2
Post Fuel Injection (mg/str)	0, 0.4, 0.5
Rate of Fuel Injection	0, 0.1, 0.2, 0.3, 0.4, 0.5

The local model can be represented as

$$y(k) = \bar{w} \cdot \bar{u}^T \tag{10}$$

where $\bar{w}$ is the weight vector and $\bar{u}$ is the set of inputs. The set $\bar{u}$ is different for each local model and contains only the most relevant inputs identified by OLS.

The overall model output is calculated by summing the contributions of every local model at leaf nodes, weighted with their validity function values. The validity functions pass their contribution to the next higher node (parent).

$$y = \Phi_1\left(\sum_i \Phi_{1i} f_{1i}(\bar{w},\bar{u})\right) + \cdots + \Phi_m\left(\sum_i \Phi_{mi} f_{mi}(\bar{w},\bar{u})\right) = \sum_j^m \Phi_j\left(\sum_i \Phi_{ji} f_{ji}(\bar{w},\bar{u})\right) \quad (11)$$

where Φ_{ji} is the validity function of leaf node local models and $y_{ji} = f_{ji}(\bar{w},\bar{u})$ are the outputs of local models.

The relevant time lag for some input signals change with engine speed. OLS is capable of figuring out the time delay automatically and includes only the relevant time delayed signals. This helps in keeping the number of input regressors small, making the model more efficient and robust.

The soot and NO_X emission model are trained using the data generated during testing. The virtual sensors are then coded in C++ and then cross-compiled and cross-linked for the dSPACE real-time platform. The virtual sensor interfaces with PUMA Open and receives engine speed, engine boost, in-cylinder injected fuel, EGR valve angle and post fuel injection at 10Hz, and predicts instantaneous soot and NO_X emission.

The test cell allows for concurrent running of physical engine with virtual driveline and vehicle over different driving schedules [7], [21]. The engine is coupled to the virtual series hydraulic hybrid driveline, and the engine-out emissions are recorded over the FTP75 driving schedule for validation. This ensures that validation and training data sets are completely different. Figure 14 and Figure 15 shows the instantaneous predicted and measured soot and NO_X emissions respectively, along with steady state model predictions. The predicted emissions show a good match with soot and NO_X measurements from fast analyzers. The model prediction over transient engine operation is far superior to steady state model predictions for soot emissions, as seen in Figure 14. The ability of the virtual sensor to capture frequent transient spikes of soot emission is particularly relevant since steady-state model significantly underpredicts concentrations during dynamic engine operation. Looking at the cumulative emissions, the steady state model under predicts soot by 60% and over predicts NO_X by 21%. The proposed emission virtual sensor provides an order of magnitude improvement, since the soot predictions are only 5% higher and NO_X 0.2% higher compared to measurements.

To explain the deficiencies of steady state models to predict transients correctly, we need to look closely into engine operation during transients. Consider the time interval around 35 sec. The engine command changes nearly instantaneously and the fuel injected follows the demand. The intake manifold pressure lags due to turbocharger inertia and the delay in boost pressure results in lower in-cylinder air-to-fuel ratio. The EGR command also changes to zero, however, residual gas dynamics have slower time scales and it takes time to purge the intake manifold. In addition, step change of load results in increased exhaust backpressure to inlet manifold pressure thereby increasing the internal residual [1], [13]. The presence of residual helps in reduction of NO_X but results in higher soot production. The combined effect of high instantaneous values of Fuel/Air ratio at the onset of the load transient [1], [13] and increased residual lead to sharp spikes of particulate concentration. The steady state emission model is only function of engine load and hence, cannot capture the transient effects. The transient model, on the other hand, can capture the effect of turbocharger inertia and EGR valve actuator dynamics on in-cylinder constituents, thereby giving superior predictions of resulting emissions.

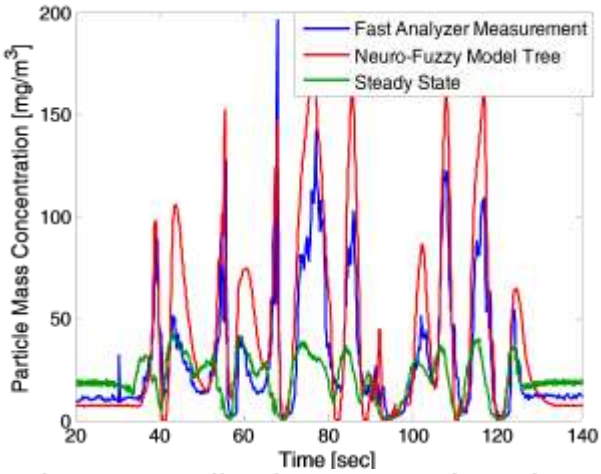

Figure 14: Predicted vs. measured transient soot emission for a series hydraulic hybrid over FTP75

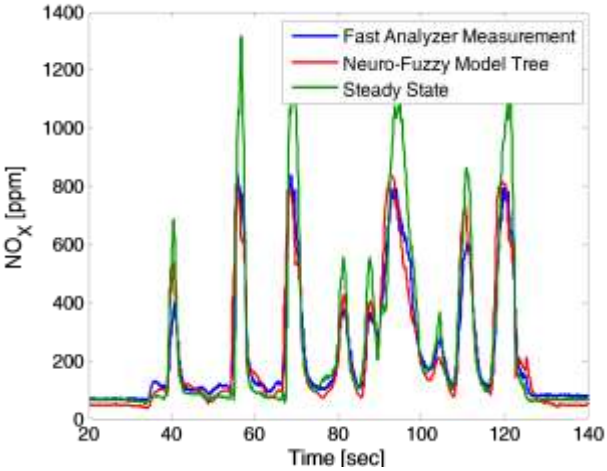

Figure 15: Predicted vs. measured transient NO_X emission for a series hydraulic hybrid over FTP75

CONCLUSIONS

The paper proposes modeling neuro-fuzzy based transient emission models for particulate matter and NO_X in a diesel engine. They accurately capture the transient dynamics of soot and NO_X emissions unlike steady state models. In particular, the new model is able to capture extreme spikes of soot emissions occurring at the onset of rapid load increases. The model is intended to run on a microprocessor in real-time and

predict engine-out soot and NO_X emissions using signals from the ECU and low-cost physical sensors. The key aspects of this modeling work are:

- Modeling relies on dividing the input space into smaller subspaces and fitting local models.
- Recurrent architecture of the model allows for capturing transient characteristics.
- Selection technique, orthogonal least squares is applied for selecting the structure of local model inputs.
- Multi-level pseudo random perturbation signal is designed specifically for characterizing the diesel engine transients.
- Virtual sensors are fast and capable of running in real-time along with the real engine.

The training and validation data is obtained from the experimental setup at the University of Michigan for transient testing of a medium duty diesel engine in the loop with virtual vehicle. Comparison of the predictions with transient measurements demonstrates very good agreement.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial and technical support by the Bosch-Rexroth Corporation, Engineering Hydraulics - North America.

REFERENCES

[1] Filipi, Z. , Hagena, J. , and Fathy, H. , 2008, "Investigating the impact of in-vehicle transients on diesel soot emissions," Thermal Science, **12**(1), pp. 53-72.

[2] Kirchen, P. , Obrecht, P. , and Boulouchos, K. , 2009, "Soot Emission Measurements and Validation of a Mean Value Soot Model for Common-Rail Diesel Engines during Transient Operation," SAE Int. J. Engines, **2**(2009-01-1904), pp. 1663-1678.

[3] Bayer, J. and Foster, D. E., 2003, "Zero-Dimensional Soot Modeling," *SAE Technical Paper 2003-01-1070*.

[4] Brahma, I. , Rutland, C. J., Foster, D. E., and He, Y. , 2005, "New Approach to System Level Soot Modeling," *SAE Technical Paper 2005-01-1122*.

[5] Dec, J. E., 1997, "A Conceptual Model of DI Diesel Combustion Based on Laser-Sheet Imaging*," SAE Transactions, Journal of Engines, **106**(970873).

[6] Galindo, J. , Bermudez, V. , Serrano, J.R. , and Lopez, J. J., 2001, "Cycle-to-cycle diesel combustion characterization during engine transient operation," SAE Transactions, Journal of Engines, **110**(2001-01-3262).

[7] Filipi, Z. , Fathy, H. , Hagena, J. , Knafl, A. , Ahlawat, R. , Liu, J. , Jung, D. , Assanis, D. , Peng, H. , and Stein, J. , 2006, "Engine-in-the-Loop Testing for Evaluating Hybrid Propulsion Concepts and Transient Emissions – HMMWV Case Study," SAE Transactions, Journal of Commercial Vehicles, **115**(2006-01-0443).

[8] Godfrey, K. , 1993, *Perturbation signals for system identification*, Prentice Hall.

[9] Braun, M.W. , Rivera, D.E. , Stenman, A. , Foslien, W. , and Hrenya, C. , 1999, "Multi-level pseudo-random signal design and model-on-demand estimation applied to nonlinear identification of a RTP wafer reactor," *Proceedings of the American Control Conference*, **3**, pp. 1573-7.

[10] Ljung, L. , 1987, *System identification : theory for the user*, Prentice-Hall.

[11] Zhu, Y. , 2001, *Multivariable system identification for process control*, Pergamon.

[12] Hagena, J. , 2008, "An experimental technique for determining cycle-resolved pre-combustion in in-cylinder composition and its application towards the understanding of diesel engine emissions during transient operation," PhD thesis Mechanical Engineering, University of Michigan,.

[13] Hagena, J. R., Filipi, Z. S., and Assanis, D. N., 2006, "Transient Diesel Emissions: Analysis of Engine Operation During a Tip-In," *SAE Technical Paper 2006-01-1151*.

[14] Johansen, T. A. and Foss, B. A., 1995, "Identification of non-linear system structure and parameters using regime decomposition," Automatica, **31**(2), pp. 321-326.

[15] Murray-Smith, R. and Johansen, T.Arne , 1995, "Local learning in Local Model Networks," *IEE Conference on Artificial Neural Networks*, pp. 40-46.

[16] Johansen, T. A. and Foss, B. A., 1997, "Operating regime based process modeling and identification," Computers & Chemical Engineering, **21**(2), pp. 159-176.

[17] Ragot, J. , Mourot, G. , and Maquin, D. , 2003, "Parameter estimation of switching piecewise linear system," *Proceedings of 42nd IEEE Conference Decision and Control*, **6**, pp. 5783-5788.

[18] Nelles, O. , 2001, *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*, Springer.

[19] Chen, S. , Billings, S.A. , and Luo, W. , 1989, "Orthogonal least squares methods and their application to non-linear system identification," International Journal of Control, **50**(5), pp. 1873-1896.

[20] Chen, S. and Wigger, J. , 1995, "Fast orthogonal least squares algorithm for efficient subset model selection," IEEE Transactions on Signal Processing, **43**(7), pp. 1713-1715.

[21] Filipi, Z and Kim, Y J, 2010, "Hydraulic Hybrid Propulsion for Heavy Vehicles: Combining the Simulation and Engine-In-the-Loop Techniques to Maximize the Fuel Economy and Emission Benefits," Oil & Gas Science and Technology - Revue de l'Institut Francais du Petrole, **65**(1), pp. 155-178.

