Published 03/28/2017
Copyright © 2017 SAE International
doi:10.4271/2017-01-0583
saecomveh.saejournals.org

Accurate Cycle Predictions and Calibration Optimization Using a Two-Stage Global Dynamic Model

Farraen Mohd Azmin, Phil Mortimer, and Justin Seabrook
Ricardo UK Ltd.

ABSTRACT

With the introduction in Europe of drive cycles such as RDE and WLTC, transient emissions prediction is more challenging than before for passenger car applications. Transient predictions are used in the calibration optimization process to determine the cumulative cycle emissions for the purpose of meeting objectives and constraints. Predicting emissions such as soot accurately is the most difficult area, because soot emissions rise very steeply during certain transients.

The method described in this paper is an evolution of prediction using a steady state global model. A dynamic model can provide the instantaneous prediction of boost and EGR that a static model cannot. Meanwhile, a static model is more accurate for steady state engine emissions. Combining these two model types allows more accurate prediction of emissions against time. A global dynamic model combines a dynamic model of the engine air path with a static DoE (Design of Experiment) emission model. The dynamic model is constructed using a Volterra series model for the EGR response and a Stochastic Process Model (SPM) for boost pressure. Both models are trained using data collected from APRBS (Amplitude Modulated Pseudo Random Binary Sequences) type tests. The static model is an SPM trained using data collected in a steady state DoE test.

The output of the global dynamic model is an accurate prediction of engine emissions with a drive cycle and calibration as model inputs. The global dynamic model is called during the calibration optimization process and the cycle cumulative results are used to control the constraints and optimize the objectives. This produces final calibration maps ready for immediate vehicle tests without test bed validation, thereby improving the efficiency of the calibration process.

CITATION: Mohd Azmin, F., Mortimer, P., and Seabrook, J., "Accurate Cycle Predictions and Calibration Optimization Using a Two-Stage Global Dynamic Model," *SAE Int. J. Commer. Veh.* 10(1):2017, doi:10.4271/2017-01-0583.

INTRODUCTION

Diesel engine calibration requires a powerful optimization process to find optimum engine settings. An optimizer takes into account engine constraints, emission limits and other objectives including map smoothness. The optimization process gets more complex when calibrating an engine with multiple models, for example cold, warm and hot static models to account for engine warm up. Calibration maps are usually optimized against an engine cycle to ensure the engine produces emission levels that meet the regulatory requirements.

One of the limitations in conventional diesel calibration is transient soot prediction. Typically, soot is estimated using a static model trained from a steady state DoE. This method often underestimates soot and may miss some significant soot "spikes" on the transient cycle. An alternative approach is to combine a dynamic air system model with a static soot model. The dynamic model, which can be a physical or empirical model, gives a more accurate air path prediction transiently, which in turn leads to more accurate transient prediction from the static soot model. Fig 1 shows the different approaches for soot predictions.

Using models for a 4-cylinder light duty diesel engine, this paper discusses the implementation of an empirical air system model and compares it with other global modelling approaches.

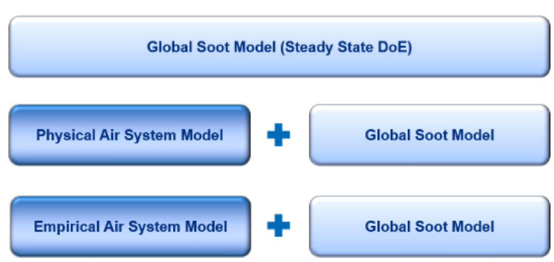

Figure 1. Approaches to transient soot prediction

OPTIMIZATION ROUTINE

Calibration optimization is an automated iterative process of adjusting calibration maps to get engine settings that meet optimization objectives and constraints. Fig 2 shows a typical optimization routine for calibrating diesel engine maps. Engine models are important in calibration optimization because they provide engine response prediction values as well as the cumulative emission values. It is obviously important that the predictive values are close to actual values to ensure the optimizer is converging to correct solution. Engine inputs and current calibration settings are fed into the engine model to get cumulative emission values, which are handled as constraints. This ensures the optimum settings meet the required emission levels. A map smoothness constraint is used to limit the gradient and curvature of the final calibration maps. Smooth transition between calibration points is important to prevent sudden changes to engine input variables which are detrimental to drivability and emissions robustness. Meanwhile, map monotonicity constraints may be included to avoid undulations with respect to speed or torque in, for example, the rail pressure map.

The intention of this research is to produce more accurate emission prediction while retaining the established calibration process. To achieve this, a dynamic model is incorporated in the optimization routine to predict instantaneous boost and EGR rate. Fig 3 shows the implementation of a dynamic model in the optimization routine. The objective with this new setup, is more accurate cumulative cycle predictions.

Besides model accuracy, prediction time also is also important when applying a dynamic model because the optimization process can take a significant amount of time to converge to a solution that satisfies all constraints.

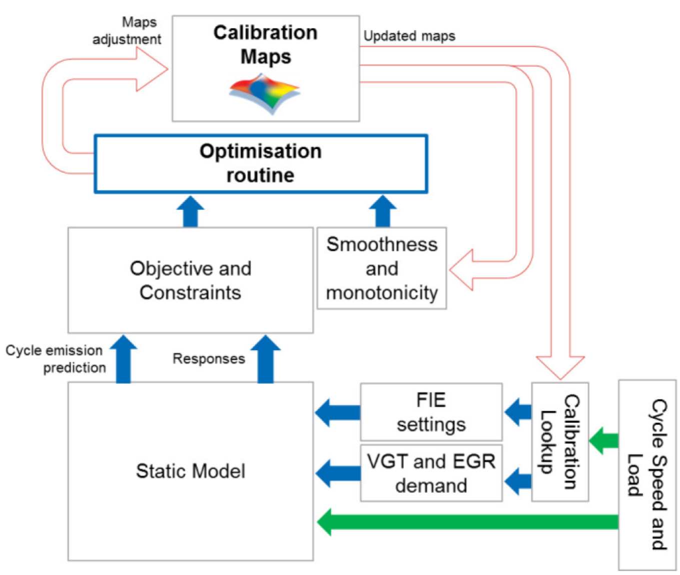

Figure 2. Optimization routine of calibrating diesel engine maps

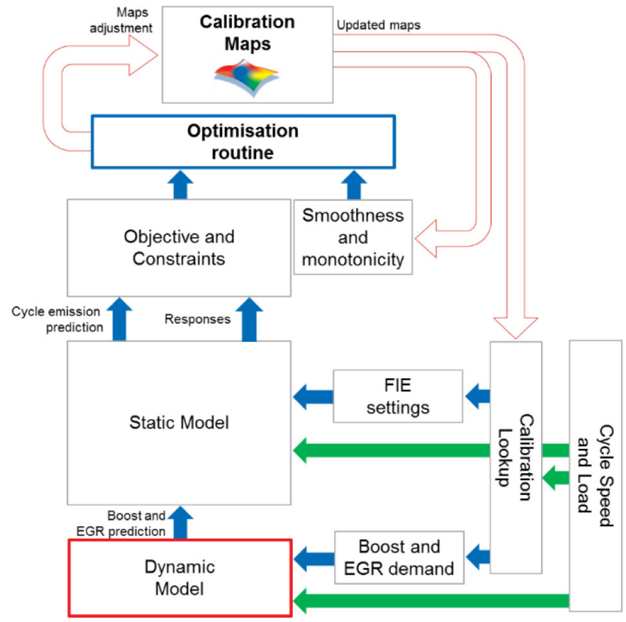

Figure 3. Updated optimization routine with dynamic model

METHODS OF CYCLE PREDICTION

Three methods of transient cycle prediction for diesel engine were introduced in Fig 1:

1. Static model
2. Static model + Physical air system model
3. Static model + Empirical dynamic air system model

In each case, the static model is a Stochastic Process Model (SPM) trained using steady state DoE data. The steady state experimental design had points spread across the entire operating region [1]. At each DoE point, the engine input variables were set and held until the engine condition was stable enough to make a measurement. This data was then used to train the SPM. Fig 4 shows the inputs/outputs of the model. This method predicts time series emissions by feeding the cycle and calibration to a static model (SPM1).

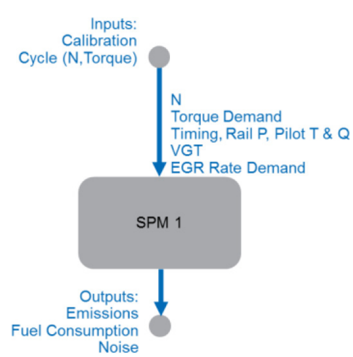

Figure 4. Schematic of SPM1 static model

The second case is not covered in detail in this paper, but results are included later for comparative purposes. The physical air system model is a WAVE-RT [2] model. WAVE-RT is a real-time 1D gas dynamics simulation package. WAVE-RT is coupled with the static SPM. In fact, two variants of the static model are employed. The DoE test, which featured wide variable ranges including boost and EGR levels only observed during transients, was used to create two models: SPM1 &; SPM2. SPM1 has the same inputs as in the first case of a static model alone. SPM2 is similar to SPM1, except it uses boost pressure instead of VGT duty as one of its 8 inputs. SPM1 is used to predict the heat release data needed for WAVE-RT and perform certain plausibility checks on the output. The physical WAVE-RT model predicts dynamic boost and EGR rate which are in turn fed into the SPM2 emissions model.

The third approach uses an empirical dynamic model to predict the engine air path. The air path prediction with more accurate dynamics is fed into a static model to get better emission predictions. This method is discussed in more detail in the next section.

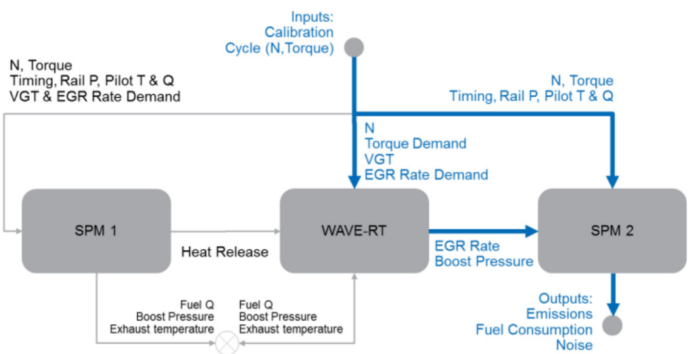

Figure 5. WAVE-RT with SPM2 static model predicting emissions

DYNAMIC MODEL

Dynamic DoE

A dynamic model must be trained with test data. Dynamic models are typically trained with quasi-random data recorded on the test bed or vehicle. Steady state DoE data cannot be used for dynamic modelling because it does not capture the dynamic behavior of a system. Random steps are needed to capture the lagged system outputs. However, DoE methods can be used to design the dynamic experiment.

An APRBS (Amplitude Modulated Pseudo Random Binary Sequence) is a random step sequence that stochastically covers a wide range of amplitudes [3,4]. For engine tests, this imposes severe demands on the dynamometer control system and hardware. Therefore, an alternative sequence based on a sinusoidal rather than square wave was utilized. This might be termed an APRSS with Sinusoidal substituted for Binary in the acronym. Earlier in-house studies have shown the models from sinusoidal data were generally better than pure binary sequences for model quality as well as ease of testing.

Fig 6 shows a section of the APRSS generated for dynamic DoE inputs. The APRSS is constructed using LHS (Latin Hypercube Sampling) and used to train the dynamic model. The APRSS was designed based on the engine operating envelope and covers the WLTC operating region (Fig 7).

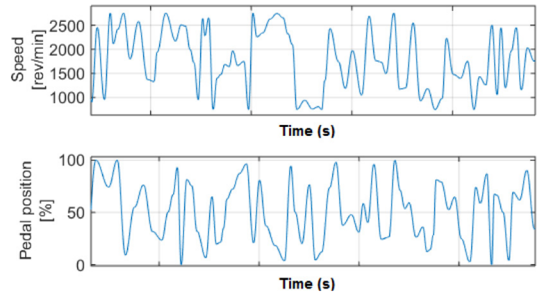

Figure 6. Dynamic DoE Sequence (Sinusoidal APRBS)

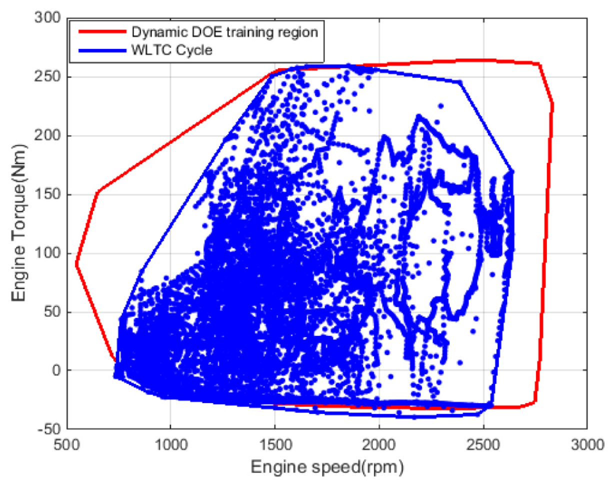

Figure 7. Dynamic DoE sequence region & WLTC points

Model Structure

The model structure contains a function kernel and model constants which must be parameterized. The model structure complexity depends on the system to be modelled. A model kernel can be anything from a polynomial model to a stochastic process model. Model types are selected based on the problem. For this diesel engine calibration experiment, SPM and Parametric Volterra Models gave the best performance.

An SPM was used for boost prediction. SPMs are parameterized using Maximum Likelihood Estimation (MLE). MLE estimates the parameters by finding the parameter values that maximize the likelihood of the sample given the model. The sample in this case is the dynamic DoE test data. The general form of an SPM is:

$$y(x) = \mu + Z(x) \tag{1}$$

Where μ is a weighted mean of the response, y, and Z(x) is a spatial process with mean 0, variance σ^2 and correlation structure, R [5]:

$$R(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(\sum_{j=1}^{k} -\alpha_j \left| x_{1j} - x_{2j} \right|^{\nu_j} \right) \tag{2}$$

Where x_1 and x_2 are points in the test matrix, X, k is the number of variables and α and ν are model parameters.

A Parametric Volterra Series Model is an expanded Taylor series model. The model can be made into a Volterra kernel and applied into a NARX model structure which has the ability to capture past values. A Volterra model without memory has the following form [6]:

$$y(t) = a_0 + \sum a_n x_n^k(t) + \sum b_{i,j} x_n^k(t-i) x_m^l(t-j) + \sum c_i\, y(t-i) \tag{2}$$

Where:

a_0 = constant term

$\sum a_n x_n^k(t)$ = static terms

$\sum b_{i,j} x_n^k(t-i) x_m^l(t-j)$ = cross term

$\sum c_i y(t - i)$ = the impulse term

Both models mentioned above can be expanded into a NARX structure. In a NARX structure, the system input is expended to contain the memory effects of inputs and outputs [7]. The memory effects are used to characterize the lagged behavior of the system. A MISO (Multiple Input Single Output) NARX can be described as follows [8]:

$$y(t) = f\big(x(t), \dots, x(t-n), y(t-1), \dots, y(t-n)\big) \tag{3}$$

Where:

y = model output

x(t-n) = delayed inputs

y(t-n) = delayed output

f = model function (SPM/Volterra)

The model structure contains a fixed number of delay terms. The delay terms are selected based on the characteristic of the system. The manual approach of selecting the delay is a difficult process. Too short a delay and the critical lagged responses are not captured. Several automatic methods were available and have been further developed by the authors.

The methods available to determine the delays are:

1. Maximum likelihood estimation
2. Sequential feature selection (built-in Matlab function)
3. PCA (Principal component analysis)

In this work, the model structure was determined using the maximum likelihood mapping method. The same model structure has since been used for other projects. This method does have a high computational overhead, but it produces the best model compared to the others.

Model Training

The static model uses steady state DoE test data for training. Inputs to the model are engine speed and load, VGT position and EGR demand plus fuel-related calibration variables (timing, rail pressure, pilot timing and quantity). The model is trained to predict engine emissions, fuel consumption and combustion noise. The training process uses a gradient descent optimization process to calibrate the model parameters. Fig 8 shows the inputs and outputs of a static model training process.

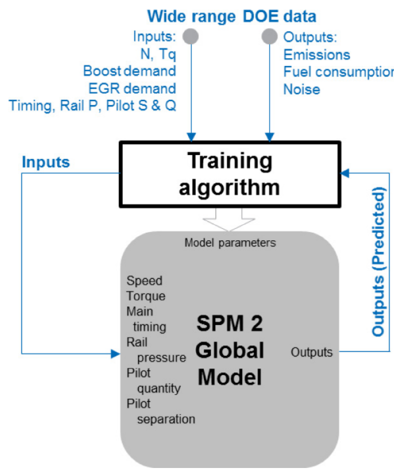

Figure 8. Static engine model training

The dynamic model is trained using dynamic DoE data. The inputs of the dynamic model require engine operating variables from the calibration data. The model is trained against the feedback values. Fig 9 shows the general flow of dynamic model training. Table 1 shows the inputs and outputs of the dynamic model and the kernel used for the response prediction.

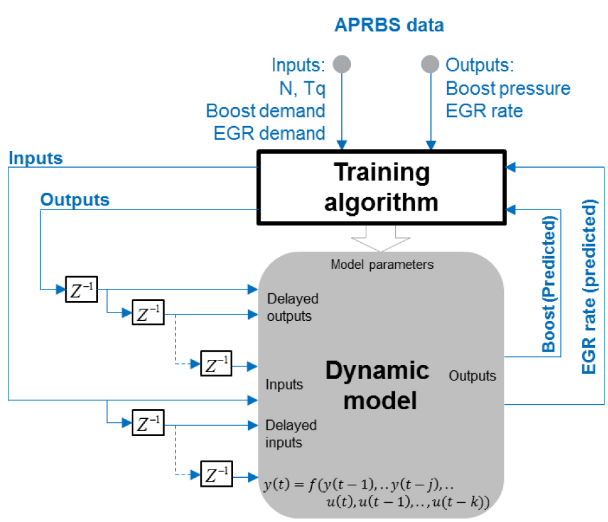

Figure 9. General view of dynamic model training

Table 1. Dynamic model inputs and outputs

	Boost prediction	EGR prediction
Inputs	Engine speed Brake torque Boost demand	Engine speed Brake torque EGR rate demand
Outputs	Instantaneous boost	Instantaneous EGR
Model kernel	SPM	Volterra Series

For boost pressure prediction, the model was trained using a fast-SPM process for boost prediction. Fast-SPM is a compact version of the SPM process that uses a fraction of the points from the full dataset. These points are extracted using a Euclidean distance method. Figure 10 shows the model parameter optimization using fast-SPM kernel.

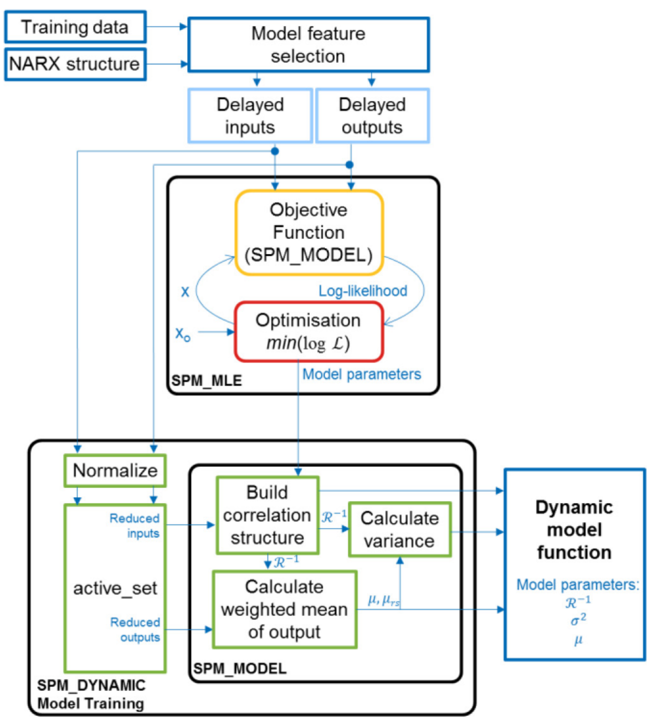

Figure 10. Dynamic model training using SPM kernel

This method shares the same modelling functions as the static method. The model parameters are then used to calculate the correlation matrix, R. μ and σ are calculated from the reduced test data and R. These model parameters are then used to construct the model function later during model prediction.

A Parametric Volterra Series model kernel was used for EGR prediction. Figure 11 shows the model parameter optimization process using a Volterra series kernel. This method uses in-house functions to calculate the regressor using the static and interaction polynomial terms. The regressor is an array of calculated polynomial terms based on provided delayed inputs. The regressor is fed into an objective function which calculates the model error. A gradient descent optimization process finds the polynomial constants for each static, cross and impulse terms for the parametric Volterra model.

The optimization objective is to find the model parameters which give the lowest model error.

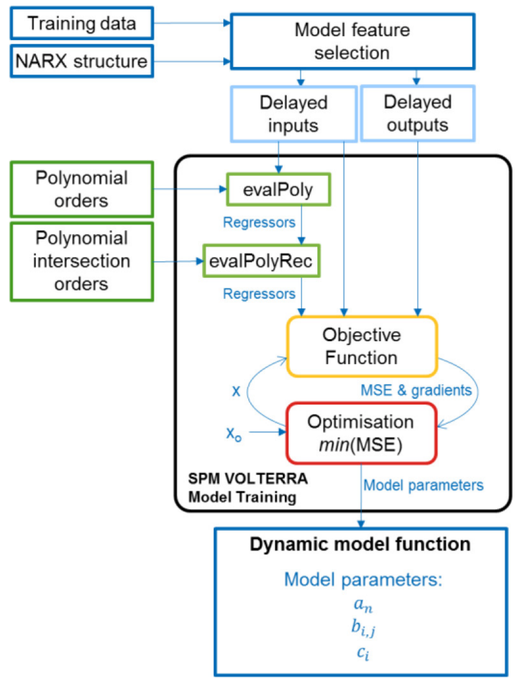

Figure 11. Dynamic model training using SPM Volterra kernel

The boost model uses an SPM model kernel and the EGR rate model uses a Volterra series model. The models are configured this way because the EGR model needs high-frequency data due to the highly nonlinear system response. The Volterra model is good at handling high-frequency data and can model a faster system. For the boost model, the response can be modelled at a lower frequency because it has fewer variations compared to EGR, and the SPM provides the more accurate boost model.

R^2 is used to evaluate the model quality. R^2 is the measure of the proportion of change in response that is accounted for by the changes in input variables. However, relying on model R^2 is not a robust way to determine the predictive capability of a model. Instead, the model is validated with a WTLC cycle. The models predict the boost and EGR rate for the cycle, as shown Fig 12.

The model validation of boost and EGR rates shows reasonably good results. The models are able to track the boost and EGR well with good R^2 values. Figs 13 and 14 shows the correlation plots for the boost and EGR rate models. The plots highlight points that do not lie within three standard deviation of the mean.

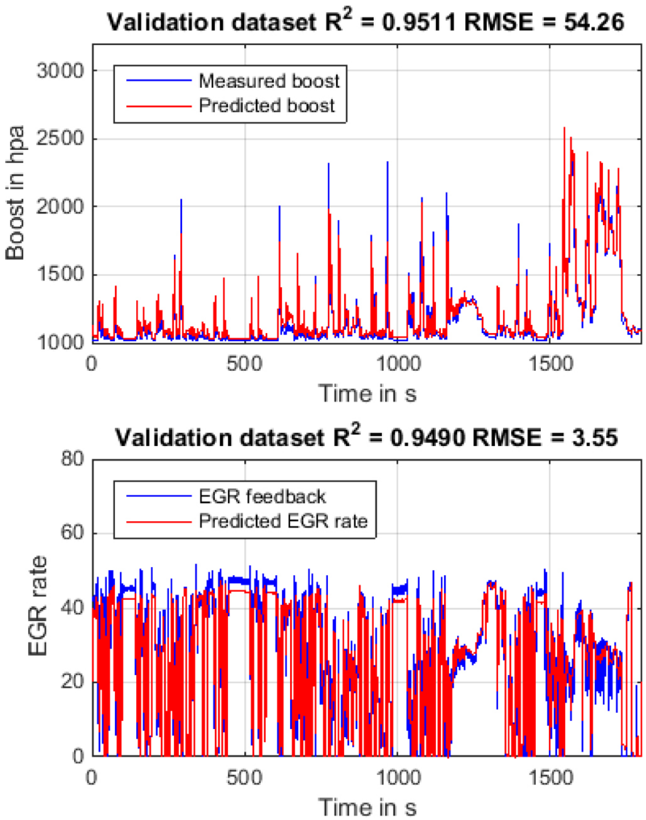

Figure 12. WLTC dynamic model validation results

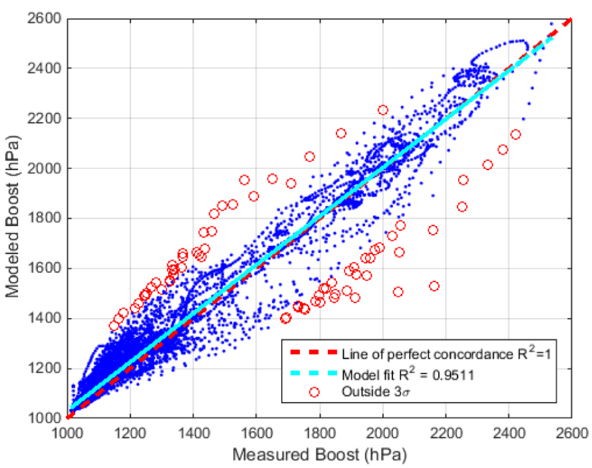

Figure 13. Boost model correlation plot

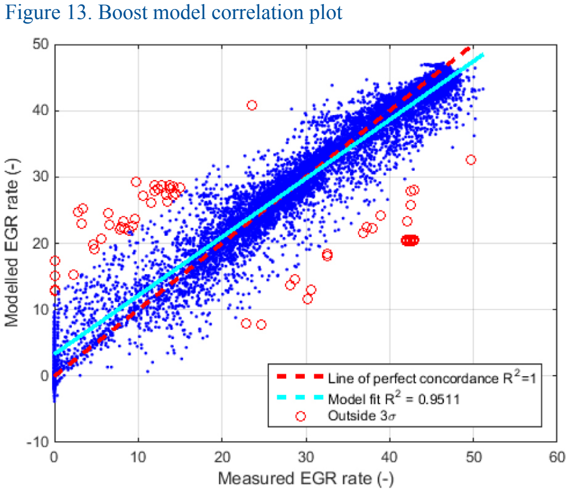

Figure 14. EGR model correlation plot

Two-Stage Global Model

For prediction, the dynamic and static models are combined to form a 2-stage global model as shown in Fig 15. The dynamic model is fed with test cycle inputs together with the delayed inputs/outputs. The predicted boost and EGR from the dynamic model provide an instantaneous prediction of the air path. This is then fed into the global model which, in turn, predicts the emissions.

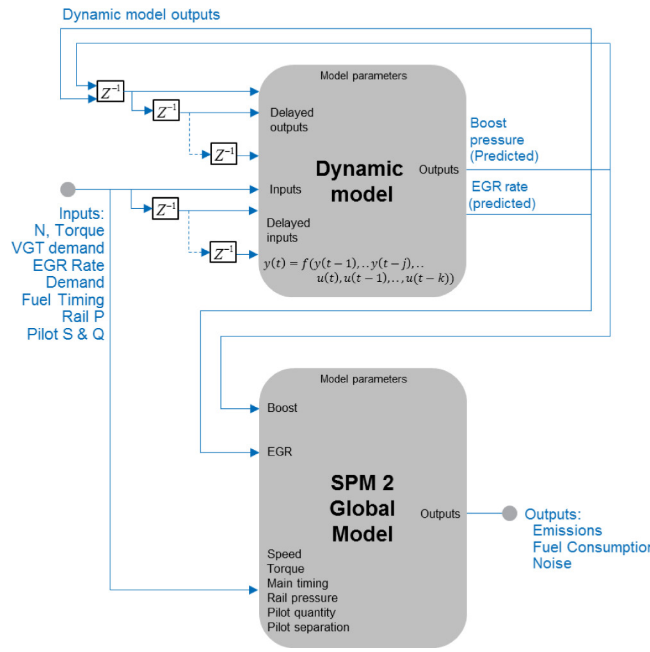

Figure 15. Prediction mode in two-stage global model

PREDICTION

Fig 16 shows the tracking of boost and EGR prediction compared to WLTC measurements. The predicted boost and EGR track well with the actual measurements. The dynamic model captures the lagged characteristics well compared to the steady state model.

Figs 17 and 18 show the Soot and NOx for all approaches. The two-stage global model managed to capture soot spikes missed by the SPM1 model. The dynamic model improved the model prediction because it managed to match air path dynamics closely to actual measurements. The NOx prediction is likewise improved.

Table 2 shows the cumulative results and the error between test and predicted emissions value. RMSE or Root Mean Squared Error describes the standard deviation of the model and validation data [9]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2} \tag{4}$$

Where:

n = number of observations

y_i = target output

x_i = input data

f = model function

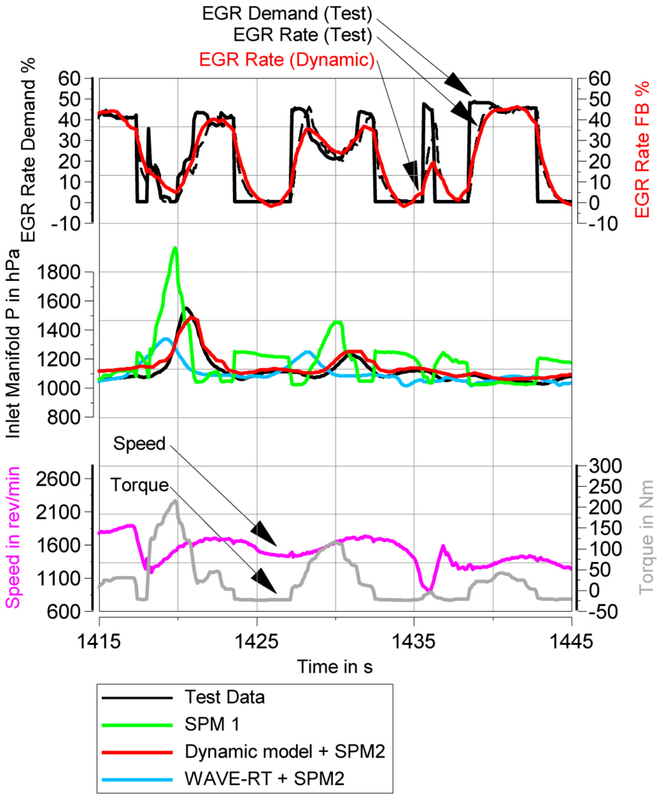

Figure 16. WLTC Boost and EGR tracking

For some responses over some cycles, the SPM1 method does not have the required accuracy since it relies on the steady state DoE data. The actual engine air path contains lagged outputs when a step input is fed to it. The SPM1 model lacks the ability to include the dynamic behavior in its prediction. In most cases, this model underestimates tip-in soot.

Cumulatively, the WAVE-RT+SPM2 model predicts soot better than SPM1 alone. This approach produces more accurate emission predictions. However, there is still some tip-in soot missed in its prediction. Judging a model by its cumulative value alone can be misleading. Looking closely at the RMSE column, it shows WAVE-RT+SPM2 has a high model error. The apparent contradiction between cycle accuracy and RMSE is because false or inaccurate predicted spikes can contribute to the cumulative values. Equally, small time alignment issues can inflate the RMSE, so it is important to consider both cumulative and instantaneous statistics. From the table, the dynamic+SPM2 offers best combination of cumulative and instantaneous model error.

Table 2. Cycle cumulative comparison

Model type	Cumulative (mg/km)		Model RMSE (g/h)	
	Soot	NOx	Soot	NOx
Test	60.3	304	-	-
SPM1	14.8	356	9.28	24.7
Wave-RT +SPM2	51.4	294	18.71	16.3
Dynamic+SPM2	49.4	289	8.01	16.0

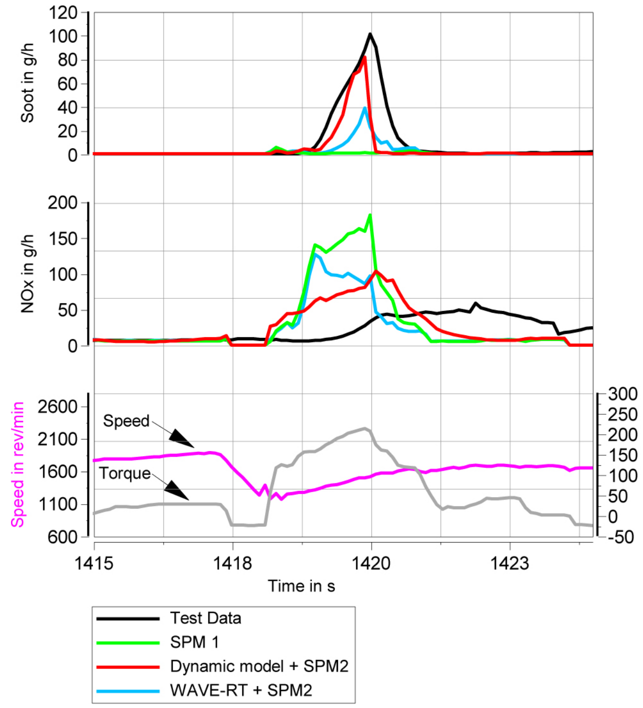

Figure 17. WLTC Soot and NOx comparison

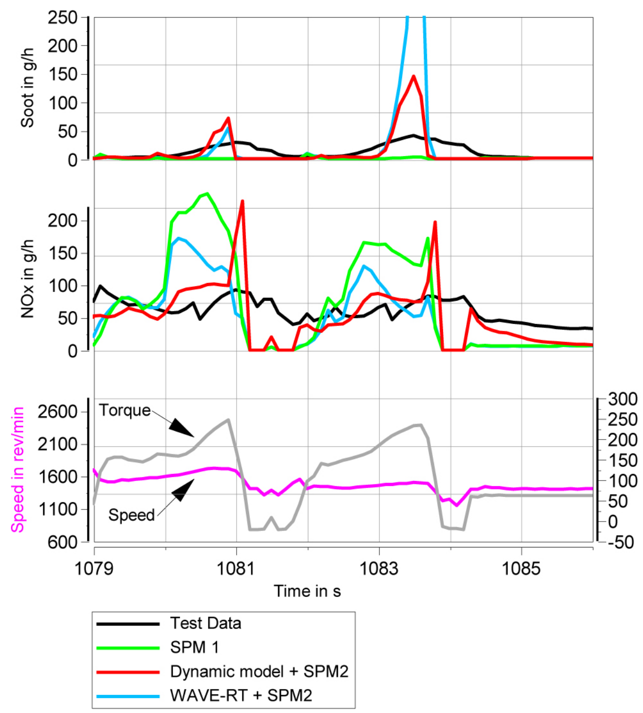

Figure 18. WLTC Soot and NOx comparison

OPTIMIZATION

The optimization is set up to minimize WLTC CO2, NOx and soot. The cumulative NOx and soot are predicted using the two-stage global models. The models are called each iteration of optimization process.

A physical SCR model was coupled to the dynamic NOx model to enable optimization to a tailpipe NOx target [10]. The optimizer was initially configured to minimize tailpipe NOx with a constraint on soot. The steady state model comprises of three different models (cold, warm and hot) with a single dynamic model applied to all cycle points (regardless of temperature). Fig 19 shows the optimisation output during the run, where the top plot shows the tailpipe NOx reducing from 115 to 76mg/km, while the constraint violations including soot were brought down to zero. The optimization process ran for 218 iterations and completed in a few hours. A further iteration loop of similar duration followed for CO2 optimisation with soot and tailpipe NOx constraints.

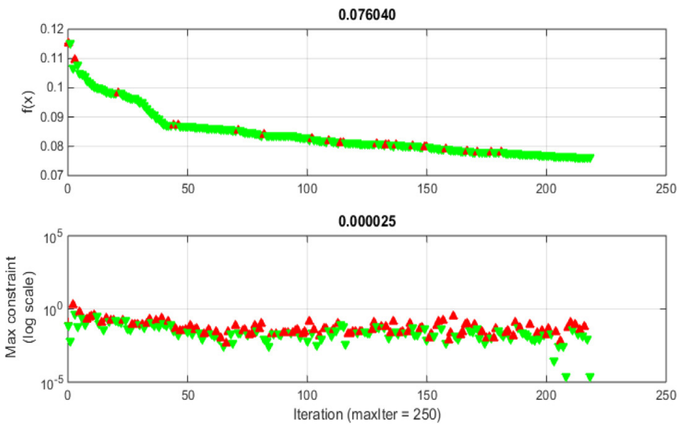

Figure 19. Global optimization with dynamic model

CONCLUSION

In general, the dynamic boost and EGR rate prediction improved soot prediction locally during tip-ins where a static model loses accuracy. Cycle cumulative results show a reduction in prediction error for the two approaches that use a dynamic model. Co-simulation of physical or empirical air system models with an empirical combustion model is an effective means to improve transient soot prediction compared using a static model approach. But of the two dynamic approaches, the empirical dynamic model coupled to the static model demonstrated the greatest improvement in emissions prediction.

Integration of these models in the optimization process produces better calibration maps. This is because more accurate cycle soot prediction is used for constraints during the optimization process. Inaccurate constraints can cause the optimization not to converge to a good solution. More accurate cycle prediction is important for calibration process because it results in fewer iterations (prediction – validation loop), produces a better calibration and is more robust for RDE.

This two-stage calibration optimization process has since been applied successfully on three series programs and is now embedded in the calibration process for diesel applications.

REFERENCES

1. Mohd Azmin, F. and Stobart, R., "Benefiting from Sobol Sequences Experiment Design Type for Model-based Calibration," SAE Technical Paper 2015-01-1640, 2015, doi:10.4271/2015-01-1640.
2. Kouba, A., Navratil, J., Hnilicka, B., and Niven, P., "Sensorless Control Strategy Enabled by a Sophisticated Tool Chain," SAE Technical Paper 2015-01-2847, 2015, doi:10.4271/2015-01-2847.
3. Ezzeddinne, M., Castro, E., and Lengellé, R., "Dynamic Design of Experiments for Engine Pollutants Emissions Modeling and Optimization," SAE Technical Paper 2008-01-2454, 2008, doi:10.4271/2008-01-2454.
4. Hafner, M., "Model Based Determination of Dynamic Engine Control Function Parameters," SAE Technical Paper 2001-01-1981, 2001, doi:10.4271/2001-01-1981.
5. Sacks J, Welch W J, Mitchell T J & Wynn H P (1989), “Design and Analysis of Computer Experiments.”, Statistical Science 1989 vol. 4 no 4 pp409-435.
6. Deng, Q. and Burke, R., "Turbocharger Dynamic Performance Prediction by Volterra Series Model," SAE Technical Paper 2014-01-2558, 2014, doi:10.4271/2014-01-2558.
7. Gutjahr, T., Kleinegräber H., Ulmer H., Kruse T., and Eckstein C.. “New Approaches for Modeling Dynamic Engine Behavior with Gaussian Processes.” In 7th Conference on DoE in Engine Development. Berlin: Expert Verlag. 2013.
8. Dickinson, P. and Shenton, A., "Rapid Transient Fuelling Calibration Method for the PFI SI Engine," *SAE Int. J. Engines* 2(1):327-334, 2009, doi:10.4271/2009-01-0583.
9. Mohd Azmin, F., Stobart, R., Rutledge, J., and Winward, E., "Using a Statistical Machine Learning Tool for Diesel Engine Air Path Calibration," SAE Technical Paper 2014-01-2391, 2014, doi:10.4271/2014-01-2391.
10. Seabrook Justin, Day Raoul, Newman Philip and Jarvis Simon, “Co-Optimisation of Engine and SCR System in an Advanced Diesel Vehicle.”, 7th Conference on Design of Experiments (DoE) in Engine Development, June 2013, Berlin, Germany.

CONTACT INFORMATION

Dr Farraen Mohd Azmin
Engine Control and Calibration
Ricardo UK Ltd
Shoreham Technical Centre, Shoreham-by-Sea,
West Sussex, BN43 5FG, UK
Tel: +44 (0) 1273 794561
Farraen.MohdAzmin@ricardo.com

DEFINITIONS/ABBREVIATIONS

APRBS - Amplitude Modulated Pseudo Random Binary Sequences

APRSS - Amplitude Modulated Pseudo Random Sinusoidal Sequences

DoE - Design of Experiments

EGR - Exhaust Gas Recirculation

FIE - Fuel Injection Equipment

LHS - Latin Hypercube Sampling

MISO - Multiple Inputs Single Output

MLE - Maximum Likelihood Estimation

NARX - Non-linear Auto Regressive network with exogenous inputs

ηCAL - Ricardo *Efficient Calibration* Global DoE Toolkit

SPM - Stochastic Process Model

RDE - Real Driving Emissions

RMSE - Root Mean Squared Error

VGT - Variable Geometry Turbocharger

WLTC - Worldwide harmonized Light Vehicles Test Cycle