

Development of Dynamic Constraint Models for a Model Based Transient Calibration Process

2011-01-0691

Published
04/12/2011Indranil Brahma
Bucknell Univ.

Copyright © 2011 SAE International

doi:[10.4271/2011-01-0691](https://doi.org/10.4271/2011-01-0691)

ABSTRACT

Model based calibration has gained popularity in recent years as a method to optimize increasingly complex engine systems. However virtually all model based techniques are applied to steady state calibration. Transient calibration is by and large an emerging technology. An important piece of any transient calibration process is the ability to constrain the optimizer to treat the problem as a dynamic one and not as a quasi-static process. The optimized air-handling parameters corresponding to any instant of time must be achievable in a transient sense; this in turn depends on the trajectory of the same parameters over previous time instances. In this work dynamic constraint models have been proposed to translate commanded to actually achieved air-handling parameters. These models enable the optimization to be realistic in a transient sense. The air handling system has been treated as a linear second order system with PD control. Parameters for this second order system have been extracted from real transient data. The model has been shown to be the best choice relative to a list of appropriate candidates such as neural networks and first order models. The selected second order model was used in conjunction with transient emission models to predict emissions over the FTP cycle. It has been shown that emission predictions based on air-handling parameters predicted by the dynamic constraint model do not differ significantly from corresponding emissions based on measured air-handling parameters.

INTRODUCTION

Model based calibration for automotive applications is becoming commonplace in the automotive industry but only for steady state calibration. Many commercial steady state calibration products are now beginning to be used by the automotive industry. There exist many barriers to developing a model based transient calibration process including the following questions: a) What kind of data is required to build transient models? b) What modeling technique would be appropriate c) Given reliable transient models, how should the optimization problem be framed? and d) How can the optimization be truly transient rather than quasi-static? The current work focuses on the last question.

The literature on engine calibration [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] focuses primarily on steady-state or quasi-static calibration processes. Related literature addressing function approximation [22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46], i.e., choice of model for model-based calibration is no different. With the exception of Atkinson et al. [16,17] and Westland et al. [18], there exist very few works in published literature that address the transient calibration problem for diesel engines with the goal of meeting emissions over regulated cycles such as the FTP cycle. This is surprising given the importance of transients on emission formation, but can be attributed to the following reasons: a) Fuel savings resulting from an optimized transient calibration is unclear and unstudied and b) The FTP cycle is seen more as a regulatory emissions hurdle while steady state duty cycles are considered important from the fuel efficiency point of view. Many manufacturers use two separate calibrations for the two purposes. An optimum transient calibration can save fuel by producing less particulate matter for the DPF to regenerate and by producing less NO_x for the SCR (if applicable) to reduce; this saves fuel and urea, respectively.

Better managed transients avoid emission spikes by preemptive smooth transitions rather than sudden surges through the air-handling system thus lowering pumping work and saving fuel. Many such transient strategies contribute to fuel savings but do so indirectly. Besides as mentioned the savings are not quantified in literature. In addition to lower fuel consumption, the environmental benefits of less ammonia slip (if applicable) and less particulates (the smallest and more harmful particles that are not trapped by the DPF) are motivations for the current work.

The current work links recent work by the author on transient emission modeling [19,20] and transient optimization of calibratable ECM tables [21] by constraining the optimization to only allow solutions that are achievable in a transient sense. Figures 1 and 2 show the results of a model based transient calibration process [21] where the goal was to reduce the engine out BSNO_x from 0.75 g/bhp-hr to less than 0.5 g/hp-hr while reducing PM to the lowest possible value. Note that the data in this figure and all others in this work have been normalized for confidentiality purposes. Without the use of dynamic constraints and statistical leverage [22] constraints (see bottom plot of each figure) the optimization reduces to a quasi-static problem and the optimized solution produces steady-state like cumulative emissions. This is clearly unrealistic and makes the optimization process meaningless.

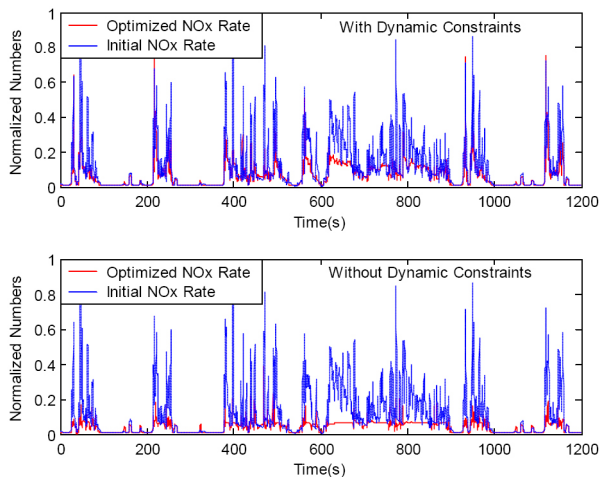


Figure 1. Transient (top) versus quasi-static (bottom) optimization of NO_x rate. The latter is unrealistic

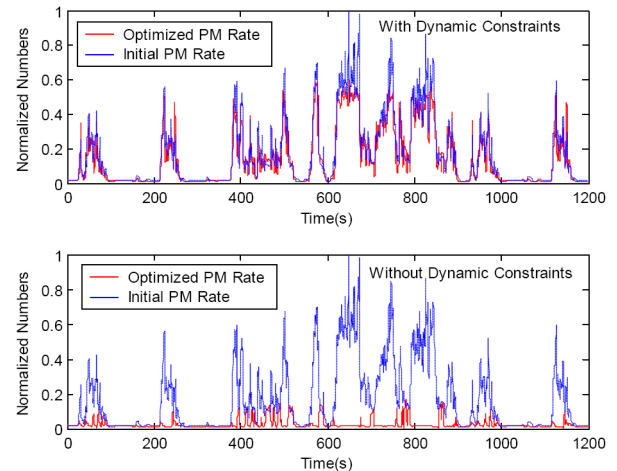


Figure 2. Transient (top) versus quasi-static (bottom) optimization of Particulate Matter. The latter is unrealistic

Before describing dynamic constraints it is important to understand the distinction between transient and quasi-static emission predictions. Figure 3 shows measured brake specific NO_x and PM emissions corresponding to transient and steady state data superimposed on each other. Approximately two thousand steady state points and data from 34 FTP cycles corresponding to about 180,000 transient data points have been plotted. This data spanned the spectrum of possible emission levels. It is evident that peak transient emissions are about five times higher than peak steady state emissions. Figure 4 shows how this result translates into FTP numbers. Plotted along with measured FTP emissions are predictions made by an empirical model trained on the steady state data, i.e., the predictions simulate quasi-static emissions. Quasi-static cumulative PM emissions were observed to be about half of transient PM emissions. The results shown by figures 3 and 4 can be understood by looking at figures 5 and 6. The distinguishing features of transient operation are the extremes of air-handling parameters. The two dimensional EGR fraction - fresh air flow space, important for NO_x formation and the fueling - in cylinder Oxygen mole fraction space, important for PM formation are much larger than corresponding steady state regions. Non air-handling engine parameter spaces are not substantially different between steady state and transient operation.

The extremes of the two air handling parameters fresh air flow (or charge flow) and EGR fraction are therefore primarily responsible for transient emissions. The in-cylinder mole fraction shown in figure 6 has also been calculated from these two parameters. These extremes are caused by the differences in sub-system response time [20]. Actually achieved charge flow and EGR fraction are significantly different from commanded flows and fractions during transients. Any transient calibration process that tries to optimize the commanded engine maps must be able to predict

actual achieved fractions and flows to be used as inputs within emission models. The models used to predict achieved fractions and flows are termed as dynamic constraint models in this work because they constrain the optimization to be dynamic instead of quasi-static. In this respect the dynamic constraint models form the core of the transient optimization process.

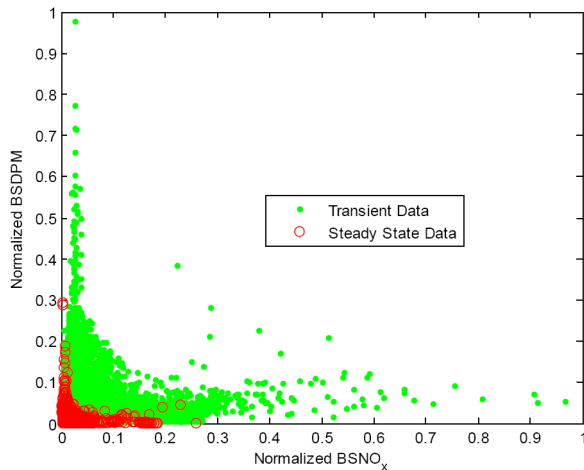


Figure 3. Peak transient emissions are about five times greater than peak quasi-static emissions

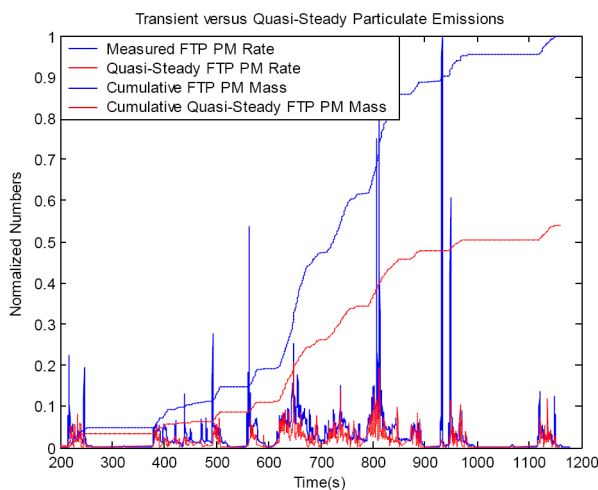


Figure 4. Cumulative transient PM is about twice cumulative quasi-static PM

The transient optimization process involves a large number of calibratable parameters to be optimized over the entire engine map. An additional reason to use a dynamic constraint model is to restrict the empirical emission model within its statistical leverage limits. The leverage limit [22] separates the multidimensional space into two regions based on the training data. Any point outside the leverage limit can be considered extrapolation where any empirical model prediction would be unreliable. Although it has been shown by the author that it is not possible to define one leverage limit [21] with correlated

transient data, constraining the optimizer with several leverage metrics and dynamic constraint models together helps prevent inevitable excursions and entrapment into completely invalid regions while searching for the optimal solution.

The transient emission models mentioned throughout this work and reported previously [21] are regression models based on engine speed, fueling, EGR fraction, charge flow, injection timing and rail pressure. Regression was chosen as the empirical model of choice based on its robustness in predicting transient emissions; previous work by the author had found neural networks to be the best modeling approach for steady state emissions and global regression to be the best approach for modeling transient emissions [19]. The dynamic constraint models predict the achieved EGR fraction and charge flow rates that are then plugged into the transient emission model inputs. The emission models are used within the optimization objective function. The emission models and the optimization process have been addressed in separate works [20,21].

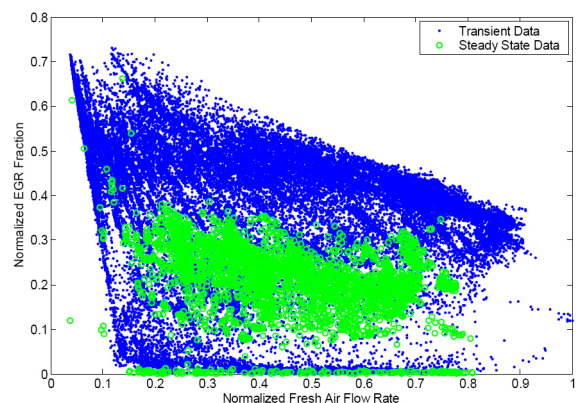


Figure 5. Transient EGR fraction - Fresh air flow space is much larger than the corresponding steady state space

All data for this work was acquired with a diesel engine equipped with electronic multi-injected high pressure common rail fuel system, a high pressure EGR loop and a variable geometry turbocharger. All data has been normalized for confidentiality reasons. The data used in this work was extracted from about 80 transient FTP and transient snap throttle cycles (used for torque response optimization) which were run as part of a manual transient optimization process. Methods developed in this work are not expected to be particular to any engine hardware or calibration.

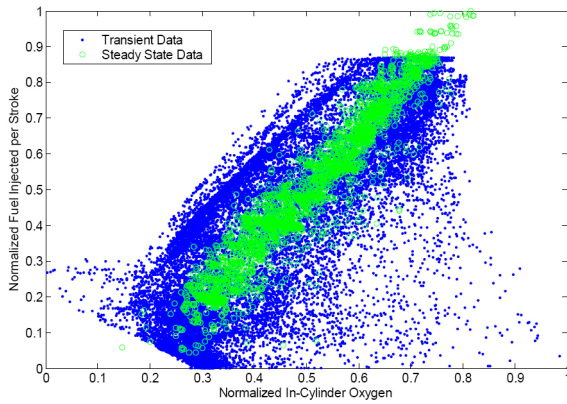


Figure 6. Transient Fueling - In Cylinder Oxygen Mole Fraction space is much larger than the corresponding steady state space

DYNAMIC CONSTRAINT MODELS

METHOD

EGR fraction and charge flow were considered as second order systems according to the following linear second order equation:

$$\ddot{y} + \alpha_1 \dot{y} + \alpha_2 y = \alpha_3 u \quad (1)$$

where y is either the achieved EGR fraction or charge flow, u is defined as the forcing function to the system that drives y . The α_i are model parameters. Since the air-handling for the particular engine used in this study is proportional derivative (PD) controlled, u is written as:

$$u = k_p (c - y) + k_d (\dot{c} - \dot{y}) \quad (2)$$

where c represents commanded quantities, i.e., either commanded EGR fraction or commanded charge flow and k_p and k_d are the proportional and derivative gains. Discretizing equations (1) and (2) with first order accuracy yields:

$$y(t) = \beta_1 y(t-1) + \beta_2 y(t-2) + \beta_3 c(t) + \beta_4 c(t-1) \quad (3)$$

where

$$\begin{aligned} \beta_1 &= \frac{\frac{2}{h^2} + \frac{\alpha_1}{h} + \frac{\alpha_3 k_d}{h}}{\frac{1}{h^2} + \frac{\alpha_1}{h} + \alpha_2 + \alpha_3 (k_p + \frac{k_d}{h})}, \\ \beta_2 &= \frac{-\frac{1}{h^2}}{\frac{1}{h^2} + \frac{\alpha_1}{h} + \alpha_2 + \alpha_3 (k_p + \frac{k_d}{h})}, \\ \beta_3 &= \frac{\alpha_3 (k_p + \frac{k_d}{h})}{\frac{1}{h^2} + \frac{\alpha_1}{h} + \alpha_2 + \alpha_3 (k_p + \frac{k_d}{h})}, \\ \beta_4 &= \frac{-\frac{\alpha_3 k_d}{h}}{\frac{1}{h^2} + \frac{\alpha_1}{h} + \alpha_2 + \alpha_3 (k_p + \frac{k_d}{h})} \end{aligned}$$

where h is the size of the time step used for discretization. All data was captured with 5HZ frequency so the time step duration (h) was chosen as 0.2 seconds for model training as well as testing.

Although the PD gains are known the second order system parameters α_i are not, without which the parameters β_i cannot be determined. The approach used in this work is to use linear regression to extract the β_i s from available system data. A prediction at any instant of time will therefore depend on the prediction at the previous two time steps, and the current and the past commanded values. Although there seem to be four parameters in the final discretized equation (3) as opposed to three parameters in the original model equation (1), equation (3) can be written as:

$$\begin{aligned} y(t) &= \beta_1 y(t-1) + \beta_2 y(t-2) + \dots \\ &\quad \beta_3 \left[c(t) - \frac{k_d}{k_p h + k_d} c(t-1) \right] \end{aligned} \quad (4)$$

However equation (3) has still been used in the current work because of uncertainties associated with PD gains in the available datasets. Both the EGR fraction and the charge flow rate are the result of EGR flow as well as fresh air flow. Hence the physical system approximated by equation (1) is the complete air-handling system including the engine block, turbocharger, charge air cooler, EGR valve, EGR cooler and charge air cooler. It is possible to represent each of those sub-systems as lumped parameter sub-models, but the

approximations required to capture turbulent, nonlinear fluid dynamics and the complexity of coupling these dynamics to the control system are two of the three motivating factors for the current approach. The third and most important motivation is to extract information from real data so that the model can be used for transient calibration that predicts accurate numbers that can be replicated on the engine.

The second order assumption was the simplest model that could handle the oscillations about commanded quantities. Figure 7 shows a representative segment of the FTP cycle where the EGR fraction oscillates around the commanded fraction. The predicted fraction using the second order model is seen to capture some but not all of the measured behavior. Note that the data has been normalized for

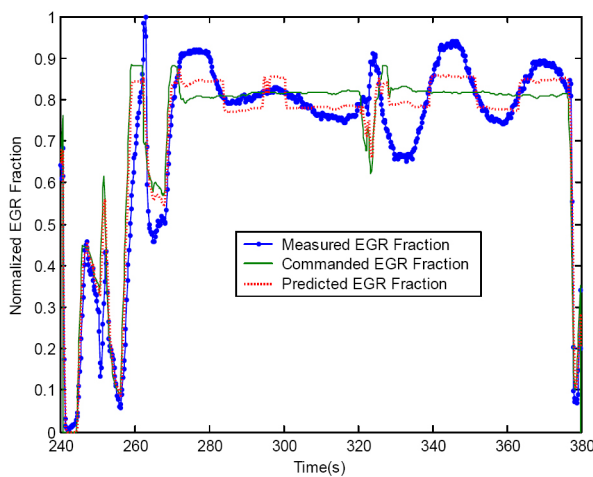


Figure 7. A second order model is required to capture at least part of the measure oscillations around commanded quantities.

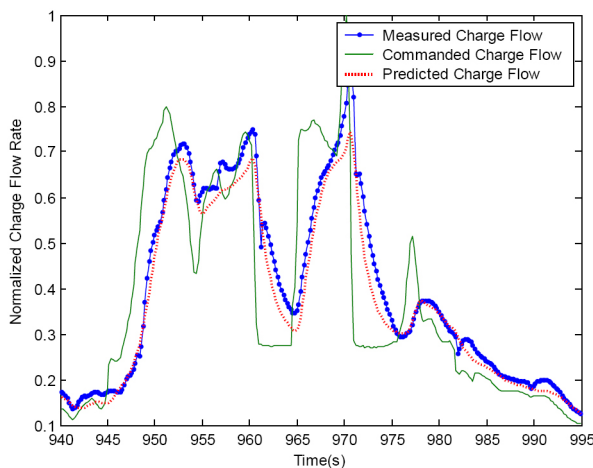


Figure 8. The time constant was calculated to be different depending on if the commanded charge flow is higher or lower than achieved charge flow

confidentiality purposes. Figure 8 shows commanded versus measured and predicted charge flow for a representative segment of the FTP cycle. It can be observed that there exists a significant response delay to commanded charge flow. This is a result of turbocharger lag and the primary cause of transient emissions. The time constant was calculated to be different depending on whether the commanded charge flow was higher or lower than the achieved charge flow. The use of a linear second order approximation is strictly true in the vicinity of linearization, however the premise of this work is that this vicinity is broad enough to encompass normal engine operation. However the physical phenomena associated with the system trying to increase the charge flow (turbocharger lag phenomena) is different from the physical phenomena associated with the system trying to reduce the charge flow (depends on the volume capacitance of the manifolds and charge air cooler). Separate coefficients for equation (3) corresponding to whether the commanded charge flow is trying to increase or decrease charge flow would therefore be consistent with the physical phenomena. Since EGR fraction depends on charge flow it would be logical to use two separate equations for EGR fraction predictions also. The two equation model is implemented according to:

$$\begin{aligned}
 y(t) &= \beta_{1+} y(t-1) + \beta_{2+} y(t-2) + \beta_{3+} c(t) + \beta_{4+} c(t-1) \\
 &\quad \text{for } c(t) \geq y(t-1) \\
 y(t) &= \beta_{1-} y(t-1) + \beta_{2-} y(t-2) + \beta_{3-} c(t) + \beta_{4-} c(t-1) \\
 &\quad \text{for } c(t) < y(t-1)
 \end{aligned}
 \tag{5}$$

where the positive and negative subscripts denote separate coefficients. Equation (3) is the simplest possible method of accommodating two distinct time constants corresponding to two separate physical phenomena.

As will be reported in the following section the results of using a linear second order approximation with two equations proved to be superior to linear first order systems as well as a second order system without separate equations.

TEST METHODOLOGY AND RESULTS

Two separate stages of testing were undertaken, the first to select the best model approximation based on agreement between measured and predicted charge flow and EGR fraction and the second to compare emission model predictions with measured versus predicted charge flow and EGR fraction. In both stages data used to train and test the dynamic constraint models came from different cycles to ensure that data was not being ‘memorized’.

STAGE ONE - MODEL SELECTION

11 FTP cycles were selected as testing data. A single snap throttle cycle was chosen as the training dataset. The training and testing datasets were therefore