

Data Driven Calibration Approach

Nahid Pervez, Ace Koua Kue, Adarsh Appukuttan, John Bogema, and Michael Van Nieuwstadt
Ford Motor Company

ABSTRACT

Designing a control system that can robustly detect faulted emission control devices under all environmental and driving conditions is a challenging task for OEMs. In order to gain confidence in the control strategy and the values of tunable parameters, the test vehicles need to be subjected to their limits during the development process. Complexity of modern powertrain systems along with the On-Board Diagnostic (OBD) monitors with multidimensional thresholds make it difficult to anticipate all the possible scenarios. Finding optimal solutions to these problems using traditional calibration processes can be time and resource intensive. A possible solution is to take a data driven calibration approach. In this method, a large amount of data is collected by collaboration of different groups working on the same powertrain. Later, the data is mined to find the optimum values of tunable parameters for the respective vehicle functions. This large scale data (terabytes) gives the engineers more samples to analyze and confidence in the calibration. A robust data analysis platform capable of handling large scale data and flexible enough to cater to different engineers' needs is required to accomplish this task. In this paper, some salient features of such a platform developed by Ford Motor Company are presented. The memory efficiency, Model-In-Loop and parallel processing capability of this platform are highlighted. The use of MapReduce programming techniques to analyze large scale calibration data is also discussed.

CITATION: Pervez, N., Kue, A., Appukuttan, A., Bogema, J. et al., "Data Driven Calibration Approach," *SAE Int. J. Commer. Veh.* 10(1):2017, doi:10.4271/2017-01-0607.

INTRODUCTION

The advent of more sophisticated features in vehicle control, tighter emission requirements and more comprehensive OBD requirements are making the control system more complex. Increasing competition in the market forces OEMs to keep refreshing their product lineup. This leads to a short timeframe for development with an increasing number of features to develop and calibrate. In order to complete this task, development teams have to rely on efficient tools to analyze data and take critical decisions based on the analysis.

With more capable data collection tools, methods and analysis techniques, it is now possible to design control systems based on the real world data collected rather than designing for worst case scenarios. This novel ideology for calibration requires a robust data analysis platform with the capability to perform complex analyses with ease. Such a data analysis platform was developed by Ford Motor Company and a data driven calibration approach is being practiced in several Ford calibration teams. To limit the scope of this study, only some features of this tool that are used extensively in diesel OBD calibration are discussed in this paper.

BACKGROUND

Modern model based design of powertrain control systems follows the typical V-cycle of a software development process (Figure 1). Calibration and fine tuning are the final steps of this process. This

task is often performed on prototype vehicles with complete subsystems. This ensures that all possible system variations can be accounted for in the development process.

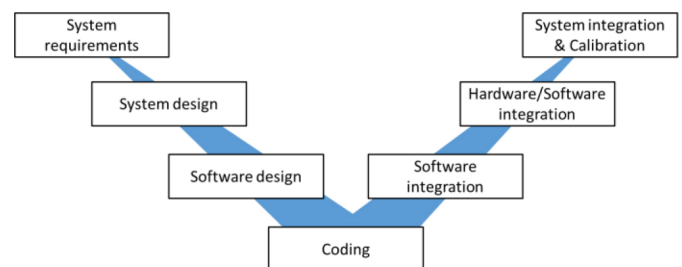


Figure 1. V cycle model of Software development

Typical calibration processes require a calibration engineer to collect data with rough values for tunable parameters. Fine tuning of these parameters requires in-depth analysis of the data collected. In most cases, complex statistical analyses have to be performed on the collected data to set values to the tunable parameters. A calibration engineer must know the limits of the control system, have test plans for extreme conditions and analyze numerous data files. For a small feature, a calibrator may have to go through a significant number of files, pull out the critical values from data, and perform statistical analyses on those data points to get a robust value for the tunable parameter.

With the small time frame available and the increasing number of features, achieving a robust calibration using manual data analysis is almost impossible with the same number of engineers working on the project. By automating this data analysis process, the calibrator can work on multiple features simultaneously. Using this process for automation of data analysis also provides a quantitative measure of success for the calibration. Innovative analysis techniques such as parametric optimization and machine learning can be used to automatically calculate some of the complex calibration parameters which are dependent on several interconnected subsystems. Additionally, advanced simulation capabilities such as Model-In-Loop and Software-In-Loop can be utilized to refine the calibration without the need for further in-vehicle data collection. This reduces the use of expensive prototype vehicles and increases the efficiency of the calibration workflow significantly.

Data Analysis Platform

The major tasks of any capable data analysis tool are [1]:

1. Read data from the native file format.
2. Perform analyses on the data set.
3. Present the information using various visualization techniques.
4. Publish the results in a versatile format.

Developing a generic calibration data analysis tool is challenging for the following reasons:

1. Calibration data file formats can be diverse depending on the tools used to collect data.
2. Different vehicle features or subsystems need different kinds of analyses. A 'one size fits all' type of solution is not suitable for data analysis.
3. Performing analyses on large scale datasets is difficult.
4. Continuous support and improvement of the tool is necessary.

Furthermore, such a tool should have the capability to implement advanced data analysis techniques such as optimization and machine learning. When designing such a tool to overcome the challenges and incorporate the requirements of calibration data analyses, it becomes evident that a complete data analysis platform with the capability to run analyses from user created scripts is a more suitable approach than a single tool. This makes the application versatile and can morph to the need of individual users.

A sample architecture of such a platform can be found in Figure 2. It contains three major layers and the data source. From the data source, data in a standard format is transferred to the data layer. The data layer can vary from a simple shared drive to state-of-the-art distributed file storage systems like Hadoop. The choice of the type of data layer used depends on the budget and also on the objective of the analytics layer. If the analytics layer is heavily focused on parallel programming applications such as MapReduce, having a distributed file storage system will boost the computational performance by a significant margin. The analytics layer is the most utilized layer in the entire architecture and all computation happens here. Complex and use-case specific large scale analyses are performed in this layer and the results

are then passed to the presentation layer for visualization and reporting. It is common practice to merge the analytics and presentation layers. Normally, analysis application providers also offer libraries that are required to present results in an easy-to-use format. This is also dependent on the type of analysis, and, often the end-user designs the output as he/she sees fit. This layer can be a web based interface as well, in which case the analytics and data layers are normally stationed in a remote high performance computing machine.

Ford has implemented a batch processing data analysis platform (dTool) which closely follows this structure. MATLAB was selected as the language used to develop this platform because of its rich numerical analysis library, capability of Model-In-Loop and Software-In-Loop simulation and ease of use. dTool can perform multiple analyses on large sets of data files efficiently and summarize the outputs of individual analyses in a report. It can implement complex machine learning algorithms competently and with ease. It is also designed to handle BigData analysis techniques such as MapReduce with or without using a distributed file system.

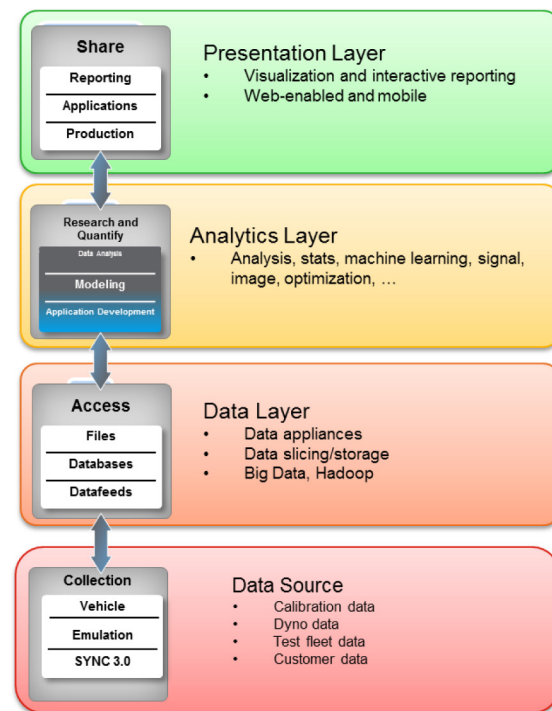


Figure 2. An example architecture of a robust data analysis tool

This platform enables Ford to look at calibration work flow from different perspectives. This new ability to perform complex analyses on large scale data paved the way for a dominant data driven calibration approach.

TRADITIONAL CALIBRATION APPROACH

When software for new monitors is developed, it needs to be calibrated using massive amounts of data from prototype vehicles at varying road and environmental conditions in order for the calibration to be robust. The traditional calibration approach would involve driving the vehicles with a rough calibration and collecting data, then analyzing the data to improve the calibration. The new and

improved calibration would then be tested and validated in-vehicle, and the robustness of the new calibration can only be validated by further data collection across different conditions. This approach is a repetitive loop (Figure 3) that requires continuous in-vehicle data collection to validate new calibrations or a deep dive analysis of the calibration across individual data files which is very time intensive. This drives the need for a more efficient calibration workflow.

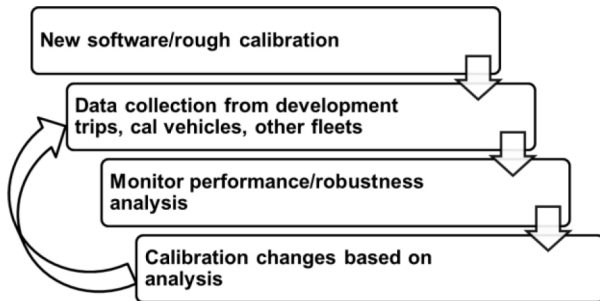


Figure 3. Illustration of a traditional calibration approach

DATA DRIVEN CALIBRATION APPROACH

In a data driven, Model-In-Loop calibration approach, data collected in vehicles from the start of a program can be used to validate the robustness of newer calibrations. This is done by simulation of the monitors' performance over thousands of past drives. This eliminates the need for data collection at later stages of a program and helps provide a much more robust calibration than can be achieved by the traditional approach (Figure 4).

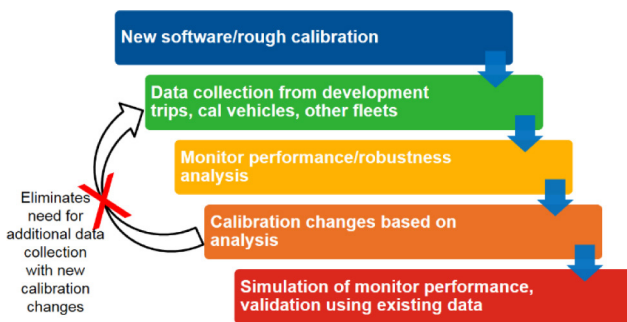


Figure 4. Illustration of a data driven, model-in-loop calibration approach

The extent of simulation and the capability of the model-in-loop approach varies depending upon the monitor. Some examples of this approach being used in OBD monitors are illustrated in the following sections.

Simple Histogram and Debounce Plot

Simple histograms and debounce plots can be created for all OBD monitors to evaluate their robustness, and thousands of drives can be processed in a relatively short time. Figure 5 demonstrates the performance of an air mass setpoint deviation monitor. The histogram plots the probability density of the separation between the monitor metric and the monitor threshold. The debounce plot below this shows the duration of debounce in seconds from each file. A debounce of (-1) indicates that the monitor did not release in that drive and a debounce of (-2) indicates that one or more required data

items were not recorded in those files. About 3000 drive files were processed in this example. This plot demonstrates that the monitor is very robust against false fail conditions.

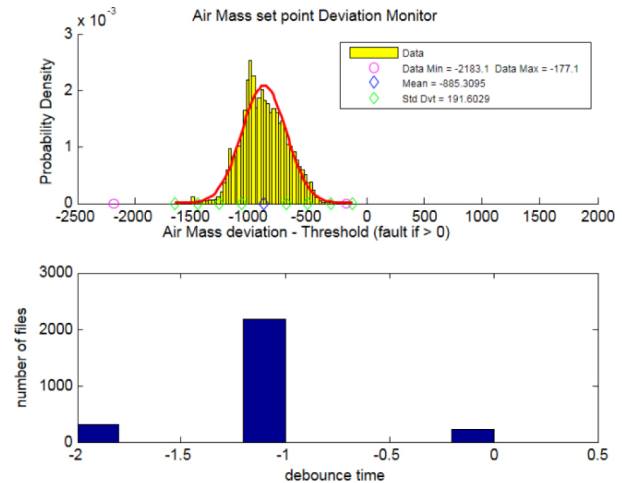


Figure 5. Sample output histogram and debounce plot for a continuous monitor

Simple Sensor Plausibility Check

In the following example, temperature sensor output values are cross checked with each other for plausibility after a soak for a certain period. If the absolute temperature difference between a sensor and all other sensors is greater than a certain threshold, a fault is reported. The output plot (Figure 6) provides important statistics on the absolute temperature differences and these statistics can be used to help set thresholds for each sensor such that a sensor that is reading beyond its normal range is detected as a faulted sensor.

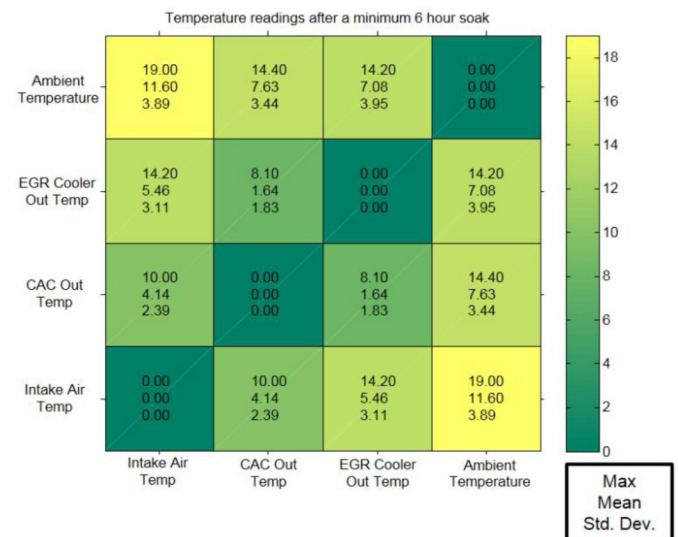


Figure 6. Statistics of a simple sensor to sensor plausibility check monitor

Intrusive Monitors

Intrusive monitor performance can only be assessed from actual data files and have limited simulation capability. Figure 7 shows the performance of an intrusive monitor designed to detect a slow response. The turbocharger vanes are moved intrusively from closed