Bayesian Statistics in Engine Mapping

Ward M.C., Brace C.J.
Shaddick G., Finch A.J. Akehurst S.
University of Bath

Ceen R.
CP Engineering

Hale T. , Kennedy G.
Cosworth Technology

In order to comply with increasingly stringent emissions standards and meet drivability requirements, modern automobile engines are equipped with an increasing number of subsystems and controlling elements. The result has been to greatly increase the calibration effort required to find the parameter settings that offer the best global compromise across the entire engine map. An increasing trend within engine testing is the application of statistical tools; these include design of experiments (DoE), Bayesian and stochastic methods, which have all proved themselves in addressing multivariate problems in other fields.

This paper examines a core subset of engine mapping termed 'spark sweeps'. Spark sweeps are necessary to identify the operating regions, which are potentially destructive to the engine, and to provide torque reduction data necessary for satisfactory transient operation. Software for automating spark sweeps is currently running on many test rigs. An example of such software is the I-CAM code developed by CP Engineering, which is used as the basis for this work.

This paper shows a further improvement in terms of time by coupling a Bayesian prior knowledge algorithm to I-CAM. As any benefits found are likely to be applicable to the wider calibration problem. This development was carried out on a turbocharged gasoline engine rig, the control of which was abstracted into the MATLAB/SIMULINK environment. This link was developed in house at the University of Bath and enables the full range of MATLAB data processing, statistical, and optimisation routines to be used online. This paper presents the Bayesian prior knowledge algorithm itself and seeks shows the improvement in efficiency of the whole spark sweep procedure, and to validate the results against the existing code.

NOTATION

MBT	Mean Best Torque
BLD	Borderline Detonation
ICAM	Integration Calibration and Mapping Component
EGT	Exhaust Gas Temperature
EMS	Engine Management Systems
ME7	Bosch ME7 Engine Management system
RL	Relative Load
ETC	Electronic Throttle Control
GDI	Gasoline Direct Injection

1.INTRODUCTION

In order to comply with increasingly stringent emissions standards and to meet drivability requirements, modern engines are equipped with an increasing number of subsystems and controlling elements. This has led to the use of torque based engine control architectures [ ], which use electronic throttle control (ETC). The result has been to greatly increase the calibration effort required to find the parameter settings that offer the best global compromise across the entire engine map. This task is set to get significantly harder in the future with the increasing manufacture of GDI and introduction of hybrid engines.

A current trend within engine testing aimed at reducing the time required for the calibration procedure, is the application of statistical tools developed for use in other fields where empirical models of systems are necessary. This paper seeks to examine Bayesian statistics which offers a way to reuse prior knowledge that is traditionally discarded by classical statistical methods such as design of experiments [ ]. Data yielded from experimentation is used to continually update a model, which initially is dominated by prior knowledge, then progressively becomes influenced by experimental results. This means that a test program can be guided in real time by the feedback of data, and hence should be flexible than the traditional approach of collecting all the data before building the model. This paper seeks to examine the effectiveness a Bayesian statistics algorithm at characterizing a particular area of engine behavior a traditional method of collecting the same data. Although the experimentation outlined in this paper is centered around a particular problem it is envisaged that the same algorithm statistics could be developed to tackle the wider engine calibration problem.

1.1.SPARK SWEEPS

The power output of a spark ignition engine is determined by the airflow through it, which is governed via the throttle valve; and in the case of boosted engines also a turbocharger. The airflow to the inlet manifold is normally determined using a hot film sensor located upstream of the throttle assembly and is processed by the ECU into a measure designated Relative Load (RL). RL is defined from 0 to 130% with Values over 100% meaning that the engine is running at wide-open throttle and that the airflow is being boosted by the turbocharger. Knowledge of ignition angle vs. torque (figure) at a fixed RL is necessary for calibration of gasoline engines with a torque based Engine Management System (EMS), because ignition timing is used as a fast control path for traction control, during rapid pedal transients and for, knock control. During a spark sweep engine speed, lambda and RL are controlled to constant values with the only variable being ignition angle.

1.2.TESTBED LAYOUT AND ICAM

The mechanical setup is shown schematically in Error! Reference source not found.. The test cell is based upon a Volkswagen group production specification 1.8 litre turbocharged gasoline engine, having a maximum power output of 135KW and a maximum torque of 210 Nm. The engine managed is via a torque based Electronic control unit (ECU). The engine is loaded by a Schenck eddy current dynamometer, capable of absorbing up to 230 kW. Control of the test cell, dynamometer and data acquisition is carried out by a CP Engineering Cadet V12 [2,3] host system shown in Error! Reference source not found.**.** The software environment of Cadet is shown in Error! Reference source not found.**,** Cadet has an ASAP3 link to the calibration tool (VS100) and also has an Ethernet link to MATLAB running on a second PC. This allows MATLAB scripts authority over ECU parameters .

ICAM (Integrated **C**alibration and **M**apping) is a software component [ ] figure # running in Cadet developed to automatically acquire spark sweep data for Torque based engine management systems. It allows torque data to be measured for any required combination of engine control parameters. The program consists of two nested procedures.

1.Inner Loop __ Spark Sweep, (Figure #). This consists of a forward spark sweep where the ignition angle is advanced stepwise until MBT or BLD is found, then a reverse sweep is carried out with the ignition being retarded until either missfire or EGT limits are exceeded. In both cases ICAM terminates the sweep immediately to prevent hardware damage. The knock indication was provided from accelerometer signal processed by the ECU and fed into the Cadet via ASAP3. Missfire was indicated by a signal representing an increased level of fluctuation in crankshaft speed. And EGT via thermocouples in the exhaust manifold.

2. Outer Loop __ predefined engine parameter settings, at which a spark sweep is to be carried out at. For example engine speed, relative load, AFR, camshaft phase.

ICAM was run in its standard form to produce baseline surfaces of the engine (figure #) with varying lambda and spark angle. 14 point Spark sweeps were carried at lambda values of 0.8, 1, and 1.2 at the following sites:

- Engine Speed (RPM): 2000
- Relative Load (%):25,30,40,50

Thus producing 4 surfaces each derived from 42 test points which were use as prior knowledge data sets for the and to test the accuracy of the models produced by the Bayesian algorithm. Each point takes 35 seconds to take consisting of a 30 second settling and 5 second averaging period, giving a total test time of for each surface of 1470 seconds (or 24 mins 30 seconds). The aim of the experimentation therefore is to obtain the same surface with a reduced number of test points.

BAYESIAN STATISTICS

The Classical or frequentist approach to statistics is to assume that nothing is known about a system until all the experiments on it are complete, and a model is constructed. Frequentist statisticians object to the use of prior information because it is subjective, depending on the personal judgment of the individual from whom the information is elicited ie. One must make a decision whether the data from previous experimentation can be reapplied. The limitations of this approach are that there is no feedback during the test program and data gathered from previous experimentation is discarded. Bayesian Statistics on the other hand allows the use of prior information, which in theory should lead to a more efficient program of experimentation as all available data about the system, is used and feedback can be provided to guide the test program whilst it is underway. A Bayesian statistics approach offers the potential of outlier detection, which can be used to flag up possible measuring device faults of other possible fault causes of spurious data.

The Bayesian algorithm (figure #) begins with the formulation of a prior model with an associated distribution of the model coefficients. The prior information is then combined with the information from the test data via the likelihood function to produce a updated posterior distribution for the model terms. The posterior distribution is simply proportional to the prior and the likleyhood, but if the prior distribution is not carefully chosen the calculation of the constant of proportionality can become difficult. For this work it was assumed that the variance of the coefficient terms was independent and normally distributed, meaning that the distributions of the posterior coefficient term will also be normal.

There are two useful statistical metrics calculated by the Bayesian algorithm each time the model is updated.

CONFIDENCE INTERVALS

The first is confidence Intervals for the updated model. These can be used as a method of assessing the quality of the current model, give define how much confidence we have in our fitted model. They provide a method of assessing convergence of the model and when to terminate testing. Generally speaking good prior information creates smaller confidence intervals earlier in the testing program and yields quicker convergence. Large model variances place restrictions on the minimum width of the interval so a specified convergence criteria might never be reached. Confidence Intervals are highly dependent on knowing the model variance accurately. If the model variance is underestimated then we can be overconfident with the model. When it is overestimated then we have less confidence in the model. This problem can be alleviated by giving the model variance a probability distribution so that it may vary as data is collected.

PREDICTIVE INTERVALS

Predictive Intervals can be used to find outlying data points provided the model variance is known precisely.
As with Confidence Intervals this can be overcome by giving the model variance a probability distribution.
Good prior information creates narrower predictive intervals earlier in the testing program and hence it can detect outliers, which is a useful feature for checking for errors in test instrumentation.

IMPLEMENTATION OF THE BAYESIAN ALGORITHM

The Bayes algorithm was implemented as a MATLAB script, which featured embedded calls to the host system to select new test points and subsequently measure the torque response of the engine online. The Bayesian technique is generic across different modeling techniques. It was found that the torque, spark lambda surfaces could be satisfactory fitted with a second order least squares fit, however for different data other models can just as easily be used. The implementation assumed that the variance of the coefficient terms was independent and normally distributed, as well as the prior and sample data from the testbed. The advantage of using a model where the likelihood, prior and posterior are all normal is that the normal distribution is time invariant. This means we do not have to combine each new data point with the existing data set. After each testpoint is selected the posterior distribution is set to the new prior and updated using the single data point only. In the event that all data was lost the posterior distribution could be used as a good prior distribution with which to continue testing. The model variance is easily obtained by carrying out the repeat points on the engine. In this case it is known that the maximum variance in the data will be when the engine is running at its leanest Lambda value and most retarded ignition angle. Hardware protection of the engine is looked after solely in Cadet, if combustion knock, misfire of EGT limits were encountered as a result of the parameter settings the host system immediately imposed safe settings. and returned a failure message to MATLAB and another point was chosen.

TESTPOINT SELECTION METHODS

Traditional engine calibration follows a fixed data selection procedure where testpoints are selected incrementally from a range of data, this is often done to prevent hardware damage for example in the case of combustion knock, or temperature limits. This contradicts the frequentist requirement of randomization which underpins the assumption that the experimental errors are independent of each other. The testbed engineer would argue that in a carefully controlled test cell environment with proper parameter settling periods and temperature controls the need for randomization is minimized. Despite this, it is sensible to introduce a degree of randomization or inferences made from the model may be invalid. Bayesian statistics offers us the chance to continuously update our model it is possible to select test points one at a time and analyze the effect they have on the model. In this way information from each test point can be maximized and experimentation can be guided in order to increase efficiency.

Three selection procedures were examined for their effect on convergence time: -

1. Random.
2. Maximum Residual.
3. Sequential.

1. Random selection was used to compare statistically ideal data selection to possible preferable techniques.

2. Maximum Residual selection takes testpoints from areas of the model where the residuals are greatest – i.e. where the difference between the fitted model and the selected data was greatest. The idea was to try and minimise the largest residual by pulling the model towards another testpoint nearby.
   Preliminary tests showed that choosing the maximum residual each time meant that the same testpoint would be sampled continuously. In order that some randomisation could be achieved it was decided that data would be sampled from a normal distribution with a mean centred on the maximum residual. The 95% range of the distribution was taken as a proportion of total range of the data. i.e the most likely single value to occur was the maximum residual.

3. Sequential selection was designed to imitate traditional data selection by taking testpoints from a pre-determined sequence.

SCALING THE INPUT VARIABLES

It was found that when the input variables, and hence the model coefficients, had varying order of magnitudes the predictive interval became extremely wide rendering it useless at outlier detection. This problem is caused because it allows squared terms to unduly dominate the calculation of the predictive interval. This can be overcome by scaling the input variables. Here, Lambda (0.8 – 1.2) and ignition angle (14 - 36) are exhibit this problem, and so the ignition angle was divided by 10 to bring it into line with the value of lambda.

RESULTS AND ANALYSIS

CONCLUSIONS AND OUTLOOK

To be able to use the Bayesian method effectively a consistent system of gathering and storing data needs to be employed. In this way we can maximise available information prior to testing. Given a set of input variables and a response it should be possible to identify a suitable order model, previous data to model as a prior distribution, values to define a distribution for the model variance and a data selection technique to maximise convergence.

It was found that the algorithm obtained a model in a reduced number of test points compared with ICAM only if the prior knowledge and model variances set to be tight.

REFERENCES

[1] CP Engineering, 2000, “Cadet V12 – Run test Manual”

[2] CP Engineering, 2000, "ICAM component manual"

[3] Mathworks, “MATLAB online help”, Mathworks

[4] A.J Finch, 1999, “An Application of Bayesian Statistics to Engine Modeling.”, MSc Thesis, Bath University

[5] D Mowell, D R Robinson, A D Pilley 1996, “Bayesian Experimental Design and its Application to Engine Research and Development”, SAE Paper 961157

[6] Bosch, 1999, “ME-Motronic Engine Management” 1st Edition Bosch

[7] Usborne, 1996, “My First Book of Statistics, suitable for under 8’s”, 1st edition Usborne.

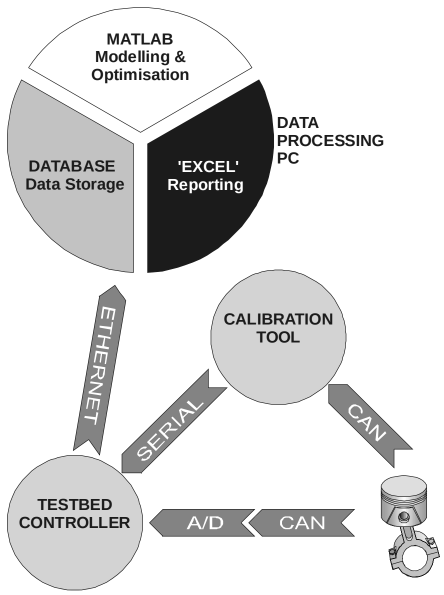


Figure 1: Testbed Network

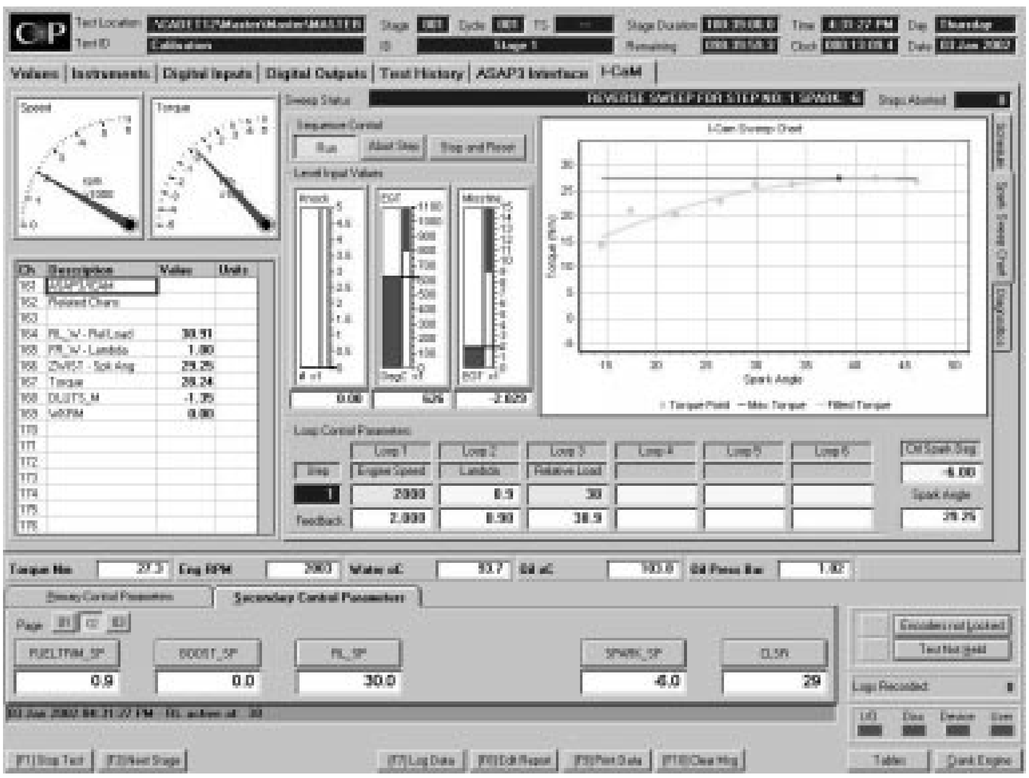
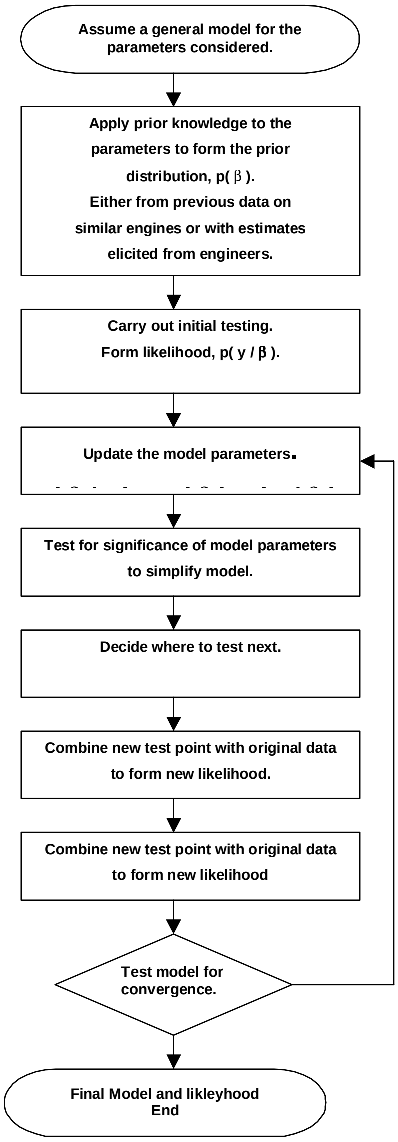
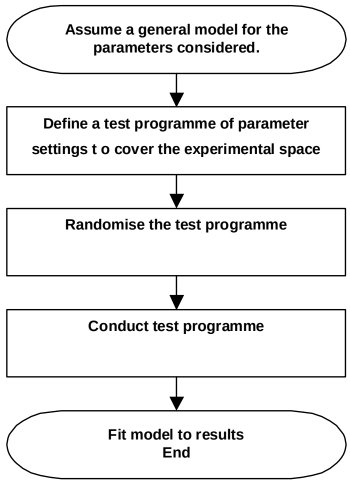


Figure 2: ICAM component, showing a Torque Spark data points with a second order fit


Assume a general model for the parameters considered.
Apply prior knowledge to the parameters to form the prior distribution, p(β). Either from previous data on similar engines or with estimates elicited from engineers.
Carry out initial testing. Form likelihood, p(y / β).
Update the model parameters.
Test for significance of model parameters to simplify model.
Decide where to test next.
Combine new test point with original data to form new likelihood.
Combine new test point with original data to form new likelihood
Test model for convergence.
Final Model and likleyhood
End


Assume a general model for the parameters considered.
Define a test programme of parameter settings t o cover the experimental space
Randomise the test programme
Conduct test programme
Fit model to results
End

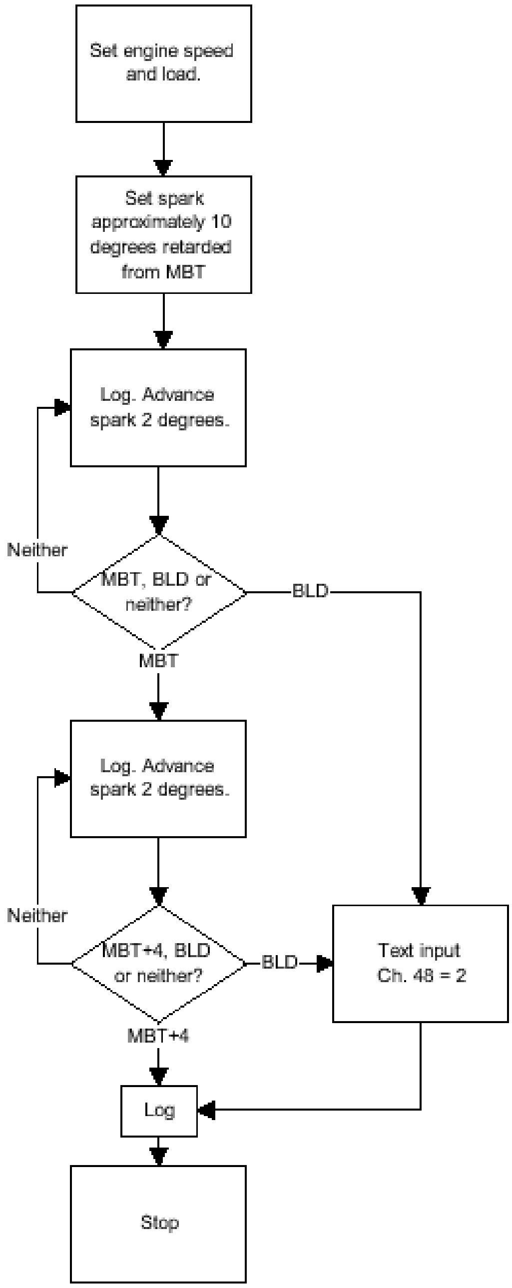


Figure 3: Innerloop Sparksweep