Cambridge Centre for Computational Chemical Engineering

University of Cambridge

Department of Chemical Engineering

Preprint ISSN 1473 – 4273

Moving Toward Establishing More Robust and Systematic Model Development for IC Engines Using Process Informatics

Andrew Smallbone [1,2] , Amit Bhave [1,2] , Andreas Braumann [2] ,

Markus Kraft [2] , Antonis Dris [3] , Robert McDavid [3]

released: 05 January 2010

[1] Reaction Engineering
Solutions Ltd. (reSolutions)
Sheraton House
Castle Park
Cambridge, CB3 0AX
UK

[2] Department of Chemical Engineering
and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge, CB2 3RA
UK

[3] Caterpillar Inc.
Applied Research Europe
Peterborough
UK

Preprint No. 92

c4e

Key words and phrases: Engine experiments, Engine modeling, Uncertainty, Optimisation, HCCI

Edited by
Cambridge Centre for Computational Chemical Engineering
Department of Chemical Engineering
University of Cambridge
Cambridge CB2 3RA
United Kingdom.

Fax: + 44 (0)1223 334796
E-Mail: c4e@cheng.cam.ac.uk
World Wide Web: `http://www.cheng.cam.ac.uk/c4e/`

Abstract

Analyzing the combustion characteristics, engine performance, and emissions pathways of the internal combustion (IC) engine requires management of complex and an increasing quantity of data. With this in mind, effective management to deliver increased knowledge from these data over shorter timescales is a priority for development engineers. This paper describes how this can be achieved by combining conventional engine research methods with the latest developments in process informatics and statistical analysis. Process informatics enables engineers to combine *data*, *instrumental* and *application models* to carry out automated model development including optimization and validation against large data repositories of experimental data. This is complemented with the inclusion of experimental error and model parameter uncertainty, to yield confidence regimes on the final model result, hence the impact of specific shortcomings of the model and/or experimental dataset can be identified in a systematic manner. A methodology for model implementation is described including an extensible *data model* for storing engine experimental data in a consistent format. Finally, a working example for an *application model* is presented through the development of a semi-empirical soot model for diesel engines.

Contents

1 Introduction

There are two major ways in which models can yield added value to any development process, 1) through education of the underlying physics in the "*real world*" by yielding data representative of "*real world*" experiments, and 2) by having some level of predictive capability. The methodology outlined in this research paper seeks to use the latest approaches in Process Informatics as a means to deliver increased predictive capability and model robustness through effective model development and design of experiments. Through this, it is anticipated that the ongoing burden of carrying out extensive experimental and validation activities can be alleviated thereby leading to reduced overall development costs and timescales.

"Process Informatics seeks to solve these problems through the integration of hardware, middleware, software, databases, and human resources, all integrated through a network" [1]. An example of a practical working system is the Process Informatics Model (PrIMe) [2] which is used for the determination of chemical kinetic rates for combustion based on an open-source database written using eXtensible Mark-up Language (XML). The objective is to reduce the uncertainty on chemical kinetic rate constants through systematic comparison with fundamental experimental data and a set of computer-based tools to process data consistently from available data sources such as the literature and ongoing research. Researchers aim to systematically identify regions in which the models are unsuccessful, then suggest to the community the most useful future experiments and thus result in more rapid development timescales [3].

The latter remarks are of particular relevance to the I.C. engine community as a whole, and to a similar extent within individual R&D groups, where similar or similar-enough research activities are regularly carried out unnecessarily at significant cost. Either because a) the researchers were not aware of the previous research, or b) they were not aware that the model solution was robust enough in these circumstances to have proven adequate. Whilst the development of a properly integrated database may well resolve a) at least in part, the latter can only be achieved when model robustness can be determined systematically, one approach would be to include uncertainties within the model result as demonstrated for chemical kinetics [4] and for granulation processes [5, 6]. With the luxury of uncertainty bounds obtained via systematic comparison with the entire relevant dataset, confidence in the robustness of the model can be assumed. Furthermore, these results can then be employed to indicate the most useful set of experimental data points to measure for further and more rapid model improvements.

As we make the transition from a data-poor to a data-rich working environment, only through careful consideration of the aspects outlined above, then new ideas for future technology development can be identified, encouraged and realised.

In this research paper, the implementation of a Process Informatics approach for future IC engine development is outlined. Details of the engineML database structure, development of a desktop and web based Graphical User Interface (GUI) and model integration are described. Finally, a simple working example is discussed for developing a model to determine exhaust soot particulates in diesel fueled engines.

2 Model Development

Conventionally, the concept of a *model* is usually considered as a set mathematic equations which describe a physical system or process, however these descriptions are usually incomplete at some level and require the adoption of "*optimisable*" parameters. Given that these parameters are coupled to the experimental data adopted in their determination, here a *model* is redefined to include the mathematic equations as well as the experimental dataset used for parametric optimisation.

Based on our new definition, a *model* can be separated into three sub-components, *instrumental models*, *data models* and *application models*. A description of these sub-components are outlined in the following sections.

2.1 Instrumental Models

Outlined in Figure 1 is a schematic of a *model*. Here the "*real world*" is represented by an engine, in order to extract out useful data for use in analysis or modeling some observations of the "*real world*" must be made. Observations are usually carried out using electronic experimental devices which are used to produce point data such as brake power, fuel consumption, exhaust emission etc., as well as time dependent data such as in-cylinder pressure measurements etc. This conversion is through storing the data and converting them to something useful via an *instrumental model*, this conversion carries some uncertainty as it is often based on a set of parameters determined from correlations, hence the error associated must be stored and its influence considered whenever adopted.

2.2 Data Models

A *data mode*l is a description of how data is stored, data models are usually inconsistent depending on the apparatus, the engineer‘s personal preference etc. Furthermore, the data can be poorly labeled and thus redundant without its corresponding experimentalist. This limits its useful lifetime and increases the perceived uncertainty and accessibility of the data. Ultimately, the potential to exploit the knowledge held within the data to facilitate other relevant development activities is reduced. Finally, due to ongoing advancement of data storage technologies, these data are sensitive to loss via computer hard drive failure, use of storage formats which are no longer supported e.g. zip drive, accidental neglect (and deletion), or even the thought that older experimental measurements are irrelevant. Given the financial costs associated will the acquisition of each data point and the knowledge of the "*real world*" held within them, any losses of these kinds are inexcusable.

2.3 Application Models

An application model is the set of mathematical equations which describe the physical processes which occur in the "*real world*". This aspect of the *model* is often the main

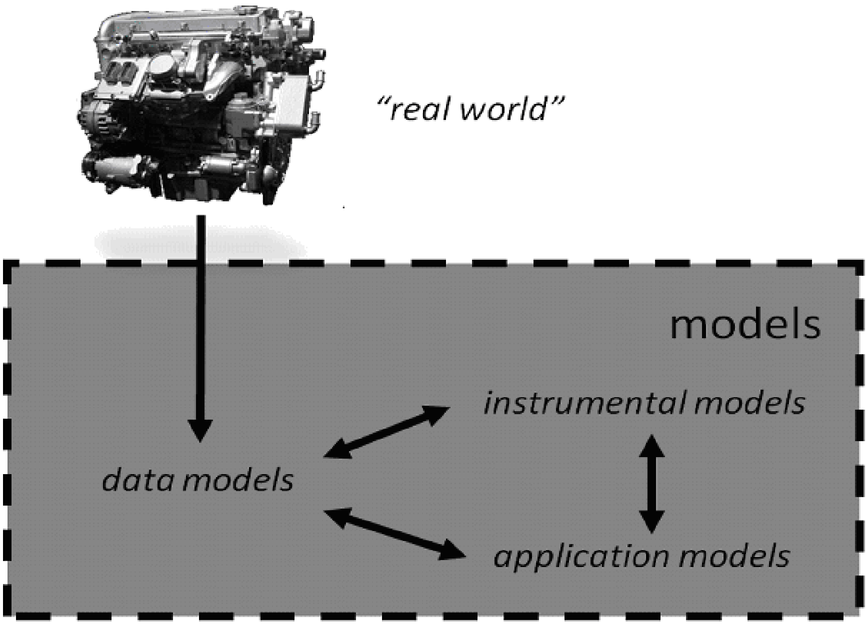


Figure 1: *Overview of the sub-components of a model.*

focus of the development process for example by extending the underlying physics from 1D to a 3D computation.

2.4 Model Validation and Optimisation

To senior engineers, the most significant aspect of model development is the validation and optimisation phase as success or failure dictates the distribution of future model development resources. Hence, measures of model success must be considered more systematically by determining both the robustness and accuracy of the model when compared to a wide range of experimental data. It is important to recognise that models have sets of parameters that require optimisation against experimental data, it is therefore necessary to embrace this fact and try to control the influence that the experimental data has on the model predictions itself. Here we distinguish five potential sources of inaccuracy of a model:

1. Errors associated with the numerical solution method of a mathematical model;
2. Errors that are based on the missing chemical or physical insight into the process that is to be modeled;
3. Errors based on the inaccuracies of the data used in a model;
4. Errors that arise because there are conflicting data for a number of parameters;
5. Errors that are associated with an inadequate experimental dataset.

If one wants to make models more robust one needs to address all of the above issues. In the past scientists and engineers have concentrated on the first two aspects. The third aspect is mainly looked at by statisticians but has not had a large impact on the way automotive engineers use their models. The fourth aspect is used by the machine learning community to a) help identify erroneous experimental/model data, and b) carry out further analysis to identify the source and thus improve the second aspect.

The final aspect is relevant to every model during the optimisation and validation, the scope of the experimental dataset. Examinations of model accuracy are a common research activity in the engine community, however these are often limited to a handful of experimental data points acquired during testing of an engine designed to meet the latest set of emission targets. The result is a limited dataset coupled with an application model which contains a large number of parameters such as CFD due to "*overfitting*", the final validation demonstrates an adequate model result compared to experiments. However, when these same model parameters are applied to other experimental data, large model errors are often reported. Indeed, any model should be optimised against whole datasets of engine data, engine types, etc. and model robustness properly assessed.

Conventionally, a metric of the accuracy of a model is often comparing the model result with the corresponding experimental data. Model accuracy is often loosely defined by assessing the deviation between these data and considering any experimental uncertainty. A metric to describe the robustness of the model should also be defined by addressing the uncertainties associated with the optimisation of model parameters against experimental data, as such an uncertainty is associated with each parameter depending upon how well known it is or the confidence and thus ultimately on the final model result. This will yield both a measure of the model accuracy as well as a measure of the robustness.

To apply these concepts to an automated model development process three major developments are required, 1) a consistent *data model* for storing and sharing experimental and model data, 2) a graphical interface to visualise the data/model inputs, 3) for *application models* to be extended to include parametric uncertainties, and 4) for the adopted optimisation routine to include experimental and parametric uncertainty.

The sections which follow detail the methods employed in applying these concepts to the engine development process.

2.5 A Data Model: EngineML

The eXtensible Markup Language (XML) is used as the fundamental code of the engine Markup Language (engineML). It is felt important that any such database structure should be built for the long term, hence XML is considered the most favourable because it has a tree-like structure, is both human and machine readable with many programming languages carrying the I/O libraries. The structure is fully extensible allowing for future engine developments in experimental or model systems to be easily added. It is also has an open standard and is platform independent making it a timeless format and thus ideal for Process Informatics approaches [2]. In addition, an XML schema can be adopted to ensure consistency between files created from multiple users and multiple programs which is important in large collaborative research activities such as engine development.

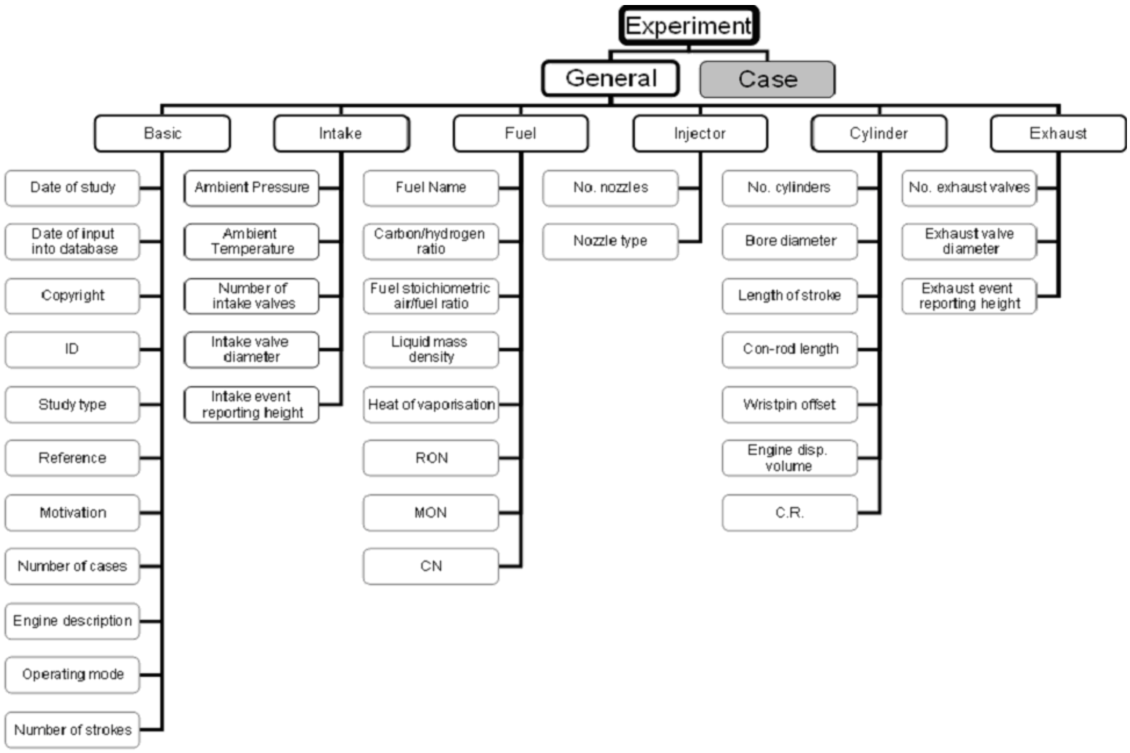


Figure 2: *General properties of engineML database structure.*

An example of the engineML structure is presented in Figures 2 and 3, data is separated into General data (independent of operating point) and Case data (dependant on the operating point), each are then divided into Basic, Intake, Injection, Fuel, Cylinder and Exhaust for simple referencing. The extensibility of the structure means that if the database does not contain a component that the user requires, it can be added without compromising the structure of the infrastructure (the additional data is compatible with the GUI too).

In order to ensure the information in the repository is adopted with full confidence by engineers, whom in many cases are not necessarily involved in the original data collection/processing, each entry into the repository must have a well defined name, value and unit. In addition, for data to be useful in the long term it is critical to properly define the apparatus and measurement devices as this ensures a comprehensive record of the experiment is held with the experimental measurements. Furthermore this information is of great importance for model developments which include error and uncertainty propagation as described in the example in this paper.

In the engineML structure, the following information can be defined if required;

- Name of data (e.g. spark timing)
- Short name of data (e.g. spk)
- Description of the data to define it precisely (Timing of the ignition spark signal)

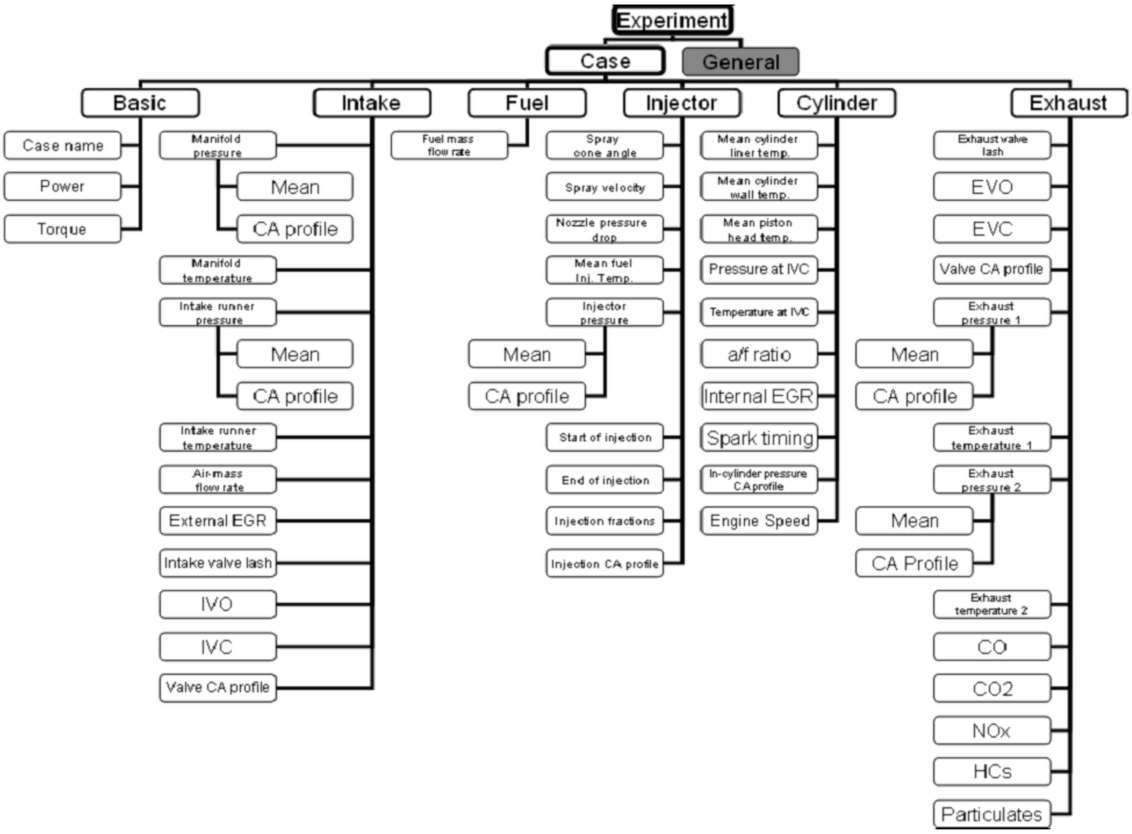


Figure 3: *Case properties of engineML database structure.*

- Value of data(e.g. +/-20)
- Uncertainty of data (e.g. 1, linked to the apparatus)
- Equipment used to take data (e.g. signal measurement device)
- Measurement location (Engine management system)
- Units (e.g. degrees)
- Relative unit (e.g. bTDC or aTDC)
- Data type (e.g. a crank angle)
- Data structure (e.g. point or time dependent)

2.6 Database Entry, Manipulation and Visualisation

For ease of transfer between data acquisition systems, modeling and visualisation tools, engine related data are often stored (or easily converted into) in MS Excel or text files. In order to convert large quantities of engine related data acquired prior to this development a conversion program, named "text2engineML" was developed. This tool can be applied, to rapidly convert engine data from spreadsheet and text file format into engineML.

To facilitate the storage, manipulation and visualisation of the engine data stored in the engineML format, a Graphical User Interface (GUI) has been developed in Java. A screenshot of the GUI is shown in Figure 4. Using this tool it is possible to open, input and save data stored in engineML. For consistency, the data are presented out in a similar structure to the engineML sheets and can be navigated using tabs which display the General and multiple Case data sheets.

It is possible to visualise the point data in tables and examine the time (or crank angle) resolved data in both table and graphical form. In addition, a tree structure can be used for the selection of the required operating points the user may wish to use in the model.

2.7 Implementation, Optimisation and Development

A schematic showing a summary of the steps involved in the model implementation is presented in Figure 5. In order to fully exploit the potential of the engineML *data model, application models* were developed such that they could extract data relevant to their initialisation directly from the repository, for example engine speed, injection timing etc. The latest set of parameters were available via an engine parameter repository stored in a shared location. This file contained the parameters, some of which were common to a number of other *application models*. Once the input file was generated, the *application model* was executed and its results stored in the repository using the engineML *data model*, due to the adoption of a consistent storage format this meant that they could be routinely and automatically compared against experimental data.

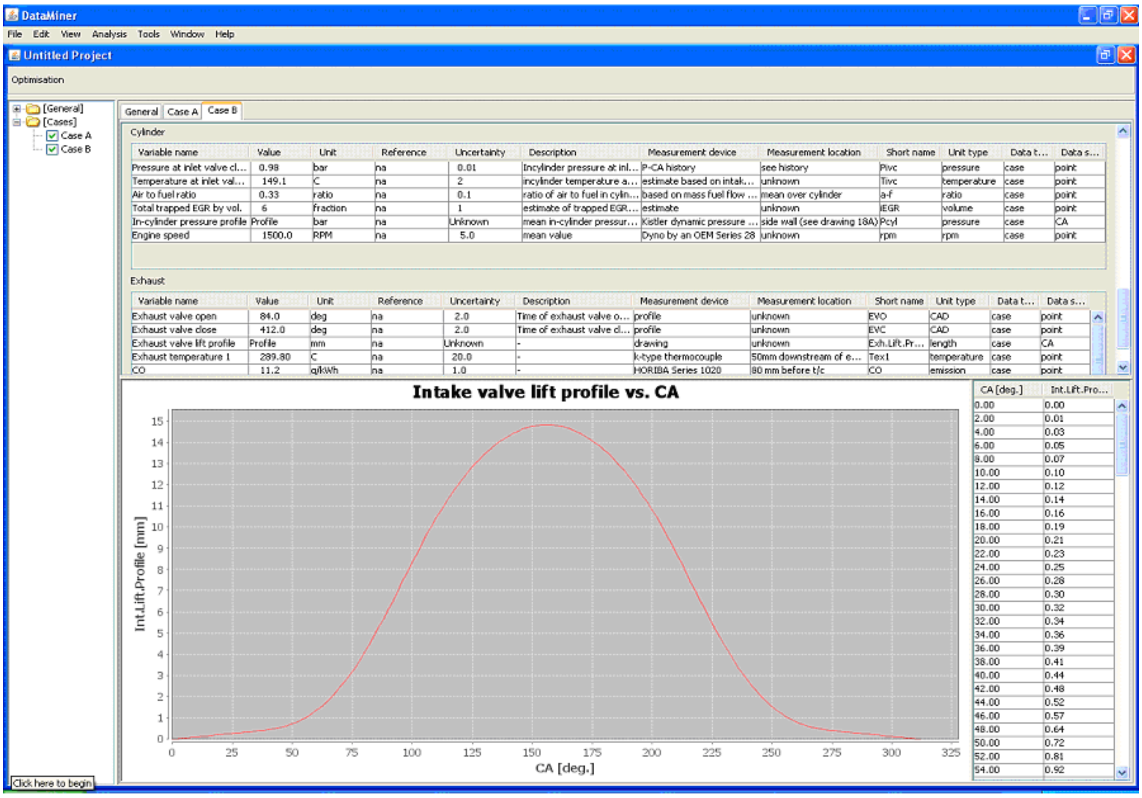


Figure 4: *Screenshot of the graphical interface displaying the data stored in an engineML sheet.*

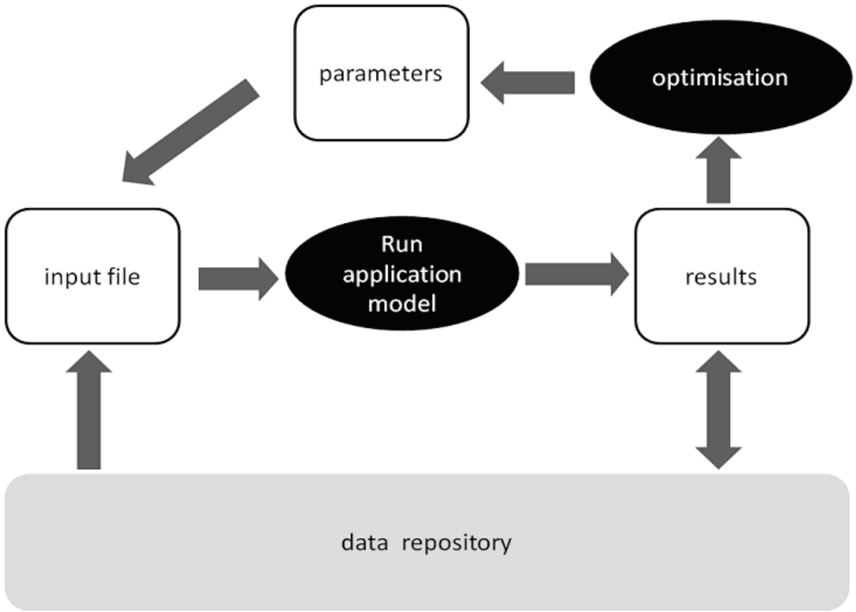


Figure 5: *Schematic of the model implementation phase.*

Next a routine was set up for carrying out parametric optimisation, the repository was accessed and used for automated validation.

2.8 Model parameter optimisation including experimental error and uncertainty propagation

In order to formally assess the robustness of a model, the uncertainty of model parameters much be identified. Furthermore, in the process of model optimisation the influence of experimental error must also be considered [3–6]. In order to include these aspects into a model, experimental observation, η^{exp} is characterized by two parameters, the measured value, η_0^{exp} and its uncertainty, σ^{exp} [7]. Hence, the experimental observation is written as,

$$\eta^{\mathrm{exp}} = \eta_0^{\mathrm{exp}} \pm \sigma^{\mathrm{exp}} \,. \tag{1}$$

The model response, η will be required for the optimisation. This model response is a function of the model parameters, x

$$x = (x_1, \ldots, x_k) \,, \tag{2}$$

so that

$$\eta = \eta(x) \,. \tag{3}$$

Due to experimental uncertainty, σ^{exp} more than one set of x^* will lead to agreement between experimental observation, η^{exp} and model response, η. In order to resolve this, some uncertainty is allowed within the description of each model parameter, x

$$x = x_0 + c\xi \,, \tag{4}$$

with x_0 as the base value, an uncertainty factor, c and the random variable, ξ. As a result, x becomes a random variable. The random variable, ξ shall be a standard normally distributed. Hence the model response, η can now be written as,

$$\eta(x) = \eta(x_0, c, \xi) \,. \tag{5}$$

Due to the dependency on the random variable, ξ. the model response also becomes a random variable. However, the model prediction should be represented by just one value and its associated uncertainty. Taking the expectation of the model response and the variance respectively, results in the model prediction, $\overline{\eta}$

$$\overline{\eta}(x_0, c) = E\left[\eta(x_0, c, \xi)\right] \,. \tag{6}$$

and the corresponding model uncertainty, σ

$$\sigma(x_0, c) = \sqrt{\text{var}(\eta(x_0, c, \xi))} \,. \tag{7}$$

It is not always straightforward to derive $\overline{\eta}$ and σ for many *application models* as they are described by differential equations or are simply too complex to derive analytically. Hence here we have adopted a model response surface for local ranges of the unknown parameters, x. Hence to identify these ranges, the first task is to approximate the parameter, x_0 by carrying out a conventional optimisation by solving the objective function, Φ_1

$$\Phi_1(x_0) = \sum_{i=1}^{N} \left[\eta_i^{\text{exp}} - \overline{\eta_i}(x_0)\right]^2 \,, \tag{8}$$

with $i = 1, \ldots, N$ being the number of experimental observations using an initial estimate of these parameters (either using a sequencing algorithm or by some prior knowledge of the parameters) and then carrying out an minimisation to obtain the best set of parameters, x_0^*.

$$x_0^* = \underset{x_0}{\text{argmin}}(\Phi_1(x_0)) \,. \tag{9}$$

This minimisation was carried out adopting a Levenberg-Marquardt non-linear least squares minimisation algorithm. A linear response surface optimisation is then performed around the point, x_0^* to optimize the result further and to estimate the uncertainties in the parameters and in the model response. A simple linear model response, $\eta(x)$ can be approximated by,

$$\eta(x) = \beta_0 + \sum_{k=1}^{K} \beta_k x_k \,, \tag{10}$$

with β_0 and $\beta_k(k = 1, \dots, K)$ being the parameters of the response surface. In this study, these were determined numerically by evaluating the model response using a finite differences method. This was carried out by fitting the linear model parameters β_0 and $\beta_k(k = 1, \dots, K)$ to the model response at x_0^* and those obtained for 1% deviations in each parameter.

After substituting Equation (4) into Equation (10) the model response becomes

$$\eta(x_0, c, \xi) = \beta_0 + \sum_{k=1}^{K} \beta_k (x_0 + c\xi)_k \,. \tag{11}$$

The mean value, $\overline{\eta}$ and the uncertainty, σ are obtained after taking the expectation and the variance respectively

$$\begin{aligned} \overline{\eta}(x_0) &= E\left[\eta(x_0, c, \xi)\right] \\ &= \beta_0 + \sum_{k=1}^{K} \beta_k x_{0,k} \,, \end{aligned} \tag{12}$$

and the model uncertainty

$$\begin{aligned} \sigma(c) &= \sqrt{\mathrm{var}(\eta(x_0, c, \xi))} \\ &= \sqrt{\sum_{k=1}^{K} \beta_k^2 c_k^2} \,. \end{aligned} \tag{13}$$

This means that the most likely model solution is within

$$\eta(x) = \overline{\eta}(x_0) \pm \sigma(c) \,. \tag{14}$$

The task is then to find the optimal values of both x_0 and c using an objective function. This time the aim of the objective function is 1) obtain a prediction as close to the experimental observation as possible, and 2) find the uncertainty of the prediction itself. The objective function, Φ_2 takes the following form

$$\Phi_2(x_0, c) = \sum_{i=1}^{N} \left(\left[\eta_i^{\mathrm{exp}} - \overline{\eta_i}(x_0)\right]^2 + \left[\sigma_i^{\mathrm{exp}} - \sigma_i(c)\right]^2 \right) \,, \tag{15}$$

and is minimized with respect to both x_0 and c starting with the initial set of parameters, x_0^* and corresponding adopted uncertainties, c.

$$(x_0, c)^* = \underset{x_0, c}{\operatorname{argmin}}(\Phi_2(x_0, c))\,. \tag{16}$$

The minimisation was carried out adopting a Levenberg-Marquardt non-linear least squares minimisation algorithm with bounded constraints [8].

3 Example for Application to Engine Model Development

3.1 Data repository

A repository was developed using the engineML data model in collaboration with Caterpillar Inc., a summary of these data are presented in Table 1. Experimental engine data from six engines and a total of over 400 operating points were included in the repository.

Table 1: *Experimental engine data repository summary*

	Number of operating points	**Description of data**
Engine A	79	Design of experiments project
Engine B	384	Major design of experiments project covering significant parametric sweeps
Engine C	17	Individual engine development
Engine D Engine E Engine F	23	Individual engine developments

3.2 Soot model development

In order to demonstrate the effectiveness of the proposed methodology in establishing more robust model developments an example of an *application model* for soot was employed. Soot modeling was adopted as the ideal example problem for three major reasons, 1) as the experimental measurement uncertainty is high relative to the magnitude of the observation at 0.05g/kW-hr per standard deviation, 2) it is widely modeled by the automotive community using empirical methods and, 3) since it is an empirical expression any computation effort associated with its solution during the optimisation is small. An *application model* for soot based on fundamental principles would be expected to include the physics to describe in-cylinder thermodynamic and combustion coupled with the chemistry of soot formation and nanoparticle interaction and growth. Due to this level of complexity, the soot emission modeling (and many other equivalent application models for emissions, autoignition and flame propagation) still adopted widely by the engine development community are based on simple empirical relationships developed in the 1970s and 1980s. In this study, as an example a "*representative expression*" similar to the *application model* described by Plee et al. [9], has been adopted. The expression for the adopted "*representative expression*" is outlined below;

$$\mathrm{soot}[g] = A\,\overline{S_p}^{B}\phi^{C}\exp\left(\frac{D}{T_f}\right)\,, \tag{17}$$

where $\overline{S_p}$ is the mean piston speed, ϕ is the global equivalence ratio, T_f is the adiabatic flame temperature based on the composition of the air-fuel mixture and trapped/recirculated exhaust gas. A, B, C and D are parameters. Here the adopted parameters hold little real underlying physical meaning (such as chemical kinetic rate parameters). As such, any constraints for feasible ranges are not applicable other than orders of magnitude. However the methodology can be employed such that parameters are constrained to such ranges, which is ideal for less empirical adoptions e.g. chemical kinetics [3]. The initial parameters, A, B, C and D and their single standard deviation uncertainties were set to 500 ± 1000, 1.8 ± 10, 2 ± 10 and -1000 ± 2000 and the procedure outlined above was carried out.

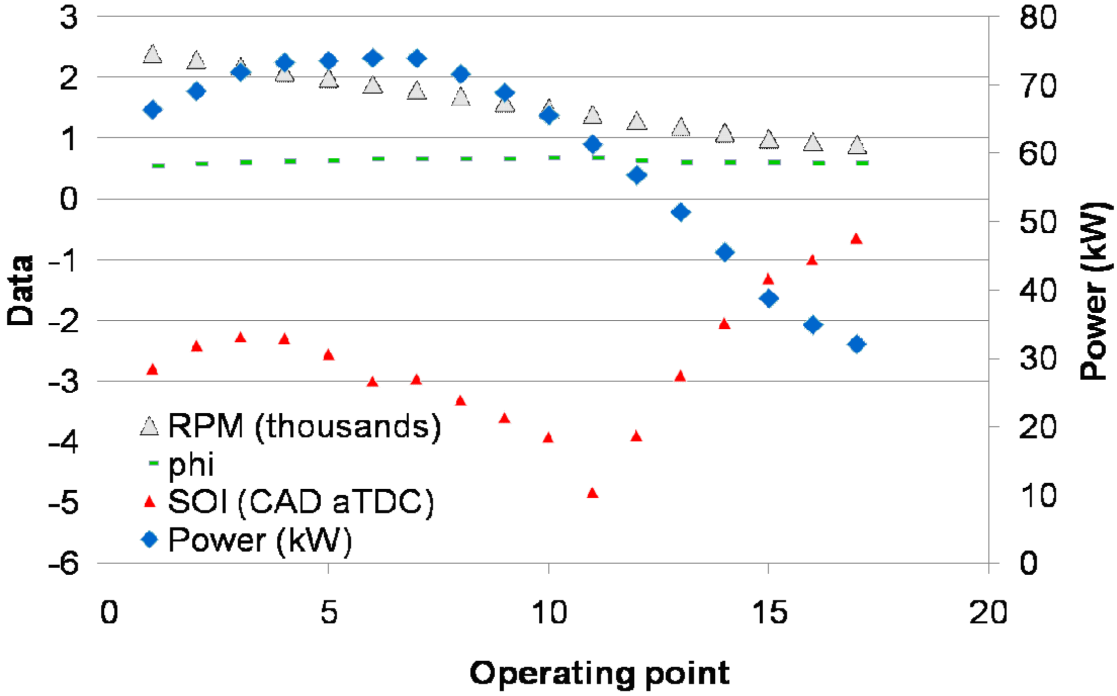
3.2.1 Small experimental datasets

Initially, to examine the influence of adopting a small number of datasets for model parametric optimisation, the data from Engine C was employed for the experimental data. Presented in Figure 6(a) are the key experimental data of RPM, ϕ (equivalence ratio), SOI (Start Of Ignition) and engine power, for the 17 operating points. Presented in Figure 6(b), is the soot formation rate for the corresponding operating points with the experimental uncertainty.

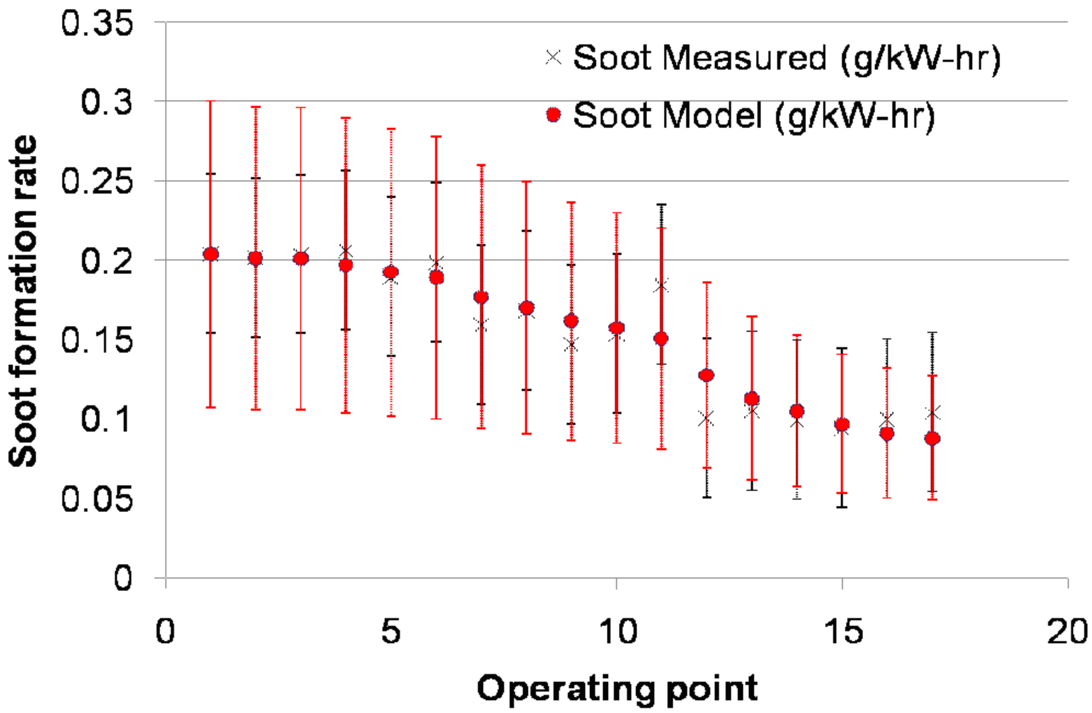
The optimisation of the "*representative expression*" was carried out using the experimental and model parametric uncertainty minimisation approach outlined above and in more detail elsewhere [3–5]. The optimisation was carried out systematically, initially with 8 operating points (the minimum was eight as there are four model parameters and four parametric uncertainties) and then by increasing the number of operating points in the optimisation to seventeen. The result of the optimisation is presented in Figure 7.

For each parameter, as the number of operating points in the optimisation were increased, the parametric uncertainty associated with that parameter was reduced. This indicated that in this case, the model and the experimental data were consistent and the model proved a sufficient interpretation of the response of the engine.

The estimated parameter values and their uncertainties can be used for the calculation of the model predictions and their uncertainties (Figure 6(b)). Using conventional model optimisation methods, the mean model response can be compared directly with the experimental results and here, in all cases the response is to well within the experimental uncertainty. By adopting the proposed methodology additional information is now available in terms of the model parametric uncertainty. In this case, by adopting the objective function (Eqution(15)) during the optimisation, the model uncertainty has been reduced to the same order of magnitude as the experimental uncertainty. Due to the exponential term the



(a) Key experimental data



(b) Measured and modeled soot formation rates, error bars represent 1 standard deviation

Figure 6: *Experimental data for Engine C.*

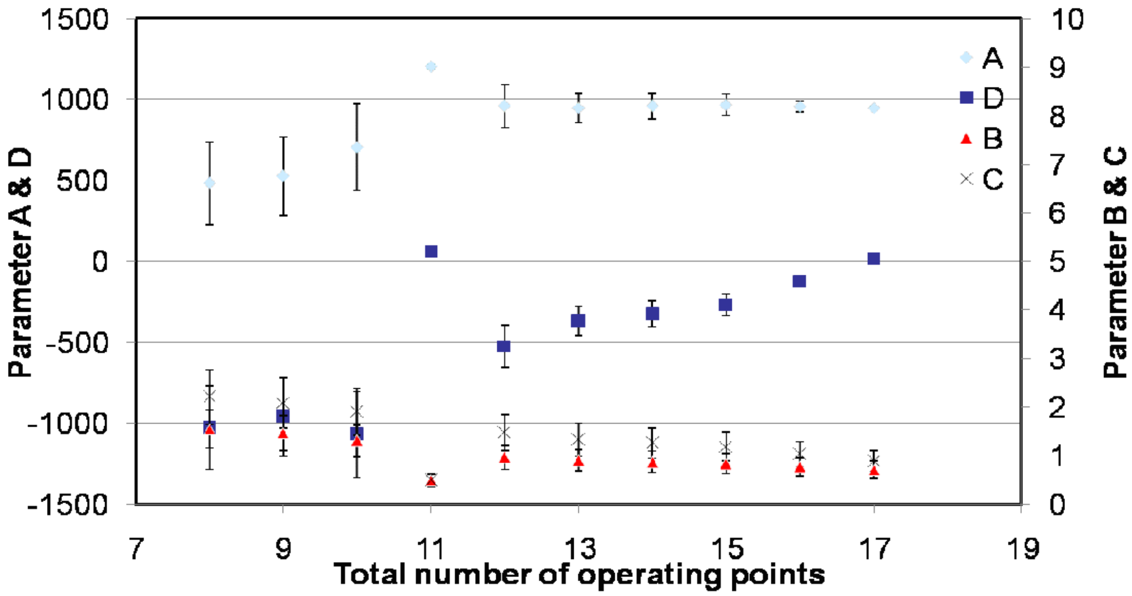


Figure 7: *Obtained model parameters and corresponding uncertainty for an increasing quantity of included operating points. Error bars represent 1 standard deviation.*

uncertainty proved greater for those cases with greatest soot formation rates. Operating point eleven proved to be an interesting data point, here the optimisation procedure was forced to readjust to a new set of parameters (cf. Figure 7). A hypothesis for this might be that this point was either a) an outlier, or b) included some additional physical component which had not been observed in the previous included data points. The experimental key variables presented in Figure 6(b) demonstrate that point 11 has the earliest start of injection, an aspect not included directly in the model. Furthermore, in the parameter estimation with all 17 data sets, parameter C carried the greatest uncertainty, suggesting that perhaps if further developments were to be carried out to the application model, that inclusion of the physics associated with fuel stratification and mixture preparation would most likely prove most effective.

3.2.2 Larger experimental datasets

The total number of operating points used in the optimisation were increased to include the first 36 operating points from Engine A. The result of this optimisation on the parameters is presented in Figure 8, here Parameters A and B remained relatively constant, whilst C and D found new values, however the most significant aspect was the increase in associated parametric uncertainty for B and D. Whilst these observations could be associated with the adoption of a linear response surface (rather than using quadratic equivalents), this suggests that this model is an inadequate interpretation of these two engines. A hypothesis for these observations might be through further development in terms of the mixing (through mean piston speed being proportional to turbulence [10] and chemistry (being approximated by the adiabatic flame temperature term) terms.

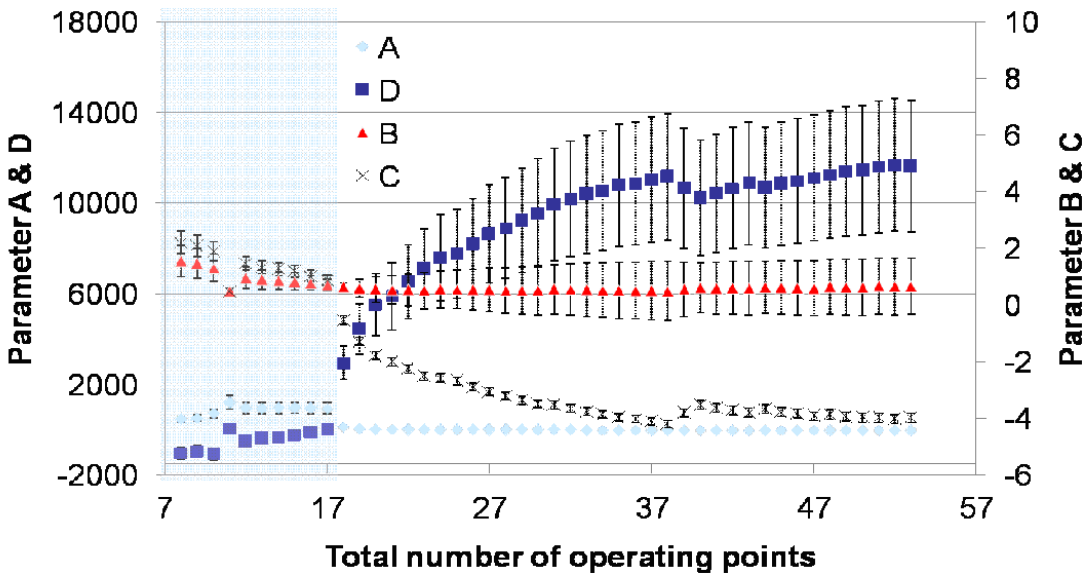


Figure 8: *Obtained parameters and corresponding parametric uncertainty for an increasing quantity of experimental data, the highlighted region represents the data presented in Figure 7. Error bars represent 1 standard deviation.*

Presented in Figure 9 are the final model predictions for the two datasets employed so far. In most cases, when the model uncertainty was also considered, the model predictions were to within the observed experimental uncertainty. Operating Points 39 and 40 were well outside the model and experimental uncertainties, as such further tuning of model parameters would not yield improved results in these cases. Knowing this, it is then possible to focus research efforts to determine whether the source was i) due to erroneous experimental error, or ii) due to some inadequacy in the underlying physical assumptions.

When this same model was optimised against the whole engine dataset of more than four hundred operating points as outlined in Table 1, a 91% model accuracy, that is when model and experimental uncertainty ranges overlapped was obtained, Figure 10. However, the average model uncertainty, presented in Figure 11, was ten times greater than the experimental uncertainty, that is that these model assumptions can only be fitted to these data if large uncertainty ranges are associated with the parameters. This re-enforces the notion that the "*representative expression*" can be used as an adequate simplified interpretation of small datasets but is not robust enough to be applied to all engines.

3.3 Further model development

The reSolutions Soot Post Processor Model (rSPP) is an ongoing development of a system level soot emissions *application model* much like the "*representative model*". It was developed using the approach outlined above by seeking to improve model robustness against databases of engine data by identifying the parameters with greatest uncertainty and selectively improving those assumptions in the model which were neglected but actu-

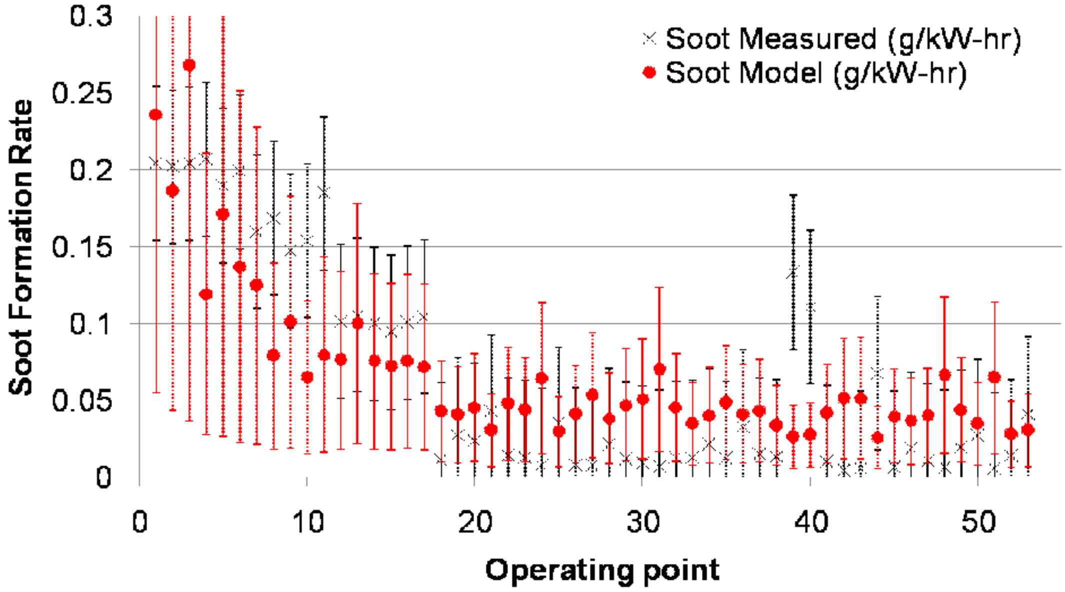


Figure 9: *Measured and modeled soot formation rates for the two data sets (all data of engine C and first 36 points of engine A).*

ally proved critical to the soot formation process. As new experimental data are measured, these are used to extend the data repository and thus the knowledge retained in the model.

The rSPP model was developed by identifying six operating regimes in which particular chemical-thermo-physical processes dominate. The experimental data are separated into these regimes, and within each, an optimisation was carried out to determine key parameters. The resulting model accuracy and uncertainties for the whole repository presented in Table 1 are presented in Figures 10 and 11.

The model performance varied depending on the operating regime, however an overall model performance of 83% accuracy was determined for the whole engine database. Most importantly, the resulting model parametric uncertainty proved significantly better than the '*representative expression*" and around half that of the experimental uncertainty. Hence if compared on a *like-for-like* basis, the "*representative expression*" would only yield a 28% accuracy percentage. This demonstrates the rSPP model to be a sufficient interpretation of the underlying physics of soot formation in the six engines and over four hundred operating points adopted in this study.

4 Discussion

The present state-of-the-art in engine model development requires a more formal approach, as there are too many complex models and too much data for engineers to identify the most appropriate and the most effective tool for their application. Furthermore, a culture of "*re-optimize the model against small datasets*" rather than improving the physics of the model and the validation process often undermines the long term objective of de-

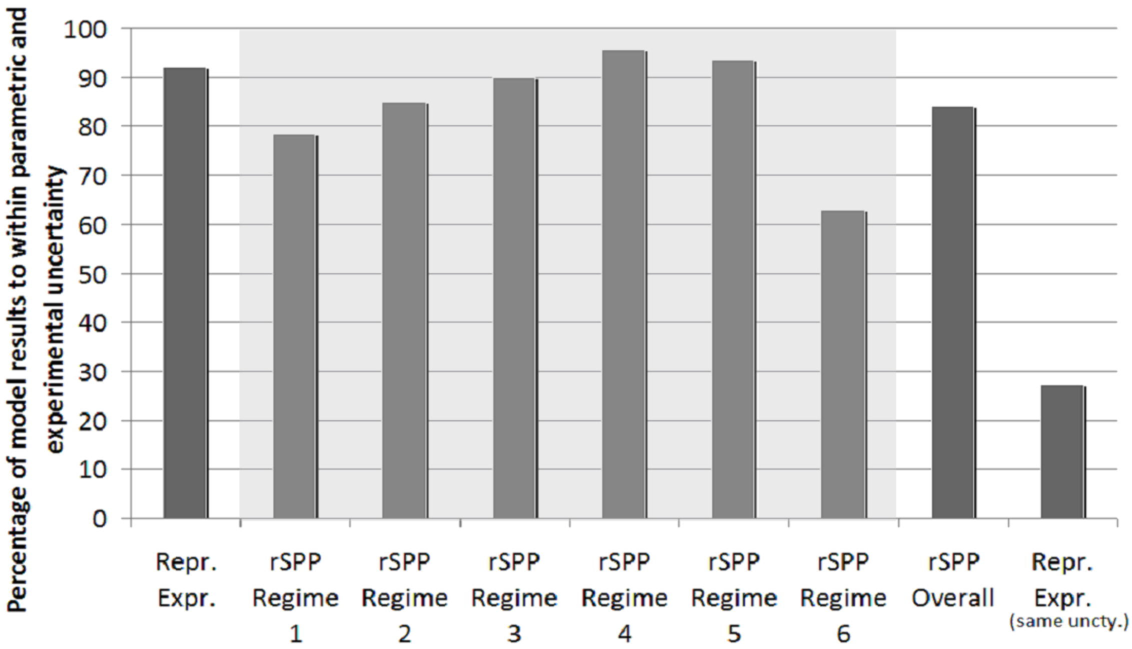


Figure 10: *Percentage of the data from each model within experimental error and model uncertainty (Repr. Expr. is the "representative expression").*

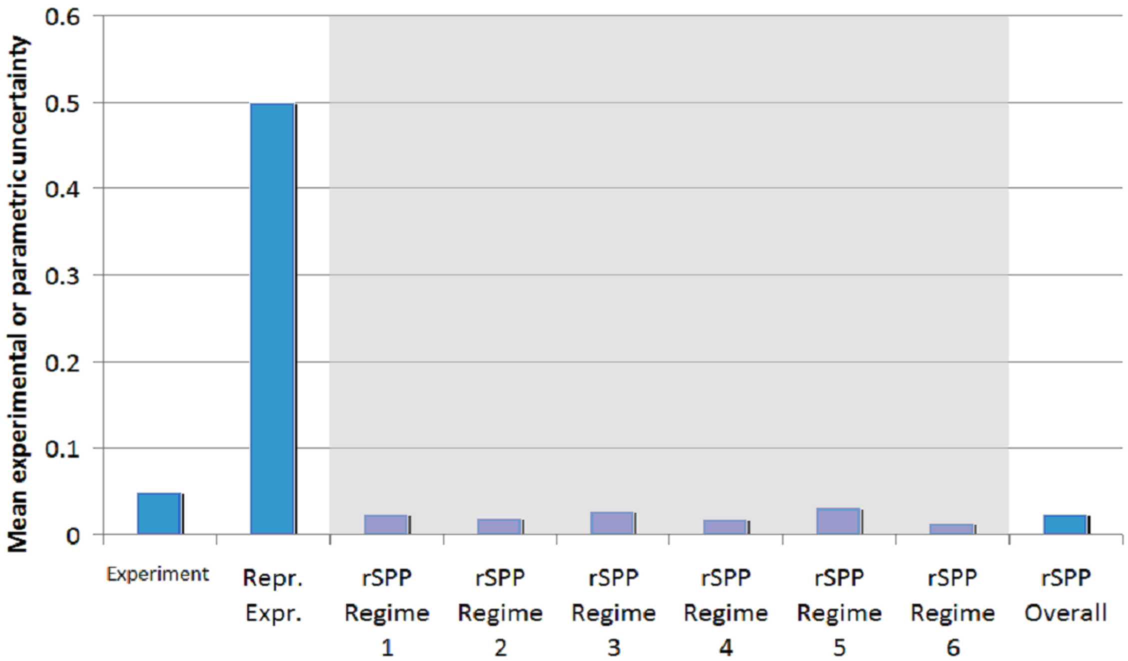


Figure 11: *Resulting average experimental and model uncertainties (Repr. Expr. is the "representative expression").*

livering useful predictive and robust models.

The methodology outlined in this research demonstrates the adoption of the "*representative expression*" is adequate to describe small datasets, however they not robust enough to deal with the wide range of variables that exist in larger datasets. By creating a *data model* such as engineML, experimental data is retained whilst still being easily accessible to models, modelers and experimentalists in the long term. These data can easily be included in any future model parameter optimizations and thus *application models* such as the "*representative expression*" can be examined formally and systematically. By carrying out systematic model testing, decision makers can easily identify those aspects of engine models or modeling which require further development and those which do not, they can then allocate the resources to the most critical path. This will ultimately lead to a new set of more robust models such as the rSPP.

The next step is to use the response of these models to enable intelligent design of experiments by determining the next "*most useful*" set of experiments to be carried out. That is to identify a number of experimental operating points which would yield the most effective contribution towards the model development process by reducing the parametric uncertainty - this approach would enable for the experimental process to be streamlined thus reducing the development timescales and costs.

5 Summary

A methodology for the development of more robust engine models by including the latest developments in process informatics and statistical analysis was outlined. To achieve this, the following technologies were developed,

1. An extensible *data model* named "engineML" for rapid data access and storage.
2. A tool to convert existing datasets to engineML
3. A GUI tool for visualisation and manipulation of the database.
4. A technique to determine model robustness via carrying model and experimental uncertainty into the optimisation procedure was outlined.

Using the infrastructure, an example of a "*representative expression*" was validated and analyzed against a small, a larger and the complete database. This process demonstrated that *application models* of its kind can easily be fitted to small datasets but carry inadequate robustness to deliver equivalent results across a larger dataset. New insights into sources of model uncertainty were discussed.

The methodology was then employed to deliver a robust model for system level soot modelling applications.

Acknowledgements

Technology Strategy Board (TSB) and One NorthEast, UK (Validation of Complex Systems (VoCS) grant programme).

References

[1] V. Venkatasubramanian. Drowning in data: Informatics and modeling challenges in a data-rich networked world. *AIChE Journal*, 55(1):2–8, 2009. doi:10.1002/aic.11756.

[2] M. Frenklach et al. *http://www.primekinetics.org*.

[3] M. Frenklach, A. Packard, P. Seiler, and R. Feeley. Collaborative data processing in developing predictive models of complex reaction systems. *International Journal of Chemical Kinetics*, 36(1):57–66, 2004. doi:10.1002/kin.10172.

[4] D. Sheen, X. You, H. Wang, and T. Lovas. Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion. *Proceedings of the Combustion Institute*, 31:1:535–542, 2009. doi:10.1016/j.proci.2008.05.042.

[5] A. Braumann and M. Kraft. Incorporating experimental uncertainties into multivariate granulation modelling. *Chemical Engineering Science*, 65:1088–1100, 2010. doi:10.1016/j.ces.2009.09.063.

[6] A. Braumann, P. L.W. Man, and M. Kraft. Parameter estimation in a multidimensional granulation model. *Powder Technology*, 197:196–210, 2010. doi:10.1016/j.powtec.2009.09.014.

[7] J.R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, 2nd edition, 1997.

[8] M. Lourakis. Levenberg-marquardt nonlinear least squares algorithms in C/C++. 2004. `www.ics.forth.gr/~lourakis/levmar/`.

[9] T. Ahmad and S.L. Plee. Application of flame temperature correlations to emissions from a direct-injection diesel. *SAE Technical Paper Series*, (831734), 1983.

[10] J.B. Heywood. *Internal Combustion Engine Fundamentals*. McGraw Hill, 1988.