The future of computational modelling in reaction engineering

BY MARKUS KRAFT* AND SEBASTIAN MOSBACH

Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK

In this paper, we outline the future of modelling in reaction engineering. Specifically, we use the example of particulate emission formation in internal combustion engines to demonstrate what modelling can achieve at present, and to illustrate the ultimately inevitable steps that need to be taken in order to create a new generation of engineering models.

Keywords: computational modelling; reaction engineering; automation

1. Introduction

The use of computational modelling in engineering has gained increasing momentum over the past three decades. The main cause for this has been the massive advance in computer power. Industry is making use of computational models to increase the speed of technical development, which is an important factor in the overall competitiveness of a company. Environmental considerations, such as global warming and pollution reduction, demand constant development of new products for an ever changing set of constraints defined by the market and government. It is clear that computational modelling will play an increasingly significant role in this optimization process, as it reduces the cost of development. Here, we will focus on the use of computational models describing particulate formation in internal combustion engines (ICEs) as an example. Soot formation in engines is not only environmentally relevant, but exemplary for numerous other particulate processes in chemical engineering. Furthermore, the encountered modelling challenges and solutions that are described in the following apply to many other areas as well. Although there has been a great deal of progress in applying computational methods to engineering problems, there are some important difficulties with the current approach, which prevent further progress.

The purpose of this paper is to discuss how computational modelling is evolving in the near and not so near future. Starting with the present, we describe the state of the art of modelling soot formation in ICEs across multiple length scales. Using

*Author for correspondence (mk306@cam.ac.uk).

One contribution of 18 to a Triennial Issue ‘Visions of the future for the Royal Society’s 350th anniversary year’.
this example we discuss the limitation of this approach by defining what a model is on a more formal level. We proceed to show how statistics, standardization of data and models, data collaboration over the World Wide Web, and the automation of experiments will lead to a new generation of more robust models. We argue that these developments will eventually lead to robot engineers that support model building.

2. Modelling across multiple length scales

In this section, we show how multi-scale modelling has been implemented for ICEs and also suggest improvements. One can separate the problem into the micro- or molecular, meso- or nanoparticle, and macro- or continuum scales.

(a) Complex reaction networks

Gas-phase combustion chemistry can be investigated using detailed chemical models sometimes called mechanisms. These mechanisms are collections of species and reactions, i.e. detailed descriptions on the molecular level. Such a model for a fuel can contain up to several thousand reactions whose rates are given by the so-called Arrhenius expression containing at least three constants. There are two major difficulties with this approach. Firstly, the determination of the rate constants, and secondly, despite the size of the models, there may still be a number of important pathways missing. The latter problem can be addressed by generating the reaction networks automatically and simultaneously reducing them (Green 2007).

(b) Quantum chemistry and statistical mechanics

The reaction rate constants can be obtained from either parameter fitting to specially designed experiments, or directly from quantum calculations. Using results from these calculations, the rate constants and thermodynamic properties of molecules can be calculated using statistical mechanics and (variational) transition state theory. A number of software packages are available offering various degrees of accuracy and functionality. The most popular ones are based on density functional theory and can be used to calculate the geometry and electronic energy of a molecule. Transition states can be calculated from the same packages by searching for saddle points on the potential energy surface. Figure 1a shows a transition state of an oxygen molecule reacting with a polycyclic aromatic hydrocarbon (PAH) molecule (Celnik et al. 2009b).

(c) Molecular dynamics and kinetic Monte Carlo simulation

For larger molecules and clusters of molecules, quantum calculations become unfeasible. Instead, molecular potentials can be fitted to the potential energy surfaces of individual atoms. The most commonly used of these is the Lennard-Jones potential. These potentials can then be used in various ways, often providing excellent approximations to electronic structure calculations. Figure 1b shows the result of a basin hopping algorithm with which an energetically favourable structure of 50 coronene molecules has been calculated.
Review. Models in reaction engineering

Figure 1. Multi-scale modelling at micro- and meso-scales. (a) Oxygen reacting with PAH molecule; (b) PAH cluster, scale bar 1 nm; and (c) soot aggregate.

(Totton et al. 2010). Particles of this size play an important role in the formation of soot. The dynamics of small molecular clusters can be described using molecular dynamics (MD) techniques, where the equations of motions are solved directly. For large clusters, coarse graining of the molecular potentials is necessary. The chemical growth of large molecules or particles, i.e the chemical reactions with the surrounding gas, can be modelled with a kinetic Monte Carlo simulation (Raj et al. 2009).

(d) Population balance

The dynamics of the particle population is modelled on a meso-scale using a multivariate population balance equation, which is coupled to the gas-phase chemistry. In the past, each particle used to be described simply by mass and surface area. However, progress in the numerical treatment of population balance equations has made it possible to extend this description considerably to include a large number of internal variables, such as the chemical composition and the spatial structure of an aggregate (Celnik et al. 2009a). Figure 1c shows a soot aggregate as an example.

(e) Fluid dynamics

Transport of the gas and particulate phase on the macro-level is determined by conservation equations for all chemical components, momentum and energy. In an ICE, the flow is non-stationary and turbulent, and the geometries are complex, which means that a direct numerical simulation of the problem is currently not practicable. Computational fluid dynamics (CFD) software packages employ a variety of turbulence models, which can be solved for a complex geometry. A major challenge is that in modern engines fuel is injected into the engine cylinder in liquid form as spray. This process is only poorly understood, which means that spray models tend to be empirical to a large extent and require fitting to the specific conditions that are to be modelled. Figure 2a shows a CFD simulation of a spray injection in a premixed charge compression ignition (PCCI) engine (Cao et al. 2009).
Figure 2. Multi-scale modelling at macro-scale. (a) Spray in an ICE; and (b) soot PSD in a DISI engine.

(f) Reaction engineering models

Presently, it is not possible to include detailed fuel and particle models into CFD simulations without making significant simplifications. For this reason, classical models from reaction engineering are combined with CFD. These models make some strongly simplifying assumptions on spatial distributions to render the corresponding mathematical equations numerically tractable. Multi-zone models as well as stochastic reactor models (SRMs) have been widely used: for example, Mosbach et al. (2009) used an SRM with a detailed chemistry and soot model. Figure 2b shows the result of an SRM calculation that includes a detailed chemistry model for a real fuel coupled to a detailed model for soot particles to predict emissions in a direct injection spark ignition (DISI) engine. The boundary conditions for SRMs can be obtained from CFD simulations.

3. General description of a model

The previous section illustrated how a variety of models on different scales can be combined into one model for an ICE. One important purpose of a model in engineering is to inform a user about the properties of an industrial application and to make some forecasts about the system based on how the model behaves when certain settings are changed. Ultimately, this can be used to improve upon the current design or settings. Although multi-scale modelling has been a partial success, there are a number of open problems. In order to analyse these problems, it is helpful to formalize the notion of a model.

(a) Characterization of a (multi-scale) model

The schematic in figure 3 shows important elements of a comprehensive multi-scale model, which is called model $k$ to indicate that it can be considered part of a sequence of models in which each model aims to improve upon its predecessor. The model itself comprises application models (AMs), instrumental models (IMs) and data models (DMs), together with a choice of parameter estimation methodology and numerical methods. An AM is directly inspired by the physics underlying the

Phil. Trans. R. Soc. A (2010)
phenomenon we are interested in. For example, application model one, AM($i_1$), concerns one of the molecules in the chemical mechanism, AM($i_2$) describes the fuel chemistry as a whole, AM($i_3$) describes soot particle inception, AM($i_4$) describes the heat transfer, and so on. AMs may be nested in other AMs, each containing a set of parameters. As the parameters are not known exactly, it is necessary to specify an uncertainty for each parameter. This can either be a deterministic range, an interval in which the parameter must lie or a probability density. Each AM is formulated in terms of mathematical equations, mainly ordinary or partial differential equations, but also stochastic processes, as in the case of an SRM or the population balance equation. The numerical method for each AM also needs to be specified, including the chosen numerical parameters. AMs need validation and, for this reason, they are fitted to experimental data. In the case of an ICE, various experimental configurations are used for validation: engine tests for validating the final engine model, laminar flame and shock tube studies for validating the chemistry submodels, etc. In a flame experiment, a typical quantity of interest is the concentration of a chemical species. Using laser techniques, fluorescence signals are measured. This data are described by a data model, which specifies the experimental apparatus and settings, the raw data and the systematic error associated with the measurement. From the raw data, the information of interest is then derived using an instrumental model. For example, a light intensity obtained from an extinction measurement is used to calculate soot volume fraction from Mie theory using assumptions on the refractive index of soot particles. The instrumental models, as the application models, are based on physical insight and formulated in terms of mathematical equations. The model $k$ is finally specified by the mathematical fitting techniques and the objective function that are used to specify parameters of the AMs by minimizing the objective function, i.e. the ‘difference’ between AMs and IMs.

(b) Problems with the present approach

Ideally, when one moves from model $k$ to model $k+1$, one expects an improvement in predictive power. At present, this is unfortunately not guaranteed, partly because the improvement cannot be measured. The lack of
predictive power is mainly caused by incomplete information and non-systematic model building. As pointed out earlier, the number of parameters frequently far exceeds the number of experimental data available for validation. This means that the model can almost always be fitted to experimental results. Therefore, falsification of model assumptions and true model development is not possible. In the literature, raw data are almost never published, which means that the instrumental model and the data model cannot be separated and conflicting assumptions in AM and IM may exist. Also, the IM is often not fully specified, which makes comparison between AM and derived experimental data impossible. Detailed description of the experimental set-up, to the extent that the initial and boundary conditions for the AMs can be fully specified, is often missing in journal articles. The lack of error propagation constitutes another common problem. This problem occurs because errors in raw data are not propagated through to the IM and are not incorporated into the objective function that is used to fit AMs to IMs. Even when experiments and AMs are specified, the numerical treatment of the mathematical equations is often unsatisfactory. In the engineering community, there are many examples where the numerical method is intermingled with the model formulation, which makes reproduction of the results very difficult.

(c) How can these problems be solved?

In the next three sections, we shall introduce the techniques that are necessary to alleviate some of these problems. The World Wide Web, modern statistics, automated experiments and, most of all, massive computing power will play a key role.

4. The role of statistics

Statistics and operational research offer a number of concepts that are useful in model formulation, improvement and selection. In statistics, both the frequentist and the Bayesian approaches can be adopted. In Bayesian statistics, the parameters are viewed as random variables.

(a) Model reduction

As described earlier, the model is normally very complex and requires numerical solution, which may render a statistical approach prohibitive because of limits in CPU time available. Moreover, the number of parameters is usually very high. In combustion models, there are frequently more than 1000 parameters. For both these reasons, model reduction is almost always required. There are many techniques with which such a reduction can be achieved. In particular, reducing the chemical submodel is necessary. Sensitivity and flux analysis are just two techniques that have been applied in the past.

(b) Approximation of the model response

Even a reduced model is often too computationally expensive to be used in optimization algorithms, for example, as part of a parameter estimation technique. For this reason, the model response has been approximated by either
a tabulation procedure or a locally fitted response surface, also called response surface methodology (RSM). Mosbach et al. (2008) have developed a tabulation method for an engine model that makes use of cubic natural splines. This technique is suitable if the number of unknown parameters is relatively low, typically less than five. For larger numbers of parameters, the RSM can be combined with Latin hypercube sampling or low-discrepancy sequences to obtain an approximated model response in hypercubical subdivisions of the parameter space. In chemical engineering applications, linear models have been used to construct response surfaces (Feeley et al. 2004). A more general approach is kriging, which is an interpolation technique for random fields. The random fields are used to obtain an approximation to the model response accounting for the covariance between the response point and the design points.

(e) Parameter estimation

In the frequentist approach, one assumes that the model response is distributed according to a known probability distribution with unknown parameters. For a given set of experimental data, which are also assumed to be realizations of random variables, representing statistical and systematic error in the measurement, optimization yields new estimates for the parameters. In the Bayesian framework, the parameters themselves are random variables. Starting from a prior distribution of the unknown parameter, which can incorporate any additional knowledge, Markov chain Monte Carlo techniques can be used to obtain empirical posterior densities of the model parameters. These posterior distributions can then be further analysed. For example, using kernel marginalization, one can calculate credible regions such as the highest probability density regions for parameters and model responses. For a combustion problem, a somewhat similar approach has been successfully used in conjunction with model reduction by Sheen et al. (2009). Braumann et al. (submitted) have used the Bayesian approach to identify parameters of a multi-dimensional population balance model for granulation. An alternative, deterministic approach has been developed by Frenklach and co-workers (Feeley et al. 2004), who formulated a concept of dataset consistency based on bounds on the model prediction error as well as on the parameters. Considering the sensitivity of the consistency measure with respect to the data error bounds and the a priori assumptions on the parameters can then help to find possible outliers in the data.

(d) Design of experiments

All the above approaches can be used to suggest new experiments to reduce the uncertainty in the model parameters and therefore increase the predictive power of a model. In the classical frequentist experimental design, a new design can be found, for example, by maximizing the Fisher information matrix, which is a measure of how much information one has about the unknown parameters. Different designs are obtained by using different maximization criteria; for example, D-optimality maximizes the determinant of the Fisher information matrix, i.e. it minimizes the volume of the ellipsoid confidence regions for the unknown parameters. Similarly, the approach by Frenklach and co-workers mentioned above (Feeley et al. 2004) can be used to determine the sensitivity of
uncertainties in model predictions with respect to uncertainties in the underlying experimental data, as well as the sensitivity of dataset consistencies with respect to experimental results and \textit{a priori} information. These sensitivities enable identification of data entries that most inhibit the consistency of a dataset, and hence provide suggestions as to which measurement, if refined, would result in significant uncertainty reduction. In the Bayesian framework, an experimental design can be obtained by maximizing a utility function with respect to all experimental designs. However, the computation of the optimal experimental design requires the maximization of two nested integrals, which can be computationally very expensive.

\textit{(e) Model discrimination}

All the above approaches can be used for model discrimination. The aim is to falsify a model based on a set of experimental data. For example, Man \textit{et al.} \textit{(in press)} have extended the approach by Sheen \textit{et al.} (2009) to resolve conflicting parameter estimates in multivariate population balance models.

5. Role of the World Wide Web

\textit{(a) Incomplete information}

In the past, experimental data have been taken from the literature, which means that data were extracted manually from published articles. This has a number of disadvantages, which crucially influence the quality of a model. The foremost problem is that the information available to the model builder is incomplete. So far, experimentalists have almost always reported data that have been derived from raw data. The instrumental models used as well as the assumptions made in these models are, if at all, only implicitly stated. Furthermore, uncertainties associated with the raw data and detailed error propagation analysis of the instrumental model are often missing. A further problem is that the experimental set-up is not described in sufficient detail to fully specify the initial and boundary conditions necessary to find a numerical solution of the model. This situation is improving as many scientific journals now include additional information on the Web along with the electronic version of the corresponding article. However, this information is often in proprietary or software-specific standards and therefore requires access to specific software tools, which in many cases are not available as open source. But even if this information is provided, a researcher may not have access to expensive journals in which the experimental data including supplementary data have been published. While this can be circumvented by contacting the author directly in many cases, the doctorate students who carried out the experimental work may not have stored the raw data of their experiments and their laboratory log books may not be available.

\textit{(b) Metadata for models}

In order to overcome these problems, the following issues need to be addressed: standardization, accessibility and automation. Computer science has developed a number of concepts that may be used to progress on these issues. Ideas based
on the ‘Semantic Web’ seem to be particularly promising and have already been adopted by the science and engineering community. The World Wide Web Consortium (W3C), an organization that defines Web standards such as HTML, XML, RDF, SPARQL, etc., has envisioned the Semantic Web as a web of data that enables computers to process and understand information on the Web.

Extensible Markup Language (XML) aims to provide new standards for encoding documents and organizing their contents over the Internet. XML allows the flexible development of user-defined document types. It provides a core set of standards that developers can use to create their own. For this reason, XML is suitable for providing a framework for specific standards of data and models that occur in the engineering community. As XML is becoming a global standard, parser libraries are available in a large number of programming languages, e.g. Fortran, C, C++, Java, Python, etc. This allows data to be exchanged between different software packages regardless of programming language. In the combustion community, first applications of XML data representation have emerged. The PrIMe Data Model (PDM; Frenklach 2007) is a proposed standard based on XML technology, which aims to standardize many aspects related to modelling combustion chemistry using detailed chemical kinetics. The Computational Modelling Group has developed an XML specification for ICE data called EngineML (Smallbone et al. 2010). All these XML representations have in common that each piece of data must comply with an ‘XML schema’, which specifies the data model and acts as a grammar for XML datasets. Validation tools exist that can be used to check whether a specific piece of data is consistent with a given standard.

Although XML is suitable for representing standardized data, it does not explicitly relate data entries to each other. In order to query the XML data, the data have to be processed by specific software. To use XML datasets in any application, one needs to understand the particular XML representation. In the ‘Web data’ as described by W3C, the Web resources are well described by their relationship to one another. In the Semantic Web, this issue can be addressed through Resource Description Framework (RDF). RDF uses the concept of ‘entity–relationship’ based on the idea of making a statement about the data and their relations, so the Web data are not only human-understandable but also machine-understandable. RDF formalizes the free-form hierarchical structures of XML into triplets of subject, relationship and object. Thus, in order to obtain a piece of information, a search query can be performed by traversing the graph formed by the collection of triplets, using an RDF query language, such as SPARQL. The open-source software openRDF provides a framework for storing, inferring from and querying RDF data, allowing a human user to work with the RDF repository. Work is in progress to extend the use of RDF beyond EngineRDF to the PrIMe Data Models as well as to the output of the quantum chemistry package GAUSSIAN (QuantumML and QuantumRDF). Although XML and RDF schemas check the consistency of data, they do not check whether or not the relations contained in RDF are meaningful and consistent. To address this problem, the W3C has introduced OWL (Web Ontology Language), a language that is built upon RDF and RDF schemas. It is designed to rigorously define the description of things and their relationships. The Web ontology also provides the ability to
validate RDF documents beyond schema validation. Ontologies for EngineRDF and QuantumRDF are under development. The ultimate aim is to create ModelML and ModelRDF, which represent the full model $k$ as described in the previous sections.

6. Synthesis: automatic experimentation

In the previous section, experimental design was discussed based on a hierarchy of detailed physical models. The natural progression for these experiments is to be conducted automatically. In biology, this has already been achieved when ‘Adam’, the first robot scientist, was created by King et al. (2009). Adam can conduct high-throughput individually designed microbial batch growth experiments measuring growth curves (phenotypes) of selected microbial strains (genotypes) growing in defined media (environments). However, the experimental design is a data-driven approach and does not use scientific knowledge in the form of physical or chemical models. In chemical engineering, automation of experiments is commonplace from bench to industrial scale. For example, in Cambridge, a tank reactor has been automated using the industrial plant scale Siemens PCS7 system to control a chemical reaction (Selmer et al. 2007). In this experiment, the user interface is fully accessible over the Internet. Using a computer to provide the necessary input of initial and boundary conditions for the experiment represents no extra technical hurdle.

A complete representation of the model $k$ in XML as outlined above allows model improvement to be fully computer-driven. Sensitivity analysis can indicate which application models are important with respect to the variables of interest. The computer can design experiments to produce additional data to reduce the variance in the important parameters using, for example, Bayesian techniques. If the experiment is set up, this step can be fully automated, i.e. a robot can carry out the experiments. Adam generates knowledge in functional genomics using scientific reasoning, allowing it to form hypotheses, design experiments to test these and finally to carry out the experiments. The approach outlined in this paper differs in that the models used by the computer include physical insight as opposed to being entirely data-driven. This is because the computer can set up quantum calculations to calculate the properties of chemical species and reactions and employ conservation equations to calculate what happens on an industrial scale. This new system will be connected to the Internet, giving it access to a plethora of Web applications. Using Semantic Web technology, it will be able to take advantage of the latest experimental results from around the globe. It could conceivably even have access to real-time industrial operating conditions and results. This methodology will lead to a step change in the relationship between scientists, engineers and business professionals. Real economic decisions can be made by companies on the basis of reliable, comprehensive and cheap models.

The authors are grateful to all past and present members of the Computational Modelling Group and acknowledge support from the EPSRC, the Royal Society, the Royal Academy of Engineering, Churchill College Cambridge and the Weierstrass Institute for Applied Analysis and Stochastics in Berlin, Germany.

Phil. Trans. R. Soc. A (2010)
References


Cao, L., Bhave, A., Mosbach, S., Su, H., Kraft, M., Dris, A. & McDavid, R. M. 2009 Influence of injection timing and piston bowl geometry on PCCI combustion and emissions. *SAE Int. J. Engines* 2, 1019–1033. (Also available as SAE Paper No. 2009-01-1102.)


AUTHOR PROFILES

Markus Kraft (left)

Markus Kraft is a professor at the Department of Chemical Engineering and Biotechnology at the University of Cambridge and a Fellow of Churchill College. He leads the Computational Modelling Group that currently consists of 23 PhD students and postdocs. He studied applied mathematics at the University of Kaiserslautern and completed his Dr. rer. nat. in chemistry at the same University in 1997. He and his research group have been working for several years on combustion problems, the formation of inorganic and organic nanoparticles, granulation, engine modelling and the numerics of Monte Carlo methods at the Universities of Kaiserslautern, Karlsruhe, the Weierstrass Institute for Applied Analysis and Stochastics in Berlin, and for the past 10 years at Cambridge University.

Sebastian Mosbach (right)

After his studies of physics and computer science at the University of Kaiserslautern, Germany, supported by the Studienstiftung des Deutschen Volkes, Sebastian Mosbach completed his Certificate of Advanced Study in Mathematics at the Department of Applied Mathematics and Theoretical Physics at the University of Cambridge. He then joined the Computational Modelling Group at the Department of Chemical Engineering and Biotechnology in Cambridge and obtained his PhD in 2006. Since then, he has become a Senior Research Associate and was awarded a Raymond and Beverly Sackler Research Fellowship at Churchill College, Cambridge, in 2008.