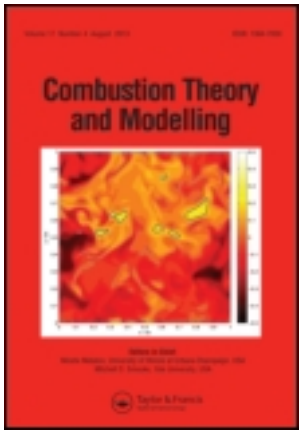


Combustion Theory and Modelling
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tctm20

Bayesian analysis of syngas chemistry models

Kalen Braman[a], Todd A. Oliver[b] & Venkat Raman[a]
[a] Department of Aerospace Engineering & Engineering MechanicsUniversity of Texas at Austin, Austin, TX 78712 USA
[b] Institute for Computational Engineering and SciencesUniversity of Texas at Austin, Austin, TX 78712 USA
Published online: 20 Sep 2013.

To cite this article: Kalen Braman, Todd A. Oliver & Venkat Raman (2013) Bayesian analysis of syngas chemistry models, Combustion Theory and Modelling, 17:5, 858-887, DOI: 10.1080/13647830.2013.811541

To link to this article: http://dx.doi.org/10.1080/13647830.2013.811541

Combustion Theory and Modelling, 2013
Vol. 17, No. 5, 858–887, http://dx.doi.org/10.1080/13647830.2013.811541

Bayesian analysis of syngas chemistry models

Kalen Braman[a*], Todd A. Oliver[b] and Venkat Raman[a]

[a]*Department of Aerospace Engineering & Engineering Mechanics, University of Texas at Austin, Austin, TX 78712, USA;* [b]*Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA*

(*Received 11 December 2012; accepted 11 April 2013*)

Syngas chemistry modelling is an integral step toward the development of safe and efficient syngas combustors. Although substantial effort has been undertaken to improve the modelling of syngas combustion, models nevertheless fail in regimes important to gas turbine combustors, such as low temperature and high pressure. In order to investigate the capabilities of syngas models, a Bayesian framework for the quantification of uncertainties has been used. This framework, given a set of experimental data, allows for the calibration of model parameters, determination of uncertainty in those parameters, propagation of that uncertainty into simulations, as well as determination of model evidence from a set of candidate syngas models. Here, three syngas combustion models have been calibrated using laminar flame speed measurements from high pressure experiments. After calibration the resulting uncertainty in the parameters is propagated forward into the simulation of laminar flame speeds. The model evidence is then used to compare candidate models for the given set of experimental conditions and results. Additionally, the technique MUM-PCE, an interesting uncertainty minimisation method for kinetics models, has been compared to the Bayesian method for this application to the prediction of syngas laminar flame speeds. This comparison shows the importance of model form error and experimental error representations in the uncertainty quantification context, for these choices significantly affect uncertainty quantification results.

Keywords: chemical kinetics; syngas combustion; uncertainty quantification; Bayesian statistical methods; laminar flame speed simulation

1. Introduction

As numerical modelling becomes critical for the computational design of practical combustion devices, there is an increasing focus on the reliability of models. Specifically, an estimate of the error incurred by the different models is sought. Given that models are necessarily imperfect, a quantitative estimate of the uncertainty associated with the representation of the underlying physics needs to be obtained. This process is termed uncertainty quantification (UQ). Many different techniques for assessing uncertainty exist [1–12]. In this work, the Bayesian description is used to assess uncertainty. Here, the lack of knowledge or the imprecision of the model is expressed probabilistically. Learning, through experimentation or improved modelling, then manifests as a change in the probabilities. In this work, the Bayesian approach is used to analyse a syngas combustion model.

The purpose of the chemistry model is to provide reaction rates that will then be used in the spatial transport equation for species mass fractions to predict measurable

*Corresponding author. Email: kalen@utexas.edu

quantities (e.g., mass fraction profile, burning velocity, etc.). The model itself is built for a certain range of operating conditions (e.g., high pressure or fuel-lean conditions). For the purpose of this discussion, it is assumed that the chemistry model contains all necessary pathways to describe the combustion process. In this sense, the model error arises from the inability to determine the Arrhenius rate coefficients precisely. Experiments or quantum chemistry simulations could then be used to increase the accuracy of the rate coefficients. However, both these approaches are themselves imperfect. Experiments contain measurement errors and may also be subject to inherent variability from one run to the next. Quantum chemistry calculations utilise a variety of simplifying assumptions (e.g., the Born–Oppenheimer approximation, transition-state based rate computation) that will lead to errors in the potential energy surface as well as chemical rate coefficients. Consequently, the reaction rate coefficients in a chemistry model cannot be known to arbitrary precision.

Consider a chemistry model consisting of N parameters, where this number consists of pre-exponential factors, activation energies, and all model parameters used to define the entire set of reactions. None of these N parameters is known exactly, but often are given a range or a distribution. These ranges for the individual parameters combine to form a hypervolume of possible parameter settings in N-dimensional space. It is important to note that the different parameters are correlated, and that varying one without adjusting the rest is not meaningful. Using data, it is then possible to reduce this volume and understand parameter correlations. This process is termed learning. While different UQ approaches all have this basic philosophy regarding the parameters, they vary in the methodology used to represent the relative plausibility of points inside the hypervolume as well as in the approach to using data for reducing the volume of the plausible region in this high-dimensional space.

In the area of combustion chemistry, three different approaches have been formulated. Najm and coworkers [7–9] use a polynomial chaos expansion approach, where the parameters are assumed to be random variables. Such application allows the simulation results to be treated probabilistically with the uncertainty in the model parameters propagating to those results. Frenklach and coworkers [1–6] propose an approach called 'data collaboration', where estimates of the bounds of the parameters are obtained using semi-linear programming. This approach focuses on obtaining a feasible set, which is defined as the hypervolume that reproduces the experimental observations within the associated experimental uncertainty. In this sense, estimates of the lower and upper bounds of the parameter values are obtained without using a probabilistic description. Further, a response-surface approximation is used which deserves further explanation. The experimental data used to develop the uncertainty estimates typically contain laminar burning velocity or autoignition time. In the response-surface method, the experimental measurement is represented as a quadratic function of the parameters. Since the true relation between the outputs and the parameters is highly nonlinear and may involve spatial transport as well, this reduced-order representation itself will introduce some error. Nevertheless, Frenklach and coworkers have demonstrated in numerous studies [1–6] that this approach is very robust, and well-suited for the problem of chemistry modelling.

The third approach is that of Sheen and Wang [10–12], which is billed as an uncertainty minimisation rather than a quantification technique. This method combines a simplified form of the Bayesian approach with the response-surface approximation. In this approach, the parameters are adjusted such that the uncertainty predicted by the chemistry model for the experimental data used in calibration is reduced. Note that the goal of uncertainty minimisation is different from that of determining the hypervolume of possible parameter values. Further, while the authors themselves do not relate their technique to the Bayesian

technique, we show in this work that this approach is a reduced form of the Bayesian technique that invokes several simplifying assumptions. First, the PDF of the parameters is indirectly assumed to be a Gaussian distribution. Second, the response-surface approach is used to approximate the relation between the parameters and the model output. Thirdly, in order to maintain the Gaussian nature of the PDF after the learning process, the response surface is further approximated using a linear function. Consequently, this approach has the potential to introduce large errors in the estimation of uncertainty. For instance, Russi and Frenklach [13] explore the effects of assuming specific shapes for the parameter hypervolume. In particular, they note the substantial error introduced by assuming high-dimensional ellipsoids.

With this background, the purpose of this study is to explore the use of the Bayesian UQ approach for the modelling of syngas combustion chemistry. Syngas, which is a mixture of carbon monoxide and hydrogen, has become central to the use of integrated gasification combined cycle (IGCC) systems for carbon capture and sequestration [14, 15]. Due to the presence of hydrogen, syngas mixtures exhibit wider flammability limits and higher reactivity. Interestingly, chemistry models for syngas combustion have considerable difficulty predicting combustion characteristics (e.g., laminar flame speed) at high pressures for a range of syngas compositions [16]. Application of UQ to this system is of immediate interest, not only in characterising the uncertainties but in determining the root cause of this prediction problem. This study evaluates the Bayesian approach, with particular focus on kinetics parameter calibration and evidence-based chemistry model comparison.

2. Methodology: Bayesian analysis

The starting point for the Bayesian approach is a chemistry model with a specific set of reactions and associated rate parameters. It is assumed that some information about these rate parameters is available and is termed as prior information. The purpose of the Bayesian approach is to improve the estimates of these rate parameters based on hitherto unused experimental data. Since changing the rate parameters for a single reaction will affect the overall performance, all the rate parameters are linked to one another and are simultaneously updated. This process is termed as a global calibration approach.

In this work, a global calibration approach based on Bayesian statistics is used. Bayesian probability analysis is a well-developed field that is increasingly used to pose and solve inverse problems in many areas of science and engineering [17–21]. While some examples have appeared involving chemistry and combustion [22–24], the approach is not widely used. Thus, Sections 2.1 and 2.2 describe the Bayesian formulations of the calibration and model comparison problems, respectively. Finally, Section 2.3 briefly describes the statistical algorithms used to compute results in this work.

2.1. *Calibration*

In the Bayesian interpretation of probability, model parameters such as kinetic rate coefficients are treated as random variables. Since the 'true' value of the parameter is unknown, one represents what is known about the 'true' value using probability. Then, the state of knowledge about a parameter value is represented by the PDF of the corresponding random variable. For instance, if the parameter were known with complete certainty, the corresponding PDF would be a delta function. Given these PDFs, one updates one's knowledge to account for new data by updating the parameter PDFs according to Bayes' theorem

[17, 19]. Specifically, given two quantities, x and y, Bayes' theorem states that

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)}, \tag{1}$$

where $p(x|y)$ is the probability distribution of x, conditioned on a specific value of y, and similarly for $p(y|x)$.

In order to apply this theorem to chemistry modelling, let $\boldsymbol{\theta}$ denote the vector of uncertain kinetics parameters being calibrated. This vector will be the combination of activation energies, pre-exponential factors, and any other parameter that appears in the chemistry model. Let $\mathbf{d}$ denote the experimental data used for calibration (e.g., flame speed). Bayes' theorem implies that

$$p_{\text{post}}(\boldsymbol{\theta}|\mathbf{d}) = \frac{p_{\text{prior}}(\boldsymbol{\theta})\,\pi(\boldsymbol{\theta};\mathbf{d})}{\int p_{\text{prior}}(\boldsymbol{\theta})\,\pi(\boldsymbol{\theta};\mathbf{d})\mathrm{d}\boldsymbol{\theta}}. \tag{2}$$

In (2), p_{prior} denotes the prior PDF, which quantifies available information about the parameters that is independent of the data, and p_{post} denotes the posterior PDF, which quantifies the state of knowledge about the parameter values after incorporating the information in the data. These PDFs are connected through the likelihood function, $\pi(\boldsymbol{\theta};\mathbf{d})$. The likelihood function quantifies the level of agreement between the model and the data for specific values of the parameters. It is the function mapping the parameters $\boldsymbol{\theta}$ to the PDF associated with the observed data. That is,

$$\pi(\boldsymbol{\theta};\mathbf{d}) = p_{\text{like}}(\hat{\mathbf{d}}|\boldsymbol{\theta})|_{\hat{\mathbf{d}}=\mathbf{d}}, \tag{3}$$

where $\hat{\mathbf{d}}$ is the variable representing the observed quantity. Given only the model parameter values, the value that will be observed in an experiment differs from the model prediction due to inadequacies in the chemistry model (model error) as well as the observation process (experimental error). The PDF p_{like} represents the state of knowledge regarding these errors. When p_{like} is evaluated at the actual observed values $\mathbf{d}$ and considered as a function of $\boldsymbol{\theta}$, it becomes the likelihood function, π.

Thus, Bayes' theorem provides a probabilistic approach for extracting information about parameters from experimental data. Furthermore, the posterior obtained from one calibration can be used as the prior for a subsequent calibration problem if more experimental data become available. In this sense, the Bayesian approach provides a naturally self-consistent process for learning based on all available information.

The prior and likelihood must be constructed to represent the state of knowledge before the data are obtained. Rigorous approaches for specifying these forms is the subject of ongoing research. Here, simple common forms are used to illustrate the process.

Specifically, two forms of prior PDF are used: uniform and Gaussian. A uniform prior assigns an equal prior probability density to any parameter value within its bounds, which are selected to span the expected possible range of the parameter. Alternatively, a Gaussian assigns higher prior density near the mean. Here, each Gaussian prior is assigned a mean equal to the parameter's nominal value. The standard deviation for each parameter is estimated from the uncertainty factor listed in the literature. Furthermore, the parameters are assumed to be independent in the prior.

The likelihood function is defined based on the description of the error resulting from the application of the chemistry model and the experimental error. For instance, if d_i refers

to an experimental measurement at certain conditions, and x_i denotes the simulated value of this measurement (using the chemistry model), the two could be related by the following additive error model:

$$d_i = x_i + \epsilon_i, \tag{4}$$

where ϵ_i refers to the total error due to both experimental and modelling errors. The reasoning for the choice of combined model and experimental error is elucidated in Section 3.3. When d_i and x_i are both known, (4) defines ϵ_i. However, to use such a model in a calibration, one must construct a model for ϵ_i that does not depend on d_i. In this situation, the state of knowledge is expressed probabilistically. A simple model that is often used is to assume that ϵ_i are independent, identically distributed (i.i.d.) zero-mean Gaussian random variables. In this case, the likelihood function is written as

$$\pi(\boldsymbol{\theta}; \mathbf{d}) = \frac{1}{(2\pi\sigma^2)^{N_d/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N_d}(d_i - x_i)^2\right], \tag{5}$$

where N_d is the number of data points and σ denotes the standard deviation of ϵ_i. The i.i.d. assumption for ϵ_i implies that errors are independent even for data points that are close together in scenario space – i.e. that the errors in the laminar flame speeds are independent even at nearly the same pressure and equivalence ratio. While this assumption is not realistic, it allows calibration of the model parameters in a simple setting. The development of more realistic covariance structures is left for future work. Even in this simple setting, σ is generally not known a priori, which is the case here. In this situation, σ is treated as a hyperparameter to be calibrated along with the kinetic model parameters.

A multiplicative error model is also investigated here. A multiplicative error model applies to cases in which the error is proportional to the output of the model, as can be the case for a range of conditions for which the model result varies widely in magnitude, or for cases in which the output must retain the same sign [25]. For the multiplicative error form used here, the model and experimental values are related by

$$d_i = x_i \exp(\epsilon_i). \tag{6}$$

Again assuming ϵ_i are Gaussian and i.i.d. leads to the following likelihood:

$$\pi(\boldsymbol{\theta}; \mathbf{d}') = \frac{1}{(2\pi\sigma^2)^{N_d/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N_d}(d'_i - x'_i)^2\right], \tag{7}$$

where $d'_i = \log d_i$ and $x'_i = \log x_i$.

It should be noted that similar modelling choices are inherent to any UQ process, not just the Bayesian methodology. All results are contingent on these choices. While the descriptions used here represent convenient choices, it should be recognised that a more in-depth analysis of the details of the chemistry models and experimental data would likely lead to more complex models. Such UQ model development is beyond the scope of the current work.

2.2. *Model comparison*

Bayes' theorem can also be used as a basis for model comparison. This comparison methodology is used here to evaluate the relative merits of competing chemistry models. Just as PDFs are used to characterise the state of knowledge of kinetics parameters, probability can be used to characterise knowledge about which model in a given set is best.

To define the procedure, let $\mathcal{M} = \{M_1, \ldots, M_K\}$ denote a set of K candidate models M_i that one wishes to compare. The parameters of each model are calibrated using a Bayesian update based on the data $\mathbf{d}$, as described in Section 2.1. Rewriting (2) with explicit dependence on the model gives

$$p_{\text{post}}(\boldsymbol{\theta}_i|\mathbf{d}, M_i) = \frac{p_{\text{prior}}(\boldsymbol{\theta}_i|M_i)\,\pi(\boldsymbol{\theta}_i, M_i; \mathbf{d})}{\displaystyle\int p_{\text{prior}}(\boldsymbol{\theta}_i|M_i)\,\pi(\boldsymbol{\theta}_i, M_i; \mathbf{d})\mathrm{d}\boldsymbol{\theta}_{\text{i}}}, \tag{8}$$

where $\boldsymbol{\theta}_i$ indicates the parameters for the ith model.

In the model comparison problem, the task is to rank the models according to which is best, given the data. In the Bayesian framework, this ranking is determined by the posterior probability of the models. As in the calibration problem, the posterior distribution is determined from Bayes' theorem:

$$P_{\text{post}}(M_i|\mathbf{d}, \mathcal{M}) = \frac{P_{\text{prior}}(M_i|\mathcal{M})\,\pi_{\text{evid}}(M_i; \mathbf{d})}{\sum_{k=1}^{K} P_{\text{prior}}(M_k|\mathcal{M})\,\pi_{\text{evid}}(M_k; \mathbf{d})}. \tag{9}$$

In (9), $P_{\text{prior}}(M_i|\mathcal{M})$ is the prior probability of the model M_i. That is, it is the probability assigned to M_i based on information that is independent of the data. Often, there is very little such information. In this case, a uniform prior, i.e. $P_{\text{prior}}(M_i|\mathcal{M}) = 1/K$, is appropriate. Then, the relative posterior probability is determined entirely by $\pi_{\text{evid}}(M_i; \mathbf{d})$, which is known as the evidence function. The evidence function measures the consistency of the model and the data considering the entire parameter space. It is given by the following integral over parameter space:

$$\pi_{\text{evid}}(M_i; \mathbf{d}) = \int p_{\text{prior}}(\boldsymbol{\theta}_i|M_i)\pi(\boldsymbol{\theta}_i, M_i; \mathbf{d})\mathrm{d}\boldsymbol{\theta}_{\text{i}}. \tag{10}$$

Thus, the evidence for the model M_i is just the normalisation constant (the denominator) in the calibration expression (8).

To gain insight into the Bayesian model comparison, note that this process can be thought of as a natural formalisation of Ockham's razor. To see this interpretation, it is helpful to write the evidence in a different form. Following Muto and Beck [26], the log-evidence can be decomposed into two terms:

$$\begin{aligned}\log\left(\pi_{\text{evid}}(M_i; \mathbf{d})\right) = &\int \log\left(\pi_{\text{like}}(\boldsymbol{\theta}_i, M_i; \mathbf{d})\right)\, p_{\text{post}}(\boldsymbol{\theta}_i|\mathbf{d}, M_i)\,\mathrm{d}\boldsymbol{\theta}_{\text{i}} \\ &- \int \log\left(\frac{p_{\text{post}}(\boldsymbol{\theta}_i|\mathbf{d}, M_i)}{p_{\text{prior}}(\boldsymbol{\theta}_i|M_i)}\right) p_{\text{post}}(\boldsymbol{\theta}_i|\mathbf{d}, M_i)\mathrm{d}\boldsymbol{\theta}_{\text{i}}.\end{aligned} \tag{11}$$

The first term is the posterior expectation of the log-likelihood which measures how well the model is able to fit the data, averaged over the posterior PDF for the parameters. The second term is the relative information entropy (or Kullback–Leibler divergence [27]) between the posterior and the prior PDFs. It measures the information about the parameters that is gained from the data. For two models that fit the data equally well, as measured by the first term in (11), the model that requires the least tuning, as measured by the information gain, is preferred. For more details and discussion, see Jaynes [19, Chapter 20] and Muto and Beck [26, Section 4].

2.3. *Statistical algorithms*

While the posterior PDF in (8) and posterior probability in (9) are simple to write down, evaluating and using these expressions is computationally challenging. For example, computing statistics using the posterior PDF (9) and computing the evidence (10) both require the evaluation of high-dimensional integrals. To evaluate such integrals, stochastic simulation methods have been developed in which the posterior PDF is represented by samples and integrals are approximated by Monte Carlo methods using those samples. In work presented here, an advanced stochastic simulation method, referred to as the 'adaptive multi-level algorithm' [28, 29] is used. This algorithm involves the use of a staged Bayesian update in which the posterior is gradually approached across a sequence of intermediate updates. The main idea is to construct and then sample a sequence of intermediate distributions between the prior and the posterior. Specifically, at the ℓth intermediate 'level', one uses Markov chain Monte Carlo to sample

$$\pi_{\text{int}}^{(\ell)}(\boldsymbol{\theta}; \mathbf{d}) \propto p_{\text{prior}}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}; \mathbf{d})^{\tau_\ell},$$

where $0 \leq \tau_\ell \leq 1$. Clearly, when $\tau_\ell = 0$, $\pi_{\text{int}}^{(\ell)}$ is the prior, and when $\tau_\ell = 1$ it is the posterior. After sampling at level ℓ, a selection of the samples generated are used to begin Markov chains for the $(\ell + 1)$th intermediate distribution, where $\tau_{\ell+1} > \tau_\ell$. Complete details of the algorithm can be found in [29]. This algorithm is implemented in the QUESO library [30], which was used to generate all of the results shown in Sections 4 and 5.

3. Application of the Bayesian approach to high-pressure syngas combustion modelling

For application in hydrogen-rich gas turbines for power generation, the chemistry models have to be calibrated at high pressure and low equivalence ratio conditions. Here, three different chemistry models are used to demonstrate the Bayesian approach to model calibration, model selection, and error propagation.

3.1. *Synthesis gas chemical kinetics models*

Three kinetics models from Davis *et al.* [31], Li *et al.* [32], and Sun *et al.* [33] are considered here. For simplicity, these models will be referred to as DM, LM and SM, respectively, from here forward.

DM involves 14 species and 30 reactions. The Arrhenius reaction rate parameters were compiled from recent kinetics experiments as well as the GRI-Mech 3.0 model. The Arrhenius reaction rate pre-exponential coefficient was optimised for 19 of those reactions based upon experimental results of laminar flame speeds, peak mole fractions of

low-pressure burner-stabilised flames, fuel consumption rates in a turbulent flow reactor, and ignition delay times behind reflected shock waves. The LM involves 14 species and 31 reactions. Reaction rate parameters for this model were compiled by calibration to data from formaldehyde oxidation in a flow reactor, from a hydrogen–oxygen model [34], and by calibration to other recent experimental data. The SM involves 15 species and 33 reactions. The reaction rate parameters were compiled from recent literature, including ab initio calculations for specific reactions. The three models have an identical list of species involved, except for the SM, which also involves CH_2O.

3.2. *Kinetic parameters used in Bayesian calibration*

The Bayesian methodology can incorporate the calibration of an arbitrary selection of model parameters. In combustion, the pertinent parameters include those from the kinetics model such as the Arrhenius pre-exponential coefficient, temperature exponent, activation energy, and third body efficiencies, as well as those from transport and thermodynamic models. For the context of this study, only Arrhenius pre-exponential coefficients in each syngas kinetics model have been incorporated in the Bayesian update. Of course, in general, the set of parameters need not be limited only to the pre-exponential coefficients.

Additionally, the method of Sheen and Wang has been incorporated in this study for comparison to the Bayesian method. The Sheen and Wang method commonly restricts itself to pre-exponential parameters. That method, in order to minimise computational cost, incorporates a list of pre-exponential parameters selected by a sensitivity study for each experimental target rather than all pre-exponential parameters. In the cases shown in Section 5, after the sensitivities of the flame speed to the kinetics parameters, $\partial\eta/\partial \ln k_j$, were calculated, the sensitivity-uncertainty index (SUI) [35], $C_j(\eta) = |(\partial\eta/\partial \ln k_j) f_j|$, was calculated for each parameter. In this expression, η is the simulated flame speed, k_j is the parameter and $f_i = \log_{10}(k_j^0/k_j^{\min}) = \log_{10}(k_j^{\max}/k_j^0)$ is the multiplicative uncertainty factor, which relates the literature-listed uncertainty in each rate parameter. In the expression for the uncertainty factor, the superscript '0' refers to the nominal value of the parameter and 'min' and 'max' refer to two standard deviations away from the nominal. These indices give a relative ranking of the sensitivity of the flame speed calculations to each parameter weighted by the uncertainty in that parameter. The sensitivity threshold was set as 10% of the maximum SUI for each experimental target.

3.3. *Premixed laminar flame experiments and simulations*

The Bayesian update of the model parameters incorporates information from experimental data with results from corresponding simulations. For this update process, laminar premixed flame experiments were chosen. The high pressure flame speed data of Sun *et al.* [33] serves as the experimental data set. These data were selected because the conditions approximate those expected in ground-based combustors. Specifically, experiments conducted at equivalence ratios of 1.2 or lower and at pressures of 5 and 10 atm were used for calibration. Flame speed data for a pressure of 20 atm was used for predictive comparison. The $CO{:}H_2$ ratios tested were 1, 3 and 19. These conditions and the observed flame speed values are listed in Table 1. Note that the experimental error in the flame speed is not reported by [33] and thus, for simplicity, the experimental error is lumped with the model error in ϵ_i during calibration, as discussed in Section 2.1. Simulations of premixed laminar flames were completed with the CHEMKIN PREMIX routines [36]. The standard CHEMKIN routines for thermodynamic and transport properties [37] were also used.

Table 1. Experimental data [33] used in calibration (top group of values) and comparisons (lower group).

p (atm)	$CO:H_2$	ϕ	η (cm/s)
5.0	1.0	0.8	79.4
5.0	1.0	1.0	113.6
5.0	1.0	1.2	136.1
5.0	3.0	0.8	52.6
5.0	3.0	1.0	64.3
5.0	3.0	1.2	81.3
5.0	19.0	1.2	40.6
10.0	1.0	0.6	27.8
10.0	1.0	0.8	68.7
10.0	1.0	1.0	101.2
10.0	1.0	1.2	128.0
10.0	3.0	0.6	25.9
10.0	3.0	0.8	43.9
10.0	3.0	1.0	64.1
10.0	3.0	1.2	79.3
10.0	19.0	0.8	21.2
10.0	19.0	1.0	27.7
10.0	19.0	1.2	39.1
20.0	3.0	0.8	36.3
20.0	3.0	1.0	56.3
20.0	3.0	1.2	68.7
20.0	19.0	1.0	29.2
20.0	19.0	1.2	36.7
40.0	19.0	1.0	30.1

The Bayesian process need not be restricted to a single type of experiment. Other experimental results such as autoignition times may be used in the process. For this procedure, laminar flame speeds were selected since they provide a fundamental basis for many turbulent combustion models.

4. Bayesian calibration results

This section details the results of the use of the Bayesian methodology for the UQ of syngas kinetics models. The three kinetics models are calibrated, uncertainty is propagated, and the model results are compared in Section 4.1. Next, the Bayesian evidence is used to compare the relative plausibility of each kinetics model in Section 4.2. Then, the form of the prior distributions is discussed in the context of the DM in Section 4.3, and the form of the error model is discussed in the context of the LM in Section 4.4.

4.1. *Bayesian calibration and uncertainty propagation*

The Bayesian uncertainty analysis first updates the model parameters according to Equation (2) with the algorithm described in Section 2.3. This calibration leads to an updated level of knowledge regarding the uncertain parameters in the kinetics model as provided by the joint posterior distribution. All Arrhenius pre-exponential coefficients from the three kinetics models, 37 for the DM, 36 for the LM, and 44 for the SM, are updated in this fashion using the additive error model and Gaussian priors. The use of the additive error

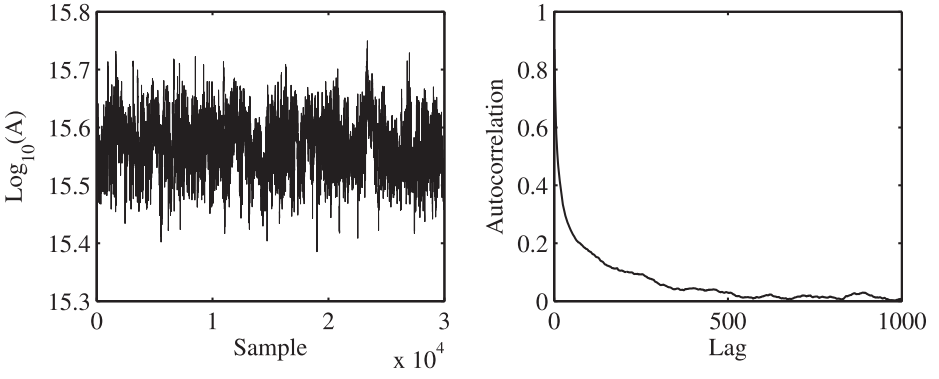


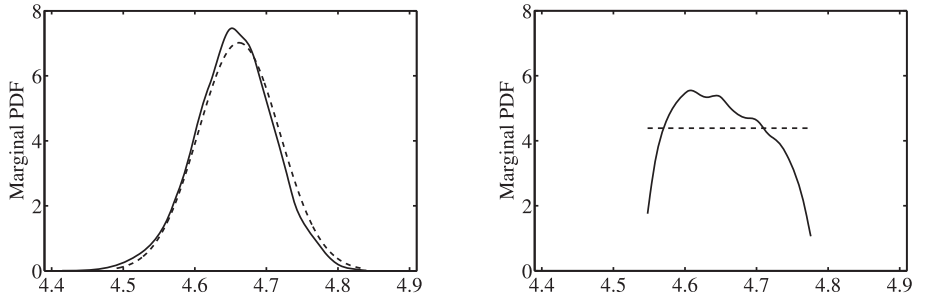
Figure 1. Subsection of sampling results (left) and autocorrelation for the $O + H_2 \rightleftharpoons H + OH$ pre-exponential coefficient in the LM (right).

model also introduces an additional parameter, the error model variance hyperparameter, leading to 38, 37 and 45 updated uncertain parameters for the DM, LM and SM models, respectively. The data used for the calibration are the laminar flame speed data described in Section 3.3.

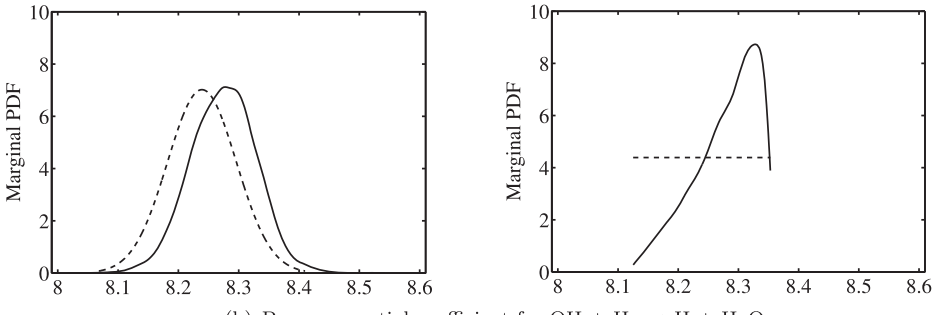
The final joint and marginal parameter distributions are constructed from the full dimensional sampling results. Figure 1 shows a section of the raw chain, as well as the autocorrelation, for the parameter with the slowest decaying autocorrelation. The chain itself shows reasonable mixing, and the autocorrelation, although long lasting, dies out at approximately 500 samples. For each test case, the full computation took approximately 15,000 processor hours, using 3.33 GHz compute cores with 2 GB of memory per core. The majority of the computational time is spent computing the flame speed. Each sample flame speed calculation takes approximately one to three seconds depending on the number of iterations required to reach convergence. The multi-level algorithm computes approximately 300,000 accepted samples on the final level.

The left side of Figure 2 shows a selection of the marginal posterior PDFs obtained by the stochastic multilevel algorithm for selected pre-exponential parameters plotted with their Gaussian prior distributions (the right side of this figure will be discussed in Section 4.3). Note that the posterior PDF for a multi-parameter calibration is a joint PDF of all the parameters. Here, only the marginal PDFs are shown. The shift from the prior distribution to the posterior shows some of the information gained by calibrating the kinetics model parameters to the data set. For some parameters, the distribution remains nearly the same. For example the posterior distributions of reactions $O + H_2 \rightleftharpoons H + OH$ and $H_2O_2 + OH \rightleftharpoons HO_2 + H_2O$ shift little from the priors. For other parameters, such as $OH + H_2 \rightleftharpoons H + H_2O$ and $CO + OH \rightleftharpoons CO_2 + H$, the most likely value shifts or the shape of the distribution changes. In particular the distribution for the parameter of $CO + OH \rightleftharpoons CO_2 + H$ shifts significantly lower and becomes slightly tail heavy and more peaked. Thus certain parameters are informed more by the Bayesian update. Such results imply either that the chosen experiments contain more information regarding those shifted parameters or that the information in the experiments has already been encapsulated in the prior distribution.

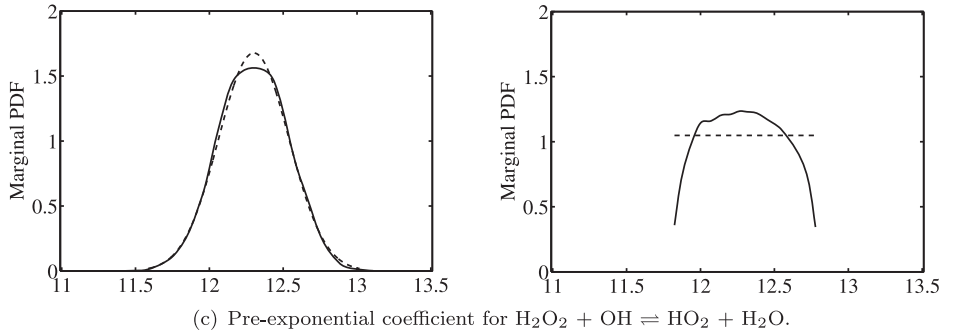
In order further to investigate the information provided regarding the parameters by the calibration data, Figure 3 shows the Kullback–Leibler (KL) divergence between the marginal posterior and prior PDFs for all parameters in each model. Most parameters have a small divergence, indicating that those parameters have gained little from the calibration.



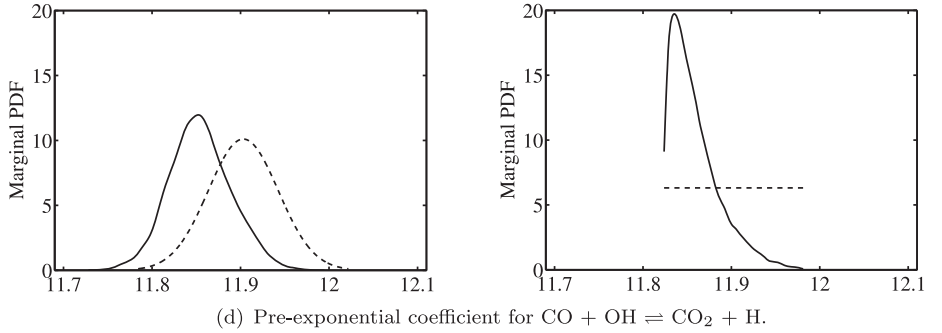
(a) Pre-exponential coefficient for $O + H_2 \rightleftharpoons H + OH$.



(b) Pre-exponential coefficient for $OH + H_2 \rightleftharpoons H + H_2O$.

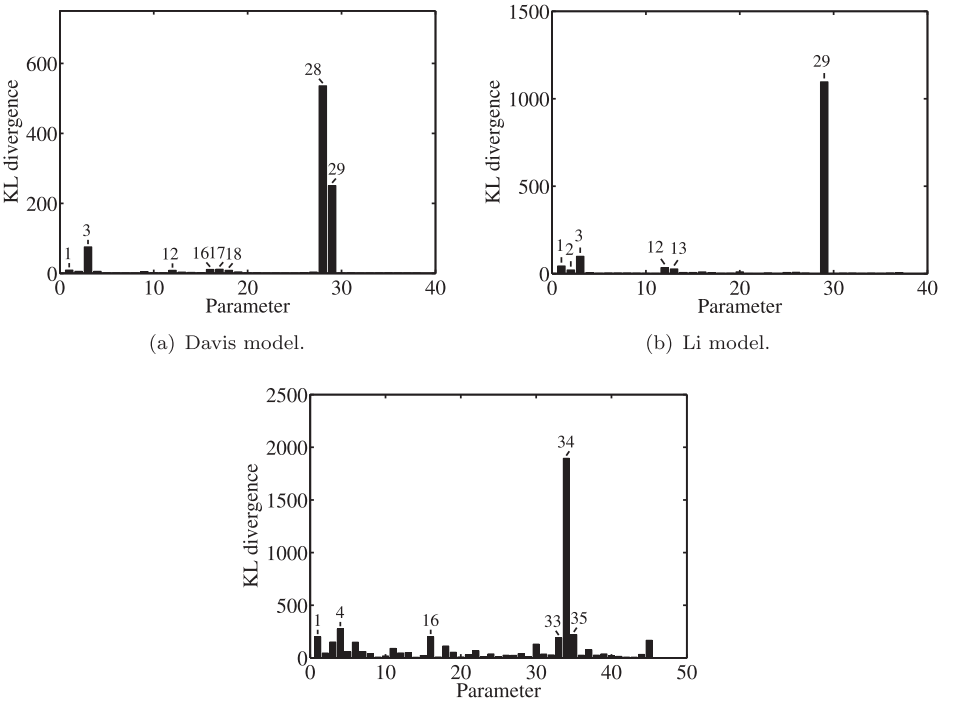


(c) Pre-exponential coefficient for $H_2O_2 + OH \rightleftharpoons HO_2 + H_2O$.



(d) Pre-exponential coefficient for $CO + OH \rightleftharpoons CO_2 + H$.

Figure 2. Prior and final marginal PDFs of Arrhenius pre-exponential parameters for the DM with Gaussian priors (left) and uniform priors (right).



(a) Davis model. (b) Li model.

Figure 3. Kullback–Leibler divergence for every parameter of each model, with the parameters within the top 10% of divergence for each model labelled.

This result is not surprising given that the number of parameters is greater than the number of data points, and thus not every parameter is identifiable using only this data. Further, only parameters that significantly affect the computed laminar flame speed will be constrained by the Bayesian update. Thus, low KL divergence indicates that, over the range of the prior, the parameter does not greatly influence the laminar flame speed for the calibration cases. However, some of the parameters, in particular the pre-exponential coefficients for $CO + OH \leftrightharpoons CO_2 + H$ and its duplicate reactions in all three models, have relatively large divergence. These parameters have gained the most from the Bayesian update, indicating that the chosen experiments provide information for the calibration of those parameters.

While the marginal posterior PDFs are informative, one of the important aspects of UQ methods like the Bayesian approach is the joint calibration of the parameters. In essence, kinetic parameters cannot be tuned individually without degrading performance. The best representation of the experimental data is achieved when all parameters are calibrated simultaneously. The simultaneous calibration of the entire parameter set updates not only individual parameter distributions as described above, but also the correlation between parameters. The posterior joint-PDF of the parameters provides insight as to how a change in a single parameter value changes the other parameters. For instance, Figure 4 shows the two-dimensional joint-PDF of the pre-exponential factor for reactions $H + OH + M$ $H_2O + M$ and $HO_2 + H$ $OH + OH$. This joint-PDF itself is a marginal PDF obtained from the full multi-dimensional PDF of all calibration parameters. Moreover, the domain of likely values has a complex shape, and is not limited by the presumed shapes for the prior PDFs.

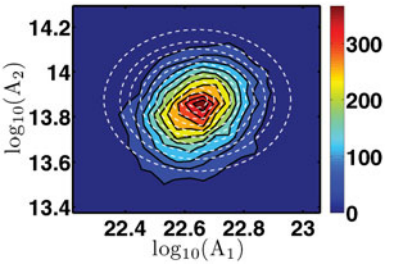


Figure 4. Posterior two-dimensional marginal joint PDF for the pre-exponential parameters of reactions H + OH + M ⇌ H_2O + M (1) and HO_2 + H ⇌ OH + OH (2) for the DM with Gaussian prior shown as dashed lines.

Figure 5 shows slices of a three-dimensional PDF, indicating that only a narrow range of parameter values is optimal in reproducing the experimental data.

Once the parameters have been calibrated and the samples of the posterior joint PDF obtained, this information can be propagated to determine the uncertainties in predictions of the laminar flame speed. Here, this is illustrated using the 'pushed forward posterior' method. Specifically, for the jth posterior sample of the kinetics model parameters $\boldsymbol{\theta}_j$, the corresponding laminar flame speed $\eta_j = \eta(\boldsymbol{\theta}_j)$ is computed using the chemistry model. In total, 18,432 laminar flame speeds were computed for each condition, leading to an ensemble of flame speed values that can be used to estimate the flame speed distribution. Note that this procedure is equivalent to marginalising the posterior over the hyperparameter σ and propagating the resulting distribution for the kinetics parameters alone. This allows one to see the uncertainty in the predictions given by just the posterior uncertainty in the calibrated kinetics parameters.

The left side of Figure 6 shows the laminar flame speed computations for two different syngas mixtures at 10 atm pressure. The experimental data for these conditions are listed in the top group of values in Table 1 and formed a portion of the data set used to calibrate the kinetics models. For comparison, the flame speed computed with the original (nominal) model parameter values and the experimentally observed values are also shown. Since the calibration process produces a PDF for the uncertain parameters, the propagation step produces a PDF for the flame speed results. The plot shows the 95% confidence interval bounding a grey-scale plot of probability density. It is seen that the original models predict

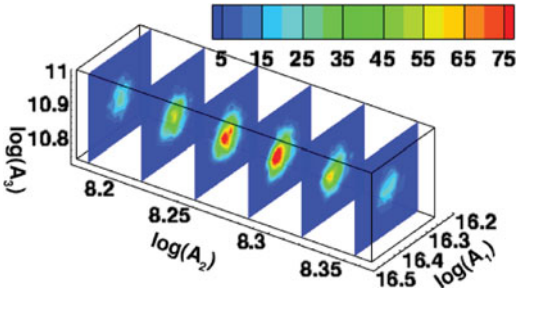


Figure 5. Posterior three-dimensional marginal joint PDF for the pre-exponential parameters of reactions H + O_2 ⇌ O + OH (1), OH + H_2 ⇌ H_2O + H (2), and CO + OH ⇌ CO_2 + H (3) for the DM with Gaussian prior.

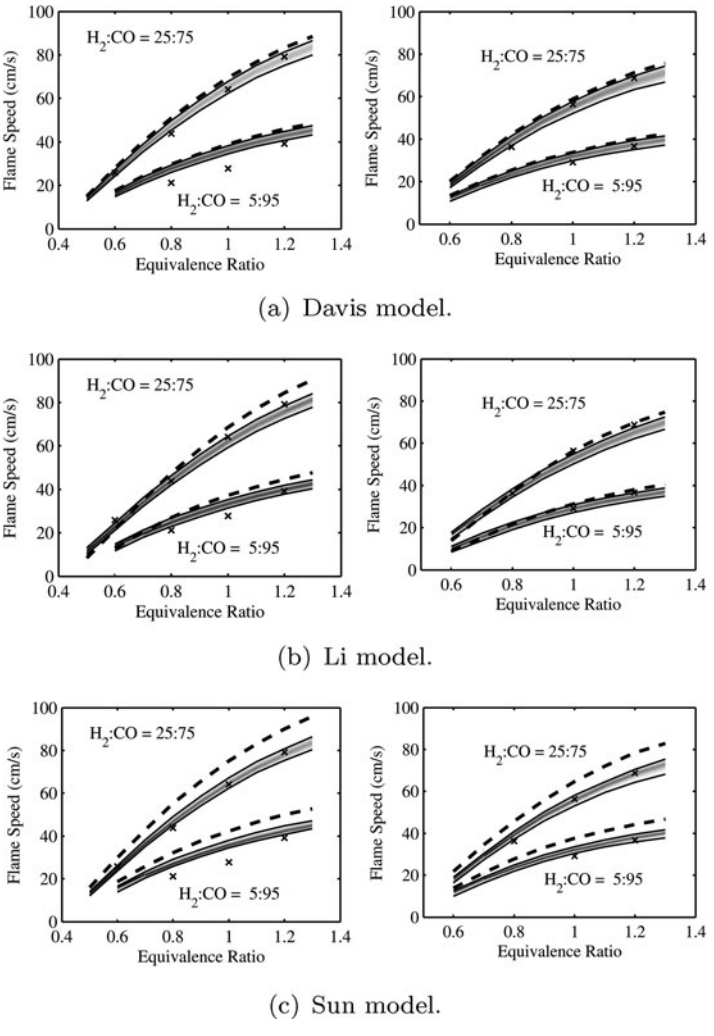


Figure 6. Flame speed results for (left) 10 atm and (right) 20 atm, where grey intensity indicates probability between the 95% confidence interval bounds, dotted lines represent results using pre-calibrated parameters, and symbols (x) represent experimental results [33].

higher flame speed compared to the calibrated models for the same conditions. Overall the uncertainty in the predictions is roughly 10% over the range of equivalence ratios considered.

The right side of Figure 6 shows the flame speed computations for an operating pressure of 20 atm, also computed using the pushed forward posterior method. Note that the experimental data for this case are not part of the calibration data set. Thus, this calculation examines the effect of the kinetics parameter uncertainty in extrapolation. The uncertainty in the results increases compared to the 10 atm case, with the 95% confidence interval bands producing a wider spread of flame speeds. Overall, the SM seems to have gained the most in terms of accuracy, with the flame speeds computed from the original model being

highly inaccurate for all equivalence ratios considered. The LM produces the least spread in the simulations, indicating lower uncertainty in the parameters considered.

For both the 10 and 20 atm computations, the maximum a posteriori (MAP) estimates of the flame speed in many cases differ from the experimental measurements. In particular, the 10 atm results with a H_2:CO ratio of 5:95 show a significant disparity from measurement. Even with the kinetic parameter uncertainty propagated through the simulations, the uncertainty spread of the flame speeds does not capture the experimental results. In order to explore the simulation uncertainty further, the posterior predictive propagation method has also been employed. In this method, the samples of the hyperparameter σ are used in addition to the kinetics parameters. Specifically, for the jth sample $(\boldsymbol{\theta}_j, \sigma_j)$, the predicted flame speed η_j is implied by the chosen likelihood function. For example, using the additive model from (4) leads to

$$\eta_j = \eta(\boldsymbol{\theta}_j) + N\left(0, \sigma_j^2\right),$$

where $\eta(\boldsymbol{\theta}_j)$ is the flame speed given by the chemical model with kinetics parameters $\boldsymbol{\theta}_j$, and $N(0, \sigma_j^2)$ is the zero-mean Gaussian with standard deviation σ_j. Thus, in this method, both the parameter uncertainty and the combined model/experimental error term explicitly contribute to the uncertainty of the simulations. With this propagation method, the 95% confidence interval spread encompasses the data. This result can be seen in Figure 7 which shows the posterior predictive flame speed results for 10 and 20 atm for all three models.

Given a properly chosen prior, the above results indicate one of two conclusions: either model form error or experimental error must explain the discrepancy. This can also be seen by examining the marginal posterior for σ directly, as shown in Figure 8.

Note that the MAP estimate of σ is a significant fraction of the predicted laminar flame speeds, particularly for lower equivalence ratios. For instance, at 20 atm with equivalence ratio 0.8 and H_2:CO ratio 5:95, the ratio of MAP σ to the experimental flame speed is approximately 25% for all three models. Even in the best case, the ratio is nearly 4%. This result indicates that the chemical models used here may not be rich enough to reproduce the calibration data. While this could be due to experimental error, given the large magnitude of σ, it seems likely that it is at least partially due to model form error. One way to enrich the current models would be to calibrate additional parameters, such as activation energies, third body efficiencies, and thermodynamic parameters, that have been assumed to be perfectly known in the current calibration.

4.2. *Bayesian evidence model comparison*

The three different syngas chemistry models form a model set. Thus, these models can be compared using the Bayesian framework described in Section 2.2. Different choices made in setting up the Bayesian calibration will affect the evidence results, including the choice of prior. This dependence is natural given that the evidence depends not only on how well the data are fitted but also on how much information is extracted from the data, as measured by the change from prior to posterior. Here, the comparison is made with Gaussian priors used for every parameter in each model. For brevity, we show results only for the additive error likelihood. However, a similar process could be used to compare the chemistry models using different prior and likelihoods and/or to compare the different prior and likelihood forms.

Table 2 shows the logarithm of the evidence and the posterior probability for each chemistry model. Using this measure, the LM is strongly preferred, with a posterior probability of 0.9696. Recalling the decomposition of the evidence shown in (11), this result implies

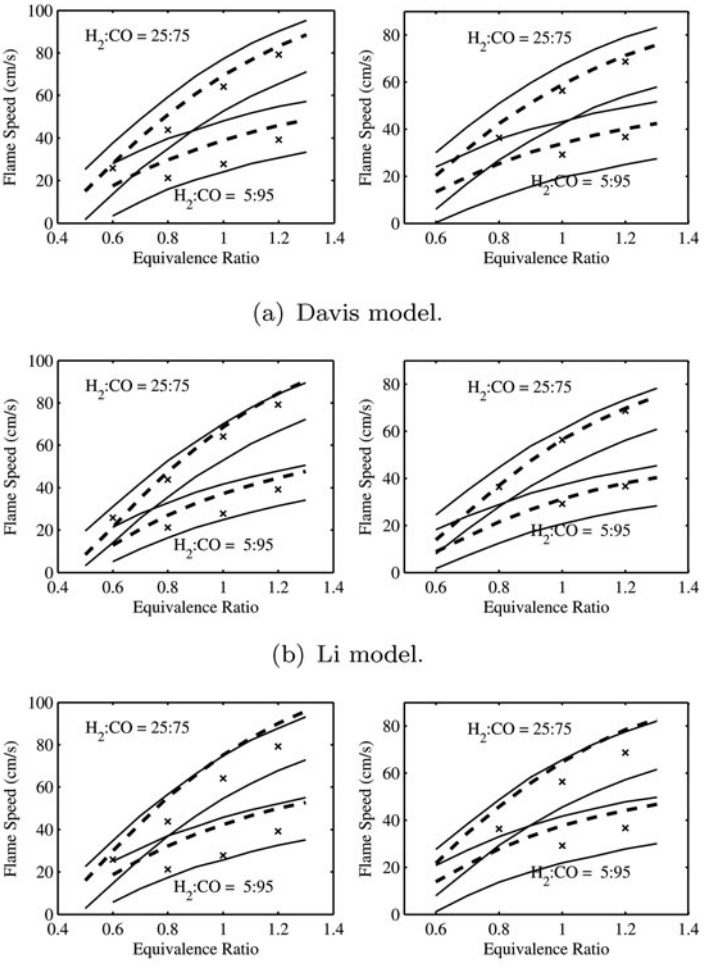


Figure 7. Posterior predictive flame speed results for (left) 10 atm and (right) 20 atm, where the solid lines indicate 95% confidence interval bounds, dotted lines represent results using pre-calibrated parameters, and the symbols '×' represent experimental results [33].

that the LM is able to fit the data well without extracting so much information from the data that the information gain term overwhelms the data fit term. Thus, the LM provides the best combination of data fit and minimal tuning of the parameters from their initial priors for this set of laminar flame speed data. This result does not imply that this kinetics model is the best model, but only that it is the best model of this set at reproducing the data used with minimal fitting.

4.3. *Prior selection*

Common choices for priors when little prior information exists include the uniform prior and the Gaussian prior [38]. The uniform prior assigns an equal probability to the parameter over a range of values bounded by a minimum and maximum value. However, care must be

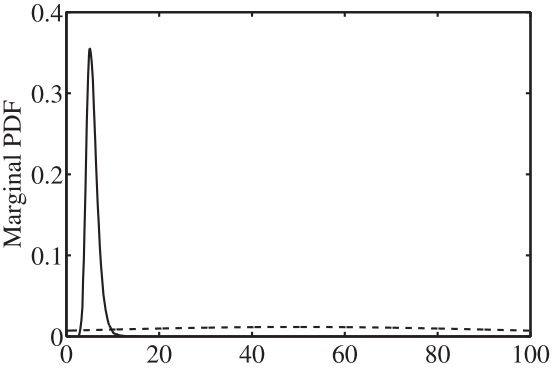


Figure 8. Marginal PDF of the error variance hyperparameter σ for the DM with prior shown as a dashed line.

taken since the posterior in the case of a uniform prior will not be able to span beyond the minimum or maximum of the prior. If a region of parameter space is given zero probability by the prior, the posterior will have zero probability in that region as well. The Gaussian prior is a common choice when given information about a parameter's mean and variance. This choice can be motivated by considering Shannon's information entropy, which can be interpreted as a measure of uncertainty of a random variable. One can show that, given only mean and variance, the Gaussian distribution maximises this entropy [39]. Similarly, the uniform prior provides the highest information entropy given only a minimum and maximum. Here, the results using uniform priors and Gaussian priors for the DM are compared and discussed.

Figure 2 shows sample parameter distributions for the Davis model given Gaussian priors in the left column and uniform priors in the right column. For the case of reactions $O + H_2 \rightleftharpoons H + OH$ and $H_2O_2 + OH \rightleftharpoons HO_2 + H_2O$, some information is learned regarding the parameters when the uniform prior is applied. The posteriors have obtained peaks, but are not highly peaked. For reactions $OH + H_2 \rightleftharpoons H + H_2O$ and $CO + OH \rightleftharpoons CO_2 + H$ when applying the uniform prior, the posterior distributions become more highly peaked; however, the distributions also abut against the bounds of the prior distribution. This effect shows that the bounds on the prior can limit the posterior distribution. As mentioned earlier, where the prior distribution has zero probability, the posterior must have zero probability. As a result the posteriors for the parameters in those two reactions are limited and do not reach the parameter values attainable if not bounded. When the Gaussian prior is applied, the posterior distributions shift for those two pre-exponential coefficients, with the distribution for $CO + OH \rightleftharpoons CO_2 + H$ becoming slightly more peaked. Most importantly, the support of

Table 2. Log evidence and posterior probability for all three models with additive error form and Gaussian prior.

Model	Log evid.	Post. prob.
DM	−61.6433	0.0127
LM	−57.3094	0.9696
SM	−61.3102	0.0177

the Gaussian priors allows the posterior distributions to attain the distributions as informed by the data.

4.4. *Error model comparison*

The Bayesian methodology provides for the utilisation of different error models, which are incorporated in the likelihood function. Prior knowledge regarding how the error will develop for a particular problem can be integrated into the form of the likelihood. In the results above in Section 4.1, the additive error model has been used. In the following, the additive and multiplicative error models, detailed in Section 2.1, are compared for the LM.

The different error models result in different posterior parameter distributions. Samples of the one-dimensional marginal distributions are plotted in Figure 9. For some of the parameters, in particular the coefficient for $HO_2 + H \rightleftharpoons H_2 + O_2$, the multiplicative error model leads to a more peaked distribution with a different MAP estimate.

Furthermore, the propagated results for the multiplicative error form display a broadening of the flame speed uncertainty in comparison with the additive model as seen in Figure 10. The multiplicative treatment of the error model, although resulting in less certainty in the results, is not necessarily a poor choice of error model. A reduced uncertainty does not imply that the additive error model is the correct model for a particular application.

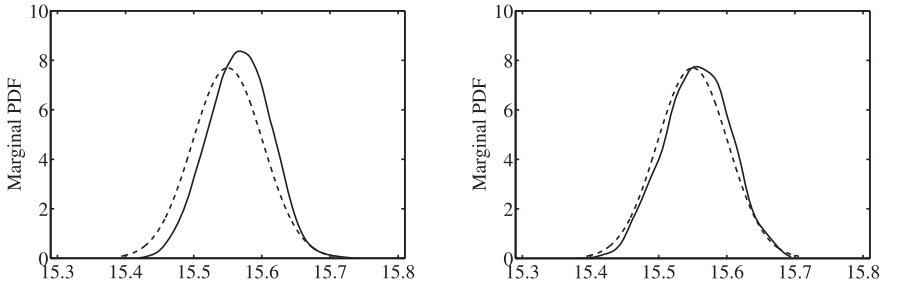
The additive error model weights all model error equally amongst the separate experimental conditions. Thus, for conditions which result with flame speeds on the order of 100 cm/s, a relative error of say 5% will affect the model error term more drastically than conditions with a flame speed on the order of 10 cm/s. With a multiplicative model error term, which scales with the magnitude of the result, the error at each condition is weighted more naturally so that the conditions with higher flame speeds do not dominate the error from conditions with lower flame speeds. Such a treatment is not unknown in combustion applications, and it has been shown to be beneficial when modelling chemical kinetics [2].

Table 3 shows the logarithm of the evidence and the posterior probability for the Bayesian update using additive and multiplicative error models. Much like comparing different kinetics models, the evidence may be used to compare results from the same kinetics model, but with different applied error models. The two cases with different applied error models then become two different model classes within the set of models. Using the evidence measure, the multiplicative error model is strongly preferred, with a posterior probability of essentially one. Similar to the result for the kinetics model comparison, this result implies that the multiplicative error model provides the best combination of data fit and minimal tuning of the parameters for this set of laminar flame speed data.

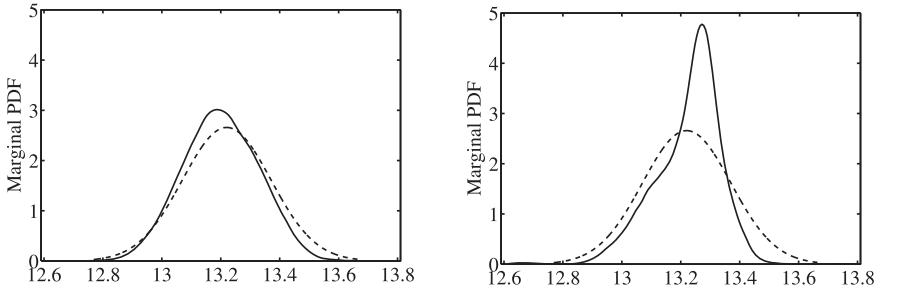
Other error models could potentially provide a better match. Parameterisfing an error model proportional to some experimental parameters such as pressure may prove to fit the data better. The downside is that these parametric relations may not be known a priori and could require a significant number of parameters.

Table 3. Log evidence and posterior probability for the LM with additive and multiplicative error forms and Gaussian prior.

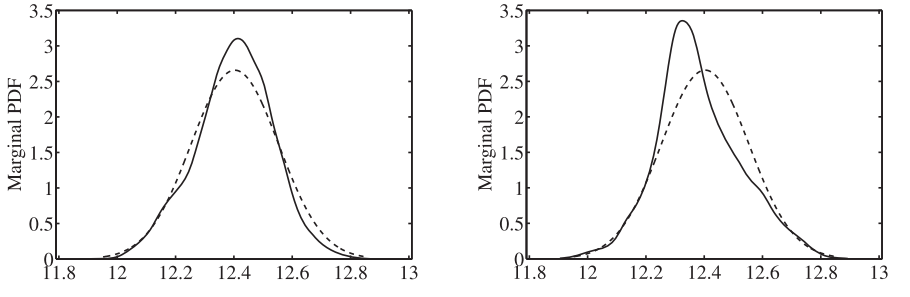
Error model	Log evid.	Post. prob.
Additive	−57.31	0.0
Multiplicative	6.236	1.0



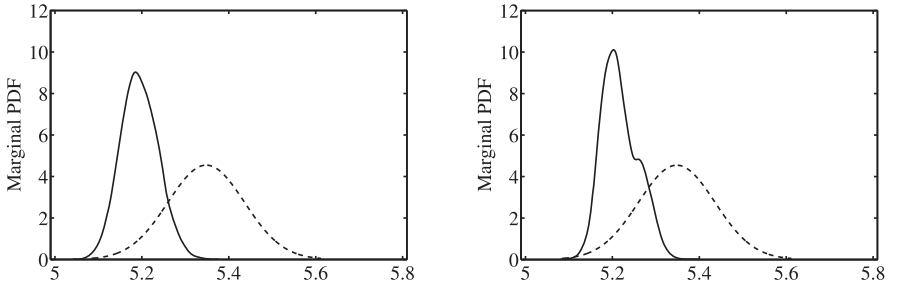
(a) Pre-exponential coefficient for $H + O_2 \rightleftharpoons O + OH$.



(b) Pre-exponential coefficient for $HO_2 + H \rightleftharpoons H_2 + O_2$.



(c) Pre-exponential coefficient for $CO + O_2 \rightleftharpoons CO_2 + O$.



(d) Pre-exponential coefficient for $CO + OH \rightleftharpoons CO_2 + H$.

Figure 9. Prior and final marginal PDFs of Arrhenius pre-exponential parameters for the LM with additive error (left) and multiplicative error (right).

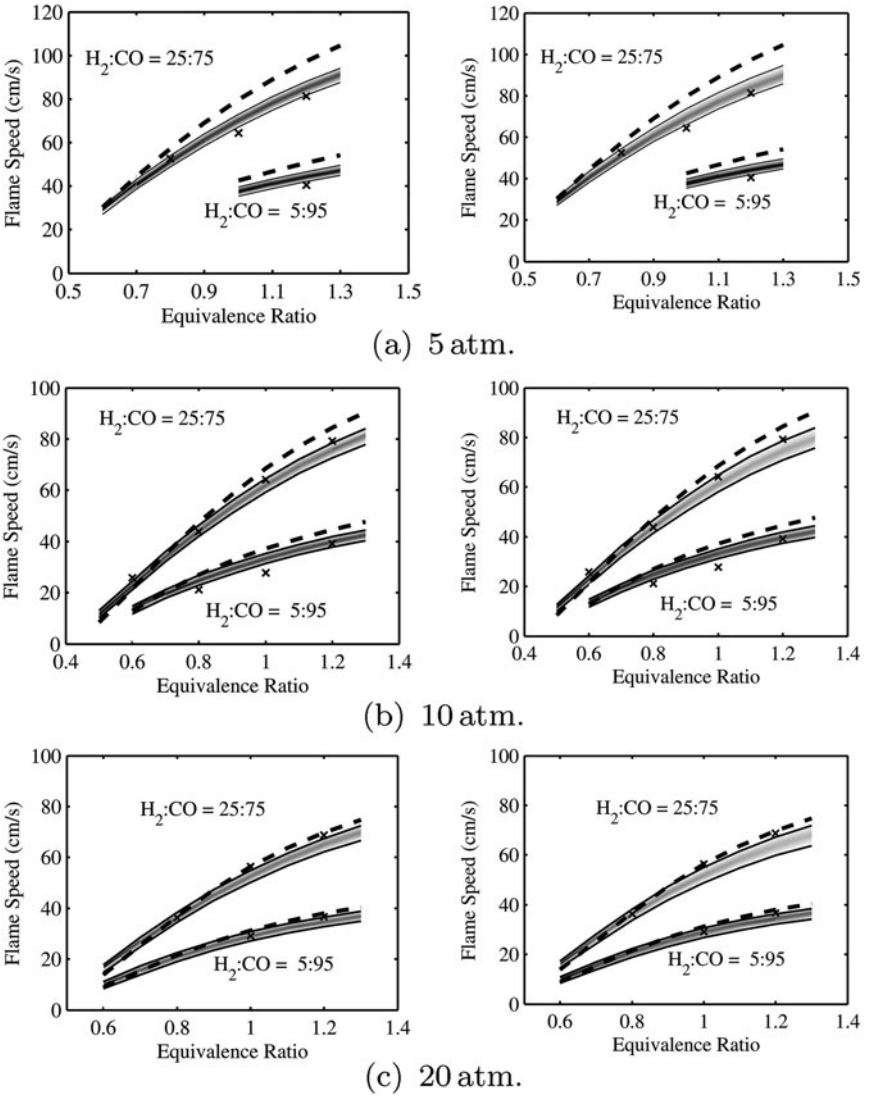


Figure 10. Flame speed results for the LM with additive error (left) and multiplicative error (right), where grey intensity indicates probability between the 95% confidence interval bounds, dotted lines represent results using pre-calibrated parameters, and the symbols '×' represent experimental results [33].

5. MUM-PCE and Bayesian approach comparison results

The method of Sheen and Wang [10–12], referred to as MUM-PCE, is an interesting technique that combines aspects of the Bayesian formulation with regression-type error minimisation to produce updated parameters. In this section, the specific assumptions made in this technique are analysed in order better to understand the relative importance of model and error forms on the calibration process. In Appendix 6, it is shown that the MUM-PCE is a simplified Bayesian technique that utilises the following simplifications: (1) the target model is replaced by a surrogate model, (2) the surrogate model is assumed

to be linear with regard to the parameters, and (3) the error in the predictions appears only from experimental uncertainty. In other words, if the experimental error were to reduce to zero, the parameters would be tuned to capture the experimental data with no resultant uncertainty in either the parameters or the simulation results.

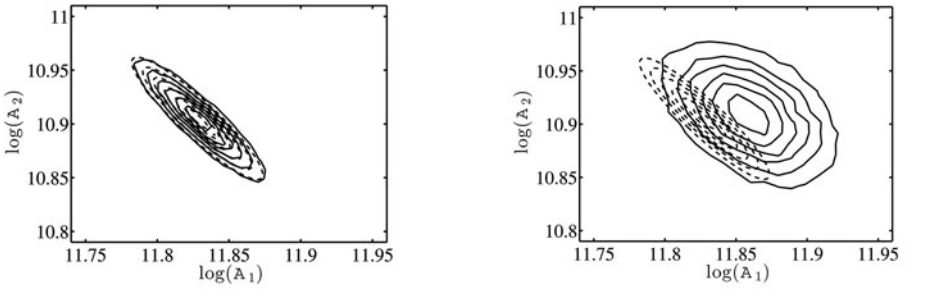
During the parameter update process with MUM-PCE, a surrogate model, the quadratic response surface model, is assumed for the flame speed. This assumption varies from the direct use of the target model in the Bayesian methodology. Furthermore, the surrogate model is linearised with respect to the model parameters when determining the covariance between parameters. This represents a further step from the Bayesian method. Next, treating the error in the parameter update process as strictly from experimental uncertainty neglects any error inherent in the kinetics model itself. This assumption applied to the Bayesian methodology is equivalent to assuming an additive error in which no additional error variance hyperparameter is employed. Applying these assumptions to the sampling-based Bayesian method results in nearly identical results in comparison to MUM-PCE results. Removing those assumptions one at a time reveals how the assumptions affect the outcome.

Applying the above described assumptions incrementally leads to a set of tests of the modified Bayesian method. These tests are categorised by choice of target model, linearised response surface, full response surface, or true flame speed model, and by choice of model error, either additive error with fixed $\sigma = \sigma_{\text{obs}} = 2.0$ or additive error with the inferred error variance hyperparameter. The following sections detail the effects of removing the Sheen and Wang assumptions from the modified sampling-based Bayesian method. These test cases were all performed using a single kinetics model, the DM, for brevity. Section 5.1 describes the posterior distributions which result from the incremental testing, and Section 5.2 describes the resultant flame speed distributions from the uncertainty propagation.

5.1. *Marginal parameter distributions*

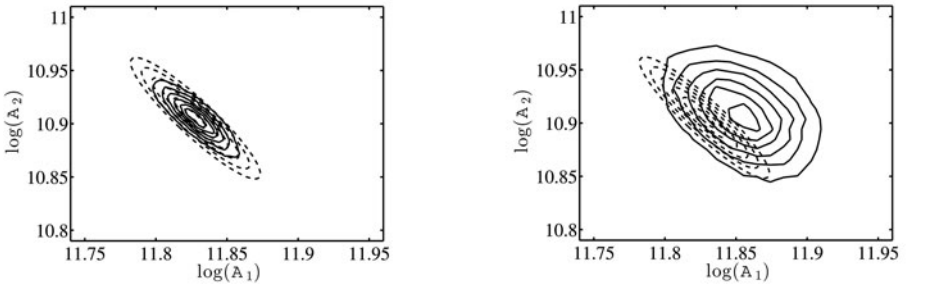
Figure 11 shows two-dimensional plots of the CO + OH $\rightleftharpoons$ CO_2 + H and of the duplicate CO + OH $\rightleftharpoons$ CO_2 + H pre-exponential parameters for the different options. The first reaction is referenced as R1 with parameter A_1, and the second reaction is referenced as R2 with parameter A_2. When the Bayesian update is performed with the linearised response surface and fixed additive error, hereafter referred to as the baseline and shown in plots labelled with (a), the solution closely matches that of the MUM-PCE method. With the modification to the full response surface shown in the plots labelled (c), the solution changes from the baseline. While the parameter MAP values remain essentially the same, the shape, size and alignment of the probability contours change. These changes can manifest themselves as a modification of the multivariate Gaussian covariance, as well as a shift away from a multivariate Gaussian to an arbitrary distribution. Such an outcome shows that the use of linearised response surfaces affects the shape and size of the probability surfaces, here two-dimensional marginal distribution contours.

When the error standard deviation hyperparameter is inferred, i.e. updated along with the kinetics parameters in the Bayesian update, the shape, size, alignment and location of the probability contours changes. These changes are shown in plots labelled (b) and (d). The MAP parameter values show a significant shift and the sizes of the probability contours show a significant increase. The MAP parameter value shift is due to the shift in the region of the error parameter space. In Figure 12, which shows kinetics parameters versus the error term, the corresponding MAP kinetics parameter values for each error value are plotted with dashes. Following the dashed line towards an error of $\sigma_{\text{obs}} = 2.0$



(a) Bayesian with linearised response surface and error term specified.

(b) Bayesian with linearised response surface and error term inferred.



(c) Bayesian with quadratic response surface and error term specified.

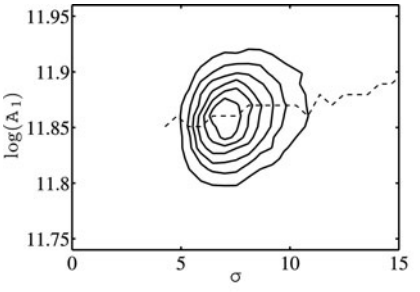
(d) Bayesian with quadratic response surface and error term inferred.

Figure 11. Bayesian posterior (solid) and MUM-PCE posterior (dashed) two-dimensional marginal PDFs of R1 and R2 pre-exponential coefficient A for the DM with two parameters varied.

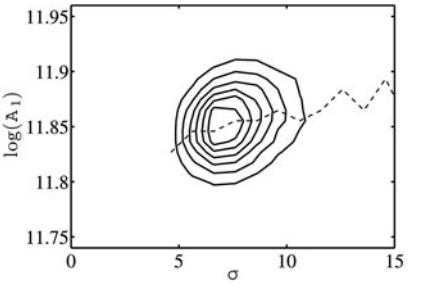
shows the shift in the MAP kinetic parameter value, and thus that the results from an error of $\sigma = 2.0$ can be extrapolated as a solution for the parameter space with the inferred error term hyperparameter. Regarding the expansion of the probability contours, including the error model standard deviation as a hyperparameter to be inferred reduces the certainty with which the kinetics parameters are determined. The modelled flame speed results vary from the corresponding experimental results with a standard deviation greater than that assumed for the MUM-PCE solution. With the MUM-PCE assumed experimental error standard deviation, the kinetics parameters are determined with more certainty than the modelled results themselves indicate. Thus, the kinetics parameters are determined with more certainty than warranted.

5.2. *Propagated flame speed distributions*

Figure 13 displays the flame speed propagation results. The results for the Bayesian cases vary when switching from fixed to inferred error standard deviation. With the linearised response surface and inferred error standard deviation shown in Figure 13(c), the uncertainty is greater as compared with fixed error standard deviation shown in Figure 13(b). Similar



(a) Bayesian with linearised response surface and error term inferred.



(b) Bayesian with quadratic response surface and error term inferred.

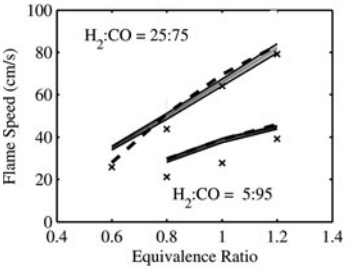
Figure 12. Bayesian posterior (solid) two-dimensional marginal PDFs of the error standard deviation and R1 pre-exponential coefficient A for the DM with two parameters varied, with a line of MAP value (dashed).

results are displayed with the quadratic response surface shown in Figures 13(e) and 13(d). Such results imply one of two conclusions. First, using the error model standard deviation as a hyperparameter in the inference problem leads to larger uncertainty in the kinetics parameters and consequently the flame speed solutions. Second, specifying the standard deviation of the error term unnecessarily increases certainty in the kinetics parameters and consequently the flame speed solutions.

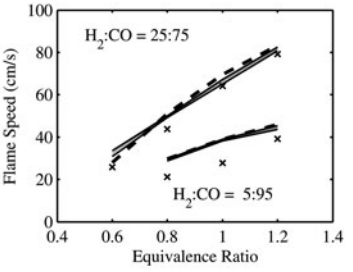
The flame speed results for the full Bayesian case shown in Figure 13(f) exhibit a closer match to the experimental data than the response surface cases with inferred error σ. This result indicates that calibration and propagation with the true rather than surrogate model likely provides more accurate flame speed results in the parameter space regions farther from the nominal parameter values where the linearisation assumptions break down. Table 4 lists the mean, RMS and maximum error between the response surface and Chemkin-determined flame speeds for the two experimental conditions leading to the largest and smallest error. The table lists those errors for flame speeds determined from parameter values chosen on an evenly spaced grid within the parameter space bounded by two standard deviations from the mean parameter values. Additionally, the table lists the error at the MAP estimate for the parameters. Although for some experimental conditions the error is low, for others the error reaches magnitudes of more than 10 cm/s near the boundaries of the response surface. Looking at the MAP values, the response surface errors reach a magnitude as high as 5.1 cm/s. Such errors potentially can lead to errors in the determination of the nominal parameter values and parameter covariances.

Table 4. Mean and RMS error ε in cm/s between the response surface and Chemkin-determined flame speeds for the experimental conditions which lead to the largest and smallest error.

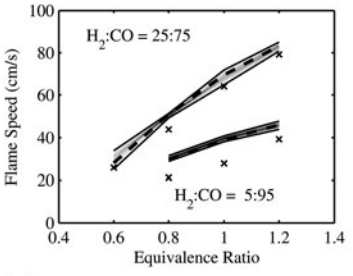
Expt	Mean ε	RMS ε	Max ε	RS MAP ε	Bayesian MAP ε
8	2.2	2.1	10.8	5.1	2.1
9	0.1	0.1	0.2	0.1	0.0



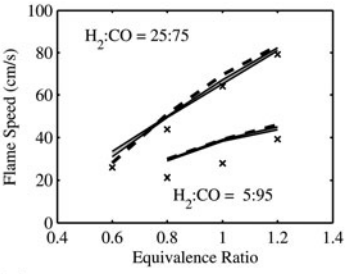
(a) R1 and R2 varied, MUMPCE.



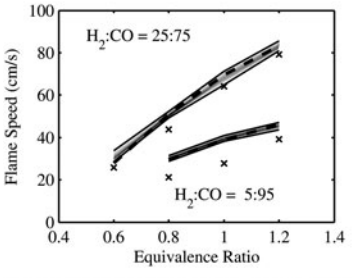
(b) R1 and R2 varied, linearised response surfaces with σ specified.



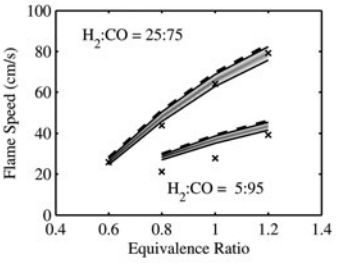
(c) R1 and R2 varied, linearised response surfaces with σ inferred.



(d) R1 and R2 varied, quadratic response surfaces with σ specified.



(e) R1 and R2 varied, quadratic response surfaces with σ inferred.



(f) R1 and R2 varied, full Bayesian with σ inferred.

Figure 13. Flame speed results at 10 atm for the Davis *et al.* [31] model, where grey intensity indicates probability between the 95% confidence interval bounds, dotted lines represent results using the nominal parameters, and the symbols 'x' represent experimental results [33].

6. Conclusions

Bayesian methods provide a powerful framework for quantifying uncertainty in syngas combustion applications. Bayesian calibration, given a set of experimental data to compare against, both improves calibration of the syngas chemistry models as well as providing updated distributions for the parameters of those models. Those distributions can then be propagated forward into simulations of laminar flame speed to determine the uncertainty in their predictions. The same framework can also be used to rank a set of candidate models.

The application of this approach to existing experimental data and a select group of chemistry models was conducted here. It was found that the model due to Sun [33] undergoes

significant changes in order to capture the experimental data. All three models showed some increase to uncertainty in the results at higher pressures, even when calibrated using experimental data at 10 atm, confirming prior analyses that suggest increased sensitivity of results to model coefficients at higher pressures. The functional form and support of the prior PDF was found to have an important effect on the calibration results. Specifically, the bounds on the uniform prior set based on previous uncertainty estimates can be too narrow to allow a good calibration. Additionally, the error model affects the Bayesian calibration and forms an integral part of the calibration process.

MUM-PCE has also been applied to the same syngas flame speed simulations. MUM-PCE has been shown to be a simplified version of the Bayesian method after the application of several assumptions. Furthermore, in comparison with corresponding Bayesian results, the effects of those assumptions are shown to affect the update of the parameter hypervolume shape and parameter distributions significantly, as predicted by data collaboration analysis [13]. This indicates that modelling of uncertainty itself will alter the results.

The Bayesian calibration process provides a rigorous mechanism for incorporating new data in order to improve existing models. The availability of such tools is important for engine design for two different reasons. The use of uncertainty estimates provides a better characterisation of the state of knowledge, and allows for robust decision making. Further, these estimates also provide information about the most important models and model parameters, allowing better allocation of limited resources.

Acknowledgements

This material is based on work supported by the Department of Energy (National Nuclear Security Administration) under Award Number DE-FC52-08NA28615. In addition, one of the authors (VR) was partially supported by DOE-NETL grant DE-FE0007107 with Steven Richardson as the program monitor.

References

[1] M. Frenklach, *Systematic optimization of a detailed kinetic model using a methane ignition example*, Combust. Flame 58 (1984), pp. 69–72. Available at http://dx.doi.org/10.1016/0010-2180(84)90079-8.

[2] M. Frenklach, H. Wang, and M.J. Rabinowitz, *Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method – combustion of methane*, Prog. Energy Combust. Sci. 18 (1992), pp. 47–73. Available at http://dx.doi.org/10.1016/0360-1285(92)90032-V.

[3] R. Feeley, P. Seiler, A. Packard, and M. Frenklach, *Consistency of a reaction dataset*, J. Phys. Chem. A 108 (2004), pp. 9573–9583. Available at http://dx.doi.org/10.1021/jp047524w.

[4] R. Feeley, M. Frenklach, M. Onsum, T. Russi, A. Arkin, and A. Packard, *Model discrimination using data collaboration*, J. Phys. Chem. A 110 (2005), pp. 6803–6813.

[5] P. Seiler, M. Frenklach, A. Packard, and R. Feeley, *Numerical approaches for collaborative data processing*, Optim. Eng. 7 (2006), pp. 459–478. Available at http://dx.doi.org/10.1021/jp076861c.

[6] T. Russi, A. Packard, R. Feeley, and M. Frenklach, *Sensitivity analysis of uncertainty in model prediction*, J. Phys. Chem. A 112 (2008), pp. 2579–2588. Available at http://dx.doi.org/10.1021/jp076861c.

[7] M.T. Reagan, H.N. Najm, R.G. Ghanem, and O.M. Knio, *Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection*, Combust. Flame 132 (2003), pp. 545–555. Available at http://dx.doi.org/10.1016/S0010-2180(02)00503-5.

[8] M.T. Reagan, H.N. Najm, B.J. Debusschere, O.P. Le Maître, O.M. Knio, and R.G. Ghanem, *Spectral stochastic uncertainty quantification in chemical systems*, Combust. Theory Model. 8 (2004), pp. 607–632. Available at http://dx.doi.org/10.1088/1364-7830/8/3/010.

[9] M.T. Reagan, H.N. Najm, P.P. Pébay, O.M. Knio, and R.G. Ghanem, *Quantifying uncertainty in chemical systems modeling*, Int. J. Chem. Kinetics 37 (2005), pp. 368–382. Available at http://dx.doi.org/10.1002/kin.20081.

[10] D.A. Sheen, X. You, H. Wang, and T. Lovas, *Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion*, Proc. Combust. Inst. 32 (2009), pp. 535–542. Available at http://dx.doi.org/10.1016/j.proci.2008.05.042.

[11] D.A. Sheen and H. Wang, *The method of uncertainty quantification and minimization using polynomial chaos expansions*, Combust. Flame 158 (2011), pp. 2358–2374. Available at http://dx.doi.org/10.1016/j.combustflame.2011.05.010.

[12] D. Sheen and H. Wang, *Modeling high-pressure $H_2/CO/O_2$/diluent mass burning rates with the method of uncertainty minimization using polynomial chaos expansions*, in *7th US National Technical Meeting of the Combustion Institute*, 20–23 March 2011, Atlanta, GA, Vol. 1, The Combustion Institute, Pittsburgh, PA, 2011, pp. 185–208.

[13] T. Russi, A. Packard, and M. Frenklach, *Uncertainty quantification: Making predictions of complex reaction systems reliable*, Chem. Phys. Lett. 499(1) (2010), pp. 1–8.

[14] I. Wender, *Reactions of synthesis gas*, Fuel Process. Technol. 48 (1996), pp. 189–297.

[15] H. Herzog and N. Vukmirovic, *CO_2 Sequestration: Opportunities and Challenges*, in *7th Clean Coal Technology Conference*, June 1999, Knoxville, TN, 1999.

[16] M.P. Burke, M. Chaos, F.L. Dryer, and Y. Ju, *Negative pressure dependence of mass burning rates of $H_2/CO/O_2$/diluent flames at low flame temperatures*, Combust. Flame 157 (2010), pp. 618–631.

[17] R.T. Cox, *The Algebra of Probable Inference*, Johns Hopkins University Press, Baltimore, MD, 1961.

[18] R.P. Christian, *The Bayesian Choice*, Springer, New York, 2001.

[19] E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, 2003.

[20] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer, New York, 2005.

[21] D. Calvetti and E. Somersalo, *Introduction to Bayesian Scientific Computing*, Springer, New York, 2007.

[22] H.N. Najm, B.J. Debusschere, Y.M. Marzouk, S. Widmer, and O.P. Le Maître, *Uncertainty quantification in chemical systems*, Int. J. Numer. Meth. Engrg 80 (2009), pp. 789–814. Available at http://dx.doi.org/10.1002/nme.2551.

[23] X. Huan and Y. Marzouk, *Optimal Bayesian experimental design for combustion kinetics*, in *49th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, Orlando, FL, 4–7 January 2011, Paper AIAA 2011-0513, American Institute of Aeronautics and Astronautics, Reston, VA, 2011.

[24] K. Miki, S.H. Cheung, E.E. Prudencio, and P.L. Varghese, *Bayesian uncertainty quantification of recent shock tube determinations of the rate coefficient of reaction $H + O_2 = OH + O$*, Int. J. Chem. Kinetics 44 (2012), pp. 586–597.

[25] K. Miki, M. Panesi, E. Prudencio, and S. Prudhomme, *Probabilistic models and uncertainty quantification for the ionization reaction rate of atomic nitrogen*, J. Comput. Phys. 231 (2012), pp. 3871–3886.

[26] M. Muto and J.L. Beck, *Bayesian updating and model class selection of hysteretic structural models using stochastic simulation*, J. Vib. Control 14 (2008), pp. 7–34.

[27] S. Kullback and R.A. Leibler, *On information and sufficiency*, Ann. Math. Statist. 22 (1951), pp. 79–86.

[28] S.H. Cheung and J.L. Beck, *New Bayesian updating methodology for model validation and robust predictions based on data from hierarchical subsystem tests*, EERL Report No. 2008-04, California Institute of Technology, 2008.

[29] E.E. Prudencio and S.H. Cheung, *Parallel adaptive multilevel sampling algorithms for the Bayesian analysis of mathmatical models*, Int. J. Uncertainty Quant. 2 (2012), pp. 215–237.

[30] E. Prudencio and K.W. Schulz, *The parallel C++ statistical library QUESO: Quantification of Uncertainty for Estimation, Simulation and Optimization*, in *Workshop on Algorithms and Programming Tools for Next-Generation High-Performance Scientific Software (HPSS)*, Bordeaux, France, 2011.

[31] S.G. Davis, A.V. Joshi, H. Wang, and F. Egolfopoulos, *An optimized kinetic model of H_2/CO combustion*, Proc. Combust. Inst. 30 (2005), pp. 1283–1292. Available at http://dx.doi.org/10.1016/j.proci.2004.08.252.

[32] J. Li, Z. Zhao, A. Kazakov, M. Chaos, F.L. Dryer, and J.J. Scire, *A comprehensive kinetic mechanism for CO, CH_2 O, and CH_3 OH combustion*, Int. J. Chem. Kinetics 39 (2007), pp. 109–136.
[33] H. Sun, S. Yang, G. Jomaas, and C. Law, *High-pressure laminar flame speeds and kinetic modeling of carbon monoxide/hydrogen combustion*, Proc. Combust. Inst. 31 (2007), pp. 439–446. Available at http://dx.doi.org/10.1016/j.proci.2006.07.193.
[34] J. Li, Z. Zhao, A. Kazakov, and F.L. Dryer, *An updated comprehensive kinetic model of hydrogen combustion*, Int. J. Chem. Kinetics 36 (2004), pp. 566–575.
[35] J. Warnatz, *Resolution of gas phase and surface combustion chemistry into elementary reactions*, Symp. (Int.) Combust. 24 (1992), pp. 553–579.
[36] R. Kee, J. Grcar, M. Smooke, and J. Miller, *A Fortran program for modeling steady laminar one-dimensional premixed flames*, SAND85-8240, Sandia National Laboratory, Albuquerque, NM, 1985.
[37] R. Kee, G. Dixon-Lewis, J. Warnatz, M. Coltrin, and J. Miller, *A Fortran computer package for the evaluation of gas-phase, multicomponent transport properties*, SAND86-8246, Sandia National Laboratory, Albuquerque, NM, 1986.
[38] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, Chapman & Hall, Boca Raton, FL, 2003.
[39] S. Goldman, *Information Theory*, Prentice-Hall, New York, 1953.
[40] D. Miller and M. Frenklach, *Sensitivity analysis and parameter estimation in dynamic modeling of chemical kinetics*, Int. J. Chem. Kinetics 15 (1983), pp. 677–696. Available at http://dx.doi.org/10.1002/kin.550150709.

Appendix A. Bayesian representation of the MUM-PCE approach of Sheen and Wang [10–12]

The basic approach of Sheen and Wang solves a limited form of the Bayesian inverse problem, and thus follows many of the ideas of Bayesian uncertainty quantification. Given a set of experimental results, the method solves for a nominal set of parameters and the covariance values amongst them. Such parameter determination is used to calculate the resultant uncertainties which are reduced in relation to uncertainties resulting from the initial values and variances for all parameters. We provide this discussion as a means of evaluating model parameter calibration processes by their effect on the posterior distribution.

A.1. *The Sheen and Wang method*

In the Sheen and Wang method, the kinetics parameters involved are the pre-exponential coefficients k_i. Each parameter is normalised to a range of -1 to 1 with

$$x_i = \frac{\ln k_i / k_{i,0}}{\ln f_i},$$

where $k_{i,0}$ is the pre-calibrated value for k_i and f_i is the multiplicative uncertainty factor. The method proceeds by treating the set of parameters $\boldsymbol{x}$ as a set of random variables with the following polynomial chaos expansion:

$$\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{i=1}^{M} \boldsymbol{\alpha}_i \xi_i + \sum_{i=1}^{M} \sum_{j=1}^{M} \boldsymbol{\beta}_{ij} \xi_i \xi_j + \cdots,$$

where ξ_i are standard random variables, typically treated by the method as standard normal random variables, $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_{ij}$ are the expansion coefficients, and M is the number of random variables used in the expansion. This expansion is simplified to the following in order to

allow the analytical simplifications of the method:

$$\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{i=1}^{M} \boldsymbol{\alpha}_i \xi_i.$$

The parameters $\boldsymbol{x}$ form a multivariate Gaussian distribution with mean $\boldsymbol{x}_0$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha}$.

The Sheen and Wang method replaces the target combustion models with surrogate models, quadratic response surfaces, the use of which has been dubbed 'solution mapping' [40]. The response surfaces are generated from selected simulations of the target combustion model. The selected simulations are intended to provide response surface support over the range of necessary simulation conditions. The response surfaces consist of the nominal modelled value $\eta_{r,0}$ and sets of coefficients a_i and b_{ij} for each experimental condition. The modelled value η_r for each experiment r is calculated as

$$\eta_r(\boldsymbol{x}) = \eta_{r,0} + \sum_{i=1}^{N} a_i x_i + \sum_{i=1}^{N}\sum_{j\geq 1}^{N} b_{ij} x_i x_j,$$

where N is the number of optimised parameters.

Determination of the parameters in the Sheen and Wang method involves two steps. Step one is the optimisation of the parameter nominal values. This step minimises the objective function Φ, which is the sum of the least squares difference between the target experimental results, and the simulation results and a parameter weighting term

$$\Phi(\boldsymbol{x}) = \sum_{r=1}^{n} \left(\frac{\eta_r(\boldsymbol{x}) - \eta_r^{\mathrm{obs}}}{\sigma_r^{\mathrm{obs}}} \right)^2 + \sum_{k=1}^{N} 4x_k^2, \tag{A1}$$

where n is the number of experimental targets. Step two is the determination of the parameter covariance matrix. The objective function Φ is employed now in the construction of the PDF of a multivariate Gaussian of the parameters

$$p(\boldsymbol{x}) = A \exp(-0.5\Phi(\boldsymbol{x})), \tag{A2}$$

where A is a normalisation constant. The distribution has its mean as the optimal parameter values $\boldsymbol{x}^*$ and has covariance matrix $\boldsymbol{\Sigma}$. Linearising the response surface about the optimal parameter values leads to an expression for the covariance matrix:

$$\boldsymbol{\Sigma} = \left[\sum_{r=1}^{n} \frac{1}{(\sigma_r^{\mathrm{obs}})^2} (\boldsymbol{b}\boldsymbol{x}_0^*\boldsymbol{x}_0^{*\mathrm{T}}\boldsymbol{b} + \boldsymbol{a}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{a}^{\mathrm{T}} + \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}) + 4\mathbf{I} \right]^{-1}.$$

This expression is solved for the polynomial chaos expansion coefficients $\boldsymbol{\alpha}^*$ from the multivariate Gaussian treatment of $\boldsymbol{x}$, for which

$$\boldsymbol{x}^* = \boldsymbol{x}_0^* + \sum_{i=1}^{M} \boldsymbol{\alpha}_i^* \xi_i,$$

with mean $\boldsymbol{x}_0^*$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\alpha}^{*\mathrm{T}}\boldsymbol{\alpha}^*$.

A.2. *The Bayesian derivation of the Sheen and Wang method*

The Bayesian method aims to minimise the difference between experimental results and simulation results using an objective function, the likelihood function. Given Gaussian experimental and model form error, the likelihood function starts from the following form:

$$\pi(\boldsymbol{x};\boldsymbol{\eta}^{\text{obs}}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{\eta}(\boldsymbol{x}) - \boldsymbol{\eta}^{\text{obs}}\right)^{\text{T}} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\eta}(\boldsymbol{x}) - \boldsymbol{\eta}^{\text{obs}}\right)\right],$$

where the covariance matrix $\boldsymbol{\Sigma}$ serves as the error model term. The Sheen and Wang method uses an additive error term which involves only the experimental error, which can be written as

$$\Sigma_{ij} = \begin{cases} \sigma_i^{\text{obs}} & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases}$$

Applying this assumption to the Bayesian likelihood leads to a simplification of the likelihood to the following:

$$\pi(\boldsymbol{x};\boldsymbol{\eta}^{\text{obs}}) = \frac{1}{(2\pi)^{n/2}\left(\prod_{r=1}^{n}\sigma_r^{\text{obs}}\right)^{1/2}} \exp\left[-\frac{1}{2}\sum_{r=1}^{n}\left(\frac{\eta_r(\boldsymbol{x}) - \eta_r^{\text{obs}}}{\sigma_r^{\text{obs}}}\right)^2\right].$$

The Sheen and Wang method treats the priors typically as normal distributions with zero mean and a standard deviation of 0.5, thus

$$p_{\text{prior}}(\boldsymbol{x}) = \frac{1}{(0.5\pi)^{N/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(2x_i)^2\right].$$

Substituting the prior and likelihood function into the equation for the posterior parameters (2) results in

$$p(\boldsymbol{x}) = A \exp\left[-\frac{1}{2}\sum_{r=1}^{n}\left(\frac{\eta_r(\boldsymbol{x}) - \eta_r^{\text{obs}}}{\sigma_r^{\text{obs}}}\right)^2 - \frac{1}{2}\sum_{i=1}^{N}(2x_i)^2\right],$$

where A is again a normalisation constant. Therefore, with the additive experimental error model and normal priors, the Bayesian posterior is equivalent to the Sheen and Wang assumed parameter probability (A2) with objective function (A1).

Proceeding from the expression for the parameter joint distribution, the Sheen and Wang method employs additional assumptions which allow analytical calculation of the distributions rather than by Monte Carlo sampling. These simplifications reduce the complexity of the solution method and allow the uncertainty results to be calculated rapidly; however, they also lead to restrictions on the flexibility of results. The following list outlines four of the primary simplifications of the Sheen and Wang method.

(1) The target combustion models are replaced by surrogate models.
(2) The parameter distributions are expressed as first order polynomial chaos expansions.
(3) The surrogate models are linearised during calculation of parameter covariance and thus for resultant uncertainty.

(4) The error model involves only additive Gaussian experimental error.

The Sheen and Wang method replaces the target combustion models with quadratic response surfaces which act as surrogate models. While simplifying the combustion model to a mere algebraic relationship of the kinetics parameters, the response surface substitution requires that the model response follow a smooth quadratic response to changes in those parameters. The random variable treatment of those parameters is accomplished with first order polynomial chaos expansions. This simplification requires that the parameters be distributed in a multivariate Gaussian distribution. However, this assumption combined with the algebraic relationship of the response surfaces allows analytical simplification to the final equations for nominal parameter values and covariances. The calculation of the model parameter covariances involves the linearisation of the response surfaces about the nominal parameter values. This simplification allows further analytical simplification for the parameter covariances. Additionally, the error model in this method involves only the experimental error, which is equivalent to a constant additive error model in the Bayesian framework, where it manifests itself in the likelihood function. The error term is distributed as a zero mean Gaussian, which maintains the simplicity of the parameter covariance calculation. Furthermore, this error term precludes the possibility of including model form error which may be present in the modelling of the target combustion experiments in the uncertainty quantification process.

The result of these simplifications is a method that limits model parameters and resultant model uncertainties to multivariate Gaussian distributions. Nevertheless, the method follows the general methodology of the Bayesian framework. A physical phenomenon (say, laminar flame speed for syngas combustion) is approximated by a mathematical model which involves model (kinetics) parameters. A set of experimental data (experimental flame speeds) is assembled to update the model parameters. Based upon the difference between the experimental and modelled results (flame speeds), the parameter nominal values and covariances are updated, and the posterior quantity (flame speed) uncertainty is determined.