Statistical inference of uncertainties in elementary reaction rates of chemical mechanisms

By N. Kseib, J. Urzay AND G. Iaccarino

1. Motivation and objectives

Uncertainties in elementary reaction rates of chemical mechanisms have traditionally been reported by experimentalists in terms of constant-temperature prefactors to account for observed sparsity in measurements (Konnov 2008; Hong, Davidson & Hanson 2011*b*). These uncertainty prefactors have been used in uncertainty quantification and optimization of chemical mechanisms by lumping all the uncertainties in the exponential prefactors (Reagan *et al.* 2003; Davis *et al.* 2005). The development of methodologies to correctly represent the chemical reaction-rate measurement uncertainties is a necessary step towards quantitative predictions of complex combustion systems, as depicted in figure 1.

The method presented in this study provides a useful framework for quantifying uncertainties in both exponential prefactors and activation energies. That the uncertainty in the activation energy needs to be addressed is directly observed from experimental data, in which scattering in the slope of the logarithm of the reaction rate is noticeable. Additionally, it is found below that, for elementary steps with high activation-energy barriers, which are ubiquitous in combustion systems, the uncertainties in the activation energy and exponential prefactor are highly correlated.

The method is exemplified for deriving probabilistic measures of the uncertainties in the Arrhenius parameters in recently measured chemical rates of the improved hydrogen-oxygen mechanism by Hong *et al.* (2011*b*). Gaussian noise is added to each experimental data point of the corresponding elementary chemical-reaction rate. The variance of the noise is taken from uncertainty prefactors reported in experiments. The joint posterior density function of the chemical parameters is explored using a Markov-chain Monte-Carlo algorithm. In addition, an accurate representation of the randomness in the input parameters is obtained in terms of a parametrization based on independent random variables. This representation could be used in future investigations to calculate, using polynomial-chaos expansions, chemically induced uncertainties in, for instance, ignition-delay times and strains of extinction, or in the quantity of interest of more general reacting-flow computations.

2. Formulation

Consider an elementary step

$$\sum_{i=1}^{N} \nu'_{i,j} \mathcal{R}_i \rightleftharpoons \sum_{i=1}^{N} \nu''_{i,j} \mathcal{R}_i \tag{2.1}$$

in a chemical mechanism composed of $j = 1, 2...M$ reactions, with $\mathcal{R}$ the chemical symbol of species i, N the number of species, ν'_{ij} the stoichiometric coefficient of the reactant i

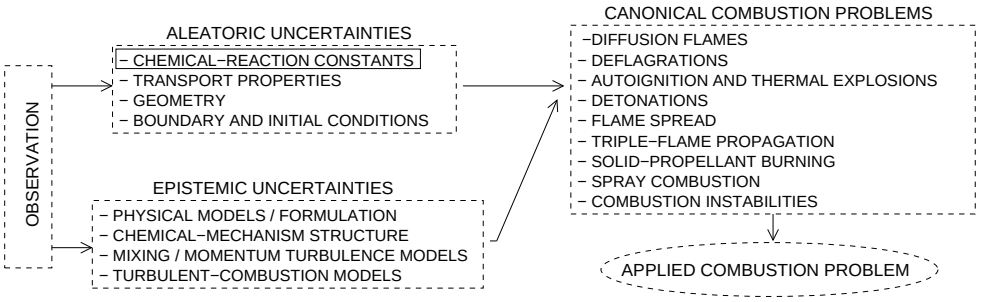

FIGURE 1. Road-map schematics of uncertainty quantification in combustion systems. The present study only tackles the problem of observation-based statistical inference of aleatoric uncertainties in chemical-reaction rates, as indicated by the solid-line rectangle.

in the step j on the reactants side, and ν''_{ij} the stoichiometric coefficient of the reactant i in the step j on the products side. The net rate of reaction of the elementary step (2.1) is given by

$$\omega_j = k_{f,j} \prod_{i=1}^{N} \left(\frac{\rho Y_i}{W_i} \right)^{\nu'_{ij}} - k_{b,j} \prod_{i=1}^{N} \left(\frac{\rho Y_i}{W_i} \right)^{\nu''_{ij}}, \tag{2.2}$$

which represents the number of moles produced or consumed per unit volume per unit time. In this formulation, Y_i and W_i are, respectively, the mass fraction and molecular weight of species i. Additionally, the constants $k_{f,j}$ and $k_{b,j}$ are, respectively, the forward and backwards specific reaction-rate constants of the reaction j, which are related by the equilibrium constant $\mathcal{K}_j$ as $\mathcal{K}_j = k_{f,j}/k_{b,j}$. In the Arrhenius approximation, the reaction-rate constants follow the expression

$$k_{f,j} = A_j T^{n_j} \mathrm{e}^{-E_{a,j}/\mathrm{R}^0 T}, \tag{2.3}$$

where A_j is the exponential prefactor, n_j is the temperature exponent, T is the temperature, R^0 is the universal gas constant, and $E_{a,j}$ is the activation energy measured in cal mol^{-1} units. In this model, $A_j T^{n_j}$ is proportional to a collision frequency that depends on the effective collision cross-section, and $\exp(-E_{a,j}/\mathrm{R}^0 T)$ is proportional to the fraction of gas molecules moving with a kinetic energy larger than the activation energy.

Although the analysis given below can be extended to reaction rates that show a stronger temperature dependence than equation (2.3), for illustration in this section attention is confined to equation (2.3) for a particular reaction j. In logarithmic form, (2.3) becomes

$$f_m(\boldsymbol{S}) = \ln\left[(k_f(\boldsymbol{S}, T_m)\right] = \ln A + n \ln T_m - \frac{E_a}{\mathrm{R}^0 T_m} \tag{2.4}$$

for $m = 1, ..., K$ measured data points. In the formulation $\boldsymbol{S} = [\ln A, n, E_a]^{\mathrm{T}}$ is the vector of input parameters, and X_m are the components of the experimental-data vector $\boldsymbol{X}(T_m)$ that is composed of the experimental values measured for the reaction rate of an elementary step at each measured temperature T_m.

The noise $\boldsymbol{\epsilon}$ in the experimental data is modeled as a vector of size K, in which the components are independent Gaussian distributions of zero mean and equal standard deviation $\sigma = \frac{1}{2}\ln(\mathrm{UF})$, with UF the uncertainty factor traditionally given in the experimental literature for the entire range of measured temperatures (Konnov 2008; Hong

et al. 2011*b*). More involved noise models could be utilized here in which, for instance, temperature-dependent variances $\sigma = \sigma(T_m)$ are used. Note that an exact quantification of the noise $\boldsymbol{\epsilon}$ would involve a large number of measurements for every temperature, which are usually not available.

The objective of the analysis is to find the input-parameter vector $\boldsymbol{S}$ that brings the model into agreement with the measurements. In mathematical terms, this can be expressed as $f_m(\boldsymbol{S}) = X_m + \epsilon_m$, with f_m given by (2.4). This implies that $f_m(\boldsymbol{S}) - X_m$ has the same distribution as ϵ_m, namely a normal distribution with zero mean and variance σ^2.

It is implicitly assumed here that all the variability in $f_m(\boldsymbol{S}) - X_m$ can be attributed to the variability in the input-parameter vector $\boldsymbol{S}$, in that the temperature T_m is assumed to be deterministic since the noise in the characterization of the temperature value in experimental measurements is typically negligible. A probability-density function in the input-parameter space is sought such that the random variable $f_m(\boldsymbol{S}) - X_m$ has the same distribution as the Gaussian noise.

The method of inference for the input parameters proposed here is the Bayes' rule,

$$P(\boldsymbol{S}|\boldsymbol{X}) = cP(\boldsymbol{X}|\boldsymbol{S})P(\boldsymbol{S}). \tag{2.5}$$

In this formulation, $P(\boldsymbol{S}|\boldsymbol{X})$ is the posterior density, or equivalently, the occurrence probability of the parameter-vector $\boldsymbol{S}$ conditioned on the data set $\boldsymbol{X}$. Similarly, $P(\boldsymbol{X}|\boldsymbol{S})$ is the likelihood function, or equivalently, the occurrence probability of the measurements $\boldsymbol{X}$ conditioned on the input-parameter vector $\boldsymbol{S}$. Additionally, $P(\boldsymbol{S})$ is the prior or occurrence probability of $\boldsymbol{S}$, which is assumed to be a uniform density in the input-parameter space. Finally, c is a normalizing constant that ensures that the product of the likelihood and the prior is a probability density function.

Since both $P(\boldsymbol{S})$ and c are constants over the parameter space, we can rewrite Bayes' rule as $P(\boldsymbol{S}|X) \propto P(X|\boldsymbol{S})$, thereby setting the foundations of the Markov-chain Monte-Carlo (MCMC) method, in which the posterior density is explored by using the likelihood function

$$P(\boldsymbol{X}|\boldsymbol{S}) = (2\pi\sigma^2)^{-\frac{K}{2}} \exp\left(-\frac{||\boldsymbol{f(S)} - \boldsymbol{X}||^2}{2\sigma^2}\right). \tag{2.6}$$

Note that, since the noise vector $\boldsymbol{\epsilon}$ is Gaussian-distributed with zero mean and constant σ, the distribution (2.6) is a multivariate Gaussian function with mean vector $\boldsymbol{f(S)}$ and scalar standard deviation σ. When $\ln A$, E_a and n are considered as random parameters, the dimensionality of the distribution function (2.6) is 3.

In the MCMC method, sampling in the vector-parameter space $\boldsymbol{S}$ is performed by using Gaussian-distributed increments $\boldsymbol{\delta_n} = \boldsymbol{S_{n+1}} - \boldsymbol{S_n}$ in such a way that high-probability regions in the likelihood function are identified. Since the constant c is unknown in (2.5), it is expedient to relate the ratio of the likelihood function to the posterior-density function $P(\boldsymbol{X}|\boldsymbol{S_{n+1}})/P(\boldsymbol{X}|\boldsymbol{S_n}) = P(\boldsymbol{S_{n+1}}|\boldsymbol{X})/P(\boldsymbol{S_n}|\boldsymbol{X})$ by using Bayes' rule. In this way, the exploration of the likelihood function is equivalent to the exploration of the posterior density.

After N_c chain steps, with N_c sufficiently large for obtaining meaningful statistics, the states $\{\boldsymbol{S_1}, \boldsymbol{S_2}, ..., \boldsymbol{S_{N_c}}\}$ constitute samples from the posterior. In practice, the first $N_1 < N_c$ samples from the set are removed, since these correspond to the so-called burn-in period in which the chain can undertake multiple different paths to arrive at high-likelihood zones. However, the MCMC method converges in the region of maximum likelihood, thereby maximizing the likelihood function in equation (2.6), which is equiva-

lent to minimizing the norm $||\boldsymbol{f}(\boldsymbol{S})-\boldsymbol{X}||$; therefore, the burn-in period can be eliminated by starting the chain with the vector-parameter $\boldsymbol{S_1}$ that minimizes this norm, which here is obtained by a least-squares procedure.

From the MCMC method, the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\mathrm{cov}(\boldsymbol{S})$ can be calculated, with $\boldsymbol{\mu}$ and $\mathrm{cov}(\boldsymbol{S})$ defined as

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{\ln A} \\ \mu_{E_a} \\ \mu_n \end{bmatrix}, \quad \text{and} \quad \mathrm{cov}(\boldsymbol{S}) = \begin{bmatrix} \sigma^2_{\ln A} & \mathrm{cov}(\ln A, E_a) & \mathrm{cov}(\ln A, n) \\ \mathrm{cov}(\ln A, E_a) & \sigma^2_{E_a} & \mathrm{cov}(E_a, n) \\ \mathrm{cov}(\ln A, n) & \mathrm{cov}(E_a, n) & \sigma^2_n \end{bmatrix}. \tag{2.7}$$

Using the MCMC statistics in (2.7), an expansion of the parameter vector $\boldsymbol{S}$ is sought here in terms of a vector-base of functions composed of independent and standard Gaussian random variables $\boldsymbol{\xi} = [\xi_1, \xi_2, \xi_3]^{\mathrm{T}}$, namely

$$\boldsymbol{S} = \boldsymbol{\mu} + \boldsymbol{\alpha}\boldsymbol{\xi}, \tag{2.8}$$

where α is a matrix chosen to be of the form

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \alpha_{1,3} \\ 0 & \alpha_{2,2} & \alpha_{2,3} \\ 0 & 0 & \alpha_{3,3} \end{bmatrix}. \tag{2.9}$$

Applying now the covariance operator to equation (2.8) gives

$$\mathrm{cov}(\boldsymbol{S}) = \boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathbf{T}}. \tag{2.10}$$

Since $\mathrm{cov}(\boldsymbol{S})$ is a known symmetric matrix, equation (2.10) yields six equations with six unknowns, namely the six components of $\boldsymbol{\alpha}$ in (2.8). The solution to equation (2.10) is

$$\begin{gathered} \alpha_{1,1} = \sigma_{\ln A}\sqrt{1-\rho^2_{\ln A,n} - \frac{(\rho_{\ln A,E_a} - \rho_{\ln A,n}\rho_{E_a,n})^2}{1-\rho^2_{E_a,n}}}, \\ \alpha_{1,2} = \frac{\sigma_{\ln A}(\rho_{\ln A,E_a} - \rho_{\ln A,n}\rho_{E_a,n})}{\sqrt{1-\rho^2_{E_a,n}}}, \quad \alpha_{1,3} = \sigma_{\ln A}\rho_{\ln A,n}, \\ \alpha_{2,2} = \sigma_{E_a}\sqrt{1-\rho^2_{E_a,n}}, \quad \alpha_{2,3} = \sigma_{E_a}\rho_{E_a,n}, \quad \alpha_{3,3} = \sigma_n, \end{gathered} \tag{2.11}$$

which represent the coefficients of the expansion (2.8). In the formulae (2.11), the symbols $\rho_{i,j} = \mathrm{cov}(i,j)/(\sigma_i\sigma_j)$ denote the correlation coefficients. The uncertainty representation (2.8) and equations (2.11) can be used to propagate uncertainties in actual combustion calculations using established methodologies, such as polynomial chaos techniques (Reagan *et al.* 2003), which require independent random inputs.

This method is demonstrated below for elementary steps of practical importance such as the hydrogen-oxygen chain branching reaction and the hydrogen-peroxide thermal decomposition.

3. Examples

3.1. *Chain branching reaction* $\mathrm{H} + \mathrm{O_2} \leftrightharpoons \mathrm{OH} + \mathrm{O}$

The chemical reaction $\mathrm{H} + \mathrm{O_2} \leftrightharpoons \mathrm{OH} + \mathrm{O}$ is the most important elementary step for hydrogen-oxygen systems and for most of hydrocarbon-air flames. In this reaction, hydroxyl and oxygen radicals are formed from collisions between hydrogen atoms and oxygen molecules. For the collisions to result in chemical conversion, the kinetic energy of

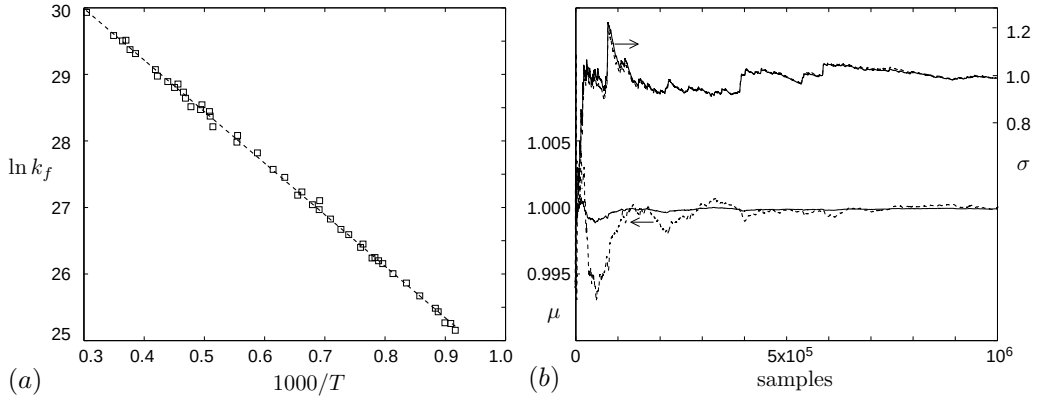

FIGURE 2. (a) Least-squares fit approximation (dashed line) of experimental data of Hong *et al.* (2011*b*) (squares). (b) Convergence of the normalized first and second statistical moments for $\ln A$ (solid lines) and E_a (dashed lines)

the colliding reactants needs to be larger than the activation-energy barrier of the reaction. When a sufficiently large pool of radicals exists, radical recombination occurs in exothermic chain-termination reactions that release most of the chemical energy in hydrogen-oxygen explosions.

For this reaction, the values $\mu_n = 0$ and $\sigma_n = 0$ in (2.11) are assumed for simplicity, in accordance with earlier experimental work (Hong *et al.* 2011*b*) in which negligible temperature dependence of the collision frequency was observed. The treatment of n as a random parameter in these examples is deferred to future studies. When A and E_a are expressed in $\mathrm{cm^3 mol^{-1} s^{-1}}$ and cal $\mathrm{mol^{-1}}$, respectively, the values $\ln A = 32.31$ and $E_a = 16386$ are obtained from the least-squares approximation of the most recent experimental data (Hong *et al.* 2011*b*), as shown in figure 2a. These values are used for initializing the Markov chain in the vicinity of the point of maximum likelihood. For this reaction, the standard deviation of the Gaussian noise is $\sigma = \ln(\mathrm{UF})/2 = 0.047$ with an uncertainty factor UF = 1.1 (Hong *et al.* 2011*b*). The values $\mu_{\ln A} = 32.31$, $\mu_{E_a} = 15385$, $\sigma_{\ln A} = 0.10$, $\sigma_{E_a} = 306.1$ and $\rho_{\ln A, E_a} = 0.93$ are obtained from the MCMC statistics after 10^6 samples. The convergence of the first and second statistical moments, normalized with their respective values at 10^6 samples, can be observed in figure 2b. With these statistics, the coefficients in (2.11) become $\alpha_{1,1} = 0.037$, $\alpha_{1,2} = 0.093$ and $\alpha_{2,2} = 306.1$.

The resulting probability distributions of $\ln A$ and E_a are shown in figure 3(b,c), in which the solution obtained by using MCMC is compared with the expansion (2.8). A Gaussian behavior is observed for both parameters $\ln A$ and E_a. Note that this is in contrast with earlier work by Najm *et al.* (2009), where it was proposed that both $\ln E_a$ and $\ln A$ should follow Gaussian distributions; this assertion is found to be inconsistent with the present results, in that here a Gaussian noise $\boldsymbol{\epsilon}$ is assumed in the $\ln k_f = \ln k_f(\ln A, E_a)$ measurement plane.

The joint probability distributions obtained by using MCMC and the expansion (2.8) are shown in figure 3. The joint probability distribution is described to a very large extent by a bivariate joint normal distribution. Additionally, for a fixed temperature range and uncertainty factor, the decorrelation between $\ln A$ and E_a is found to be inversely proportional to the mean activation energy, in that the range of possible values of the activation energy for which the Arrhenius curve (2.1) can fit within the constant-

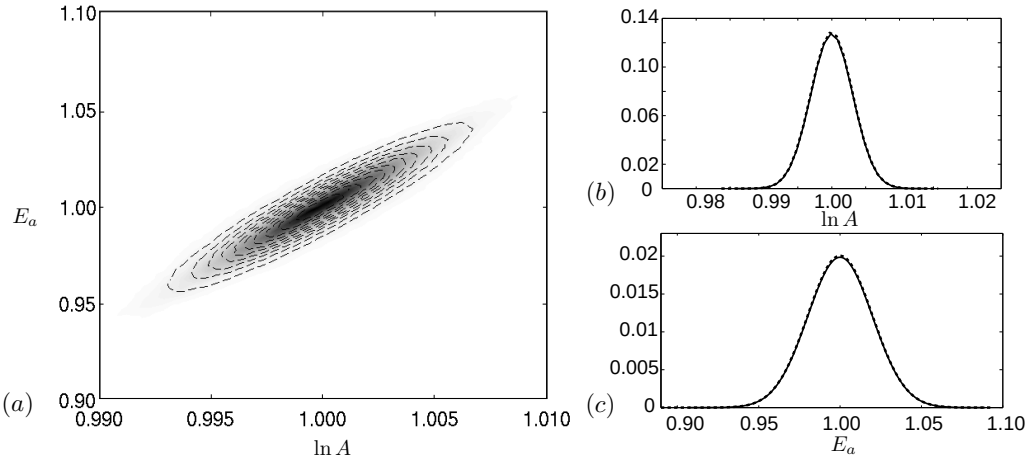

FIGURE 3. (a) Joint probability distribution between $\ln A$ and E_a. (b) Probability-density function for $\ln A$. (c) Probability-density function for E_a. In this figure, dashed lines are obtained from expansion (2.8), and solid lines and gray-scale contours denote results obtained by MCMC simulations. Additionally, the values of $\ln A$ and E_a are normalized with their corresponding mean values.

variance Gaussian-noise model is inversely proportional to the activation energy itself. It is therefore implied that $\ln A$ and E_a become correlated because of the sensitivity of the chemical reaction to temperature. These considerations are sketched in figure 4. Note that the range of possible values of the exponential prefactor A is very much reduced when the uncertainty in the activation energy is accounted for in chemical reactions with high activation energies.

In this example, because of the high values of the correlation coefficient $\rho_{\ln(A),E_a} = 0.93$ it can be shown that the expansion (2.8) simplifies to

$$\ln A = \mu_{\ln A} + \sigma_{\ln(A)}\xi_2 \quad \text{and} \quad E_a = \mu_{E_a} + \sigma_{E_a}\xi_2, \tag{3.1}$$

which is representative of perfectly correlated distributions for $\ln A$ and E_a. When $\ln A$ and E_a are normalized by their corresponding mean values and the numeric values of the standard deviations are substituted, the expansions in (3.1) become $\ln A/\mu_{\ln A} = 1 + 0.003\xi_2$ and $E_a/\mu_{E_a} = 1 + 0.019\xi_2$.

3.1.1. *Hydrogen-peroxide thermal decomposition* $H_2O_2 + M \leftrightharpoons OH + OH + M$

Conclusions similar to those obtained above could be extracted from other types of reactions for which the sensitivity to temperature is high. For instance, the thermal decomposition of hydrogen peroxide, $H_2O_2 + M \leftrightharpoons OH + OH + M$, which is responsible for engine knock and ignition at low temperatures, follows an expansion of the type (2.8), with $\mu_{\ln A} = 36.81$, $\mu_{E_a} = 42265$, $\sigma_{\ln A} = 0.54$, $\sigma_{E_a} = 1284.5$ and $\rho_{\ln A,E_a} = 0.99$, which gives $\alpha_{1,1} = 0.08$, $\alpha_{1,2} = 0.53$ and $\alpha_{2,2} = 1284.5$. Since $\rho_{\ln A,E_a} \to 1$, the model for this reaction may also be approximated by using (3.1).

4. Conclusions

A simple probabilistic representation of chemical-reaction parameters was achieved by using experimental data and Markov-chain Monte-Carlo simulations. A straightforward

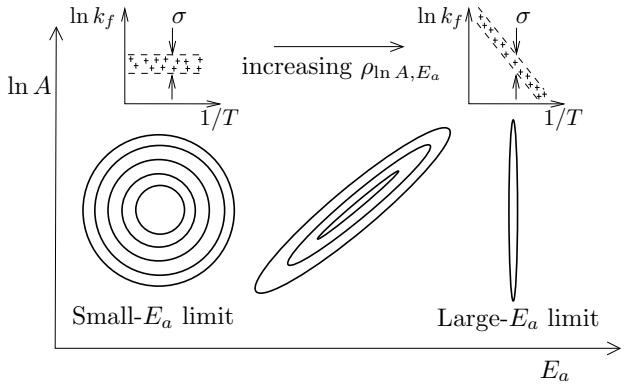

FIGURE 4. Sketch of the joint probability distribution between $\ln A$ and E_a in the limits of small and large activation energies for equal noise and equal temperature range.

expansion in terms of independent random variables was found to model accurately the simulation results. This method can be employed in uncertainty propagation analyses based on polynomial chaos expansions in terms of Hermite polynomials, because it relies on the use of independent Gaussian random variables.

The activation energy is found to be correlated with the frequency prefactor. That the uncertainty in the activation energy needs to be accounted for is observed from experiments. However, previous studies have injected all the uncertainties in the collision frequency. The inclusion of uncertainties in the activation energy reduces dramatically the range of likely values of the frequency collision. Furthermore, it is expected that, close to critical points such as flame extinction or autoignition, uncertainties in the activation energy may produce drastic variations in the solutions obtained. This is of particular importance in combustion systems, in which reaction rates are extraordinarily sensitive to the local temperature.

The technique outlined here is just a simple model of the underlying uncertainties with which reaction rates are determined experimentally. For instance, very seldom can a reaction rate be measured in experiments without making use of another chemical reactions for its determination (Hong *et al.* 2011*a*). This suggests that all uncertainties in chemical mechanisms are interwoven in a very experiment-specific manner, and that some of the regions predicted by independent random-variable models of uncertainties in multiple-reaction mechanisms are likely not physically accessible. Further research is needed to clarify these points.

Acknowledgments

This investigation has been undertaken under the framework of the Predictive-Science Academic-Alliance Program (PSAAP) for uncertainty quantification in hydrogen-fueled supersonic engines. Partial funding from the KAUST AEA is also gratefully acknowledged. The authors acknowledge encouraging discussions with Professor David Davidson.

REFERENCES

DAVIS, S., JOSHI, A., WANG, H. & EGOLFOPOULOS, F. 2005 An optimized kinetic model of H_2/CO combustion. *Combustion and Flame* **30**, 1283–1292.

Hong, Z., Davidson, D., Barbour, E. & Hanson, R. 2011*a* A new shock tube study of the $H + O_2 \leftrightharpoons OH + O$ reaction rate using tunable diode laser absorption of H_2O near 2.5 μm. *Proc. Comb. Inst.* **33**, 309–316.

Hong, Z., Davidson, D. & Hanson, R. 2011*b* An improved H_2/O_2 mechanism based on recent shock tube/laser absorption measurements. *Combustion and Flame* **158**, 633–644.

Konnov, A. 2008 Remaining uncertainties in the kinetic mechanism of hydrogen combustion. *Combustion and Flame* **152**, 507–528.

Najm, H., Debusschere, B., Marzouk, Y., Widmer, S. & Le Maitre, O. 2009 Uncertainty quantification in chemical systems. *Int. J. for Numer. Meth. Eng.* **80**, 789–814.

Reagan, M., Najm, H., Ghanem, R. & Knio, O. 2003 Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame* **133**, 545–555.