Preprint Cambridge Centre for Computational Chemical Engineering ISSN 1473 – 4273

Bayesian Error Propagation for a Kinetic Model of n-Propylbenzene Oxidation in a Shock Tube

Sebastian Mosbach [1], Je Hyeong Hong [1], George P. E. Brownbridge [1], Markus Kraft [1], Soumya Gudiyella [2] and Kenneth Brezinsky [3]

released: 03 October 2013

[1] Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge CB2 3RA
UK
Email: mk306@cam.ac.uk

[2] Department of Chemical Engineering
University of Illinois at Chicago
Chicago, IL 60607
USA
Email: sgudiy2@uic.edu

[3] Department of Mechanical and Industrial Engineering
University of Illinois at Chicago
Chicago, IL 60607
USA
Email: kenbrez@uic.edu

Preprint No. 133

Keywords: n-Propylbenzene, Bayesian parameter estimation, error propagation

Edited by
Computational Modelling Group
Department of Chemical Engineering and Biotechnology
University of Cambridge
New Museums Site
Pembroke Street
Cambridge CB2 3RA
United Kingdom

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

Abstract

We apply a Bayesian parameter estimation technique to a chemical kinetic mechanism for n-propylbenzene oxidation in a shock tube in order to propagate errors in experimental data to errors in model parameters and responses. We find that, in order to apply the methodology successfully, conventional optimisation is required as a preliminary step. This is carried out in two stages: firstly, a quasi-random global search using a Sobol low-discrepancy sequence is conducted, followed by a local optimisation by means of a hybrid gradient-descent/Newton iteration method. The concentrations of 37 species at a variety of temperatures, pressures, and equivalence ratios are optimised against a total of 2378 experimental observations. We then apply the Bayesian methodology to study the influence of uncertainties in the experimental measurements on some of the Arrhenius parameters in the model as well as some of the predicted species concentrations. Markov Chain Monte Carlo algorithms are employed to sample from the posterior probability densities, making use of polynomial surrogates of higher order fitted to the model responses. We conclude that the methodology provides a useful tool for the analysis of distributions of model parameters and responses, in particular their uncertainties and correlations. Limitations of the method are discussed. For example, we find that using second-order response surfaces and assuming normal distributions for propagated errors is largely adequate, but not always.

Contents

1 Introduction

The quantification of uncertainties in experimental as well as computational data has long been recognised as essential across all areas of science and technology. The field of combustion modelling is no exception [20, 58]. It is well-established [49, 72] that making chemical models reliable and predictive requires calculation results to be accompanied by uncertainty bounds – with greater predictive power corresponding to smaller error bounds.

Various techniques for uncertainty quantification and propagation have been used in combustion kinetic modelling. An elementary way to relate uncertainties in model parameters to uncertainties in model outputs is via local sensitivity coefficients [62, 65]. The same idea can be extended to global sensitivities in order to be able to treat strong model nonlinearities combined with large uncertainties [57, 73]. For example, global sensitivity methods have been used to propagate uncertainties from parameters in first-principles calculations to reaction rate coefficients [23]. Uncertainties of the three parameters in the Arrhenius rate law, with particular emphasis on their joint distribution and its temperature dependence, have been investigated in detail [40, 41, 66, 67]. Given that error propagation is intrinsically part of parameter estimation, frequently in the sense of optimising with respect to an objective function in one form or another, it is also natural to consider confidence regions defined by the contours of the objective function surface [21, 37].

In the Data Collaboration framework [20, 49, 51, 72], uncertainties are specified through deterministic upper and lower bounds rather than probability distributions. It employs optimisation techniques in conjunction with solution mapping [21] in order to quantify prediction uncertainties [22], rigorously measure data set consistency [16], quantitatively discriminate between multiple candidate models [17], and to conduct sensitivity analyses of uncertainties in responses with respect to experimental errors and uncertainties in model parameters [48].

Spectral uncertainty quantification is a technique based on polynomial chaos expansions [69, 70], in which uncertain quantities, such as model parameters, are represented as series of basis random variables. The essence of the method consists in determining the coefficients of the expansion, the spectral modes, which can be used to reconstruct the probability density, at least to the finite order of the truncated series. The spectral method has been introduced to combustion research, specifically in the areas of reacting flows and chemical kinetics, in the form of a post-processing step to conventional Monte Carlo analysis [45–47] in which the spectral modes are determined. The technique has been adapted, using quadratic response surfaces, to derive an analytic expression for the variance of model responses and combined with optimisation such that experimental errors can be propagated simultaneously into rate coefficients and their uncertainties [52–54].

Bayesian methods are based on a systematic probability-theoretic treatment of all involved quantities including in particular experimental data as well as model parameters, and centre around applying Bayes' theorem to update knowledge represented in the form of distributions [4]. Bayesian methods have been used in chemical kinetics for at least half a century [2], but their popularity in this and other fields [1] has increased significantly in recent years due to an influential paper by Kennedy and O'Hagan [30]. In particular, the use of Markov Chain Monte Carlo (MCMC) sampling has become wide-spread, at least

to some extent as a consequence of readily available powerful computers. For example, Bayesian uncertainty quantification using MCMC has been applied to rate coefficients of single reactions, such as $H+O_2 \rightarrow OH+O$ using shock tube data [36], and hydrogen abstraction from isopropanol using *ab initio* calculations [43]. The same methods, as well as polynomial chaos expansions, have also been employed for quantifying uncertainties in entire reaction mechanisms, focusing specifically on correlations between the three parameters in the Arrhenius law [42]. Bayesian parameter estimation also naturally extends to and is closely related to experimental design [27, 28, 39]. Although Bayesian techniques are not new as such, and some applications exist, there continues to be a need for case studies which demonstrate the performance of the methods.

The purpose of this paper is to apply a Bayesian method for parameter estimation and error propagation as a case study to a detailed chemical kinetic model for n-propylbenzene oxidation in a shock tube and report on the experience gained with the method. n-propylbenzene has been suggested as a potential surrogate for the alkylbenzene class of components of commercial aviation fuels [10, 14]. The kinetic model contains 191 species and 1127 reactions. Prior to using Bayesian methods, for reasons explained below, we perform optimisation with conventional techniques. We conduct a quasi-random global search in a parameter space spanned by 64 Arrhenius pre-exponential factors, selected by means of sensitivity analysis, and then use a set of best points found to initiate local gradient-descent optimisation from those points. A standard least-squares objective function weighted by experimental errors is utilised to assess agreement with experimental data, which comprise of 2378 individual observations of 37 species concentrations at 74 different experimental process conditions with varying pressure, temperature, and equivalence ratio. For the best point resulting from the optimisation, we apply Bayesian parameter estimation for a small number of parameters, and use an MCMC method to obtain posterior probability densities. For this purpose, we fit a surrogate model that consists of polynomials of various orders to the responses. We discuss strengths and weaknesses of the method.

2 Methodology

In this section, we briefly describe the experimental data set and combustion model used in this work, and give details of each of the steps involved in performing the error propagation, namely, sensitivity analysis, global search, local optimisation, and Bayesian parameter estimation.

2.1 Experimental data

The data set we consider here [24] was obtained using the high-pressure single-pulse shock tube at the University of Illinois at Chicago [56, 61]. Process condition variables comprise of initial temperature, initial pressure, initial composition, and reaction time. In all cases, the initial mixture is composed of n-propylbenzene, oxygen, and argon. The mixture is highly diluted, with an n-propylbenzene mole fraction of about 90 ppm

throughout, so the system can be considered isothermal to a good approximation. An overview of the conditions is given in Table 1.

Table 1: *Overview of experimental process conditions.*

Average shock pressure [atm]	Temperature range [K]	Fuel mole fraction [ppm]	Φ	Reaction time range [ms]
28	907-1551	86	0.54	1.40-2.05
51	959-1558	90	0.55	1.27-1.90
49	838-1635	90	1.0	1.21-1.95
24	905-1669	89	1.9	1.36-2.93
52	847-1640	90	1.9	1.26-1.95

Species concentrations were determined by means of gas chromatography in gases sampled from the shock tube. Measured species include O_2, CO, CO_2, aliphatic hydrocarbons such as methane, ethane, propane, acetylene, ethene, propene, propadiene, propyne, 1,3-butadiene, 1,2-butadiene, 2-butyne, vinylacetylene, diacetylene, and cyclopentadiene, aromatic hydrocarbons such as benzene, toluene, phenylacetylene, styrene, ethylbenzene, 1-propenylbenzene, 2-propenylbenzene, n-propylbenzene, indene, naphthalene, 2-ethynylnaphthalene, bibenzyl, diphenylmethane, stilbene, fluorene, and anthracene, and the oxygenated aromatics phenol, benzylalcohol, benzaldehyde, and benzofuran. Measurement errors are given for each species and range between 1.7% and 25%, with an average of about 12%.

In total, there are 2378 experimental observations of 37 species across 74 points in process-condition space. The complete data set is available online as supplementary material to a previous publication [24].

2.2 Shock tube and chemical kinetic model

The shock tube is modelled as a homogeneous, adiabatic, constant pressure reactor. The corresponding governing equations are standard and shall not be repeated here. As software to solve the equations, *k*inetics v8.0 [9] was employed.

As chemical kinetic model, a mechanism developed previously [24], extending earlier work [11, 32], was used. It contains 191 species and 1127 reactions. All species mentioned in subsection 2.1 are present in the mechanism.

2.3 Sensitivity analysis

As it is neither feasible nor necessary to tune all kinetic parameters present in a reaction mechanism simultaneously, a subset needs to be chosen. It is natural to use sensitivity analysis for this [37, 44, 59, 62, 64, 65]. The basic idea is to calculate the normalised

sensitivity coefficients of chosen responses with respect to all Arrhenius pre-exponential factors and then rank the reactions according to the modulus of their coefficients.

The normalised sensitivity coefficient of the i^{th} response η_i (evaluated at a particular point ξ in process condition space) with respect to the j^{th} model parameter θ_j is defined by

$$\frac{\theta_j}{\eta_i(\xi,\theta)}\frac{\partial\eta_i(\xi,\theta)}{\partial\theta_j}.$$

One way to determine these coefficients is through a finite difference approximation [44, 62, 65]. If we denote the vector of model parameters perturbed in the j^{th} direction by

$$\tilde{\theta}^j := (\theta_1,\ldots,\theta_{j-1},(1+r)\times\theta_j,\theta_{j+1},\ldots,\theta_n),$$

where r is a (small, positive) number representing a relative perturbation, then this can be written as

$$\frac{\theta_j}{\eta_i(\xi,\theta)}\frac{\eta_i(\xi,\tilde{\theta}^j)-\eta_i(\xi,\theta)}{(\tilde{\theta}^j-\theta)_j}=\frac{\eta_i(\xi,\tilde{\theta}^j)-\eta_i(\xi,\theta)}{r\eta_i(\xi,\theta)}. \tag{1}$$

Now, the present data set has been obtained at numerous points $\xi^{(1)},\ldots,\xi^{(N_{\text{p.c.}})}$ in process condition space. We find that the sensitivity coefficients, of any particular response with respect to any particular parameter, vary considerably between these points. There are two contributions to this: the parametric derivative, and the value of the response. If the value of the response approaches zero, which is common in this data set, the sensitivity coefficient can become unduly large, in some circumstances amplifying numerical noise to an extent that it becomes dominant. For this reason, we normalise not by the local value $\eta_i(\xi^{(n)},\theta)$ of the response, but by its maximum value $\max_n\{\eta_i(\xi^{(n)},\theta)\}$ across all points in process condition space. It is then possible to compare the resulting coefficients between different process conditions. More on the subject of dependence of sensitivities on process conditions can be found for example in [75]. Thus, instead of (1) we consider

$$S_{ij}:=\frac{\max_n\left\{|\eta_i(\xi^{(n)},\tilde{\theta}^j)-\eta_i(\xi^{(n)},\theta)|\right\}}{r\max_n\left\{\eta_i(\xi^{(n)},\theta)\right\}}. \tag{2}$$

In order to obtain a single quantity which can be applied to ranking the reactions we use the largest value $\max_i\{S_{ij}\}$ among the responses.

We note that this analysis is *local* in parameter space but *global* in process condition space in the sense that sensitivities at every considered point in process condition space are taken into account.

2.4 Global search and local optimisation

In order to quantify agreement between experiment and model, we use the least-squares objective function

$$\Phi(\theta)=\sum_{n=1}^{74}\sum_{\text{responses }i}\left[\frac{\eta_i(\xi^{(n)},\theta)-\eta_i^{\text{exp}}(\xi^{(n)})}{\sigma_i^{(n)}}\right]^2, \tag{3}$$

where n indexes the points in process condition space. For the weights $\sigma_i^{(n)}$ of the individual terms, we choose the maximum value of each response within each of the five sets of points in process condition space, multiplied by twice the percentage experimental error. The maximum is chosen rather than individual values since, by reasoning similar to the previous subsection, most responses approach zero in every set, and these points would receive a dominant weight within each set, leading to counter-intuitive results. We apply this for each set separately rather than for the entire collection of points in process condition space as the magnitude of some of the responses can vary considerably from set to set, with equivalence ratio in particular.

Optimisation of the above objective function is carried out in two stages, namely a quasi-random global search as first stage, followed by a local optimisation as second stage. For the global search, a Sobol low-discrepancy sequence [55] is employed. For the local optimisation, an implementation of the Levenberg-Marquardt [31, 34] method, which is a hybrid between a gradient-descent method and a Newton iteration, is used.

Extensive work has been carried out on uncertainty bounds of Arrhenius parameters [40, 41, 67], feasible sets in such parameter spaces [16, 22, 49, 71], and optimisation of parameters beyond the basic Arrhenius ones, such as third-body efficiencies [13]. However, in order to explore the performance of the Bayesian methodology for parameter estimation and error propagation, which is the principal aim of this paper, it is sufficient to find any local minimum with respect to whichever parameters are considered. Therefore, for simplicity, we restrict ourselves in the present work to pre-exponential factors within a hypercube. We find that optimisation is necessary, though, as direct application of Bayesian methods to the original, unoptimised mechanism proved to be too problematic mainly regarding surrogate fidelity over sufficiently wide ranges in parameter space. We point out that the global search and local gradient-based optimisation in this work are conducted using the actual model, rather than a surrogate.

2.5 Bayesian parameter estimation

In this section, we briefly summarise the Bayesian methodology [1, 30] which we have applied previously in a different context [29, 39], and extend the formulation to heteroskedastic. This further builds upon our earlier work on parameter estimation and uncertainty propagation in the areas of granulation [5–7, 33] and combustion [50].

2.5.1 Bayes' Theorem

The current knowledge about the values of the model parameters θ can be represented by a probability density $p(\theta)$, called the *prior* distribution. When additional experimental data is obtained, represented as a probability density $p(\eta^{\text{exp}}|\theta)$, then the knowledge about the model parameters can be updated, resulting in a *posterior* distribution $p(\theta|\eta^{\text{exp}})$. Bayes' Theorem states how the posterior can be calculated:

$$p(\theta|\eta^{\text{exp}}) \propto p(\eta^{\text{exp}}|\theta)p(\theta),$$

or in words, 'Posterior $\propto$ Likelihood $\times$ Prior'. In order to estimate the values of the model parameters using the posterior distribution, likelihood and prior need to be specified.

2.5.2 Likelihood

For the likelihood, *i.e.* the distribution of the experimental responses, we assume a Gaussian centred at the model response:

$$\eta^{\text{exp},(n)} = \eta\big(\xi^{(n)}, \theta\big) + \varepsilon^{(n)} \qquad \text{with} \qquad \varepsilon^{(n)} \sim \mathcal{N}_L(0, \Sigma^{(n)}), \tag{4}$$

where $\varepsilon^{(n)}$ is the vector of the measurement errors which are normally distributed with zero mean and covariance matrix $\Sigma^{(n)}$, and L denotes the total number of responses. The covariance matrix $\Sigma^{(n)}$ of the experimental errors is allowed to vary from point to point in process condition space, *i.e.* the system can be heteroskedastic. We note here that (4) does not include a model inadequacy term, which accounts for any systematic discrepancy between experiment and model [8, 26, 30].

Equation (4) implies that, as $\eta^{\text{exp},(n)} \sim \mathcal{N}_L\big(\eta(\xi^{(n)}, \theta), \Sigma^{(n)}\big)$, the probability density of observing a particular response $\eta^{\text{exp},(n)}$ in the n^{th} experiment is given by

$$p\big(\eta^{\text{exp},(n)} \big| \theta, \Sigma^{(n)}\big) = (2\pi)^{-L/2}\big(\det \Sigma^{(n)}\big)^{-1/2} \exp\Big(-\frac{1}{2}{\varepsilon^{(n)}}^{\top}\big(\Sigma^{(n)}\big)^{-1}\varepsilon^{(n)}\Big),$$

and hence, assuming independent experiments, the likelihood becomes

$$\begin{aligned} &p\big(\eta^{\text{exp},(1)}, \ldots, \eta^{\text{exp},(N)} \big| \theta, \Sigma^{(1)}, \ldots, \Sigma^{(N)}\big) = \\ &\quad = (2\pi)^{-NL/2}\Big[\prod_{j=1}^{N}(\det \Sigma^{(j)})^{-1/2}\Big] \exp\Big(-\frac{1}{2}\sum_{n=1}^{N}{\varepsilon^{(n)}}^{\top}\big(\Sigma^{(n)}\big)^{-1}\varepsilon^{(n)}\Big). \end{aligned}$$

2.5.3 Prior distributions

For the prior of θ, we consider a constant, *i.e.* uniform, distribution over a hypercube $\mathcal{C}$ which is defined as the region in P-dimensional space such that $\theta_j \in [-1, 1]$ for all $j = 1, \ldots, P$. This gives as prior probability density $p(\theta) = |\mathcal{C}|^{-1}\mathbb{1}_{\{\theta\in\mathcal{C}\}}$, where $|\cdot|$ denotes the volume of a set and $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

2.5.4 Posterior distributions

The posterior density for θ can now be obtained from Bayes' Theorem as

$$p\big(\theta \big| \eta^{\text{exp},(1)}, \ldots, \eta^{\text{exp},(N)}\big) \propto \exp\Big(-\frac{1}{2}\sum_{n=1}^{N}{\varepsilon^{(n)}}^{\top}\big(\Sigma^{(n)}\big)^{-1}\varepsilon^{(n)}\Big) \cdot \mathbb{1}_{\{\theta\in\mathcal{C}\}}. \tag{5}$$

2.5.5 Markov Chain Monte Carlo sampling

In order to derive useful information about the unknown parameters from the posterior density (5), we employ the Markov Chain Monte Carlo sampling algorithms of Metropolis-Hastings [25, 35] and Wang-Landau [68]. The collection of samples can be used to plot (marginal) distributions and obtain quantities such as 'best' parameter estimates, *i.e.* the points of highest probability density, and high probability density regions, whose bounds can serve as error bars.

The large numbers of samples typically required by these types of algorithm render the direct use of the model infeasible due to computational expense. For this reason, surrogate models are widely used in this situation. Examples for the use of such models in the field of combustion include quadratic response surfaces [15, 18, 19, 21, 37, 39, 60], third order polynomials [12], higher order orthonormal polynomials [63], cubic natural splines [38], and High-Dimensional Model Representation (HDMR) [74]. We use polynomials of arbitrary order in this work.

3 Results

3.1 Sensitivity analysis

As described in subsection 2.3, we conduct a sensitivity analysis of all 37 responses with respect to the Arrhenius pre-exponential factors of all of the 1127 reactions, across all of the 74 points in process condition space, through finite differencing. This simulation involves $74 \times (1 + 1127) = 83472$ model evaluations. We find that using a relative perturbation of $r = 1\%$ (see Eqn. (2)) represents a good trade-off between avoiding both rounding error issues and the onset of non-linearities in the model responses. This agrees with recommendations in the literature [65]. The coefficients, with their sign restored, are shown in Fig. 1 for the 64 most sensitive reactions. We find that the obtained ranking differs appreciably, though not substantially, from those carried out at single points in process condition space for individual responses [24].

3.2 Global search and local optimisation

We arbitrarily choose to include all of the reactions listed in Fig. 1 into this step. For each of the reactions, we define the lower and upper bounds of the Arrhenius pre-exponential factor to be given by the nominal value divided and multiplied by 5 respectively. We evaluate 10^4 points of a Sobol sequence, involving 7.4×10^5 model evaluations, in this hypercube.

We then run Levenberg-Marquardt optimisations starting from some of the best points, *i.e.* those with lowest objective function value. Due to the random nature of the global search, and the extreme sparsity of the points in high dimensions, it is advisable to consider not just one but several of the best points, as the best point found in the global search does not guarantee the best result overall after further optimisation. All results shown in the

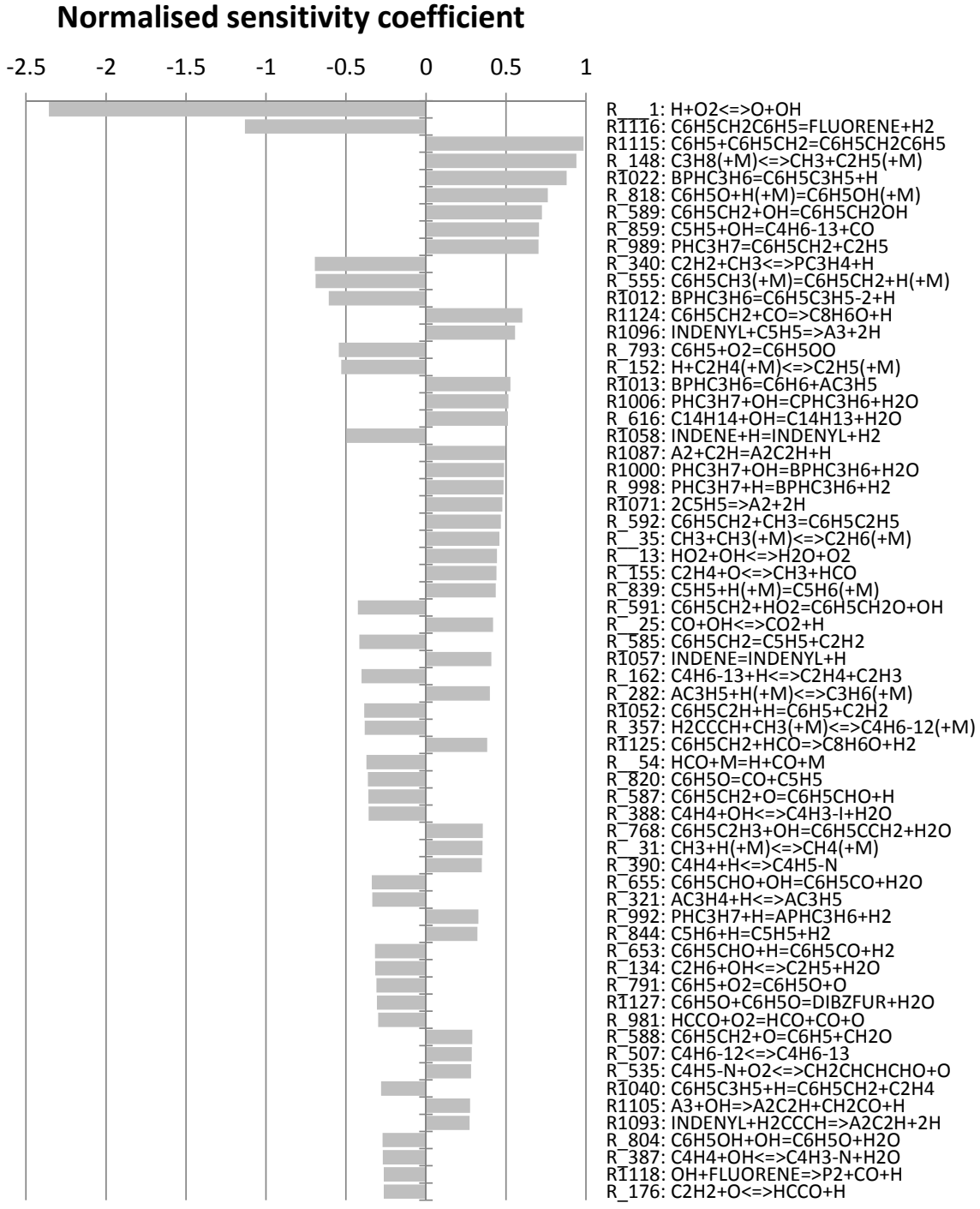
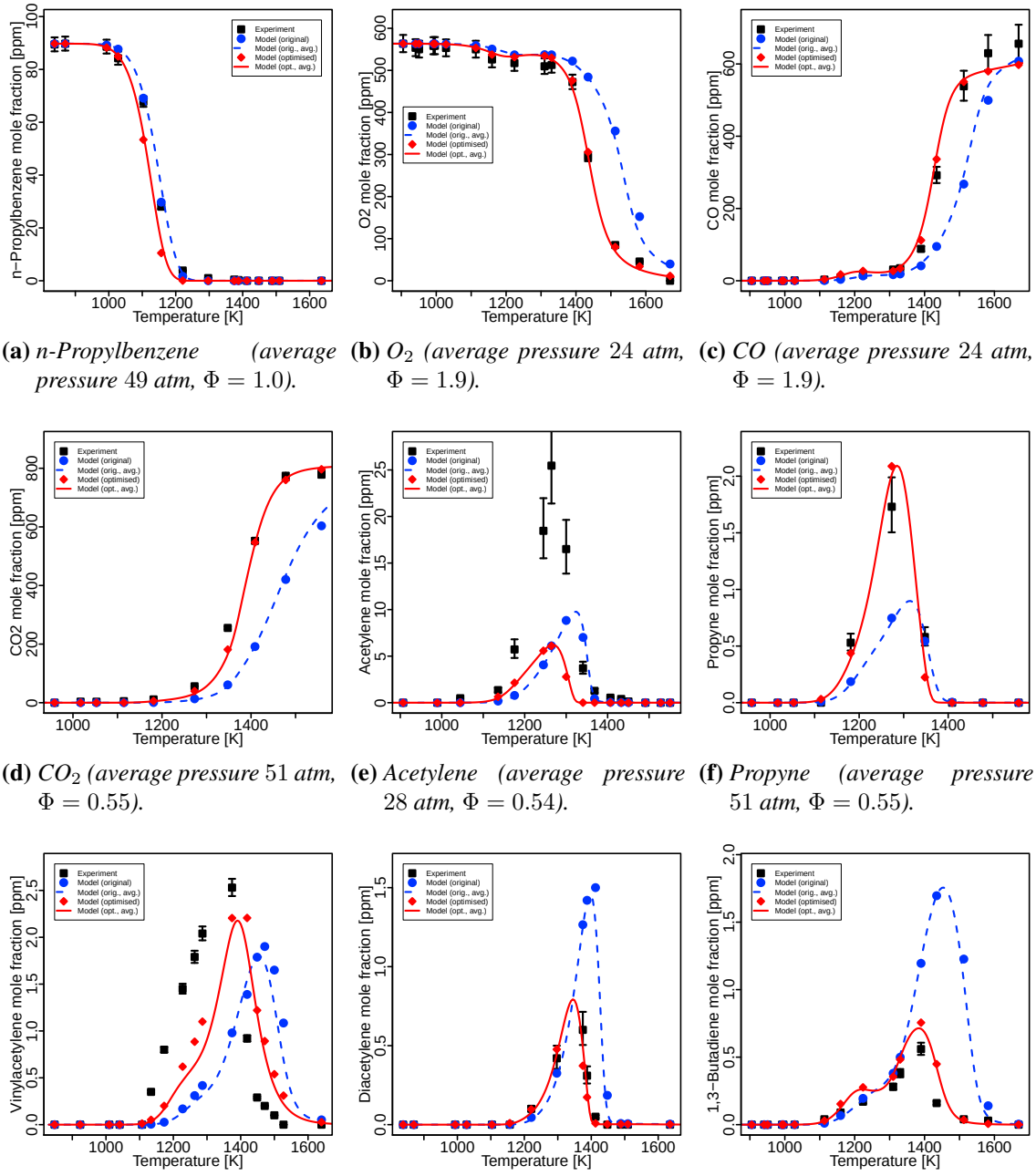


Figure 1: *Normalised sensitivity coefficients for the 64 most sensitive reactions across all considered points in process condition space.*

Table 2: *Square roots of the ratios of partial sums of objective function terms for the optimised and original mechanisms. The majority of contributions has improved (highlighted).*

Average pressure [atm]:	28	51	49	24	52
Equivalence ratio Φ:	0.54	0.55	1.0	1.9	1.9
n-Propylbenzene	1.378	1.419	4.808	1.015	0.969
O_2	1.052	0.651	0.654	0.128	0.211
CO	0.421	0.403	0.660	0.257	0.274
CO_2	0.405	0.138	0.224	0.236	0.302
Methane	2.130	2.885	0.889	0.645	0.868
Ethene	0.663	1.984	0.903	0.356	0.403
Ethane	1.875	2.375	1.809	1.214	0.728
Acetylene	1.057	1.048	1.380	1.156	1.125
Propadiene	0.777	3.582	1.898	0.437	0.510
Propyne	0.730	0.494	1.018	0.708	0.457
Vinylacetylene	0.735	0.764	0.813	0.776	0.580
Diacetylene	0.464	0.445	0.140	0.615	0.324
Benzene	0.915	0.855	1.077	0.897	0.737
Toluene	1.395	1.355	1.004	0.245	0.361
Phenylacetylene	0.076	0.170	0.349	0.139	0.168
Styrene	0.415	1.554	0.747	0.452	0.927
Cyclopentadiene	0.292	0.433	0.342	0.889	0.632
Ethylbenzene	0.481	0.252	0.333	0.360	0.244
Benzaldehyde	1.989	5.544	5.334	0.277	0.346
Phenol	0.151	0.775	0.078	0.133	0.115
1-Propenylbenzene	0.048	0.027	0.042	0.148	0.028
Indene	1.077	1.107	1.190	1.625	1.291
Naphthalene	0.630	0.598	0.806	1.083	0.995
1,3-Butadiene	0.902	1.236	0.688	0.191	0.281
Bibenzyl	0.927	1.134	0.560	0.722	0.759
Benzofuran	0.957	0.976	0.960	0.931	0.926
Diphenylmethane	0.570	0.833	0.581	0.637	0.627
Propane	0.737		0.690	0.598	
1,2-Butadiene				0.941	
2-Butyne				0.836	
2-Propenylbenzene				0.019	0.108
Benzylalcohol				0.797	0.784
2-Ethynylnaphthalene	0.676		0.669	0.929	0.922
Fluorene			0.989	0.870	1.152
Stilbene			0.428	0.554	0.505
Anthracene			0.277	0.381	0.250
Propene		4.480			3.185

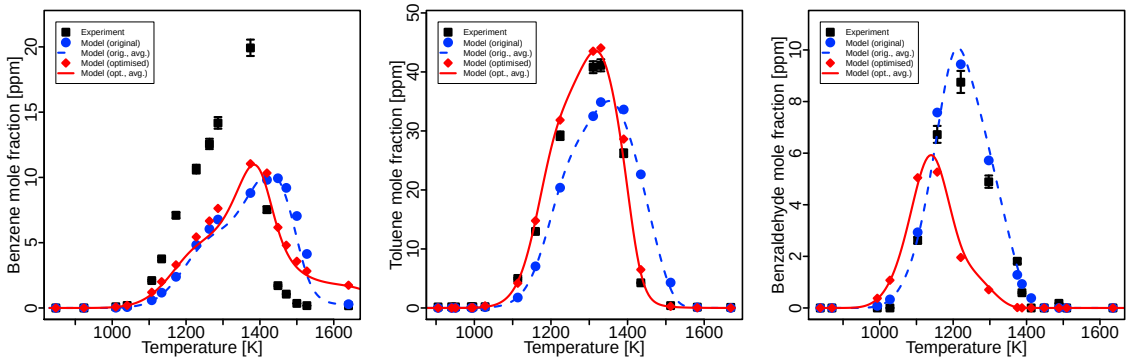


(a) *n-Propylbenzene (average pressure* 49 *atm,* $\Phi = 1.0$*).* **(b)** *O_2 (average pressure* 24 *atm,* $\Phi = 1.9$*).* **(c)** *CO (average pressure* 24 *atm,* $\Phi = 1.9$*).*

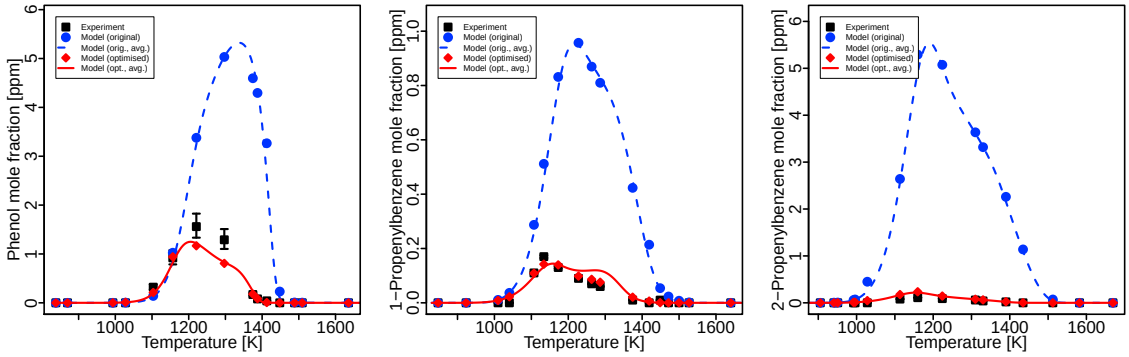
(d) *CO_2 (average pressure* 51 *atm,* $\Phi = 0.55$*).* **(e)** *Acetylene (average pressure* 28 *atm,* $\Phi = 0.54$*).* **(f)** *Propyne (average pressure* 51 *atm,* $\Phi = 0.55$*).*

(g) *Vinylacetylene (average pressure* 52 *atm,* $\Phi = 1.9$*).* **(h)** *Diacetylene (average pressure* 49 *atm,* $\Phi = 1.0$*).* **(i)** *1,3-Butadiene (average pressure* 24 *atm,* $\Phi = 1.9$*).*

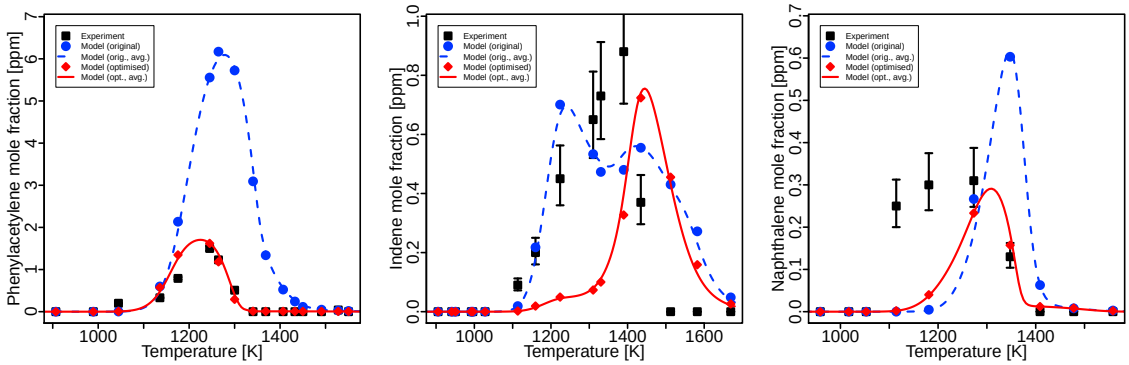
Figure 2: *Comparison of selected responses of the original as well as the optimised model with experiment. Each graph corresponds to an entry in Table 2.*



(a) *Benzene (average pressure* 52 *atm,* $\Phi = 1.9$*).* **(b)** *Toluene (average pressure* 24 *atm,* $\Phi = 1.9$*).* **(c)** *Benzaldehyde (average pressure* 49 *atm,* $\Phi = 1.0$*).*



(d) *Phenol (average pressure* 49 *atm,* $\Phi = 1.0$*).* **(e)** *1-Propenylbenzene (average pressure* 52 *atm,* $\Phi = 1.9$*).* **(f)** *2-Propenylbenzene (average pressure* 24 *atm,* $\Phi = 1.9$*).*



(g) *Phenylacetylene (average pressure* 28 *atm,* $\Phi = 0.54$*).* **(h)** *Indene (average pressure* 24 *atm,* $\Phi = 1.9$*).* **(i)** *Naphthalene (average pressure* 51 *atm,* $\Phi = 0.55$*).*

Figure 3: *Comparison of selected responses of the original as well as the optimised model with experiment. Each graph corresponds to an entry in Table 2.*

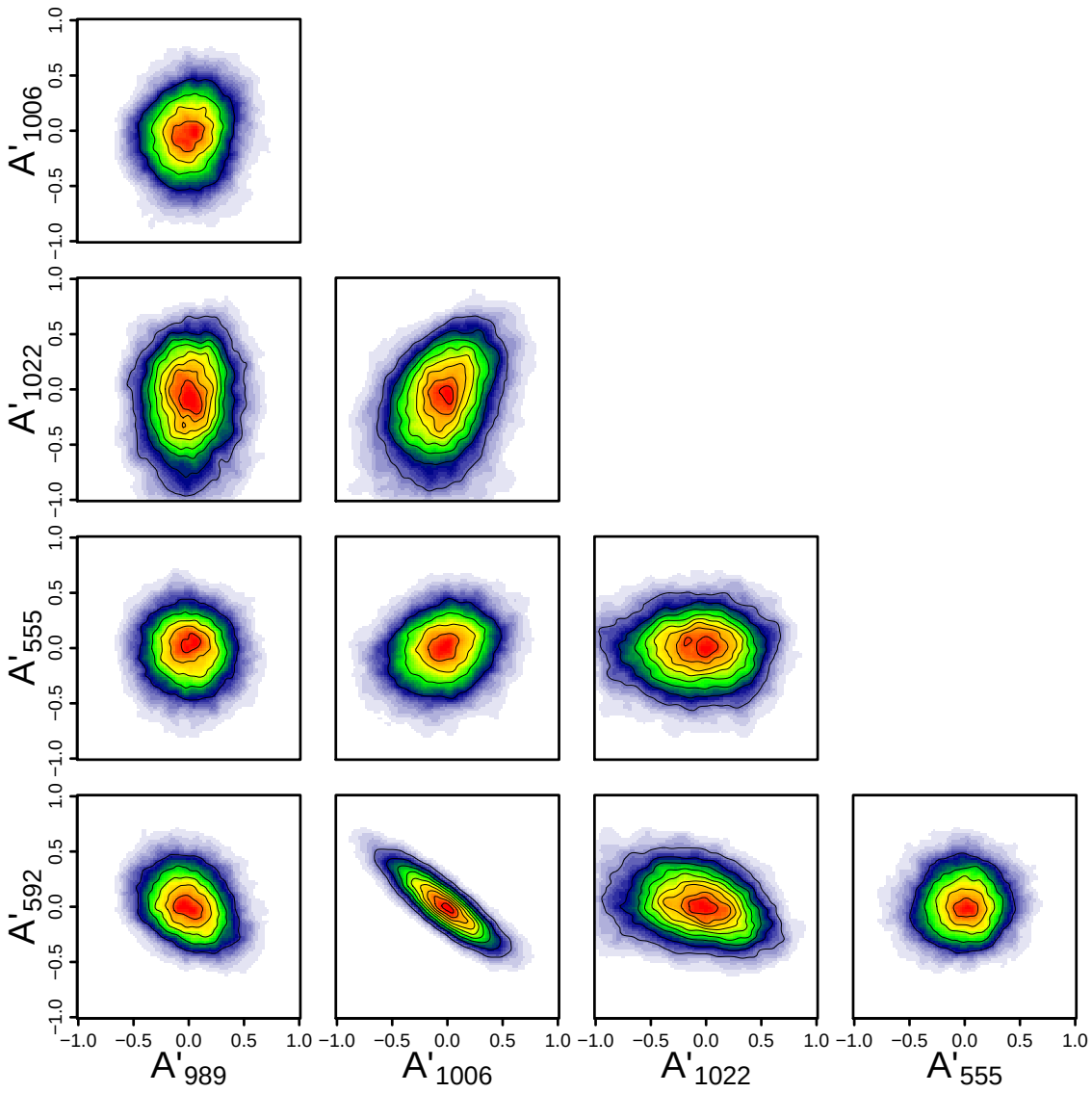


Figure 4: *Marginal posterior densities for selected pre-exponential factors. Their shape is Gaussian to a good approximation.*

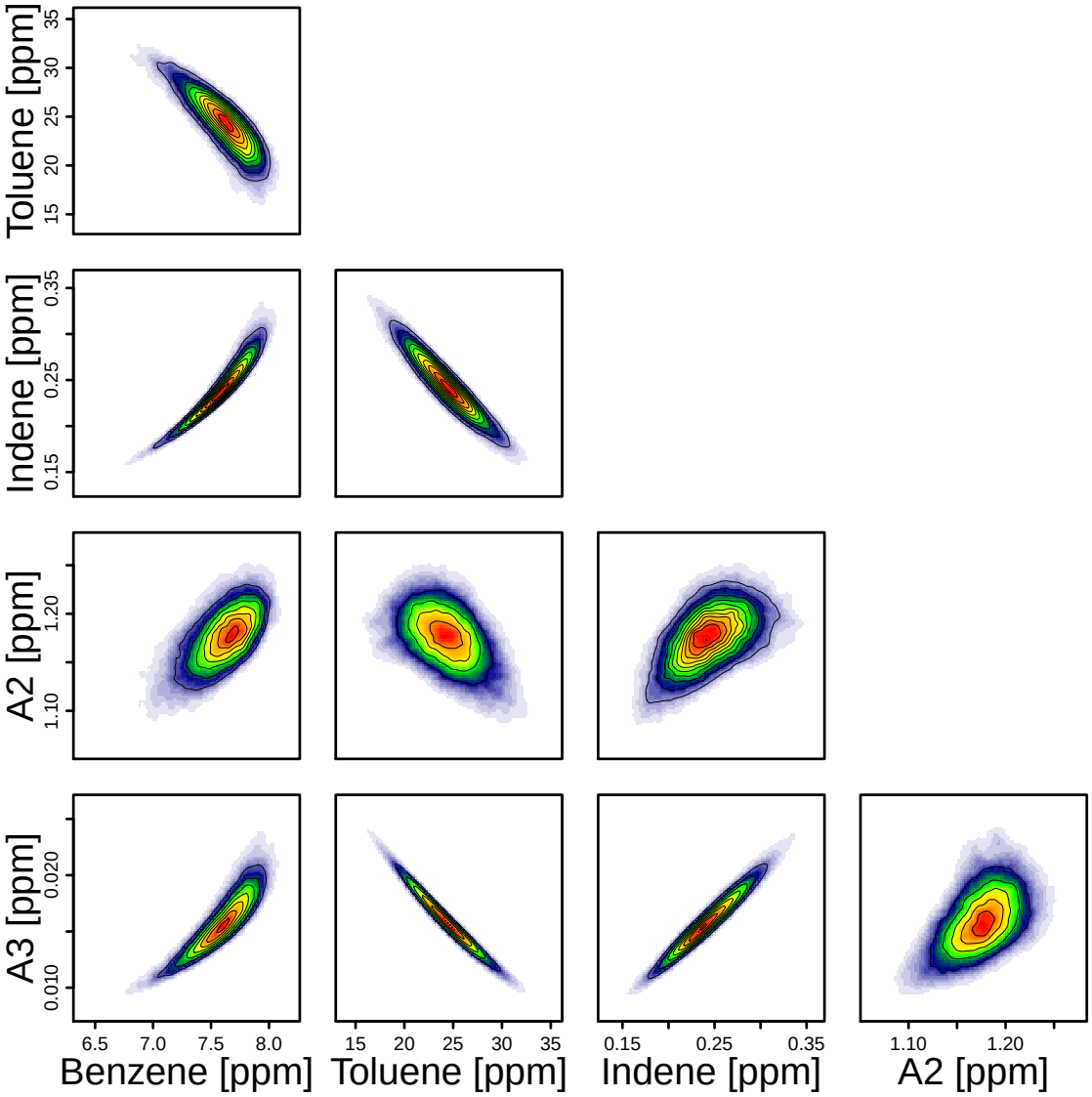


Figure 5: *Marginal posterior densities for selected responses at an average pressure of* 52 *atm,* $\Phi = 1.9$*,* $T = 1286$ *K.*

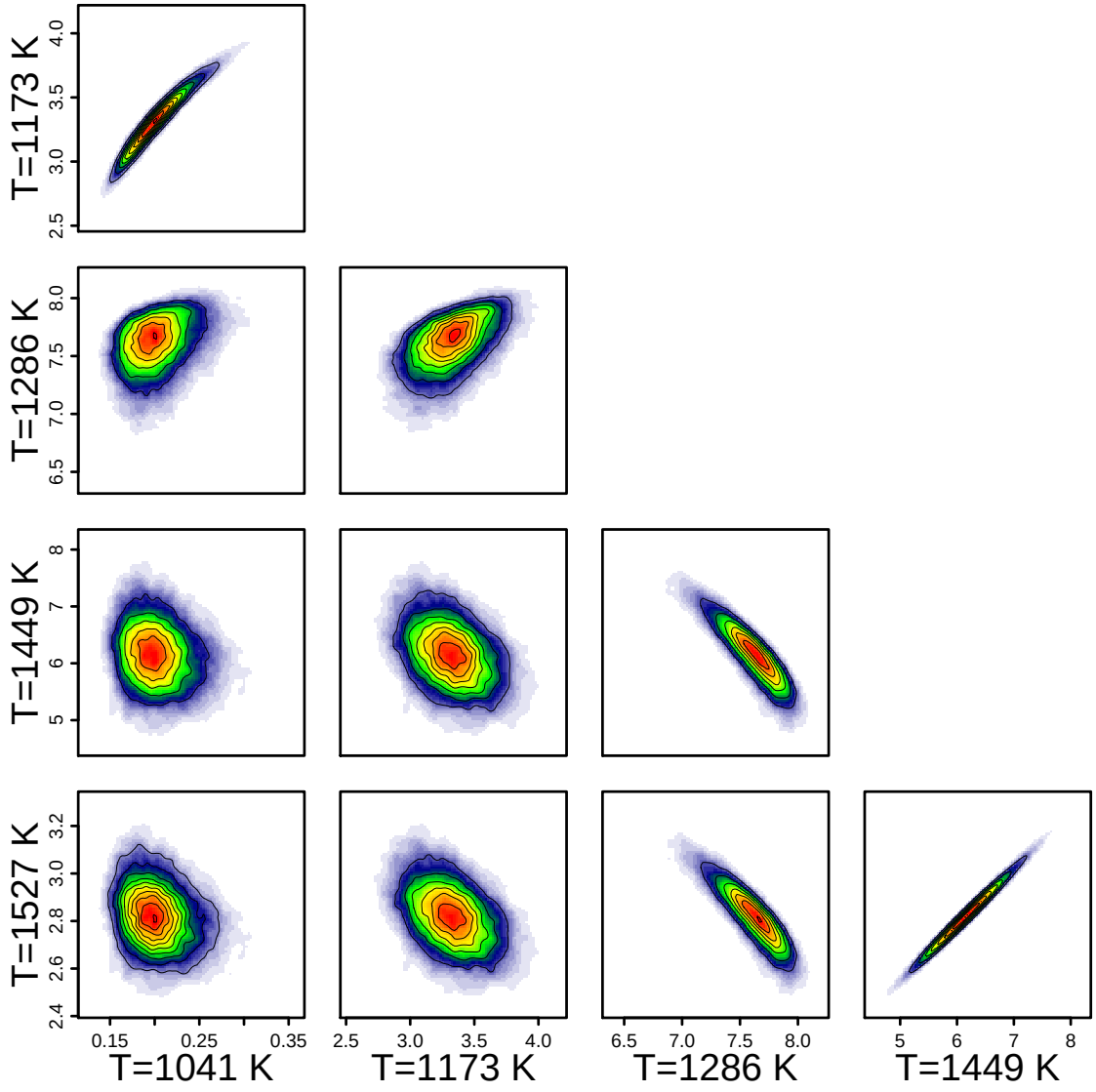
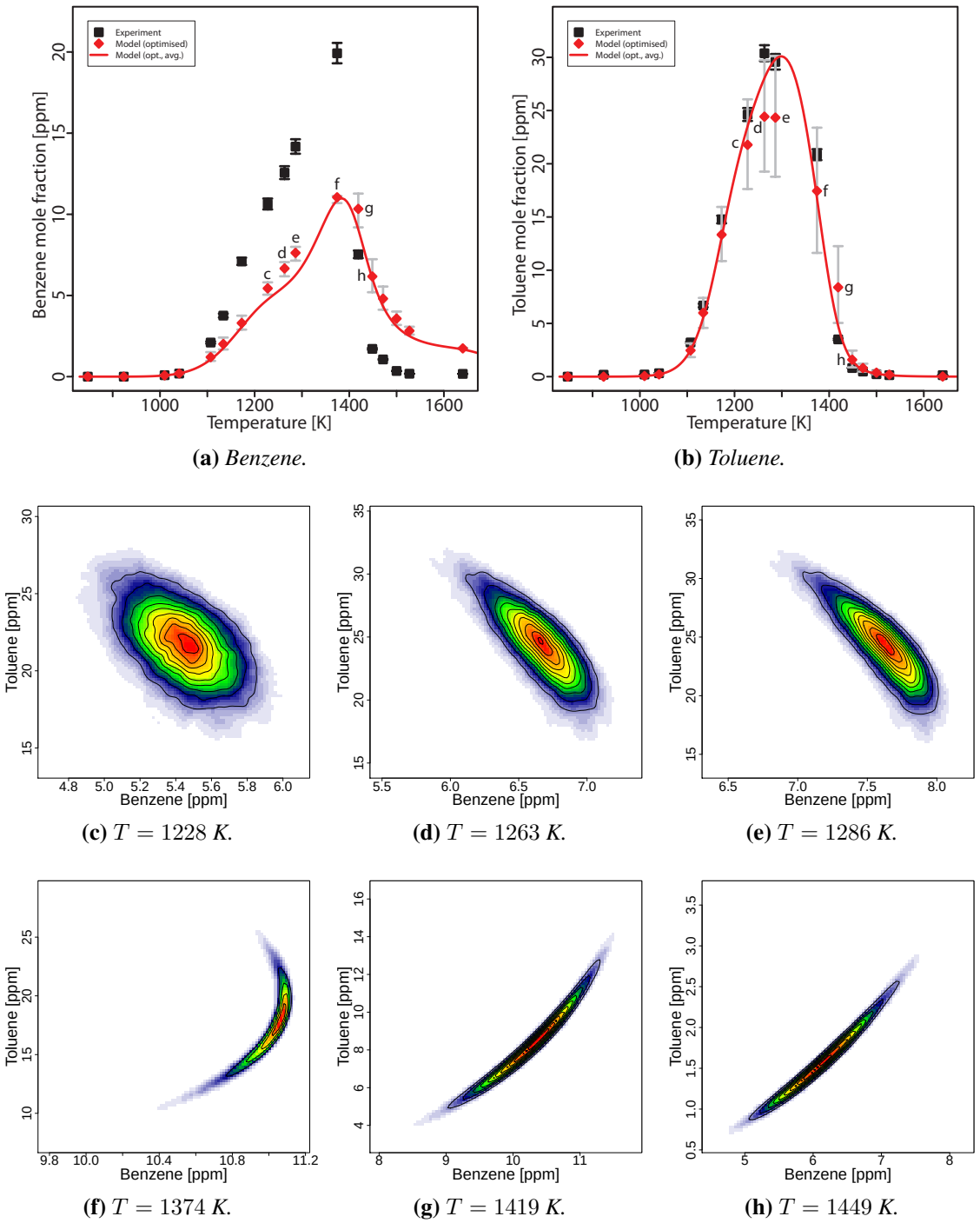


Figure 6: *Marginal posterior densities for benzene at different temperatures (average pressure* 52 *atm,* $\Phi = 1.9$*). Negative correlations such as between 1286 K and 1449 K suggest that the response is controlled by different pathways in different temperature ranges.*



(a) *Benzene.*

(b) *Toluene.*

(c) $T = 1228$ *K.*

(d) $T = 1263$ *K.*

(e) $T = 1286$ *K.*

(f) $T = 1374$ *K.*

(g) $T = 1419$ *K.*

(h) $T = 1449$ *K.*

Figure 7: *Correlations between benzene and toluene mole fractions at different temperatures (average pressure* 52 *atm,* $\Phi = 1.9$*). As the temperature increases, the correlation becomes stronger and turns from negative to positive. The error bars for the model response in Figs. 7a and 7b are the* 2σ *HPD regions derived from the one-dimensional marginal distributions.*

following are for the best point overall which was found with the Levenberg-Marquardt optimisation run initiated from the second-best Sobol point.

In order to assess in more detail the improvement or deterioration of the agreement between experiment and model, we consider the partial sums of those terms in the objective function (3) belonging to each species and each of the five sets of conditions as summarised in Table 1. Table 2 lists, for each measured species and for each of the five sets of conditions, the square root of the ratios of the value of the corresponding partial sum for the optimised mechanism and the original one. Ratios less than one, *i.e* where the agreement has improved, are highlighted. Most of the contributions – 120 out of 159, more than 75% – have improved.

Figures 2 and 3 show a comparison of selected responses of the original as well as the optimised model with experiment. Each model data point displayed is evaluated at its exact condition, *i.e.* at the experimental pressure and reaction time, whereas the curves labelled 'average' are temperature sweeps evaluated at an average pressure and reaction time. The average curves, included as a visual aid but also because they were used in previous work [24], differ somewhat from the values obtained at the exact conditions mainly due to the fact that many species concentrations change significantly over the observed time-scale. Each graph in these figures corresponds to an entry in Table 2. For example, for O_2 at an average pressure of 24 atm and and equivalence ratio of 1.9 (Fig. 2b), the sum of those terms in the objective function corresponding to the data points in the graph has decreased by a factor of $0.128^{-2} \approx 61$.

Some species have improved considerably, such as 2-propenylbenzene, 1-propenylbenzene, and phenylacetylene (Figs. 3e, 3f, and 3g). This is mostly a consequence of the fact that the value of the objective function of the original model is dominated by a few responses including 2- and 1-propenylbenzene. The reason for that is that these agree much less well with experiment, relatively speaking, than most others, and that the experimental values have small error bars, which leads to a large weight of the corresponding terms in the sum in Eqn. (3). The agreement of some responses has worsened. For example, n-propylbenzene (Fig. 2a) and benzaldehyde (Fig. 3c) are two of the worst cases. In case of n-propylbenzene, the worsening of the corresponding objective function contributions manifests itself in the fact that the decomposition curve has shifted by about 20 K towards lower temperatures. In general, worsening agreement of some model responses with experimental observations, while others improve, can be indicative of not including sufficiently many model parameters in the optimisation. While it is clear that not all of the 2378 experimental observations determine or at least constrain a model parameter, due to strong correlations in the data, the question of how many 'true' degrees of freedom this optimisation problem has is a non-trivial one [21], and we shall make no attempt to address this question here.

3.3 Bayesian parameter estimation and error propagation

In order to compute posterior densities, the covariance matrices of the (experimental) responses need to be defined. The most natural choice is

$$\sum_{n=1}^{N} \varepsilon^{(n)\top} \left(\Sigma^{(n)}\right)^{-1} \varepsilon^{(n)} = \Phi(\theta),$$

with $\Phi(\theta)$ defined by Eqn. (3), *i.e.* $\Sigma^{(n)} = \mathrm{diag}\left((\sigma_1^{(n)})^2, \ldots, (\sigma_{L^{(n)}}^{(n)})^2\right)$. The following points should be noted:

1. This choice assumes the responses are uncorrelated.
2. The fact that not every response is measured at every point in process condition space rules out homoskedasticity, *i.e.* using the same covariance matrix at every point in process condition space.
3. This highlights the close relationship between the form of the objective function and the distribution of experimental data: using a non-least-squares objective function generally implies that the experimental error can no longer be assumed to be normally distributed (see also [53]).

With the above choice, we find that the resulting errors in model parameters and responses are very optimistic – much smaller than what one would anticipate by visual inspection of the responses (Figs. 2 and 3). It is known that a main reason for this is the fact that the model discrepancy is not taken into account (in Eqn. (4)) [8, 26, 30]. In order to compensate for this, we arbitrarily multiply each of the $\sigma_i^{(n)}$ by a factor of 20. We note that rescaling the objective function by an overall (positive) constant does not affect the location of any minimum.

We select the pre-exponential factors of reactions 989, 1006, 1022, 555, and 592 for the Bayesian analysis, which are among the most sensitive ones and which are all involved in the decomposition of the fuel molecule (see Fig. 1). The chosen parameters have well-defined minima in the interior of the hypercube, in contrast to 25 of the 64 parameters whose minimum lies on the boundary. This phenomenon is characteristic of combustion kinetic systems [21], as a consequence of valley-shaped objective functions [18]. For presentational purposes, we have chosen to include only five parameters into the analysis. While some constraints on the number of parameters are imposed by the computing hardware, such as the memory required for post-processing the samples, other constraints originate from the scaling behaviour with dimension for the number of samples required for convergence of the MCMC algorithm, which needs to be assessed on a case-by-case basis. Another difficulty stems from the need for surrogates with large numbers of independent variables. In higher dimensions, it would be necessary to exploit that the number of active parameters is typically small, as a consequence of effect sparsity [3, 21].

When sampling from the posterior (5), it is critical to ensure the results are free from numerical artefacts. In particular, it needs to be established that the obtained distributions

are independent of the numbers of samples as well as the number of initial samples discarded (burn-in). Related to that, autocorrelations should be near zero, sample trace plots should be essentially time-translation invariant, and acceptance rates should be near their theoretically recommended values. Finally, we test whether two different, albeit related, sampling algorithms, namely Metropolis-Hastings and Wang-Landau, give comparable answers. These issues are described in more detail in [29, 39]. All posterior densities shown in the following have been produced with the Metropolis-Hastings method using 10^5 samples with a burn-in of 10^4.

Whenever a surrogate is used instead of the actual model, it needs to be demonstrated that the former accurately reproduces the behaviour of the latter. To this end, we test the dependence of the predictions on the order of the polynomial surrogate, which we least-squares fit to 10^3 Sobol points. Recalling that a k^{th} order polynomial in n dimensions has $\binom{n+k}{k}$ degrees of freedom, we note that, in five dimensions, this number of Sobol points is sufficient for a polynomial order of up to six without running into problems with overfitting. We find that most responses at most of the points in process condition space are captured accurately by polynomials of second order: 2226 out of 2378 are represented with an R^2 of 0.99 or higher, but the R^2 of some of the remaining 152 can fall as low as 0.57. At sixth order, only 3 fail to reach $R^2 \geq 0.99$, with the lowest being about 0.92. All results presented here are for sixth order polynomials.

Figure 4 shows two-dimensional marginal posterior probability densities for the above-mentioned five reactions. The pre-exponential factors are coded logarithmically (denoted by a prime) to a range from -1 to $+1$, where the lower bound of -1 corresponds to the nominal value divided by a factor and the upper bound of $+1$ corresponds to the nominal value multiplied by the same factor. The factors for the reactions 989, 1006, 1022, 555, and 592 are 1.6, 4.0, 2.0, 1.8, and 1.6 respectively. The pre-exponentials appear to be normally distributed to a good approximation. Some correlations can be seen, such as a negative one between reactions 592 and 1006.

As every sample in model parameter space shown in Fig. 4 is nothing but a model evaluation, one can, of course, produce analogous plots for the responses. Figure 5 shows the correlations between benzene, toluene, indene, naphthalene, and anthracene mole fractions at an average pressure of 52 atm, an equivalence ratio of $\Phi = 1.9$, and a temperature of $T = 1286$ K. We observe that deviations from Gaussian behaviour are not negligible in some cases, although assuming normal distributions should still give reasonable approximations.

Figure 6 shows the correlation of the benzene mole fraction with itself at various temperatures, for an average pressure of 52 atm and an equivalence ratio of $\Phi = 1.9$. While, as expected in general, nearby temperatures tend to be strongly positively correlated and less and less so with increasing separation, two modes can be identified. The fact that there is a negative correlation between the response at 1286 K and at 1449 K suggests that the response is controlled by different pathways in different temperature ranges.

Figure 7 shows correlations between the benzene and toluene mole fractions at various temperatures, for an average pressure of 52 atm and an equivalence ratio of $\Phi = 1.9$. The error bars for the model response in Figs. 7a and 7b are the 2σ High Probability Density (HPD) regions derived from the one-dimensional marginal distributions. As the temper-

ature increases (Figs. 7c to 7h), we observe that the correlation becomes stronger and turns from negative to positive. Also, we note that an assumption of normally distributed responses can be problematic in some circumstances, such as at $T = 1374$ K (Fig. 7f) where a crescent-shaped distribution is obtained. This is consistent with reports in the literature [2, 37]. One should bear in mind, though, that the shape of the distributions directly reflects the local geometry of the objective function surface at the point in parameter space where the analysis is carried out (see also [37]), and thereby in general depends on the location of that point, *i.e.* the chosen (local) minimum. We note, however, that this is an intrinsic property of the problem, and not of the method used to study it.

4 Conclusions

We have applied a Bayesian parameter estimation method to error propagation in a chemical kinetic model for n-propylbenzene oxidation in a shock tube as a case study. Even though the Bayesian parameter estimation method includes optimisation in the sense that the points of highest posterior density correspond to the local minima of the objective function, we found that, for the model and data considered, it was necessary to perform a conventional optimisation before applying the Bayesian method. This is essentially due to the challenges associated with producing surrogates of sufficient fidelity over large ranges in parameter space. The use of surrogates is inevitable due to large numbers of evaluations required for MCMC sampling from the posterior densities. We observed that response uncertainties are significantly underestimated, which has been noted previously and is at least partially attributable to systematic model inadequacy. Furthermore, we have found that second-order response surfaces are sufficiently accurate in general, but not always. Similarly, assuming normal distributions for propagated errors is largely adequate, but not in all cases. We hope the observations reported here may be useful to practitioners who consider using Bayesian techniques for uncertainty quantification.

Acknowledgements

This publication is made possible by the Singapore National Research Foundation under its Campus for Research Excellence And Technological Enterprise (CREATE) programme.

References

[1] G. Blau, M. Lasinski, S. Orcun, S.-H. Hsu, J. Caruthers, N. Delgass, and V. Venkatasubramanian. High fidelity mathematical model building with experimental data: A Bayesian approach. *Computers and Chemical Engineering*, 32(4-5):971–989, 2008. doi:10.1016/j.compchemeng.2007.04.008.

[2] G. E. P. Box and N. R. Draper. The Bayesian estimation of common parameters from several responses. *Biometrika*, 52(3-4):355–365, 1965. doi:10.1093/biomet/52.3-4.355.

[3] G. E. P. Box and R. D. Meyer. Some new ideas in the analysis of screening designs. *Journal of Research of the National Bureau of Standards*, 90(6):495–500, 1985. doi:10.6028/jres.090.048.

[4] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, 1973.

[5] A. Braumann and M. Kraft. Incorporating experimental uncertainties into multivariate granulation modelling. *Chemical Engineering Science*, 65(3):1088–1100, 2010. doi:10.1016/j.ces.2009.09.063.

[6] A. Braumann, M. Kraft, and P. R. Mort. Parameter estimation in a multidimensional granulation model. *Powder Technology*, 197(3):196–210, 2010. doi:10.1016/j.powtec.2009.09.014.

[7] A. Braumann, P. L. W. Man, and M. Kraft. Statistical approximation of the inverse problem in multivariate population balance modeling. *Industrial and Engineering Chemistry Research*, 49(1):428–438, 2010. doi:10.1021/ie901230u.

[8] J. Brynjarsdóttir and A. O'Hagan. Learning about physical parameters: The importance of model discrepancy. Submitted for publication, 2013.

[9] cmcl innovations. *k*inetics: the chemical kinetics model builder, version 8.0, 2013. http://www.cmclinnovations.com/kinetics/.

[10] M. Colket, T. Edwards, S. Williams, N. P. Cernansky, D. L. Miller, F. Egolfopoulos, P. Lindstedt, R. Seshadri, F. L. Dryer, C. K. Law, D. G. Friend, D. B. Lenhert, H. Pitsch, A. Sarofim, M. D. Smooke, and W. Tsang. Development of an experimental database and kinetic models for surrogate jet fuels. 45th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada. Paper No. AIAA 2007-770, 2007. doi:10.2514/6.2007-770.

[11] P. Dagaut, A. Ristori, A. El Bakali, and M. Cathonnet. Experimental and kinetic modeling study of the oxidation of n-propylbenzene. *Fuel*, 81(2):173–184, 2002. doi:10.1016/S0016-2361(01)00139-9.

[12] S. G. Davis, A. B. Mhadeshwar, D. G. Vlachos, and H. Wang. A new approach to response surface development for detailed gas-phase and surface reaction kinetic model optimization. *International Journal of Chemical Kinetics*, 36(2):94–106, 2004. doi:10.1002/kin.10177.

[13] S. G. Davis, A. V. Joshi, H. Wang, and F. Egolfopoulos. An optimized kinetic model of H_2/CO combustion. *Proceedings of the Combustion Institute*, 30(1):1283–1292, 2005. doi:10.1016/j.proci.2004.08.25.

[14] S. Dooley, S. H. Won, J. Heyne, T. I. Farouk, Y. Ju, F. L. Dryer, K. Kumar, X. Hui, C.-J. Sung, H. Wang, M. A. Oehlschlaeger, V. Iyer, S. Iyer, T. A. Litzinger, R. J. Santoro, T. Malewicki, and K. Brezinsky. The experimental evaluation of a methodology for surrogate fuel formulation to emulate gas phase combustion kinetic phenomena. *Combustion and Flame*, 159(4):1444–1466, 2012. doi:10.1016/j.combustflame.2011.11.002.

[15] B. Eiteneer and M. Frenklach. Experimental and modeling study of shock-tube oxidation of acetylene. *International Journal of Chemical Kinetics*, 35(9):391–414, 2003. doi:10.1002/kin.10141.

[16] R. Feeley, P. Seiler, A. Packard, and M. Frenklach. Consistency of a reaction dataset. *Journal of Physical Chemistry A*, 108(44):9573–9583, 2004. doi:10.1021/jp047524w.

[17] R. Feeley, M. Frenklach, M. Onsum, T. Russi, A. Arkin, and A. Packard. Model discrimination using data collaboration. *Journal of Physical Chemistry A*, 110(21): 6803–6813, 2006. doi:10.1021/jp056309s.

[18] M. Frenklach. Modeling. In W. C. Gardiner, editor, *Combustion Chemistry*, chapter 7, pages 423–453. Springer Verlag, New York, 1984.

[19] M. Frenklach. Systematic optimization of a detailed kinetic model using a methane ignition example. *Combustion and Flame*, 58(1):69–72, 1984. doi:10.1016/0010-2180(84)90079-8.

[20] M. Frenklach. Transforming data into knowledge – Process Informatics for combustion chemistry. *Proceedings of the Combustion Institute*, 31(1):125–140, 2007. doi:10.1016/j.proci.2006.08.121.

[21] M. Frenklach, H. Wang, and M. J. Rabinowitz. Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method — combustion of methane. *Progress in Energy and Combustion Science*, 18:47–73, 1992. doi:10.1016/0360-1285(92)90032-V.

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

[23] C. F. Goldsmith, A. S. Tomlin, and S. J. Klippenstein. Uncertainty propagation in the derivation of phenomenological rate coefficients from theory: A case study of n-propyl radical oxidation. *Proceedings of the Combustion Institute*, 34(1):177–185, 2013. doi:10.1016/j.proci.2012.05.091.

[24] S. Gudiyella and K. Brezinsky. High pressure study of n-propylbenzene oxidation. *Combustion and Flame*, 159(3):940–958, 2012. doi:10.1016/j.combustflame.2011.09.013.

[25] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi:10.1093/biomet/57.1.97.

[26] D. Higdon, J. Gattiker, B. Williams, and M. Rightley. Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103 (482):570–583, 2008. doi:10.1198/016214507000000888.

[27] X. Huan and Y. M. Marzouk. Optimal Bayesian experimental design for combustion kinetics. 49th AIAA Aerospace Sciences Meeting, Orlando, Florida, January 2011. American Institute of Aeronautics and Astronautics Paper AIAA 2011-0513.

[28] X. Huan and Y. M. Marzouk. Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics*, 232(1):288–317, 2013. doi:10.1016/j.jcp.2012.08.013.

[29] C. A. Kastner, A. Braumann, P. L. W. Man, S. Mosbach, G. P. E. Brownbridge, J. W. J. Akroyd, M. Kraft, and C. Himawan. Bayesian parameter estimation for a jet-milling model using Metropolis-Hastings and Wang-Landau sampling. *Chemical Engineering Science*, 89:244–257, 2013. doi:10.1016/j.ces.2012.11.027.

[30] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society B*, 63(3):425–464, 2001. Stable URL: http://www.jstor.org/stable/2680584.

[31] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathematics*, II(2):164–168, 1944.

[32] T. A. Litzinger, K. Brezinsky, and I. Glassman. Reactions of n-propylbenzene during gas phase oxidation. *Combustion Science and Technology*, 50(1-3):117–133, 1986. doi:10.1080/00102208608923928.

[33] P. L. W. Man, A. Braumann, and M. Kraft. Resolving conflicting parameter estimates in multivariate population balance models. *Chemical Engineering Science*, 65(13):4038–4045, 2010. doi:10.1016/j.ces.2010.03.042.

[34] D. W. Marquardt. An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society of Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[35] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087–1091, 1953. doi:10.1063/1.1699114.

[36] K. Miki, S. Cheung, E. E. Prudencio, and P. L. Varghese. Bayesian uncertainty quantification of recent shock tube determinations of the rate coefficient of reaction $H+O_2 \rightarrow OH+O$. *International Journal of Chemical Kinetics*, 44(9):586–597, 2012. doi:10.1002/kin.20736.

[37] D. Miller and M. Frenklach. Sensitivity analysis and parameter estimation in dynamic modeling of chemical kinetics. *International Journal of Chemical Kinetics*, 15(7):677–696, 1983. doi:10.1002/kin.550150709.

[38] S. Mosbach, A. M. Aldawood, and M. Kraft. Real-time evaluation of a detailed chemistry HCCI engine model using a tabulation technique. *Combustion Science and Technology*, 180(7):1263–1277, 2008. doi:10.1080/00102200802049414.

[39] S. Mosbach, A. Braumann, P. L. W. Man, C. A. Kastner, G. P. E. Brownbridge, and M. Kraft. Iterative improvement of Bayesian parameter estimates for an engine model by means of experimental design. *Combustion and Flame*, 159(3):1303–1313, 2012. doi:10.1016/j.combustflame.2011.10.019.

[40] T. Nagy and T. Turányi. Uncertainty of Arrhenius parameters. *International Journal of Chemical Kinetics*, 43(7):359–378, 2011. doi:10.1002/kin.20551.

[41] T. Nagy and T. Turányi. Determination of the uncertainty domain of the Arrhenius parameters needed for the investigation of combustion kinetic models. *Reliability Engineering and System Safety*, 107:29–34, 2012. doi:10.1016/j.ress.2011.06.009.

[42] J. Prager, H. N. Najm, K. Sargsyan, C. Safta, and W. J. Pitz. Uncertainty quantification of reaction mechanisms accounting for correlations introduced by rate rules and fitted Arrhenius parameters. *Combustion and Flame*, 2013. doi:10.1016/j.combustflame.2013.01.008. In Press.

[43] J. Prager, H. N. Najm, and J. Zádor. Uncertainty quantification in the *ab initio* rate-coefficient calculation for the $CH_3CH(OH)CH_3+OH \rightarrow CH_3C\cdot(OH)CH_3+H_2O$ reaction. *Proceedings of the Combustion Institute*, 34(1):583–590, 2013. doi:10.1016/j.proci.2012.06.078.

[44] H. Rabitz, M. Kramer, and D. Dacol. Sensitivity analysis in chemical kinetics. *Annual Review of Physical Chemistry*, 34:419–461, 1983. doi:10.1146/annurev.pc.34.100183.002223.

[45] M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545555, 2003. doi:10.1016/S0010-2180(02)00503-5.

[46] 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. *Combustion Theory and Modelling*, 8(3):607–632, 2004. doi:10.1088/1364-7830/8/3/010.

[47] M. T. Reagan, H. N. Najm, P. P. Pébay, O. M. Knio, and R. G. Ghanem. Quantifying uncertainty in chemical systems modeling. *International Journal of Chemical Kinetics*, 37(6):368–382, 2005. doi:10.1002/kin.20081.

[48] T. Russi, A. Packard, R. Feeley, and M. Frenklach. Sensitivity analysis of uncertainty in model prediction. *Journal of Physical Chemistry A*, 112(12):2579–2588, 2008. doi:10.1021/jp076861c.

[49] T. Russi, A. Packard, and M. Frenklach. Uncertainty quantification: Making predictions of complex reaction systems reliable. *Chemical Physics Letters*, 499(1-3):1–8, 2010. doi:10.1016/j.cplett.2010.09.009.

[50] M. Sander, R. I. A. Patterson, A. Braumann, A. Raj, and M. Kraft. Developing the PAH-PP soot particle model using process informatics and uncertainty propagation. *Proceedings of the Combustion Institute*, 33(1):675–683, 2011. doi:10.1016/j.proci.2010.06.156.

[51] P. Seiler, M. Frenklach, A. Packard, and R. Feeley. Numerical approaches for collaborative data processing. *Optimization and Engineering*, 7(4):459–478, 2006. doi:10.1007/s11081-006-0350-4.

[52] D. A. Sheen and H. Wang. Combustion kinetic modeling using multispecies time histories in shock-tube oxidation of heptane. *Combustion and Flame*, 158(4):645–656, 2011. doi:10.1016/j.combustflame.2010.12.016.

[53] D. A. Sheen and H. Wang. The method of uncertainty quantification and minimization using polynomial chaos expansions. *Combustion and Flame*, 158(12):2358–2374, 2011. doi:10.1016/j.combustflame.2011.05.010.

[54] D. A. Sheen, X. You, H. Wang, and T. Løvås. Spectral uncertainty quantification, propagation and optimization of a detailed kinetic model for ethylene combustion. *Proceedings of the Combustion Institute*, 32(1):535–542, 2009. doi:10.1016/j.proci.2008.05.042.

[55] I. M. Sobol. On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis*, 16(5):790–793, 1979. Stable URL: http://www.jstor.org/stable/2156633.

[56] W. Tang and K. Brezinsky. Chemical kinetic simulations behind reflected shock waves. *International Journal of Chemical Kinetics*, 38(2):75–97, 2006. doi:10.1002/kin.20134.

[57] A. S. Tomlin. The use of global uncertainty methods for the evaluation of combustion mechanisms. *Reliability Engineering and System Safety*, 91(10-11):1219–1231, 2006. doi:10.1016/j.ress.2005.11.026.

[58] A. S. Tomlin. The role of sensitivity and uncertainty analysis in combustion modelling. *Proceedings of the Combustion Institute*, 34(1):159–176, 2013. doi:10.1016/j.proci.2012.07.043.

[59] A. S. Tomlin, T. Turányi, and M. J. Pilling. Mathematical tools for the construction, investigation and reduction of combustion mechanisms. In R. G. Compton, G. Hancock, and M. J. Pilling, editors, *Low-Temperature Combustion and Autoignition*, volume 35 of *Comprehensive Chemical Kinetics*, pages 293–437. Elsevier, 1997.

[60] S. R. Tonse, N. W. Moriarty, M. Frenklach, and N. J. Brown. Computational economy improvements in PRISM. *International Journal of Chemical Kinetics*, 35(9): 438–452, 2003. doi:10.1002/kin.10140.

[61] R. S. Tranter, K. Brezinsky, and D. Fulle. Design of a high-pressure single pulse shock tube for chemical kinetic investigations. *Review of Scientific Instruments*, 72 (7):3046–3054, 2001. doi:10.1063/1.1379963.

[62] T. Turányi. Sensitivity analysis of complex kinetic systems. Tools and applications. *Journal of Mathematical Chemistry*, 5(3):203–248, 1990. doi:10.1007/BF01166355.

[63] T. Turányi. Parameterization of reaction mechanisms using orthonormal polynomials. *Computers & Chemistry*, 18(1):45–54, 1994. doi:10.1016/0097-8485(94)80022-7.

[64] T. Turányi. Applications of sensitivity analysis to combustion chemistry. *Reliability Engineering and System Safety*, 57(1):41–48, 1997. doi:10.1016/S0951-8320(97)00016-1.

[65] T. Turányi and H. Rabitz. Local methods. In A. Saltelli, K. Chan, and E. M. Scott, editors, *Sensitivity Analysis*, Wiley Series in Probability and Statistics, pages 81–99. John Wiley & Sons, New York, 2000.

[66] T. Turányi, T. Nagy, I. G. Zsély, M. Cserháti, T. Varga, B. T. Szabó, I. Sedyó, P. T. Kiss, A. Zempléni, and H. J. Curran. Determination of rate parameters based on both direct and indirect measurements. *International Journal of Chemical Kinetics*, 44(5):284–302, 2012. doi:10.1002/kin.20717.

[67] L. Varga, B. Szabó, I. G. Zsély, A. Zempléni, and T. Turányi. Numerical investigation of the uncertainty of Arrhenius parameters. *Journal of Mathematical Chemistry*, 49(8):1798–1809, 2011. doi:10.1007/s10910-011-9859-7.

[68] F. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters*, 86(10):2050–2053, 2001. doi:10.1103/PhysRevLett.86.2050.

[69] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4): 897–936, 1938. Stable URL: http://www.jstor.org/stable/2371268.

[70] D. Xiu and G. E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002. doi:10.1137/S1064827501387826.

[71] X. You, T. Russi, A. Packard, and M. Frenklach. Optimization of combustion kinetic models on a feasible set. *Proceedings of the Combustion Institute*, 33(1):509–516, 2011. doi:10.1016/j.proci.2010.05.016.

[72] X. You, A. Packard, and M. Frenklach. Process informatics tools for predictive modeling: Hydrogen combustion. *International Journal of Chemical Kinetics*, 44 (2):101–116, 2012. doi:10.1002/kin.20627.

[73] J. Zádor, I. G. Zsély, and T. Turányi. Local and global uncertainty analysis of complex chemical kinetic systems. *Reliability Engineering and System Safety*, 91(10): 1232–1240, 2006. doi:10.1016/j.ress.2005.11.020.

[74] T. Ziehn and A. S. Tomlin. A global sensitivity study of sulfur chemistry in a premixed methane flame model using HDMR. *International Journal of Chemical Kinetics*, 40(11):742–753, 2008. doi:10.1002/kin.20367.

[75] I. G. Zsély, J. Zádor, and T. Turányi. Similarity of sensitivity functions of reaction kinetic models. *Journal of Physical Chemistry A*, 107(13):2216–2238, 2003. doi:10.1021/jp026683h.