The Role of Correlations in Uncertainty Quantification of Transportation Relevant Fuel Models

Aleksandr Fridlyand^{1,*},
Matthew S. Johnson¹,
S. Scott Goldsborough¹,

Richard H. West²,

Matthew J. McNenly³,
Marco Mehl³,
William J. Pitz³

¹Center for Transportation Research, Argonne National Laboratory, Argonne, IL 60439, USA

²Department of Chemical Engineering, Northeastern University, Boston, MA 02115, USA

³Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

Corresponding author*:

Aleksandr Fridlyand
Energy Systems Division
Argonne National Laboratory
9700 S. Cass Avenue
Argonne, IL 60439
USA
Email: afridlyand@anl.gov
Phone: +1-630-252-3529
Fax: +1-630-252-3443

Full-length article

Abstract

Large reaction mechanisms are often used to describe the combustion behavior of transportation-relevant fuels like gasoline, where these are typically represented by surrogate blends, e.g., n-heptane/iso-octane/toluene. We describe efforts to quantify the uncertainty in the predictions of such mechanisms at realistic engine conditions, seeking to better understand the robustness of the model as well as the important reaction pathways and their impacts on combustion behavior. In this work, we examine the importance of taking into account correlations among reactions that utilize the same rate rules and those with multiple product channels on forward propagation of uncertainty by Monte Carlo simulations. Automated means are developed to generate the uncertainty factor assignment for a detailed chemical kinetic mechanism, by first uniquely identifying each reacting species, then sorting each of the reactions based on the rate rule utilized. Simulation results reveal that in the low temperature combustion regime for iso-octane, the majority of the uncertainty in the model predictions can be attributed to low temperature reactions of the fuel sub-mechanism. The foundational, or small-molecule chemistry (C_0 - C_4) only contributes significantly to uncertainties in the predictions at the highest temperatures ($T_c=900$ K). Accounting for correlations between important reactions is shown to produce non-negligible differences in the estimates of uncertainty. Including correlations among reactions that use the same rate rules increases uncertainty in the model predictions, while accounting for correlations among reactions with multiple branches decreases uncertainty in some cases. Significant non-linear response is observed in the model predictions depending on how the probability distributions of the uncertain rate constants are defined. It is concluded that care must be exercised in defining these probability distributions in order to reduce bias, and physically unrealistic estimates in the forward propagation of uncertainty for a range of UQ activities.

Key Words: Uncertainty Quantification, Kinetic Model, Gasoline Surrogate, Ignition Delay, Correlations

1.0 Introduction

The combination of predictive chemical kinetics and advanced computational fluid dynamic simulation tools is expected to play an increasingly key role in the design of future internal combustion (IC) engines. Advanced Low Temperature Combustion (LTC) concepts such as Homogeneous Charge Compression Ignition (HCCI), Reactivity Controlled Compression Ignition (RCCI), and Gasoline Compression Ignition (GCI) engines, which promise increased fuel efficiency with reduced harmful emissions, are particularly sensitive to fuel autoignition chemistry [1; 2]. To aid the design of IC engines utilizing such concepts, as well as conventional boosted, spark ignition (SI) approaches, it is critical that chemical kinetic models reliably predict important combustion properties of real fuels. These properties include the start of combustion (i.e., ignition delay), heat release rate, and engine knock. To achieve adequate confidence in the predictive capability of simulations, uncertainties in the predictions must be well characterized.

Advancements in computing power and algorithms, which permit increasingly sophisticated computer simulations, now allow uncertainties to be examined and quantified for complex problems where the range of reliability in model predictions can be determined. Recently, uncertainty quantification (UQ) techniques such as forward propagation of uncertainty, global sensitivity analysis (GSA), optimization, and uncertainty reduction methods have received a great deal of attention in the field of chemical kinetics. Previous studies have shown that variability in chemical kinetic model predictions for a variety of combustion properties can be substantial [3]. These methods have also demonstrated significant potential for improving the predictive capabilities of chemical kinetic models [3–5]. However, as an emergent form of mechanism validation and analysis, there is not yet consensus concerning best practices for critical aspects of UQ in this domain.

The first step of all aforementioned UQ methods is to assign probability distributions to describe uncertainties for all model input parameters. Forward propagation of uncertainty then provides initial estimates of uncertainty in model predictions and these results can feed into other UQ methods. However, this first step can be quite challenging and often the results reflect the analyst's subjective view of reality. Given the varied treatment in the literature, it is not a solved problem either, and the full implications of the many possible assumptions are not yet completely understood. In their recent review paper, Wang and Sheen agree [3], describing several remaining issues in UQ of combustion models. These include how to reconcile optimized mechanisms with inconsistent experiments, as well as the proper treatment of correlations that arise from the use of rate rules in detailed chemical kinetic mechanisms, where these are discussed in more detail shortly.

To date, many UQ studies have focused on the chemical kinetic models of small molecules (C_0 - C_4) at high temperatures. Arguably, the reactions included in those models are the most studied, with many independent rate constant determinations. Perhaps for this reason not much attention has been paid to the effects of correlations among reactions. However, from the few studies that have examined these relationships, the effects of neglecting correlations can be non-negligible [6–10]. This phenomenon is also well established in the broader UQ literature for cases where uncertainties are large and correlations are strong between important variables [11]. Ignoring important correlations could be considered a structural problem in how the forward propagation of uncertainty is defined, potentially leading to biased results. Wang and Sheen [3; 12] argue that such structural uncertainties can also lead to trouble in how inconsistencies between experiments and optimized models are handled, and that a rigorous analysis of uncertainty in the model is necessary to yield physically realistic results.

Correlations between chemical kinetic model parameters can manifest themselves in several different ways. One source is from the use of Arrhenius fits of measured rate constants, where the

Arrhenius parameters can be coupled. Taking these interdependences into account can be crucial when the temperature dependence of the rate constant is non-negligible [9; 13]. Uncertainty in the thermochemical parameters can also have strong correlations, and approaches such as the Active Thermochemical Tables offer means for taking these into account for accurate UQ [14]. Of particular interest to the present work are correlations due to the use of reaction rate rules and those that may arise between different branches of the same reaction.

With powerful computers having become ubiquitous, recent chemical kinetic model development approaches allow for very detailed mechanisms to be built with ease [15; 16]. The rapid growth in the size of mechanisms has significantly outpaced the production of high quality measurements and calculations for fundamental reaction rate data. This means that the majority of rate constants in large models are estimates, based on analogous reactions of molecules for which rate constants have either been measured or calculated. This methodology is behind all systematic, hierarchical model development schemes [17; 18]. While kinetic rates for two reactions cannot be exactly equal, it should be expected that a high degree of correlation exists among very similar reactions. Strong correlations between reactions with multiple branches can arise in cases when the total rate constant of a reaction may be known better than the branching ratios [8], or vice versa [10], attributed to experimental or theoretical limitations. Proper treatment of correlations arising from the use of rate rules in forward propagation of uncertainty was left as an open question in [3], and this serves as part of the motivation of the present study.

This work is part of a comprehensive effort to develop methods for uncertainty quantification and minimization of detailed chemical kinetic models of transportation-relevant fuels that are of particular interest to LTC engine research efforts. For the present investigation, we focus on the autoignition predictions of iso-octane at 20 bar, over the entire Negative Temperature Coefficient (NTC) region of the fuel, and using the Lawrence Livermore National Laboratory (LLNL) gasoline surrogate

model [19]. Iso-octane is the prototypical single-component surrogate for gasoline and is a representative iso-alkane in multi-component surrogates for gasoline [19–21]. It, along with n-heptane define the octane rating for gasoline according to ASTM International, and blends of the two are commonly referred to as Primary Reference Fuels (PRF) for gasoline. Twenty (20) bar, constant volume autoignition simulations are chosen as representative conditions of interest since it has been shown that ignition delay time predictions at ~ 20 bar and ~ 825 K strongly correlate with the Research Octane Number [21] and the Anti-Knock Index of gasoline surrogate blends [20]. The LLNL gasoline surrogate model, and its reduced and skeletal variations, are among the most widely used kinetic mechanisms for simulations of conventional SI engines. However, the same model is also used extensively in simulations of LTC engine concepts, e.g., [22]. Therefore, it is important to evaluate the uncertainty in predictions of this mechanism.

Here, the Monte Carlo technique for the forward propagation of uncertainty is applied to the detailed LLNL gasoline surrogate model in order to investigate the potential impact of correlations among reactions on estimates of uncertainty. In order to perform these simulations, we utilize a recently developed fast chemistry solver Zero-RK [23], which makes this study a sufficiently tractable problem. Several representative cases of correlations are examined. These include cases where all reactions are considered to be independent, and where reactions that use the same rate rule are considered to be perfectly correlated. Lastly the potential impact of including correlations among reactions with multiple branches is examined. The primary aim of this work is to evaluate the impact of assumptions made about uncertain model inputs on estimates of uncertainty in model predictions.

This manuscript is organized as follows. Section 2 describes the methods used to parse the detailed chemical kinetic mechanism, how the uncertainties and correlations are applied, as well details of how the simulations are performed. Section 3 discusses the results of different simulations, including the identification of sources of uncertainty in this model, as well as the impact of including correlations

due to the use of rate rules and among reactions with multiple branches. Section 4 presents some discussion of potential implications of our findings. Some concluding remarks in Section 5 and the discussion of future directions complete this manuscript.

2.0 Methods

2.1 Mechanism Identification

The mechanism employed in the present work is the gasoline surrogate model from Lawrence Livermore National Laboratory (LLNL) [19], which has 7336 reactions and 1753 species. This model has recently been updated and a detailed overview of the changes will be presented in a future publication; only a brief description is provided here. A significant modification is the incorporation of AramcoMech [24] as the foundational model, which describes the combustion chemistry of the smallest species, i.e., C_0 - C_4 . These are often the most well characterized portions of any detailed model because the included reactions and their rate constants have typically had the most independent determinations. As a result, fewer correlations attributed to the use of rate rules can be expected. However, correlations among individual reactions in the foundational chemistry sub-model can exist due to limitations in experiments and theory. Such correlations can propagate to the rest of the model, but these are not considered in detail here. Proper assignment of uncertainty probability distribution for this portion of the model requires careful analysis and literature review of each reaction, which is presently ongoing. For this reason, this portion of the mechanism is not subjected to the automated sorting procedure described next. Uncertainties for these reactions are treated simply, as described in the following. It will be shown that this treatment is sufficient for the objectives of the present study.

The rest of the model is developed hierarchically and includes chemistry for the various surrogate components, e.g., n-heptane, iso-octane, toluene. The sub-mechanisms for these species are, in large part, constructed using reaction classes, such as for the oxidation of large alkanes [17]. For many

of these reactions, rate constants derived from analogous reactions for smaller molecules are used where direct rate measurements or theoretical calculations are available. This is the simplest way in which a significant portion of the reactions can be correlated. Recently, Prager et al. [9] demonstrated on a small scale, that taking into account correlations among reactions using the same rate rule can lead to non-negligible differences in the uncertainty of the model predictions. In the present study, we seek to gauge the impact of including such correlations on uncertainty of a transportation relevant gasoline surrogate model. Due to the size of this model, automated means are required to properly parse the mechanism and identify potentially correlated reactions.

In order to identify reactions within a particular class, over 4000 reactions outside the foundational model must be considered. This is challenging for a mechanism that has been developed over many years where inconsistencies can exist, for example in species names. To overcome this barrier, the molecular structure of every species within the mechanism is first identified using a new, semi-automated importer tool [25]. This tool has been specifically developed to facilitate the comparison and decoding of CHEMKIN style mechanisms. The approach is as follows. Using an initial pool of species with known molecular structures, the structures of the remaining species are automatically determined based on reactions in which each unknown species participates. The structure is identified by the process of elimination and through comparison of standard enthalpies of formation. Each species molecular structure is confirmed manually. The product of the importer tool is a comprehensive dictionary that uniquely identifies the molecular structure of every species within the mechanism.

Using this dictionary, all elementary reactions are then automatically identified with the aid of the Reaction Mechanism Generator (RMG) software [26]. For each reaction of the mechanism, an automated script launches RMG to generate all possible reactions for a set of reactants and products. Each generated RMG reaction that matches a reaction in the mechanism (either forward or reverse) is

then used for classification. The RMG reaction class and participating molecular groups are the classifier that identifies the unique reaction rate rule. For non-elementary reactions belonging to classes of reactions not represented in the RMG database, the automated script uses additional information to classify the reactions, such as unique participating species and products with unique molecular groups. This initial step automatically and consistently identifies reactions with potential correlations and facilitates the automated assignment of uncertainty factors and probability distributions according to distinct reaction classes. In the simplest case that is investigated here, all reactions identified as using the same reaction rate rule are treated as perfectly correlated, with uncertainty probability distributions assigned according to the procedure described next.

2.2 Probability Distribution and Uncertainty Assignment

In forward propagation of uncertainty, the first step is an assignment of uncertainty probability distribution to each input variable. For this study, uncertainties in the input variables only consider the overall rate constant for each reaction, i.e., the pre-exponential factors. This is equivalent to assuming that the uncertainties are temperature independent. As discussed by Nagy and Turányi [13], temperature dependence in uncertainty parameters should be considered for the most accurate assessment of model uncertainties. Temperature and pressure dependence of the model input uncertainties, as well as the uncertainties in the thermochemistry are, however, beyond the scope of this work, and will be examined in subsequent studies. For this work, a few representative cases of uncertainty are considered, and are described in detail in the next section.

For all cases considered, a log-uniform probability is utilized to describe the uncertainty in the rate constants of individual reactions, rate rules, or the total rate constant among multiple branches. The only exception is the uncertainty for relative rate constants, i.e. branching ratios, which are subject

to a Dirichlet-type probability distribution [8], which is described in more detail shortly. The probability density function of a log-uniform distribution is given by

$$f_{\log\text{-uniform}}(X_i) = \begin{cases} 1/(\log(1/UF) - \log(UF)) & \text{for } \log(1/UF) \leq X \leq \log(UF) \\ 0 & \text{for } X_i < \log(1/UF) \text{ and } X_i > \log(UF) \end{cases} \quad (1)$$

where $X_i = \log(F_i)$, F_i is the perturbation factor for the i th reaction of an individual simulation, and UF is the Uncertainty Factor defining the limits of a particular rate constant. Random samples from this distribution are generated with the NumPy library [27] random number generator. During an individual simulation, the rate of progress of each reaction is scaled by this F_i factor.

In this investigation, the uncertainty factors considered are UF=5 for all reactions except those belonging to the foundational model (AramcoMech [24]). As discussed earlier, this is arguably the most well studied portion of the model and requires detailed evaluation of individual reactions and associated uncertainties. For this reason, modest uncertainty factor of UF=2 is utilized, and all reactions are considered to be independent, or uncorrelated. It is recognized that there are reactions within AramcoMech, which have uncertainty factors lower and higher than 2, and the danger in this simple treatment of UFs for the foundational model is that it may result in strongly biased simulation uncertainties. However, in the low temperature combustion regime, it is expected that uncertainties in the foundational chemistry are only a small contributor to the overall uncertainty in the simulation predictions. This has been identified in other work [28] and will be demonstrated to be valid here as well, as presented in Section 3.2. Future work will include a detailed evaluation of uncertainties in the foundational model, but this is beyond the scope of the present study.

In addition to correlations attributed to the use of rate rules, a source of correlations investigated here is between reactions with multiple branches. One example is the set of four hydrogen

abstraction reactions of iso-octane (iC_8H_{18}) by hydroxyl radicals (OH), leading to four different iso-octyl radicals

Correlations between such reactions can exist in cases where the total rate constant for $iC_8H_{18} + OH \rightarrow \text{products}$ is well characterized, e.g., from measurements of iso-octane depletion, but the precise branching ratio for each of the four pathways may not be. This is not an uncommon scenario, especially when only the total rate of a reaction can be measured (e.g., [21; 29]). These correlations can be best described by defining the total forward rate constant for the set of reactions, e.g., as in Eq. (2)

$$\begin{aligned} k_{tot}(T, P) &= k_a(T, P) + k_b(T, P) + k_c(T, P) + k_d(T, P) \\ &= b_a(T, P)k_{tot}(T, P) + b_b(T, P)k_{tot}(T, P) + b_c(T, P)k_{tot}(T, P) + b_d(T, P)k_{tot}(T, P) \\ &= k_{tot}(T, P)(b_a(T, P) + b_b(T, P) + b_c(T, P) + b_d(T, P)) \end{aligned} \quad (2)$$

where k_i is the respective rate constant for each channel, with subscripts a through d referring to the four different iso-octyl radicals, and b_i are the branching ratios. For Eq. (2) to be true, the sum of branching ratios must equal to 1. This ‘sum-to-one’ requirement ($\sum b_i = 1$) is what defines the correlation.

Previously, Pernot and co-workers demonstrated the impact of accounting for such correlations in UQ analysis of Titan ionosphere model predictions [6–8]. They accounted for correlations among branched reactions using the Dirichlet distribution, which is a multivariate probability distribution that

enforces the ‘sum-to-one’ requirement between variables through normalization. In the present study, we investigate the impact of accounting for correlations among reactions with multiple branches by implementing the generalized Dirichlet distribution as described by Plessis et al. [8], based on the prior work of Lingwall et al. [30]. The generalized Dirichlet is one approach to model the probability distribution and relationship between the known total rate constant and the individual branching ratios. It represents a state of knowledge where the total rate constants, the mean branching ratios, and their respective uncertainties are available. However, that scenario may not exist, and other types of probability distributions may be appropriate. Plessis et al. [8] describe several Dirichlet type distributions that can reflect different states of knowledge regarding the reacting system, ranging from the situation where all uncertainties are known, to one where almost no knowledge is available. They outline a procedure that uses a decision tree and nested distributions to describe a complex system of reactions with multiple branches. Such an approach may be adapted for detailed evaluation in the future, but is beyond the scope of the present study. Briefly, the governing equations and sampling procedure are summarized next.

The probability distribution for each branching ratio is given by the Γ distribution. The probability density function for this distribution is defined by

$$f_{\Gamma}(X_i) = \Gamma(\alpha_i)^{-1} \beta_i^{\alpha_i} X_i^{\alpha_i-1} e^{-\beta_i X_i} \quad (3)$$

with X_i the random variable, and the shape (α_i) and rate (β_i) parameters in Eq. (3) defined by

$$\alpha_i = (\mu_i/u_i)^2 \quad (4)$$

$$\beta_i = \mu_i/u_i^2 \quad (5)$$

In Eqs. (4) and (5), μ_i is the mean or nominal branching ratio, and u_i is the additive branching ratio uncertainty such that

$$b_i \sim \mu_i \pm u_i \quad (6)$$

Note that the bounds of b_i are 0 and 1, but with higher probability concentrated between $\mu_i - u_i$ and $\mu_i + u_i$.

In order to generate a sample of branching ratios, the NumPy library [27] is used to generate a random sample of the random variable X_i using the Γ distribution described by Eqs. (3)-(5). The sample of branching ratios is then determined by normalizing the sample of X_i according to

$$b_i = X_i / \sum_{j=1}^n X_j \quad (7)$$

where n is the number of product channels for a particular set of reactions. Equation (7) is what enforces the “sum-to-one” requirement of the distribution. For an individual simulation, the actual perturbation factor F_i applied to each reaction in a set is computed from

$$F_i = (b_i F_{tot}) / \mu_i \quad (8)$$

F_{tot} is the perturbation factor for the total rate constant, and is only utilized with Eq. (8) to compute the perturbation factor for each reaction of a set. F_{tot} itself is determined for each simulation by random sampling using the log-uniform probability distribution of Eq. (1).

Sets of important reactions with multiple branches are first selected for analysis based on the Spearman Rank Correlation (SRC) analysis of the Monte Carlo simulation results where all reactions are treated as independent variables (Case 1, as described in the next section). SRC coefficients are computed comparing predictions of ignition delay to perturbations of individual rate constants. Spearman Rank Correlation is a measure of statistical dependence between two variables that is able to better account for non-linear relationships compared to conventional Pearson Correlations. The values can range from -1 to 1, with 0 indicating that there is no relationship. A Spearman Rank Correlation of 1 indicates a perfect, monotonically increasing relationship, while -1 indicates a perfect, monotonically

decreasing relationship. Therefore, the higher the absolute value of this number, the more important the reaction is. The sign and relative magnitude can be interpreted in the same way local sensitivity indices are typically interpreted. This is the same screening method utilized by Hébrard et al. [28] in their GSA study of n-butane autoignition.

Reactions identified as most important are then subjected to Monte Carlos simulations accounting for correlations between multiple branches. Varying degrees of uncertainty in the branching ratios (u_i) are utilized, ranging from $\pm 10\%$ to $\pm 90\%$. These u_i values are chosen here just for exploratory purposes, representative of scenarios where the branching ratios are known well and poorly, respectively. Again, the uncertainty in total rate constant is $UF=5$ for all simulations. The temperature and pressure dependence of the branching ratios are neglected here, while the nominal values of the branching ratios, computed at a representative compressed temperature ($T=725$ K), are taken as the means (μ_i) of the Dirichlet distribution. Results of such a sampling procedure are presented in Section 3.4.

2.3 Monte Carlo Uncertainty Quantification Analysis

In Monte Carlo simulations [31], the uncertainty in model predictions is determined by random sampling from the uncertainty of all model input variables. To do so, an uncertainty probability distribution is assigned to each input variable. Section 2.2 describes our approach to assigning these probability distributions. The uncertainties are then propagated to the model predictions through a large number of simulations, where for each individual run, all of the input parameters are simultaneously perturbed. The probability distribution in the aggregate of all the individual simulations represents the uncertainty of the model predictions due to the uncertainty in the inputs. The advantage of this technique is that it can be applied to any physical model and can use any combination of uncertainty probability distributions in all input parameters, making this the ideal method for the

present investigation. A drawback is that it can require a large number of simulations to get sufficient convergence for statistical analysis, and it can therefore be a computationally expensive approach. However, utilizing the fast, adaptive Zero-RK solver [23] makes this a sufficiently affordable approach.

To examine the influence of correlations on predicted autoignition behavior, adiabatic, isochoric simulations are conducted using iso-octane as a single component surrogate, where the initial compressed pressure is 20 bar and four initial compressed temperatures are explored (675, 725, 825 and 900 K). As discussed in the Section 1, these conditions are relevant to conventional SI and LTC engine operation and span the negative temperature coefficient (NTC) region for iso-octane. They are also used in the Rapid Compression Machine Workshop Characterization Initiative [32], which will facilitate comparison in the future. All simulations are performed with stoichiometric mixtures of iso-octane/'air', where the inert diluent for the lowest temperatures (675 K and 725 K) is nitrogen, while the other two conditions use a blend of argon and nitrogen (75:25 molar ratio), as is done with experiments [32]. Ignition delay times are defined by the time of maximum pressure rise rate in each simulation.

The simulations are performed on the Cab cluster at LLNL, using the Zero-RK [23] solver. This new solver is able to achieve 1-3 orders of magnitude faster integration with large mechanisms when compared to alternative approaches. Based on an adaptive preconditioner, the speedup in computational time is achieved without changes to the underlying mechanism, allowing all accuracy to be retained. This permits a large number of simulations to be conducted for all cases, thereby ensuring sufficient convergence in the results. A typical job on the cluster runs 10,000 simulations on 8 nodes and 128 cores (Intel Sandy Bridge), with total wall-clock time of ~35 minutes.

Eight different cases of uncertainty in the input variables are considered. Case 1 is the baseline, where all reactions are considered to be independent variables, representative of a common assumption in other UQ studies [3]. The uncertainty in the foundational model is $UF=2$ and the rest of

the model is UF=5, with log-uniform probability, as described in Section 2.2. Case 2 and Case 3 are conditional simulations performed as part of sensitivity analysis in order to ascertain which portions of the model, foundational or fuel-specific chemistry sub-model, contribute the most to uncertainty. Case 4 examines the role of correlations among reactions that use the same rate rules, considering all of the reactions in the fuel-specific sub-mechanism. Cases 5-8 consider the impact of correlations among reactions with multiple branches, specifically considering reactions (R1)-(R3), which are identified later. Table 1 summarizes the details of all simulation cases considered in the present work.

3.0 Results

3.1 Baseline Uncertainty

One of the challenges in performing forward propagation of uncertainty by Monte Carlo simulations is the potentially large number of simulations required to attain adequate convergence for large models. With more than 7000 reactions, it would be too computationally taxing to explore every possible combination. However, this is not typically necessary since only a handful of reactions out of the entire mechanism are important at a particular condition. As discussed by Tomlin [33], the actual number of runs necessary to achieve reasonable convergence is driven by the number of reactions to which the simulation is sensitive, rather than the total number of reactions. The only caveat of this situation is that it is not known beforehand how many reactions are sensitive. Consequently, the required number of simulations is not known initially either. It is important therefore to establish how many simulations are necessary for a reasonable estimate of uncertainty using the Monte Carlo approach. In order to do so for this study, Case 1 is simulated at 825 K and 20 bar a total of 100,000 times. Figure 1 presents the results of these simulations, where the convergence of the sample median and different percentiles are shown.

It should be evident from Fig. 1 that after 100,000 runs, the probability distribution is reasonably converged, showing smooth behavior, where sample statistics should provide a useful measure of uncertainty. Increasing the number of simulations from 10,000 to 100,000 does not provide significantly more information regarding the shape of the distribution in the predicted ignition delays. For this reason, all subsequent runs at the other temperatures and for the different uncertainty cases only utilize 10,000 runs. Another observation that can be made from examining Fig. 1 is the large spread in the model predictions, as indicated by the interquartile range (IQR) and interdecile range (IDR) bands, which are the 25th to 75th percentiles and the 10th to 90th percentiles, respectively. A significant shift of the distribution away from the nominal model prediction is also observed, showing evidence of strong non-linear response in the model. The causes of these deviations are examined next and implications of the non-linear response are discussed in more detail later.

3.2 Sources of Uncertainty

There is a risk of introducing bias to uncertainty estimates if liberal uncertainty factors, i.e., too high or too low, are assigned to important reactions that are in fact well known. In the present study, the uncertainty in the foundational chemistry (C₀-C₄ species) is treated very simply, as described in Section 2.2. In order to ensure any bias for this work is negligible, two conditional Monte Carlo simulations are performed. For the first (Case 2), only uncertainty in the foundational chemistry is considered, and the rest of the model is left unperturbed. For the second (Case 3), the foundational chemistry is unperturbed, while the rest of the model (i.e., fuel-specific chemistry) is varied. It should be noted that these two cases constitute a simple sensitivity analysis in the present study and are not representative of any physically realistic condition.

Cases 1-3 are extended to the other three temperatures, 675, 725, and 900 K, and results of these simulations plotted in Fig. 2. For ease of comparison, the results for each case are presented using

uncertainty bands as defined in Fig. 1. Each cluster of four bands in Fig. 2 represents results of simulations at the same temperature, but shifted higher or lower in temperature for clarity. The continuous gray curve in Fig. 2 represents the nominal, unperturbed model prediction. For all subsequent discussions, the baseline Case 1 refers to the Monte Carlo simulations using the simplest assumption of uncertainty in the rate constants, where all reactions are considered to be independent. In Fig. 2, these results correspond to the red bands, which are furthest to the left in each cluster.

At all temperatures presented in Fig. 2, it can be seen that the median value for Case 2 (blue bands, second from the left in each cluster) is very close to the nominal value, and the uncertainty due to the foundational chemistry is relatively small, contributing the least to the variability observed in the baseline predictions (Case 1). Conversely, the variability in model predictions is very sensitive to uncertainties in the sub-mechanisms related to the fuel-specific chemistry (Case 3, green bands, third from the left), where these account for most of the spread seen in Case 1. This is especially noticeable at the transition from low temperature to NTC chemistry, i.e. 725 K. These observations corroborate the results of Hébrard et al. [28], who used GSA to study the LTC behavior of n-butane using a detailed mechanism. They observed that uncertainties in the foundational chemistry are only significant contributions to the variance in model predictions at intermediate and higher temperatures, i.e., greater than 1000 K. Of course, these results are in part a consequence of assuming that there are no correlations between reactions in the foundational chemistry and the low-temperature, fuel-specific chemistry. That may not be strictly the case under all circumstances and the origin of each rate rule must be carefully reviewed in a more rigorous analysis, which is beyond the scope of the present work.

As mentioned earlier, there is a significant shift of the central tendency of the model response away from the nominal over the entire NTC region, as indicated by the vertical shift of the median value. The fact that the model response is non-linear with respect to rate constant variation should not be surprising, especially given the large variability in the response. Only when the uncertainty is small, e.g.,

Case 2, does the model response appear linear and straddle the nominal predictions. For the other cases, where the response can span orders of magnitude, the central tendency of the distribution shifts. The implication of this finding is discussed in Section 4.

The present observations regarding sources of uncertainty raise an important question. If the predicted autoignition behavior of the fuel is very sensitive to fuel-specific reactions, and these are expected to have high uncertainty since they are less studied, and they potentially have strong correlations due, for instance, to the use of rate rules, what will be the impact of including correlations? As discussed by Smith et al. [11] for such scenarios, when uncertainties in important variables are high, if strong correlations are neglected, it can result in significantly different estimates of uncertainty. As such, it is necessary to determine the magnitude of the impact of including correlations among reactions in the present study on the forward propagation of uncertainty.

3.3 Correlations Between Rate Rules

As discussed in the Section 1, correlations can exist within a mechanism under conditions where reaction class formulations are employed with analogous rate rules assumed. To examine the influence of these correlations, simulations are performed here where the fuel-specific reactions are assumed to be perfectly correlated if the same reaction rate rule is used (Case 4). Reactions that utilize the same rate rule are initially identified using the procedure outlined in Section 2.1. The rate rules identified using this procedure for the top 12 most important reactions at 725 K are presented in Fig. 3. Other details of this figure are discussed in the next section. For Case 4, all other simulation parameters are the same as in Case 1. The results for these simulations are also included in Fig. 2 for comparison.

Viewing the results presented in Fig. 2, the most apparent observation is that when correlations between rate rules are taken into account, the uncertainty in the model predictions can change significantly, as indicated by the increased widths of the IQR and IDR bands, as well as the shifts in

median values in the 675-825 K range. At 900 K, the changes in the size of IQR and IDR bands and the median value are smaller. These differences can be attributed to two factors. As discussed in the previous sections, in the NTC region for this fuel, the autoignition delay is sensitive to more reactions than at other conditions. This was noted by Hébrard et al. [28] as the reason for why the extent of uncertainty is larger in the NTC region. When we introduce correlations between the rate controlling reactions in this region, Case 4, the changes in uncertainty will be more significant. As temperature increases, high temperature reactions become more rate controlling and the number of important reactions decreases. When correlations are considered in Case 4, they are only considered among low temperature reactions, therefore the effect of their introduction is less pronounced at higher temperatures. Correlations between high-temperature reactions can certainly exist, and could have a noticeable effect when considered, but the investigation of their effects is beyond the scope of the present study.

Though it should not be surprising that the probability distributions vary when correlations are considered, it is not possible to predict whether the uncertainty should increase or decrease before the analysis is performed. In their closely related study, Prager et al. [9] examined impact of correlations due to the use of rate rules with two sets of reactions. They observed that the change in the probability distribution of the model predictions due to the inclusion of such correlations depends on the magnitude of uncertainty considered, as well as the sign of the local and global sensitivity coefficients. This feature can be readily understood from the perspective of error propagation mathematics. As demonstrated in the formulations of Smith et al. [11], positive correlations will increase variance in the model predictions while negative correlations will decrease variance. Since the most important reactions typically include a mix of reactions with either positive or negative sensitivity coefficients, as well as varying degrees to which each is sensitive, the combined effects of including correlations is

difficult to predict. Smith et al. [11] provide guidelines for when correlations can be safely ignored.

Potential implications of the present results are discussed in Section 4.

3.4 Correlations Between Reactions with Multiple Branches

As introduced earlier, another possible source of correlations that has received little attention is that for reactions with multiple branches. For many systems, only the total rate constant for a set of reactions with multiple branches (e.g., R1) can be measured experimentally. For instance, the total rate constant can be determined from the consumption of the reactant, with the branching ratios left to be inferred from other experiments, or estimated from theory (e.g., [21; 29]). This is a common problem to most areas of chemical kinetics, and in a series of papers from Pernot et al. [6–8], factoring in these correlations was shown to significantly alter the prediction uncertainty. In order to determine the potential impact of including such correlations within autoignition predictions of iso-octane, the approach of Plessis et al. [8] to account for correlations between reactions with multiple branches is adapted here, as is described in detail in Section 2.2. For the exploratory purposes of this study, only a small subset of important reactions is considered for this analysis.

In order to facilitate this and identify the most influential reactions, SRC coefficients are computed for each reaction with respect to the predicted ignition delay times and perturbations of each reaction in the baseline simulations (Case 1), at 725 K (where uncertainty in predicted ignition delay times is greatest). The SRC coefficients for the top 12 reactions are presented in Fig. 3. Included at the bottom of the figure are four additional reactions, of lower importance, that will be relevant to subsequent discussion.

The Spearman Rank Correlation analysis presented in Fig. 3 reveals that the top five reactions, which contribute the most to uncertainty, all belong to sets of reactions with multiple branches. In

addition to the H-atom abstractions listed in (R1a)-(R1d) above (shown as blue in Fig. 3), two $ROO \leftrightarrow QOOH$ isomerization steps are highlighted (as red and green in Fig. 3, respectively), including

Since (R1a)-(R1d), (R2a)-(R2d), and (R3a)-(R3d) all belong to sets with multiple branches and at least one reaction of each set is indicated as being important by SRC analysis, these three sets are chosen for further analysis to determine the potential impact of including correlations among the branching ratios.

Nominal branching ratios for each channel are computed at 725 K and used as the mean of the generalized Dirichlet distribution for each reaction. Three different levels of uncertainty are considered for the branching ratios (b_i), $u_i = \pm 10\%$, $\pm 50\%$, and $\pm 90\%$ of nominal branching ratio μ_i . Histograms of the sample branching ratios for reaction (R1a) with uncertainties of $u_i = \pm 10\%$ and $\pm 90\%$ are plotted Fig. 4. The nominal branching ratio of (R1a) at 725 K is $\mu_i = 0.3$, and as should be evident in Fig. 4, the mean branching ratio stays the same irrespective of the specified uncertainty. The nominal branching ratios

for all product channels (R1a)–(R1d) range from 0.2 to 0.3, and the distributions of sampled branching ratios are similar to those presented for (R1a) in Fig. 4.

The perturbation factors F_i for each reaction considered in this analysis are computed from Eq. (8). A scatter plot compares the sampled perturbation factors for (R1a) and (R1b) in Fig. 5. This plot presents the actual perturbation factors by which rates of progress for reactions (R1a) and (R1b) are scaled in subsequent simulations. It should be noted that the total rate constant uncertainty remains unchanged at $UF=5$, with a log-uniform probability. The only difference between the two scatter plots in Fig. 5 is the extent of uncertainty considered for each branching ratio.

The ultimate degree of correlation between the perturbation factors is defined by the magnitudes of the total rate constant uncertainty, nominal branching ratio, and branching ratio uncertainty. For the scenario illustrated in Fig. 5, the nominal branching ratios are similar (~ 0.3 and ~ 0.2 respectively), and total rate constant uncertainty is unchanged. When the branching ratio uncertainty is small ($u_i=\pm 10\%$), the perturbation factors for each reaction are very tightly correlated, and the range of variation is almost entirely dictated by the total rate constant uncertainty. When the branching ratio uncertainty is high ($u_i=\pm 90\%$), the overall degree of correlation between the two reactions falls. However, even for this condition the correlation is not zero, and this leads to deviations in the estimated uncertainty compared to the baseline case due to the non-linear nature of the response. It is important to note that the overall correlation does not necessarily have to be positive. Different combinations of the aforementioned magnitudes can result in strongly positive, negative, or very weak overall correlations.

The perturbation factors plotted in Fig. 5 for reactions (R1a) and (R1b), as well as for the other two reactions in the set, which are not plotted, are used to perturb all four reactions in the simulation Case 5, while leaving all other reactions in the mechanism as independent variables, as is done for the baseline (Case 1). The same procedure is repeated for Case 6, where only reactions (R2a)–(R2d) are

perturbed, and Case 7, where reactions (R3a)-(R3d) are perturbed. For Case 8, all three sets are perturbed simultaneously, while again leaving all other reactions as uncorrelated. The results for these simulations are compared in Fig. 6.

In Fig. 6, for the scenarios where correlations among reaction branches are considered individually (Cases 5-7), the largest differences compared to the baseline simulations can be seen for Case 5. When the uncertainty in the branching ratios is $u_i=\pm 10\%$, the spread in the model predictions decreases significantly. As the uncertainty in the branching ratios is increased to $u_i=\pm 90\%$, the overall spread in the model predictions increases back to, and exceeds the uncertainty of the baseline case. For Cases 6 and 7, the results follow a similar trend, but the differences are significantly less pronounced. Finally, for Case 8, where all three sets of correlated reactions are combined, the results resemble those of Case 5, but with slightly lower uncertainty at the $u_i=\pm 10\%$ and $u_i=\pm 50\%$ levels.

The dramatic changes for Case 5 may be attributed to the fact that all four reactions (R1a)-(R1d) are important, as indicated by SRC analysis results shown in Fig. 3. In contrast, for Cases 6 and 7, only two reactions in each set (R2b), (R2c), (R3b), and (R3c) have a significant impact on the model predictions. The rest have inconsequential impacts on the predicted model uncertainty, as indicated by their much lower position of importance in Fig. 3. This means that, while the same strong correlations exist between the reactions in each set (R1), (R2), and (R3), these correlations are of little consequence when only one or two of these reactions are important in the simulations. However, the converse is true as well, if multiple reactions of a correlated set are important, these correlations will have a non-negligible effect on uncertainty estimates. What exactly the impact will be depends significantly on the strength and sign of the correlation. The trends of larger or smaller uncertainty bands are difficult to predict from prior knowledge, which is the same observation that was made by Carrasco et al. [6]. Of course, only a simple scenario for branching reaction correlations is considered here, as the level of uncertainty in the total rate constant is not changed, and the branching ratios are all assumed to have

the same uncertainty. As discussed in more detail in Section 2.2, Plessis et al. [8] proposed several alternatives to model the actual state of knowledge available about branching ratios.

Very recently, in a similarly motivated study using a small representative model, Manion et al. [10] showed that for UQ where very precise relative rate constant measurements are available, e.g., branching ratios, treating reactions as independent variables can grossly overestimate the extent of uncertainty in the predicted ignition delays. Their observations are consistent with the present simulation results for Case 5 in Fig. 4, where the total rate constant is unchanged but the branching ratio uncertainty is small ($u_i = \pm 10\%$). It would appear that the only way of determining the potential impact of properly including these correlations is through testing.

4.0 Discussion

The results of the present work lead to several broad implications for UQ of combustion models, and these are discussed next. Forward propagation of uncertainty is typically the first step in analyses focused on optimization, uncertainty reduction, and GSA. Assumptions made about the uncertain inputs, and which probability distributions they are assigned, are not often evaluated in detail and consensus for best practices does not yet exist. The present results suggest that a careful analysis of the underlying assumptions of uncertain inputs is probably necessary for most situations, and care must be taken in interpreting results of UQ studies focused on estimating uncertainty as well as optimization and uncertainty reduction.

For instance, in Bayesian-like methods such as the uncertainty quantification and minimization using polynomial chaos expansion (MUM-PCE) [12], often a three-step process is followed. The model is first optimized, in the least squares sense, against a comprehensive experimental data set, often using uniform probability distributions for the pre-exponential factors (*prior* model). A *posterior* distribution for the pre-exponential factors is then obtained through the application of Bayesian-like inference to the

optimized mechanism, considering only experimental uncertainty. A consistency check between the optimized model and available experimental data can be used to identify data sets that are inconsistent, and these can be removed from optimization targets. The implication of our results for such methods is that if they are applied to a mechanism that may have many correlations in the *prior*, the initial estimates of uncertainty as well as the resultant optimized mechanism will vary significantly if these correlations are taken into account or not. From the perspective of experiment-model consistency, if the *prior* model is properly defined, including consideration of correlations, then more or fewer experimental data sets may be found to be consistent.

In full Bayesian Inference methods, the implications of our results are similar. As shown by Braman et al. [34], depending on how the *prior* probability distribution of the rate constants is defined, the results of the inference procedure can be significantly altered. In cases where uniform probability distributions are utilized, the entirety of the *posterior* model is contained inside the *prior*. Thus, incorporating correlations in the *prior* could significantly alter the domain within which the *posterior* exists. The same was noted by Prager et al. [9], who stated that incorporating such correlations effectively reduces the degrees of freedom in the *prior*. In optimization schemes, this can reduce computational costs.

Cai et al. [35] utilized this reduction in degrees of freedom as an advantage in applying the MUM-PCE method for the optimization of reaction rate rules of a detailed n-pentane mechanism. Applying perfect correlations attributed to rate rules, reduced the dimensionality of the problem sufficiently to make application of MUM-PCE computationally feasible. However, they did not perform a detailed evaluation of this assumption, which is one of the objectives of the present work. In addition, they did not consider other interdependencies, such as relative correlations arising between competing reactions explored within the present work and that of Manion et al. [10]. The authors of this latter study suggested that ignoring such information in the forward propagation of uncertainty and model

optimization schemes can lead to physically unrealistic results. Our findings provide further support for these conclusions and demonstrate the impacts on a detailed model for a prototypical gasoline surrogate.

The results of our work also have significance to uncertainty minimization approaches alternative to Bayesian-like and regression based optimization methods. In the approach of Skodje et al. [36], forward propagation of uncertainty and global sensitivity analysis are utilized as a screening method to determine uncertain rate constants which contribute most to simulation uncertainty. These rate constants are then refined with *ab initio* calculations. Our findings suggest that depending on how correlations between reactions are initially treated, initial estimates of uncertainty and the parameters identified for refinement could vary significantly.

In the present work, significant non-linear response in the model is observed when different assumptions about the probability distributions of the uncertain rate constants are made. These results therefore suggest that diligence is required in interpreting nominal model predictions as a measure of the central tendency of the predictions when uncertainties are large. Similar non-linear response was observed by Hébrard et al. [28] for the auto-ignition delay predictions of n-butane in the NTC region. Recognizing this phenomenon, they utilize non-linear screening techniques (Spearman Rank Correlations) to identify the most influential parameters for further analysis by GSA. They recommended a sampling approach as opposed to linear propagation of uncertainty as the most appropriate method for capturing the non-linear behavior.

For the downstream users of chemical kinetics models, for example in engine simulations, the present results suggest caution should be exercised when considering uncertainty due to chemical kinetic models. While forward propagation of uncertainty and GSA has become a feasible problem to solve (e.g., [37; 38]), when chemical kinetics are invariably considered in the analysis, if the input

uncertainties for the kinetic mechanism are not treated carefully, strong and physically unrealistic biases can emerge in the estimates of uncertainty for the simulation results.

While the discussion so far has focused on reinforcing the need for careful evaluation of assumptions about uncertain model parameters, it is important to note that this remains a challenge. Often, the necessary information about rate constants, including possible correlations, is missing and the originating source cannot be reevaluated. A path forward must therefore address this challenge as well. Data mining approaches such as the importer tool [25] and the parsing scheme described in the present work offer a means to uncover and manage some of the missing information. This approach is viable for existing models, though it needs further development. Automated model generation tools, such as RMG [26], can potentially keep track of statistical information for new models. Future studies focused on developing new models and obtaining new rate constants should consider the impending need for well-characterized uncertainties and move towards statistically rigorous evaluation of data.

5.0 Concluding Remarks

In the present study, Monte Carlo techniques are used to quantify the uncertainty of autoignition predictions for iso-octane/air mixtures at LTC conditions using a detailed gasoline surrogate model. The influence of rate rule and multi-branch reaction correlations is examined in order to ascertain the impact on forward propagation of uncertainty. The fast chemistry solver, Zero-RK [23] running on LLNL's Cab cluster, along with automated model parsing tools [25; 26], are utilized in order to make this analysis sufficiently tractable. The results presented here indicate that significant uncertainty in the model's predictions of autoignition delay are inherent and that the majority of it can be attributed to uncertainty associated with the fuel-specific portion of the model, particularly the low-temperature reactions of the iso-octane sub-model. Strong non-linear response in the model is observed depending

on assumptions made about the uncertain rate constants, suggesting that care must be exercised in interpreting nominal model predictions as measures of central tendency of a model's response.

Simulations accounting for correlations between reactions using the same rate rules, as well as among some important reactions with multiple branches, show non-negligible differences in the overall model uncertainty when compared to simulations where all reactions are treated as independent parameters. However, trends are difficult to anticipate without initially performing the analysis. These results suggest that accurate forward propagation of uncertainty must include a careful treatment of potential underlying correlations in the model construction.

Ongoing efforts utilize detailed accountings of uncertainty factors for the foundational chemistry, reaction classes, and extents of correlations among reaction rate rules and important reactions with multiple branches. The effects of uncertain thermochemistry, pressure, and temperature dependence are topics of future work.

Acknowledgements

This manuscript has been created in part by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). This work is performed under the auspices of the Office of Energy Efficiency and Renewable Energy, Office of Vehicle Technology, U.S. Department of Energy, under contract number DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up non-exclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. Aspects of this material are based upon work supported by the National Science Foundation under Grant No. 1403171. The work at LLNL is supported by the U.S. Department of Energy, Vehicle Technologies Office (program managers Gurpreet Singh and Leo Breton) and performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under

Contract DE-AC52-07NA27344. The authors thank Professor Alison Tomlin (University of Leeds) and Dr. Michael Davis (Argonne) for helpful guidance. Lori Kaufman contributed to early portions of this work. The authors also acknowledge Pierre Bhoorasingh, Belinda Slakman, and Fariba Seyedzadeh Khanshan for help identifying the 1753 molecular structures in the LLNL gasoline surrogate model.

References

1. R.D. Reitz, Directions in internal combustion engine research. *Combust. Flame* 160 (1) (2013) 1–8.
2. Y. Ju, Recent progress and challenges in fundamental combustion research. *Adv. Mech.* 44 (2014) 201402.
3. H. Wang, D.A. Sheen, Combustion kinetic model uncertainty quantification, propagation and minimization. *Prog. Energy Combust. Sci.* 47 (2015) 1–31.
4. A.S. Tomlin, The role of sensitivity and uncertainty analysis in combustion modelling. *Proc. Combust. Inst.* 34 (1) (2013) 159–176.
5. A. Tomlin, T. Turányi, Investigation and Improvement of Reaction Mechanisms Using Sensitivity Analysis and Optimization. In F. Battin-Leclerc, J.M. Simmie, E. Blurock (eds). *Cleaner Combustion SE* - 16, Springer London, 2013, pp. 411–445.
6. N. Carrasco, P. Pernot, Modeling of Branching Ratio Uncertainty in Chemical Networks by Dirichlet Distributions. *J. Phys. Chem. A* 111 (18) (2007) 3507–3512.
7. N. Carrasco, O. Dutuit, R. Thissen, M. Banaszekiewicz, P. Pernot, Uncertainty analysis of bimolecular reactions in Titan ionosphere chemistry model. *Planet. Space Sci.* 55 (1–2) (2007) 141–157.
8. S. Plessis, N. Carrasco, P. Pernot, Knowledge-based probabilistic representations of branching ratios in chemical networks: the case of dissociative recombinations. *J. Chem. Phys.* 133 (13) (2010) 134110.
9. J. Prager, H.N. Najm, K. Sargsyan, C. Safta, W.J. Pitz, Uncertainty quantification of reaction mechanisms accounting for correlations introduced by rate rules and fitted Arrhenius parameters. *Combust. Flame* 160 (9) (2013) 1583–1593.
10. J.A. Manion, W.S. McGivern, The Importance of Relative Reaction Rates in the Optimization of Detailed Kinetic Models. *Int. J. Chem. Kinet.* 48 (7) (2016) 358–366.
11. A.E. Smith, P.B. Ryan, J.S. Evans, The Effect of Neglecting Correlations When Propagating Uncertainty and Estimating the Population Distribution of Risk. *Risk Anal.* 12 (4) (1992) 467–474.
12. D.A. Sheen, H. Wang, The method of uncertainty quantification and minimization using polynomial chaos expansions. *Combust. Flame* 158 (12) (2011) 2358–2374.
13. T. Nagy, T. Turányi, Determination of the uncertainty domain of the Arrhenius parameters needed for the investigation of combustion kinetic models. *Reliab. Eng. Syst. Saf.* 107 (2012) 29–34.
14. B.R. and R.E.P. and G. von L. and D.K. and A.B. and D.L. and D.M. and A.F. Wagner, Active Thermochemical Tables: thermochemistry for the 21st century. *J. Phys. Conf. Ser.* 16 (1) (2005) 561.

15. C.K. Westbrook, Y. Mizobuchi, T.J. Poinso, P.J. Smith, J. Warnatz, Computational combustion. *Proc. Combust. Inst.* 30 (1) (2005) 125–157.
16. E. Blurock, F. Battin-Leclerc, Modeling Combustion with Detailed Kinetic Mechanisms. In F. Battin-Leclerc, J.M. Simmie, E. Blurock (eds). *Cleaner Combustion SE – 2*, Springer London, 2013, pp. 17–57.
17. S.M. Sarathy, C.K. Westbrook, M. Mehl, W.J. Pitz, C. Togbe, P. Dagaut, H. Wang, M.A. Oehlschlaeger, U. Niemann, K. Seshadri, P.S. Veloo, C. Ji, F.N. Egolfopoulos, T. Lu, Comprehensive chemical kinetic modeling of the oxidation of 2-methylalkanes from C7 to C20. *Combust. Flame* 158 (12) (2011) 2338–2357.
18. E. Blurock, F. Battin-Leclerc, T. Faravelli, W. Green, Automatic Generation of Detailed Mechanisms. In F. Battin-Leclerc, J.M. Simmie, E. Blurock (eds). *Cleaner Combustion SE – 3*, Springer London, 2013, pp. 59–92.
19. M. Mehl, W.J. Pitz, C.K. Westbrook, H.J. Curran, Kinetic modeling of gasoline surrogate components and mixtures under engine conditions. *Proc. Combust. Inst.* 33 (1) (2011) 193–200.
20. M. Mehl, J.Y. Chen, W.J. Pitz, S.M. Sarathy, C.K. Westbrook, An Approach for Formulating Surrogates for Gasoline with Application toward a Reduced Surrogate Mechanism for CFD Engine Modeling. *Energy & Fuels* 25 (11) (2011) 5215–5223.
21. J. Badra, A. Elwardany, A. Farooq, Shock tube measurements of the rate constants for seven large alkanes+OH. *Proc. Combust. Inst.* 35 (1) (2015) 189–196.
22. D. Vuilleumier, H. Selim, R. Dibble, M. Sarathy, Exploration of Heat Release in a Homogeneous Charge Compression Ignition Engine with Primary Reference Fuels, SAE Technical Paper 2013.
23. M.J. McNenly, R.A. Whitesides, D.L. Flowers, Faster solvers for large kinetic mechanisms using adaptive preconditioners. *Proc. Combust. Inst.* 35 (1) (2015) 581–587.
24. W.K. Metcalfe, S.M. Burke, S.S. Ahmed, H.J. Curran, A Hierarchical and Comparative Kinetic Modeling Study of C1 – C2 Hydrocarbon and Oxygenated Fuels. *Int. J. Chem. Kinet.* 45 (10) (2013) 638–675.
25. V.R. Lambert, R.H. West, Identification, Correction, and Comparison of Detailed Kinetic Models. In 9th US National Combustion Meeting 2015.
26. C.W. Gao, J.W. Allen, W.H. Green, R.H. West, Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* 203 (2016) 212–225.
27. NumPy v1.9.0 2014. www.numpy.org
28. É. Hébrard, A.S. Tomlin, R. Bounaceur, F. Battin-Leclerc, Determining predictive uncertainties and global sensitivities for large parameter systems: A case study for n-butane oxidation. *Proc. Combust. Inst.* (35) (2015) 607–616.
29. R. Sivaramakrishnan, J. V Michael, Rate Constants for OH with Selected Large Alkanes: Shock-Tube Measurements and an Improved Group Scheme. *J. Phys. Chem. A* 113 (17) (2009) 5047–5060.
30. J.W. Lingwall, W.F. Christensen, C.S. Reese, Dirichlet based Bayesian multivariate receptor modeling. *Environmetrics* 19 (6) (2008) 618–629.
31. J.C. Helton, J.D. Johnson, C.J. Sallaberry, C.B. Storlie, Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliab. Eng. Syst. Saf.* 91 (10–11) (2006) 1175–1209.

32. International RCM Workshop. <http://www.anl.gov/energy-systems/project/international-rcm-workshop>.
33. A.S. Tomlin, The use of global uncertainty methods for the evaluation of combustion mechanisms. *Reliab. Eng. Syst. Saf.* 91 (10–11) (2006) 1219–1231.
34. K. Braman, T.A. Oliver, V. Raman, Bayesian analysis of syngas chemistry models. *Combust. Theory Model.* 17 (5) (2013) 858–887.
35. L. Cai, H. Pitsch, Mechanism optimization based on reaction rate rules. *Combust. Flame* 161 (2) (2014) 405–415.
36. R.T. Skodje, A.S. Tomlin, S.J. Klippenstein, L.B. Harding, M.J. Davis, Theoretical Validation of Chemical Kinetic Mechanisms: Combustion of Methanol. *J. Phys. Chem. A* 114 (32) (2010) 8286–8301.
37. Y. Pei, R. Shan, S. Som, T. Lu, D. Longman, M.J. Davis, Global Sensitivity Analysis of a Diesel Engine Simulation with Multi-Target Functions. SAE Technical Paper 2014.
38. J. Kodavasal, Y. Pei, K. Harms, S. Ciatti, A. Wagner, P. Senecal, M. García, S. Som, Global Sensitivity Analysis of a Gasoline Compression Ignition Engine Simulation with Multiple Targets on an IBM Blue Gene/Q Supercomputer. SAE Technical Paper 2016.

Figure Captions

Table 1 Summary of different simulation cases considered in the present investigation.

Figure 1. Convergence of the sample median and different percentiles as a function of sample size for at 825 K, with all rate constants treated as independent variables. The colored band on the right represents the interdecile range (IDR, light band), interquartile range (IQR, dark band), and the sample median (horizontal dash), used in subsequent figures.

Figure 2. Monte Carlo simulation results for 675 K, 725 K, 825 K, and 900 K, for the baseline uncertainty Case 1, as well as contributions to the uncertainty due to uncertain reactions in the foundational model (Case 2), fuel-specific chemistry (Case 3), and the impact of including correlations among reaction using the same rate rule (Case 4). See text for a detailed explanation of the figure.

Figure 3. Ranking of the importance of reactions according their contribution to uncertainty at 725 K with all rate constants considered as independent variables (Case 1). Blue – correspond to reactions (R1a)-(R1d), red – correspond to reactions (R2a)-(R2d), green – correspond to reactions (R3a)-(R3d).

Figure 4. Histograms of the sampled branching ratios of reaction (R1a) in the Case 5 simulations.

Figure 5. Sampled perturbation factors F_i for reactions (R1a) and (R1b) in the Case 5 simulations.

Figure 6. Monte Carlo simulation results for the baseline uncertainty at 725 K – Case 1, compared to simulations where correlations among branching ratios of reaction sets (R1a-d) – Case 5, (R2a-d) – Case 6, (R3a-d) – Case 7, and all three combined in Case 8. In each case, all other reactions are treated as independent variables.

Table 1 Summary of different simulation cases considered in the present investigation.

Case #	Description	Reactions perturbed
1	Baseline	All rate constants independent
2	Conditional	Only foundational chemistry varied
3	Conditional	Only fuel-specific chemistry varied
4	Rate rule correlations	All reactions using the same rate rule outside the foundational sub-model are treated as perfectly correlated, all other reactions independent
5	Branching ratio correlations	R1a-R1d considered correlated, all other reactions independent
6	Branching ratio correlations	R2a-R2d considered correlated, all other reactions independent
7	Branching ratio correlations	R3a-R3d considered correlated, all other reactions independent
8	Branching ratio correlations	R1, R2, R3 correlations combined, all other reactions independent

Figure1

Figure2

Figure3

Figure4

Figure5

Figure6